Finding the Words: Comparing the Progression of Semantic Feature Analysis and Phonological Components Analysis in Adults with Post-Stroke Aphasia

Katherine Haentjens

School of Communication Sciences and Disorders

Faculty of Medicine

McGill University, Montreal

June 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Communication Sciences and Disorders

Table of Contents

Abstract

40% - 60% of people with aphasia experience persistent word-finding difficulties into the chronic stage (starting 6 months after the stroke). Previous studies have investigated Semantic Feature Analysis (SFA) and Phonological Components Analysis (PCA), two word-finding interventions that use the generation of semantic features for SFA (e.g., category) and phonological features for PCA (e.g., first sound) to improve naming. Although the generalization to untreated items has been limited, studies have demonstrated improvements on probe word naming for both interventions. However, research concerning within-intervention effects and generalization gains to alternative contexts has been limited. This study investigated the treatment effect, generalization and presence of carry-over to different contexts for SFA and PCA probe word naming as well as their within-intervention effects in four individuals with chronic post-stroke aphasia. Baseline measures included standardized assessments and image naming tasks, where the latter was used to generate three lists (SFA treated, PCA treated, untreated), each containing 20 words that were not named correctly in at least one baseline session out of three. Six SFA and six PCA sessions were then provided concurrently to each participant, each once per week, in an alternating, counterbalanced treatment design. While only one participant experienced significant gains on SFA and PCA probe word naming, these gains were maintained four weeks post-intervention and were transferred to two types of carry-over items: different pictures of the same items and pictures of items shown in a natural context. Two participants experienced trends of improvement on probes, one for SFA and one for PCA, while one participant did not significantly improve on either set of probes. Furthermore, while generalization to untreated items did not reach significance for any participant, some generalization of gains to standardized assessment was observed. Although rarely equivalent for SFA and PCA interventions, all participants experienced at least a certain degree of within-intervention improvement over the progression of sessions on measures such as a reduction in the number of forced choices required for feature generation and/or a reduction in the number of words never named during intervention sessions. While the concept of clinical significance should be considered when choosing items to target, this study suggests that naming gains can still be accomplished at the chronic stage of post-stroke aphasia and that those gains, although present in different forms and to different degrees per individual, are not limited to probe word naming and may have the potential to transfer to daily life.

Résumé

Entre 40% et 60% des personnes ayant une aphasie éprouvent des difficultés d'accès lexical persistantes, même durant la phase chronique (au moins 6 mois après un accident vasculaire cérébral, AVC). Plusieurs études ont examiné le *Semantic Feature Analysis (SFA)* et le *Phonological Components Analysis (PCA)*, deux interventions d'accès lexical qui utilisent la génération de traits sémantiques pour SFA (ex. la catégorie) et de traits phonologiques pour PCA (ex. le premier son du mot) afin d'améliorer la dénomination. Bien que la généralisation aux stimuli non traités était limitée dans les études publiées jusqu'à maintenant, elles ont tout de même montré des améliorations sur les mesures de progrès (*probes*) pour les deux interventions. Cependant, la recherche concernant les effets obtenus pendant les séances d'intervention et la généralisation des améliorations aux contextes alternatifs est limitée. Cette étude porte sur l'effet du traitement, la généralisation et la présence de transfert des progrès à des contextes alternatifs pour la dénomination de mots pendant les mesures de progrès pour SFA et PCA, ainsi que sur les effets obtenus pendant les interventions avec quatre personnes atteintes d'aphasie après un AVC. Les mesures de ligne de base comprenaient des évaluations standardisées et des tâches de dénomination d'image. Trois listes (traitée avec SFA, traitée avec PCA et non traitée) ont été formées, chacune contenant 20 mots qui n'ont pas été nommés correctement dans au moins une séance de ligne de base sur trois. Six séances de traitement SFA et six séances de traitement PCA ont ensuite été offertes simultanément à chaque participant, une fois par semaine, dans un plan de traitement contrebalancé et alterné. Même si un seul participant a progressé d'une manière significative pour la dénomination de mots des listes SFA et PCA, ces améliorations ont été maintenues quatre semaines après la dernière séance d'intervention et ont été transférées à deux contextes alternatifs: la dénomination d'images différentes des mêmes stimuli traités et la dénomination d'images de ces stimuli représentés dans un contexte naturel. Deux participants ont montré des tendances d'amélioration pendant les mesures de progrès, l'un pour SFA et l'autre pour PCA. Un dernier participant ne s'est pas amélioré de façon significative pour les mesures de progrès, tant pour SFA que pour PCA. Bien que la généralisation aux items non traités n'était pas significative pour aucun des participants, la généralisation a été observée pour certains participants dans certaines évaluations standardisées. Même si les améliorations étaient rarement équivalentes pour les interventions SFA et PCA, tous les participants ont connu au moins un certain degré d'amélioration pendant les séances d'interventions. Ces améliorations se sont manifestées par la

réduction du nombre de choix forcés requis pour la génération de traits sémantiques ou phonologiques et/ou la réduction du nombre de mots jamais nommés pendant les séances. Bien que le concept de l'importance clinique doive être pris en compte dans le choix des stimuli à cibler, cette étude suggère que les améliorations de dénomination peuvent encore être réalisées durant la phase chronique de l'aphasie après un AVC. De plus, même si les améliorations étaient d'une forme et d'un degré différent chez les participants, elles n'étaient pas limitées à la dénomination pendant les mesures de progrès. Cela suggère que les améliorations peuvent être transférées à la vie quotidienne.

Acknowledgements

I wish to express my profound appreciation to Dr. Noémie Auclair-Ouellet, my thesis supervisor, for her valuable assistance and guidance during all phases of the thesis project, including but not limited to, the design and development of the project, the data collection, the statistical analyses and the thesis review. Without her support and feedback, this thesis would not have been possible, especially within the short timeframe allotted. Her provision of a research stipend for the duration of the two semesters of research was also extremely appreciated.

Regarding participant recruitment, I wish to acknowledge the assistance provided by Ariana Fraid, Speech-Language Pathologist (S-LP) at the Association Québécoise des Personnes Aphasiques, and Lauren Tittley, S-LP at McGill University's School of Communication Sciences and Disorders' (SCSD) outpatient teaching clinic.

Additionally, I am grateful for the assistance provided by the volunteers, Emma Feldhake and Sophie Mandl, and the research assistants, Anna Malakhova and Allison Chen, in Dr. Auclair-Ouellet's Lab. The latter found the stimuli images used in the study and completed a preliminary literature search, while the others assisted with the consensus document outlining appropriate features for the interventions and completed inter-rater reliability tasks.

I also wish to extend my thanks to the members of my internal committee, Dr. Shari Baum and Dr. Susan Rvachew, for their timely feedback and valuable comments for both the thesis proposal and final thesis as well as my external committee member, Dr. Laura Monetta, for her final revisions and suggestions regarding my master's thesis project.

Special thanks should also be provided to the graduate students in the reading group at the SCSD for their recommendations and comments regarding the logistics of the Matlab code generated for the project. The support and encouragement provided by my family and friends during the past year have also been irreplaceable.

Contributions of Authors

All chapters of the thesis were written by Katherine Haentjens, with revisions and edits provided by her supervisor, Dr. Noémie Auclair-Ouellet. Feedback and recommendations were also provided by all committee members.

1. Introduction: Rationale and Objectives

This study aimed to characterize the progression of the effects of two word-finding interventions, Semantic Feature Analysis (SFA) and Phonological Components Analysis (PCA) in individuals with chronic (at least 6 months post-stroke) aphasia. Considering the lack of consensus regarding intervention allocation based on underlying impairment (semantic versus phonological) versus remaining strengths (Neumann, 2018), both SFA and PCA protocols were completed with four participants with chronic post-stroke aphasia. The information collected during SFA and PCA probe word naming as well as the information collected during the progression of both SFA and PCA interventions during the sessions themselves (within-intervention gains) assisted in clarifying the presence of alternative gains to probe naming, the value feature generation has on naming and the progression of an individual's reliance on the therapist for feature generation.

A secondary focus of this study was to further the knowledge on generalization measures for both SFA and PCA interventions. Since response generalization gains (i.e., generalization to untreated items) and generalization gains to connected speech have been found to be inconsistent across previous studies (van Hees, Angwin, McMahon, & Copland, 2013; Coelho, McHugh & Boyle, 2000; Leonard, Rochon, & Laird, 2008), other forms of generalization were investigated. While response generalization measures were still collected through the naming of a control word list, stimulus generalization (carry-over) measures were also included. During both post-intervention and maintenance stages, participants were asked to name word list items presented in dissimilar instances (e.g., a yellow pepper versus a green pepper) and instances of the items in a natural context (e.g., a hammer alone versus a hammer in a tool box or on a work bench). Changes on standardized assessments were also examined. Investigations of other forms of generalization assisted in determining which gains have the potential to transfer to an individual's daily life.

2. Literature Review

2.1 Characteristics and View of Chronic Aphasia

One out of three stroke survivors is affected by aphasia, a language disorder affecting the receptive and expressive language abilities of an individual in both the oral and written modalities in the initial weeks following a stroke (Heart & Stroke, 2017). For 40% to 60% of these individuals, aphasia persists from the acute (less than 6 months post-stroke) to the chronic (at least 6 months post-stroke) stage (Meinzer, Djundja, Barthel, Elbert, & Rockstroh, 2005). Previously, it was assumed that language rehabilitation was most effective in the acute stage of post-stroke aphasia and that improvements in the chronic stage of aphasia were limited (Moss & Nicholas, 2006). However, this assumption may have been clouded, at least in part, by the higher proportion of spontaneous recovery exhibited by individuals in the acute stage post-stroke (Moss & Nicholas, 2006). Although spontaneous recovery decreases and gains may be seen as more modest into the chronic stage, significant improvements can still be experienced in this stage of post-stroke aphasia (Breitenstein et al., 2017). Specifically, once in the chronic stage, the amount of time post-stroke is not an effective predictor of rehabilitative language gains (Moss & Nicholas, 2006). Although research must continue to investigate specifics regarding the best practices for rehabilitation implementation and its translation to functional communication, the continuation of language intervention into the chronic stage of aphasia is warranted (Moss & Nicholas, 2006).

Aphasia is a heterogeneous condition, where people with aphasia (PWA) often present with differential degrees of severity for one or more modalities including language comprehension, language expression, reading and writing (Brady, Kelly, Godwin, Enderby, & Campbell, 2016). Word-finding difficulties, also known as anomia, are a common and pervasive symptom experienced by individuals diagnosed with all severities and subtypes of aphasia (e.g., Broca's, Wernicke's, etc.; Efstratiadou, Papathanasiou, Holland, Archonti, & Hilari, 2018; Nickels, 2002).

These difficulties are frequently experienced in daily activities at both the acute and chronic stages post-stroke (Meinzer et al., 2005). Due to its prevalence, various intervention tools have been developed for Speech-Language Pathologists (S-LP) to use during therapy to remediate elements of anomia in both the acute and chronic stage (Nickels, 2012).

2.2 Theories of Word Production & Associated Disruptions in Word Naming

Different theoretical approaches to word production exist, including discrete or serial models (e.g., Levelt, Roelofs, & Meyer, 1999) and interactive activation models (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). For both approaches, lexical access is thought to require at least two major levels of processing, semantic processing and phonological processing (Hashimoto, 2012). In the serial model, once semantic information is retrieved at the conceptual representation level, information is processed in a serial manner from the lemma (one word with semantic and syntactic specification) to the lexeme, where information regarding the word's phonology, metric structure and stress patterns is added (Levelt et al., 1999; Lambon Ralph, Sage, & Roberts, 2000). Conversely, in the interactive activation model, information is assumed to travel simultaneously and bidirectionally across three levels, the semantic processing (semantic features), the lexical selection (selection of a lexical unit) and the phonological processing (phoneme) levels (Dell et al., 1997; Lambon Ralph et al., 2000).

Activation within the semantic level is assumed to proceed by the automatic spreading of activation (Collins & Loftus, 1975). This theory indicates that concepts are represented as a series of interconnected nodes. When a concept is activated, its activation spreads to semantically related or associated nodes in the lexicon. The amount of activation spread is dependent on the strength of the secondary concept's relationship with the originally activated node. The connections between nodes are also considered to be bidirectional (Collins & Loftus, 1975). For example, the concept *tree* would strongly spread to concepts such as *trunk* and *branches*, whereas it would less strongly activate concepts such as *shade* or *squirrel*. Comparably, in the reverse direction, if *branches, roots, trunk* and *leaves* were activated, *tree* would be more strongly activated than *flower* (Collins & Loftus, 1975; Gravier et al., 2018). During the naming process, information from the most activated node at the semantic processing level would be the signal transmitted towards the phonological processing level (Collins & Loftus, 1975; Gravier et al., 2018; Hashimoto, 2012).

Lambon Ralph et al. (2000) completed experimental manipulations with two participants presenting with classical anomia post-traumatic brain injury (TBI). Manipulations attempted to increase or to reduce the degree of an individual's anomia. They found that repetition priming with a three-trial lag as well as presenting a target word's first phoneme could temporarily decrease their participant's anomia (Lambon Ralph et al., 2000). Alternatively, when presented with an incorrect phonemic cue, with a semantically related, but incorrect word, or when items were grouped by semantic category, anomia could be temporarily increased (Lambon Ralph et al., 2000). According to the authors, results were most consistent with Dell et al.'s (1997) interactive activation model, due to its inclusion of feedback connections between levels. Bidirectional interactions were deemed necessary to maximize the feedback between semantic and phonological cues (e.g., when trying to name *tiger*, obtaining the first phoneme cue [t] would be fed back to eliminate all animals with another first phoneme) to assist in eliminating unwanted lexical items and raising the target word's activation above its selection threshold. Additionally, their results indicated the need for bidirectional interactions between levels to continuously maintain the target word's association to its meaning (i.e., to the semantic level) even during later stages of processing.

Despite differences regarding the models of information processing for lexical access (serial or interactive activation models), both models globally assume the need for semantic processing (word meaning), phonological processing (word form) and a connection between these two levels (Hashimoto, 2012). Thus, deficits at either of these levels, at both of these levels or in the connection between these levels could result in word-finding difficulties (Hashimoto, 2012). Errors or paraphasias produced by patients with aphasia are often used to help determine the level of impairment. Semantic paraphasias (closely related items in word meaning, such as *cat* for *dog*) are commonly associated with impairments at the semantic processing level, while phonemic paraphasias (closely related items in word form, such as *cook* for *look*) are commonly associated with impairments at the phonological processing level or at the connection between the semantic and phonological processing levels (Hashimoto, 2012; van Hees et al., 2013; Gravier et al., 2018).

Since naming requires at least partially preserved semantic processing, at least partially preserved phonological processing and a functional connection between these processes, different language interventions have been developed to target either the semantic or the phonological processing levels of word naming (Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1985; Boyle & Coelho, 1995; Leonard et al., 2008). Traditional semantic interventions often include word-picture matching, the generation or verification of semantic features and contextual priming tasks, whereas phonological interventions often include repetition and phonemic cueing tasks (Efstratiadou et al., 2018; Nickels, 2002). However, while interventions were developed separately, it is now more explicitly acknowledged that both processes are involved in picture naming for either intervention (Meteyard & Bose, 2018). Phonological cueing has also been suggested to be more effective than semantic cueing since it is effective for a greater number of individuals and since it can be effective for individuals with both semantic and phonological impairments (Meteyard & Bose, 2018; van Hees et al., 2013). Assuming an interactive activation model of processing, Meteyard & Bose (2018) suggest that phonological cues allow for information to be fed back to picture recognition processes so as to support naming.

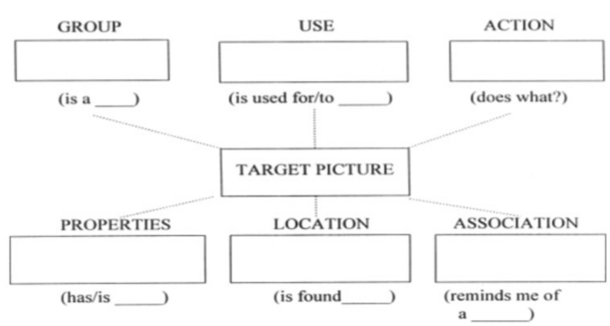
2.3 Feature Generation Interventions

Semantic Feature Analysis (SFA) and Phonological Components Analysis (PCA) are two comparable and commonly implemented word-finding interventions. The former was initially introduced in the 1990s (Massaro & Tompkins, 1994). PCA was then developed off of SFA in the 2000s (Leonard et al., 2008). Each intervention is presented separately in the following section.

2.3.1 Semantic Feature Analysis (SFA)

SFA was first developed to assist in the remediation of anomia for object word naming in traumatic brain injury (Massaro & Tompkins, 1994). The intervention was then expanded for use with individuals with post-stroke aphasia to target the semantic processing level of object word naming (Boyle & Coelho, 1995). SFA involves the generation of semantic features through the completion of feature analysis charts (Figure 1) so as to assist in activating a target word. The six semantic features of *category/group* ("It is a/an ___"), *use* ("What do you use it for?"), *action* ("What does it do?"), *properties* ("This has/is ___"), *location* ("Where do you find it?) and *association* ("It makes me think of ___") are typically included in the SFA feature analysis chart.

Figure 1: Layout illustrating SFA and the required features (Coelho et al., 2000)



Although variations exist, the original SFA protocol (Boyle & Coelho, 1995) involves presenting an image of a difficult-to-name word to an individual and asking them to name it. Irrespective of their success, a clinician assists the individual in generating each semantic feature, one at a time. When a client is unable to self-generate a feature, the clinician typically provides one in their place in both verbal and written forms. Features are also written on the feature analysis chart for future consultation. Following the completion of the feature analysis chart, the features are reviewed by the clinician and the individual is asked to (re)name the word. If unsuccessful, the individual is asked to repeat the correct word after the clinician. Once the entire process is completed for an item, another word can be targeted. As the individual becomes more familiar with the SFA protocol, their reliance on the clinician for prompting and for feature generation is expected to decrease (Neumann, 2018; Boyle & Coelho, 1995).

The hypothesis regarding SFA's method of action is imbedded in the theory of spreading activation at the semantic level (Collins & Loftus, 1975). It is suggested that spreading activation is impaired in aphasia and that lexical entries are often underspecified, making it difficult to consistently retrieve words (Gravier et al., 2018). Since spreading activation cannot automatically add the specification required for a lexical entry to reach its threshold of activation, SFA was thought to "manually" complete the specification process for target words through the process of feature generation. As more features are generated during SFA, distinguishing features (e.g., *bark* and *plays fetch*) emerge so as to encourage the selection of the appropriate word (e.g., *dog*) from other potential, initially appropriate lexical entries (e.g., *cat*). As the "facilitated" spreading of activation continues, activation is thought to eventually converge and become highest on the target word, allowing it to be correctly selected (Hashimoto, 2012; Gravier et al., 2018). Furthermore, it was thought that gradually strengthening and activating semantic representations through practice with SFA could assist in naming other words within a similar semantic category as well as assist in restoring the semantic network to a pre-aphasia state (Boyle & Coelho, 1995).

Although initially developed under this restorative framework, where SFA was thought to restore the semantic network, limited and inconsistent response generalization of treated items to untreated items has led investigators to consider SFA as more compensatory in nature (DeLong, Nessler, Wright, & Wambaugh, 2015; Wambaugh, Mauszychi, Cameron, Wright, & Nessler, 2013). Instead of restoring the impaired network, the process of completing the feature analysis chart may teach a PWA to use this process as a self-cueing strategy to assist in word naming. With practice, the self-cueing strategy could be used when confronted with difficult-to-name words outside of therapy (Efstratiadou et al., 2018).

Variations regarding feature generation for the purposes of confrontational word naming have been developed. While the standard protocol for object word naming is most typically employed (Boyle & Coelho, 1995), the methodological rationale underlying decisions such as the choice and order of features to be generated is unclear (Hashimoto, 2012). Modifications to the number of features employed have been studied (Hashimoto, 2012; Mehta & Isaki, 2016), where for example combining the *action* and *use* features in SFA has led to similar gains as compared to using all six features (Hashimoto, 2012). Furthermore, while the original protocol had clinicians automatically providing a feature to the PWA when unable to do so on their own (Boyle & Coelho, 1995), some studies alternatively required clinicians to ask the PWA to choose a feature from a limited, predetermined list of options (Leonard et al., 2008). In 2018, Gravier et al. supported this adaptation to the SFA protocol by identifying that the active participation of individuals during the feature generation process, even if only in the form of binary forced choice decisions between items containing dissimilar features, is integral to SFA's success. The role of active participation and engagement during naming tasks has also been suggested to allow for deeper levels of processing for both semantic and phonological cues (Bose, 2013; Meteyard & Bose, 2018).

In addition to the adaptations mentioned, variations to the protocol have expanded SFA for the purposes of the confrontational naming of action words (Wambaugh, Mauszycki, & Wright, 2013). For action word naming, certain features included in the protocol differed from the traditional SFA features so as to be relevant for action words. The modified features included: *function* (instead of *action/use*), *description* (instead of *properties*), *context* (instead of *location*) and *other/personal* (instead of *association*). The *group/category* feature was maintained, while the *special features* feature (e.g., *special features* of *dog* could be that it *barks* or *wags its tail)* was added. However, limited data is available on this expansion (Wambaugh et al., 2013).

2.3.2 Phonological Components Analysis (PCA)

In 2008, Leonard et al. proposed a significant variation to the SFA protocol, where SFA's principles were adapted to target the phonological level in the form of PCA. In PCA, the same protocol of feature generation was applied, with the exception of the inclusion of phonological, instead of semantic features. The five phonological features typically included for PCA are: *first sound* ("What sound does the word start with?"), *number of syllables* ("How many parts/beats does the word have?"), *rhyme* ("What does this word rhyme with?"), *final sound* ("What sound does it end with?") and *first sound associate* ("What is another word that starts with the same sound?").

While successful outcomes following the PCA procedure have led to its continued implementation (van Hees et al., 2013), individuals with aphasia have been reported to have difficulty with regards to the generation of certain phonological information (Lambon Ralph et al., 2000). Specifically, Lambon Ralph et al. (2000) indicated that their participants had difficulty self-generating the first letter/sound of target words, but were able to accurately identify a word's syllable length. Despite these difficulties with the self-generation of phonological cues, Meteyard & Bose (2018) indicated that when phonological and semantic cues were provided to participants by an experimenter, phonological cues were more beneficial than semantic cues for naming success. Similarly, Pease & Goodglass (1978) indicated that the most effective cues for naming were providing a PWA with the word's first sound and with completion sentences. In addition, a combination of semantic and phonological cues has been reported to provide the most benefit for individuals with milder anomia severities (Pease & Goodglass, 1978).

Regarding PCA's proposed mechanism of action, Leonard et al. (2008) indicated that PCA was developed and modelled off of SFA. Integral to the hypothesis of SFA's mechanism of action is the spreading activation model (Collins & Loftus, 1975), where activated concepts are spread to closely related concepts so as to (re)build semantic networks and to assist in retrieving the word that reaches a certain threshold of activation (Boyle & Coelho, 1995). However, PCA is not explicitly reported to follow the same concept of spreading activation and the literature does not specify another hypothesized mechanism of action for PCA (Leonard et al., 2008). Since phonology and its connection to semantics are largely arbitrary (Lambon Ralph et al., 2000), depending on the theory of word production considered (serial versus interactive activation models), the completion of spreading activation at the phonological level could act as a greater detriment than aid. Assuming a serial model of word production, where information is processed linearly from semantics to phonology (Levelt et al., 1999), the spreading of activation to words containing similar phonemes could result in the selection of an incorrect target word. Without feedback to the semantic level, this incorrect word choice would not be able to be rectified (Lambon Ralph et al., 2000). For example, if spreading activation occurred following the selection of the first phoneme [l], where many words starting with this phoneme were activated (e.g., *lemon, lime, loss,* etc.), *lemon* may be selected instead of *lime*. Without a feedback loop to the semantic level to re-indicate that a green, sour fruit was expected, a similar but incorrect word could be retrieved.

2.4 Theoretical Comparisons between SFA and PCA

With respect to serial models of naming, SFA functions in the same direction as normal word processing (from semantics to phonology), while PCA functions in the opposite direction. In

SFA, the generation of semantic features can assist the naming process since the semantic information can naturally spread from lemma to lexeme levels (Boyle & Coelho, 1995). However, in PCA, the generation of phonologically assistive features would only be generated at the final stage in the model and thus would not be fed back to the semantic processing level and would not be able to revise the potentially incorrectly chosen lemma at the start of lexical retrieval (Leonard et al., 2008; Levelt et al., 1999).

With regards to the interactive activation model (Dell et al., 1997) and SFA feature generation, while information would still travel in the forward direction from semantics to phonology, information could additionally be fed back to the semantic level once phonological information was added so as to revise the lexical unit initially chosen (Dell et al., 1997). With this model assumed for PCA, even if spreading activation activated all words identified as starting with a particular phoneme as per the PCA protocol (e.g., all words beginning with [d]), the information would be fed back and revised at the semantic processing level. The interaction between semantic and phonological cues could revise the originally activated lexical selections to assist in successful naming (Dell et al., 1997; Lambon Ralph et al., 2000). For example, if the phoneme [d] was fed back to the semantic processing level, it could reduce the number of originally activated animal names (e.g., *dog, cat and horse*), such that *dog* would be the lexical unit selected (Lambon Ralph et al., 2000). Thus, even if generating certain phonological features such as the first sound remained difficult (Lambon Ralph et al., 2000), the ability to generate any phonological feature (e.g., syllable length) during the PCA protocol could serve not only to activate the phonological processing level, but it could also activate the feedback loop to a concept's semantic representation (Meteyard & Bose, 2018). The bidirectional process of activation could then continue until the target word was retrieved (Dell et al., 1997; Lambon Ralph et al., 2000).

2.5 Applied Comparisons between SFA and PCA

Since PCA was developed, a number of studies have completed both SFA and PCA with the same participants to investigate their relative effectiveness (van Hees, 2013; Neumann, 2018; Hashimoto, 2012; Sadeghi, Baharloei, Moddarres, Zadeh, & Ghasisin, 2017). Considering the types of semantic versus phonological features that SFA and PCA target and considering the processing levels discussed, assumptions were developed regarding the profiles of impairment most theoretically suited for SFA and PCA. However, impairments experienced by a PWA are rarely clear-cut (Kasselimis, Simos, Peppas, Evdokimidis, & Potagas, 2017). An individual will rarely have perfectly intact semantic processing or perfectly intact phonological processing. Instead, they often experience relative strengths and relative weaknesses. For example, when a PWA is classified as having impairments at the semantic processing level, but intact phonological processing, this indicates that despite the presence of more minor phonological difficulties, phonological processing is considered to be the PWA's relative strength (Efstratiadou et al., 2018). Furthermore, although the literature highlights that SFA targets semantic processing and PCA targets phonological processing, this distinction is not perfectly exact since each intervention incorporates at least some element of the other processing level (Meteyard & Bose, 2018). SFA still requires the verbal production of a word (phonological task) and PCA still utilizes an image stimulus during the task, requiring semantic processing (Leonard et al., 2008). With these caveats in mind, it was originally thought that an intervention targeting a defective process could remediate that specific level of impairment (Wambaugh, Linebaugh, Doyle, Martinez, Kalinyak-Fliszar, & Spencer, 2001). One hypothesis was that SFA would assist in the remediation of anomia for individuals with preserved phonological functions, but semantic impairments, whereas PCA would be most suitable for individuals with preserved semantic functions, but phonological impairments

(Boyle & Coelho, 1995; Hashimoto, 2012). Conversely, as benefits have been observed when interventions have targeted a relatively spared level of processing, the alternative hypothesis was also proposed, that an intervention could employ an individual's relative strengths to improve another level of impairment (van Hees et al., 2013; Neumann, 2018). Here, SFA would assist in the remediation of impairments at the phonological processing level when semantic processing is intact, whereas PCA would target impairments at the semantic processing level when phonological processing is intact. However, results have inconsistently supported either prediction (Leonard et al., 2008; Neumann, 2018). Neither therapy's success is consistently associated with the allocation of an intervention based on a particular pattern of deficits or based on residual strengths.

When both SFA and PCA have been compared within a single study, dissociations in the effects of treatment have been reported. For example, in van Hees et al. (2013), two participants with primarily semantic deficits significantly improved following PCA, but not SFA, whereas two participants with primarily phonological deficits significantly improved following both SFA and PCA, but improvements were better maintained with SFA. The four participants presenting with impairments in the connection between semantic and phonological processing experienced differential results, where one participant did not improve following either SFA or PCA, one participant experienced gains only following PCA, one participant experienced comparable gains and degrees of maintenance for both interventions, and a final participant experienced comparable gains from both interventions, but with maintenance only obtained for PCA (van Hees et al., 2013). Due to positive results following either prediction of treatment allocation (impairment-based and residual-strength based) in van Hees et al. (2013) and in other studies such as Neumann (2018), it has been suggested that a PWA can benefit from either SFA and/or PCA interventions, irrespective of their pattern of underlying deficits. Thus, the allocation of treatments based on underlying deficits may not be the most effective predictor of intervention success (Meteyard & Bose, 2018).

Aside from underlying strengths and/or deficits, another important predictor of intervention success is aphasia severity. In a systematic review of SFA intervention studies, Efstratiadou et al. (2018) suggested that SFA may be best suited for, and can lead to larger gains with, individuals with fluent subtypes of aphasia and moderate or mild degrees of severity. Following an intervention study with PWA at the chronic stage that did not use SFA or PCA, Breitenstein et al. (2017) also proposed that baseline stroke severity was a significant predictor of intervention success, where individuals with a milder aphasia severity experienced greater gains in communication effectiveness. Additionally, Pease and Goodglass (1978) indicated that overall cueing (mix of semantic and phonological cues) was most effective for milder anomia severities.

2.6 Primary SFA and PCA Outcome Measures

SFA and PCA investigations primarily involve analyses of the degree of improvement on probes, namely item naming opportunities within a fixed amount of time that are separate from the intervention sessions themselves. Despite more recent examinations regarding the correlation between initial naming success during the SFA protocol and probe performance (Wambaugh, Mauszycki, & Wright, 2014), the degree of naming gains experienced when a reduced number of SFA features were used (Hashimoto, 2012) and adaptations to the feature generation process (Gravier et al., 2018), within-intervention gains are rarely an avenue of inquiry for SFA and PCA studies (Maddy, Capilouto, & McComas, 2014; Efstratiadou et al., 2018).

While improvements on probe naming are a valuable indicator of intervention success, not all participants experience gains under these conditions (van Hees et al., 2013; Neumann, 2018; Leonard et al., 2008). For these participants in particular, it may be beneficial to explore if any gains were experienced during the intervention sessions themselves, specifically if the feature generation protocols were of benefit or not. While probe word naming, in particular when provided with a short amount of time per word, investigates immediate word retrieval improvement, within-intervention naming improvements could investigate an individual's naming under more ideal circumstances (e.g., more time, provision of cues, etc.). Considering naming success under both circumstances is necessary for daily life (e.g., conversations require increased immediate access to words as compared to ordering food at a restaurant, where it is common to ask the waiter for more time to make a decision), the investigation of gains for both contexts is warranted.

2.7 SFA and PCA Generalization

Other areas of investigation and comparison between SFA and PCA outcomes involve their potential for generalization. Common generalization outcome measures involve gains from treated words to untreated words (response generalization), improvements on certain standardized assessment measures and improvements on discourse measures (van Hees et al., 2013; Coelho et al., 2000; Leonard et al., 2008).

With regards to response generalization, while the generalization of naming gains from treated to untreated items has been an area of focus in the majority of naming intervention studies, different degrees of success have been reported (van Hees et al., 2013; Coelho et al., 2000; Leonard et al., 2008; Neumann, 2018). Success has ranged from non-existent (van Hees et al., 2013) to moderate generalization (Neumann, 2018). A potential factor explaining the inconsistent presence of response generalization involves the degree of practice effects, namely effects obtained when an individual is repeatedly exposed to the untreated word list. DeLong et al. (2015) indicated that increased exposure to untreated items, for example during repeated probes, was related to the degree of generalization. However, even when the presence of practice effects is controlled in a more rigorous manner, inconsistent response generalization per individual and per intervention has been reported (van Hees et al., 2013; Neumann, 2018).

When investigating factors that could contribute to increased generalization from treated to untreated items, different suggestions have been proposed. In an SFA intervention study, a comparison between the use of a large number of treated stimuli ("many exemplars condition") versus the use of a smaller number of stimuli ("few exemplars condition") was completed within a single participant (Coelho et al., 2000). Coelho et al. (2000) indicated that only a small number of treated items ("few exemplars condition") were required to achieve generalization to untreated items and that including a greater number of treated stimuli did not promote further degrees of generalization. In a separate study conducted by Gravier et al. (2018), the relative contribution of different predictors of treatment response, namely dose (number of intervention sessions and the number of trials per session), dose form (the investigated task, specifically the number of participant-generated features during the SFA protocol) and dose frequency (number of trials in a certain amount of time) were investigated for SFA interventions. Dose form predicted success most effectively, where the more features generated on average per trial by a PWA, the greater their naming success for both treated and untreated items at the post-intervention phase and at the maintenance phase one month later. Another important factor involved in response generalization is active participation during feature generation. The PWA might generate their own features (as opposed to having them supplied by the S-LP). Alternatively, the PWA might choose the correct option between two dissimilar features provided by the clinician. These strategies to promote active participation have been reported to increase the degree of naming success for treated and untreated items (Gravier et al., 2018). Gravier et al. (2018) also proposed that this process leads to a deeper processing and strengthening of the semantic system as a whole, further promoting the possibility of generalization to untreated items (Gravier et al., 2018).

Variable results have been reported for improvements on standardized assessments including the Western Aphasia Battery – Revised (WAB-R; Kertesz, 2006) and the Boston

Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983) as well as for the transfer of naming gains to discourse measures, where only a subset of participants have experienced gains and for only a subset of measures (Neumann, 2018; Coelho et al., 2000; Peach & Reuter, 2010; Rider, Jarris Wright, Marshall, & Page, 2008; van Hees et al., 2013). Since intervention gains following SFA and PCA are generally item-specific (Neumann, 2018; van Hees et al., 2013), it would not be expected for SFA and/or PCA naming gains to transfer to these generalization conditions unless the standardized assessments of interest already included a targeted word or the discourse task encompassed a topic that would naturally provide the opportunity for a treated word to be used (Hashimoto & Frome, 2011). For example, the Cookie Theft task in the Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2000) would not be expected to improve unless items such as *cookie, mother,* and *stool* were specifically targeted in the investigation sessions.

While generally not included in past SFA or PCA studies, following the ROMA consensus statement that recommended core outcomes that should be completed in aphasia treatment studies (Wallace et al., 2018), the inclusion of measures of emotional wellbeing and measures of quality of life are now recommended. The inclusion of a certain standard set of assessments is thought to result in a more homogeneous measurement of outcomes and thus a greater comparative potential between aphasia studies. For the former, the General Health Questionnaire (GHQ; Goldberg & Williams, 1991) was specifically recommended, whereas the Stroke and Aphasia Quality of Life Scale (SAQOL-39; Hilari, Byng, Lamping, & Smith, 2003) was recommended to assess the latter.

2.8 Carry-Over to Different Picture Presentations

Despite SFA and PCA's inconsistent success with response generalization, but successful outcomes with stimulus generalization for alternative word-finding interventions (Howard, Patterson, Franklin, Orchard-Lisle & Morton, 1985), a very limited number of SFA or PCA studies have investigated the carry-over of naming improvements for targeted items to dissimilar instances of the trained items (e.g., a green apple versus a red apple) and/or to instances of the trained item in a natural context (e.g., an apple presented on a white background versus an apple in a grocery store). While not using the specific SFA or PCA protocols, Howard et al. (1985) completed semantic and phonological interventions with twelve individuals where one set of items was treated, but probes were presented for both the original item set as well as a for a set of different images of the treated items. Considering gains were present for both sets of items, the study determined that improvements were not picture specific and that carry-over to dissimilar instances of trained items was possible (Howard et al., 1985). To the best of our knowledge, only Hashimoto and Frome (2011) have investigated the generalization of naming gains for trained items to presentations of the items in a natural context. They used a modified version of the SFA protocol where three instead of six features were targeted with a single PWA. Instead of presenting the generalization images one-by-one, "items were presented all together in color photographs in a naturalistic context. [The participant] was then asked to name as many items in the photograph as she could" (Hashimoto & Frome, 2011, p. 464). Results indicated that naming accuracies during the generalization task (13-63%) were within the same range as naming accuracies during treatment (0-100%), thus supporting stimulus generalization to natural contexts in their participant.

Extending beyond response generalization measures to continue to determine the degree of carry-over present for SFA and PCA may further support the functional benefits and transferable improvements of these therapies to a PWA's life. Not only is it important to be able to name different words (i.e., response generalization), but it can be argued that the ability to consistently name the same word in many contexts (i.e., carry-over effects) is equally essential for a PWA's communication success (Webster, Whitworth, & Morris, 2015). While the quantity of words able to be named is valuable for communication success, another common goal is for an individual to be able to use these words in contexts other than in the clinician's office.

Although success has been experienced by PWA with regards to the carry-over of probe word naming gains from treated items to different presentations of the same items, training different representations of the same items may provide more harm than benefit. Hoffman, Clarke, Jones, & Noonan (2015) indicated that there can be a cost to specifically training multiple representations of items. In their study with three patients with semantic dementia, they reported that one participant over-generalized the use of the trained items by incorrectly using the trained word label while naming similar objects. Although individuals with post-stroke aphasia may utilize alternative training objectives (i.e., finding strategies to access vocabulary they already know; Efstradiadou et al., 2018) as compared to individuals with semantic dementia that need to relearn lost vocabulary (Hoffman et al., 2015), training PWA with semantic difficulties on multiple representations of items could result in similar detrimental effects.

While training different representations of the same items may not be warranted (Hoffman et al., 2015), probing a word's generalization to different representations of the same image (e.g., other depictions of the item on white backgrounds and depictions of the item in a natural setting) are warranted (Howard et al., 1985; Hashimoto & Frome, 2011). Especially considering the limited generalization of treated items to untreated items in SFA and PCA research, the investigation of the degree of carry-over effects of trained items to different contexts may further assist in achieving the goal of having anomia rehabilitation increase the quantity and quality of words accessible to a PWA during daily life (Hoffman et al., 2015). Since naming in daily life requires the recognition and access of items in different formats and since items are not represented in isolation in daily life, additional investigations of both forms of carry-over improvements (i.e., dissimilar instances of the trained items and instances of the trained item in a natural context) are necessary.

2.9 Item Presentation Format

Images are often the stimuli of choice to elicit naming in word-finding interventions. While black and white line drawings, in particular images from Snodgrass and Vanderwart (1980), used to be the standard method of image presentation for naming interventions with PWA (Boyle & Coelho, 1995; Wambaugh et al., 2001; Coelho et al., 2000), the addition of colour to these same items was later found to significantly increase naming accuracy and speed (Rossion & Pourtois, 2004). However, the addition of texture to line drawings could also be viewed as artificial and yield unwanted complexity (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010).

More recently, the use of realistic image depictions (namely photographs as opposed to line drawings) has been found to further improve the naming of manipulable objects (Salmon, Matheson, & McMullen, 2014). The presence of complexity in photo stimuli has also been suggested to be more natural since pictures more closely resemble objects in real life (Brodeur et al., 2010). Furthermore, lexical retrieval is enhanced when additional details are included, since details are thought to facilitate an individual's recognition of the picture they are asked to name (Cuetos, Aguado, Izura, & Ellis, 2002; Meteyard & Bose, 2018). Although for action word naming, Blankestijn-Wilmsen et al. (2017) also indicated that PWA had significantly stronger naming abilities for action words presented in dynamic (videos) rather than static (pictures) conditions.

2.10 General Considerations on Efficacy: Statistical versus Clinical Significance

A final concept integral to the discussion on intervention-based research is the difference between statistically significant results and clinically significant results. As alluded to in previous sections on generalization and carry-over, the ultimate goal of any word-finding intervention should be to enhance a PWA's communication success (Webster et al., 2015). Since statistical significance is dependent on sample size and focuses on hypothesis testing, its relation and applicability to meaningful, clinical results is not always direct. Disparities between clinical and statistical significance may be observed, where an intervention's clinical relevance or the feasibility of an intervention in clinical settings may be unrealistic or may be overlooked (e.g., the number of hours of intervention, the intensity of the intervention, etc.; Armijo-Olivo, 2018). Although the magnitude of effects can be calculated in the form of effect sizes (Beeson & Robey, 2006), it is recommended that research on intervention tools complete additional investigations on a person's meaningful and/or functional improvement following the intervention (Armijo-Olivo, 2018). As Armijo-Olivo (2018) puts it "A clinically relevant intervention is the one whose effects are large enough to make the associated costs, inconveniences, and harms worthwhile." (Armijo-Olivo, 2018, p. 176). While presented within the context of research on physical therapy, the same concept of clinical significance can be applied to communication-based interventions.

Within the context of communication, Breitenstein et al. (2017) completed intensive (at least 10 hours per week) speech and language interventions for at least three weeks with 158 individuals with chronic aphasia. The study used a randomized, open-label, blinded-endpoint, multicentre, stratified-by-centre, waiting-list-controlled and parallel-group design (Breitenstein et al., 2017, p. 1529). The primary outcome measure included was the Amsterdam-Nijmegen Everyday Language Test A-scale (ANELT), a tool that measures verbal communication effectiveness in ten scenarios experienced in daily life. When compared to a control group for which intervention was deferred for three weeks, a statistically significant mean increase of three points was obtained on the ANELT following interventions. Breitenstein et al. (2017) argued that even if three points appears minimal, even a one point change on the ANELT should be considered meaningful since it could essentially correspond to a change from no relevant information provided to at least the minimum requirements for the communication scenario. However, in a reply to the article, Sakamoto, Higuchi, Tsuda, Tanimoto, & Kami (2017) raised the concern that the minimum change on the ANELT required for a clinically significant change has not yet been determined. In addition, since the final ANELT score is calculated across all ten scenarios and ranges between 10 and 50 (Breitenstein et al., 2017), it is unclear whether any scenario(s) in particular would have actually improved. Although treatment was feasible (e.g., no withdrawals during the interventions) and statistically significant results with medium to large effect sizes were obtained (Breitenstein et al., 2017), the enthusiastic conclusions made by the authors may be tempered in the light of clinical significance, especially considering the costs associated with intensive interventions (i.e., some rehabilitation settings did not have the means to participate).

3. Present Study

Since prior SFA and PCA research has mainly tracked probe word naming success for treated words and the presence of response generalization to untreated words, the present study investigated the progression of SFA and PCA interventions within sessions (e.g., ability to name the item prior to feature generation, during the feature generation, etc.) as well as their degree of carry-over. Considering that patients with different impairment and residual strength profiles can experience gains with either or both SFA and PCA interventions and that studies comparing both interventions may be beneficial to clarify their mechanism of action, both interventions were administered in the present study. Measures of probe word naming within a short time frame of 3 seconds and measures of gains obtained within the intervention sessions themselves were collected. These situations represent naming gains with regards to the immediate retrieval of the word versus when participants were provided with time for, and access to compensatory strategies, respectively. Considering not all individuals in previous studies have experienced gains on probe word naming, this study investigated the trajectory of improvement within intervention sessions to assist in determining if other forms of gains can still be experienced (e.g., reduction in the number of words never named), the value of feature generation (i.e., does it appear to actually assist in naming) and if any features are particularly important for treatment success.

As the inconsistent generalization of gains from SFA and/or PCA treated items to untreated items has been reported, investigations regarding a person's ability to consistently name a word in many contexts is increasingly essential. While naming a large quantity of words is necessary for communication success, being able to consistently name words, especially highly important words (e.g., one's allergies), is at least equally valuable. For this reason, investigations into the degree of carry-over for two contexts, namely naming dissimilar instances of the items (e.g., a green apple versus a red apple) and naming instances of the item in a natural context (e.g., an apple presented on a white background versus an apple in a grocery store) were completed. The investigation of gains on certain standardized assessments was also completed, where the ROMA consensus statement (Wallace et al., 2018) was used to guide the selection of these additional outcome measures, including assessments such as the WAB-R.

Obtaining information related to different avenues of improvements and generalization other than probe word naming and response generalization can assist in furthering the understanding of the benefits of SFA and PCA. If additional gains supplementary to probe word naming can be experienced following SFA and/or PCA interventions and if these gains have the capacity to transfer to alternative item presentations, then SFA and/or PCA may be increasingly valuable clinical tools for rehabilitation purposes at the chronic stage of post-stroke aphasia.

4. Hypotheses

4.1 Standard Outcomes: Probe Word Naming

Consistent with previous studies, treated and untreated items were probed for naming success. Considering the traditional method of assigning an intervention based on underlying impairments has been reconsidered due to improvements also being experienced when interventions were assigned based on residual strengths (van Hees et al., 2013), both SFA and PCA were completed with all four participants. Lists of different words matched for initial naming success, length, frequency and semantic categories were used. Discussions regarding each interventions' success based on each participant's profile (aphasia severity, semantic processing and phonological processing) will be provided later. With the exception of van Hees, McMahon, Angwin, de Zubicaray, & Copland (2014), who utilized a probe naming time of 3 seconds, previous studies typically included probe naming times ranging from 5 seconds (Best, Herbert, Hickin, Osborne, & Howard, 2002) to 30 seconds (Neumann, 2018). While longer probe naming durations provide individuals with the opportunity to utilize compensatory strategies for naming, in an attempt to probe the degree of restoration to the lexical network as a result of the interventions, a shorter probe naming time of 3 seconds was chosen for the present study. The following hypotheses are included for probe word naming for SFA and PCA treated and untreated word lists:

4.1.1 Hypothesis 1: A participant's naming is expected to improve following the therapies, but gains may be different for SFA and PCA for a given participant, depending on factors such as an individual's baseline aphasia severity.

4.1.2 Hypothesis 2: Despite inconsistent results in the literature, limited generalization to the untreated list is expected. If generalization from the treated lists to the untreated list is observed, it is expected to be of a smaller magnitude (smaller effect size) than gains observed on treated lists.

4.2 Expanded Outcomes: Within-Intervention Naming and Feature Generation

Considering the potential for supplementary avenues of improvement as compared to probe word naming, progress during and across intervention sessions was examined. As the sets of SFA and PCA words were repeated each week, it was predicted that with practice, participants would gain familiarity with the targeted items and the intervention protocols. With increased familiarity, it was hypothesized that the protocols and the naming task would gradually become easier (e.g., increased autonomy in SFA and PCA feature generation, increased frequency of naming success, etc.). Specifically, the two following within-intervention hypotheses were included:

4.2.1 Hypothesis 3: The frequency of forced choice decisions required during the feature generation process is expected to decrease over the progression of intervention sessions, while a participant's autonomy for generating features will increase.

4.2.2 Hypothesis 4: The frequency of successfully naming a target item during SFA and PCA interventions will increase over the progression of therapy. Specific investigations regarding the frequency of initial naming (naming an item prior to feature generation) and the frequency of items never named within an intervention session were completed.

4.3 Generalization and Carry-Over

Considering inconsistent generalization to untreated items has been reported, alternative forms of generalization were investigated in the present study. Since naming words consistently across different contexts can be argued to be equally important to naming many different words, investigations regarding a participant's naming success on carry-over of items (i.e., items presented in dissimilar picture depictions and items presented with natural context backgrounds) and a participant's improvements on standardized assessments such as the WAB-R were completed. In addition to investigating other avenues of meaningful change following the interventions, changes on measures of quality of life and quality of communication were also investigated. Hypotheses related to generalization and carry-over are as follows:

4.3.1 Hypothesis 5: Successful naming of carry-over items (other exemplars of the targeted items and targeted items depicted in natural contexts) is expected. However, overall accuracy is expected to be lower than for original items, as shown by smaller effect sizes when comparing baseline naming of original items to post-intervention naming of carry-over items, than when comparing baseline naming of original item to post-intervention naming of original items.

4.3.2 Hypothesis 6: Considering the limited amount of change reported in the literature for measures of standardized assessments and connected speech, limited gains are expected on standardized assessments and on measures of quality of communication and quality of life.

5. <u>Methodology</u>

5.1 Study Design

Considering that the formation of authentic groups is challenging for such heterogeneous populations as individuals with post-stroke aphasia, a case series comprised of four participants was implemented in the present study.

With regards to the intervention sessions, a within-subject alternating counterbalanced treatment design was employed, where two of the four participants were randomly assigned to begin with SFA therapy and two were randomly assigned to begin with PCA therapy. Participants completed one SFA and one PCA session each week.

5.2 Procedure

The study was completed over 18 sessions during October-January of 2018-2019. Participants completed a total of six assessment sessions and twelve intervention sessions at McGill University according to the scheduled outlined in Table 1.

Table 1. Description of the study progression

Baseline Assessments	Intervention Sessions	Post-Intervention Assessment	Maintenance Assessment
Two 2hr assessment sessions completed within the same week	Twelve 1hr intervention sessions over 6 weeks: Six PCA and six SFA sessions (alternating, counterbalanced treatment design)	Two 2hr assessment sessions completed within the same week immediately following the intervention block	Two 2hr assessment sessions completed within the same week one month after the post-intervention assessment

5.3 Participants

5.3.1 Recruitment

The project underwent ethical review and was approved by McGill University's Institutional Review Board (study A06-B25-18A).

A combined total of four participants (2 male, 2 female) with post-stroke aphasia were recruited by Speech-Language Pathologists from McGill University's School of Communication Sciences and Disorders' teaching clinic and from the Association Québécoise des Personnes Aphasiques (AQPA). Inclusion criteria comprised: men and women between the ages of 25-70 presenting with chronic aphasia (at least 6 months post-stroke), the presence of word-finding difficulties and individuals who are native English speakers or had used English as a dominant language for at least 15 years preceding their stroke. Exclusion criteria included: the presence of other neurological conditions that may affect language abilities (e.g., dementia, TBI, etc.), uncorrected problems of vision and/or hearing, and severe motor speech difficulties or comprehension difficulties that would prevent from participating in interventions.

All participants provided written informed consent prior to beginning the study. All participants, with the exception if P01, provided written informed consent to be video recorded. All participants completed all assessments and treatments planned in the study.

5.3.2 Socio-demographic Characteristics

Participants ranged in age from 43-67 (mean 55.75; SD 11.18) and were between 18-153 months post-left cerebrovascular accident (CVA) (mean 81.5; SD 59.69) at the beginning of the study. Three participants experienced ischemic strokes and one participant experienced a hemorrhagic stroke. Information pertaining to the specific lesion site was not available. All participants were multilingual, with knowledge of English and French as well as at least one additional language. A description of demographic details per participant can be found in Table 2.

Table 2. Demographic and lesion information

Participant	P01	P02	P03	P04
Gender	M	M	F	F
Age (years)	63	67	50	43
Years of Education	17	17	13	18
Languages Spoken	Multilingual (3 languages)	Multilingual (5 languages)	Multilingual (3 languages)	Multilingual (3 languages)
Months Post-Stroke	153	105	50	18
Lesion Side and Type	Left Ischemic	Left Ischemic	Left Ischemic	Left Hemorrhagic

Although three participants attended conversation groups for aphasia throughout the progression of the study (P01, P02 and P03), none of the participants were receiving individual speech and language services at any time while participating in the study. Only participant P02 reported having infrequently used SFA in prior speech and language therapy sessions, where the SFA training was completed in a language other than English at the time.

5.4 Assessments

5.4.1 Standardised Assessments and Language Profiles

Following Wallace et al.'s (2018) review of aphasia outcome measures, five measures were recommended to be included in aphasia research: language, communication, patient-reported satisfaction with treatment and impact of treatment, emotional well-being as well as quality of life. Some specific assessment tools were recommended, including the Western Aphasia Battery Revised (WAB-R; Kertesz, 2006) for the language measure, the General Health Questionnaire (GHQ-12; Goldberg & Williams, 1991) for emotional-well-being and the Stroke and Aphasia Quality of Life Scale (SAQOL-39; Hilari et al., 2003) for quality of life. From the recommended measures, the WAB-R and the SAQOL-39 were included in the present study.

During the first two-hour baseline assessment session, participants completed a hearing screening to confirm that they presented with average hearing (ability to hear 500Hz, 1000Hz, 2000Hz and 4000Hz tones at 40dB or lower in both ears). Subsequently, participants completed Part 1 of the Western Aphasia Battery Revised (WAB-R; Kertesz, 2006) to classify their aphasia type and severity according to the aphasia quotient (AQ). The Pyramids and Palm Trees Test (PPTT; Howard & Patterson, 1992) and the word and non-word reading subtests from the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1996) were completed to evaluate each participant's non-verbal semantic and phonological reading abilities respectively. The Quality Communication Life Scale (QCL; Paul-Brown, Frattali, Holland, Thompson, & Caperton, 2003) assessed a participant's functional communication. Furthermore, information on their prior knowledge of the therapies was gathered. This information also assisted in appraising which techniques, if any, participants use when struggling to access a word prior to and following the intervention.

During the second two-hour baseline assessment session, participants completed certain sections of the Boston Diagnostic Aphasia Examination (BDAE; Goodglass, Kaplan, & Barresi, 2000), namely the word, nonsense word and sentence repetition tasks as a secondary measure of their phonological abilities and the Cookie Theft (oral and written) tasks as a measure of connected language production. The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983) was also completed as a measure of word retrieval ability. The Stroke and Aphasia Quality of Life Scale (SAQOL-39; Hilari et al., 2003) as well as measures of verbal fluency for fruits and vegetables (Acevedo et al., 2000), items found in a supermarket (Jurica, Leitten, & Mattis, 1988) and words starting with F, A, S (Tombaugh, Kozak, & Rees, 1999) were completed to assess an individual's quality of life post-stroke and lexical retrieval abilities respectively. While originally included in the protocol, the General Health Questionnaire (GHQ; Goldberg & Williams, 1991) was not completed due to a lack of time to administer the questionnaire with more severe participants. The apraxia supplemental section of the WAB-R (Kertesz, 2006) was completed with the two participants presenting with Broca's aphasia on part 1 of the WAB-R (P02 and P03).

A subset of these measures, namely the WAB-R, the BNT, the verbal fluency tasks, the SAQOL-39 and the QCL were reassessed at post-intervention and at maintenance to determine if the interventions had an effect on these measures.

The baseline assessments provided a general representation of each participant's profile, regarding each participant's difficulties within the scope of their aphasia type and severity, information related to semantic and/or phonological difficulties and the extent of word finding deficits. A description of each participant's profile can be found following the baseline assessment results provided in Table 3.

Table 3. Baseline assessment results

Baseline Assessments	P01	P02	P03	P04
Hearing Screening	Pass	Pass	Pass	Pass
WAB-R				
Part 1				
Spontaneous Speech (/20)	18	11	9	19
Auditory Verbal Comp. (/10)	9.3	8.1	7.4	10
Repetition (/10)	7.2	4.8	2.7	9
Naming & Word Finding (/10)	8.4	4.2	2.2	7.8
Aphasia Quotient (/100)	85.8	56.1	42.6	91.6
Aphasic? (<93.8)	yes	yes	yes	yes
Classification	Anomic	Broca's	Broca's	Anomic
Severity	Mild	Moderate	Severe	Mild
Part 2				
Apraxia (/10)	NA	9.5	8.8	NA
PPTT				
Score (/52)	49	43.5	49	47
Percentage Correct	94	83.7	94	90
Significant? (<90%)	No	Yes	No	No
BNT				
Scores				
Spontaneously Given Correct Resp.	43	10	2	24
Stimulus Cues Provided	4	17	1	0
Correct Resp. Following Stimulus Cue	1	0	0	0
Phonemic Cues Provided	16	15	58	36
Correct Resp. Following Phonemic cue	5	2	20	9
Total Number Correct	44	10	2	24
z-score	-2.02	-9.41	-13.3	-10.93
Paraphasias				
Phonological	21	7	12	4
Verbal	5	5	4	10
Neologistic	1	1	2	2
Multi-Word	0	0	0	0
Perceptual	0	0	0	0
Non-responses	7	24	30	10
BDAE				
Repetition				
Repetition: Single Words (/10)	3	6	5	9
Words Percentile	10th	20th	10th – 20th	70th
Repetition: Nonsense Words (/5)	0	2	0	5
Nonsense Percentile	0th	30th	10th	100th
Repetition : Sentences (/10)	1	1	0	7
Sentences Percentile	30th	30th	0th	60th
Cookie Theft				
Verbal Description[a]				
Number Content Units (c.u.)	17	6	2	17
z-score Content Units	0.18	-1.80	-3.40	0.18
Time to Complete Task (min)	2.16	1.85	1.22	1.51
Total Number of Syllables	70	14	22	156
Speaking Rate (syllables/min)	32.41	7.57	18.03	103.31
z-score Speaking Rate	-2.53	-2.83	-2.28	-0.50
Speaking Efficiency (c.u/min)	7.87	3.24	1.64	11.26
z-score Speaking Efficiency	-1.67	-1.20	-1.59	-1.14

Narrative Writing				
Narrative Writing Total (/11)	7	NA	3	9
Narrative Writing Percentile	70th	NA	10th – 20th	80th
PALPA				
Spell.-Sound Reg. Reading: Reg (/30)	13	18	1	29
z-score	-84.8	-59.8	-144.8	-4.8
Spell.-Sound Reg. Reading: Exception (/30)	15	10	1	18
z-score	-40.14	-53.64	-77.97	-32.03
Nonword Reading (/24)	3	3	NA	20
z-score 3 letters	-5.3	-5.3	NA	0.32
z-score 4 letters	-13.7	-13.7	NA	0.26
z-score 5 letters	-5.08	-5.08	NA	-1.74
z-score 6 letters	-6.65	-6.65	NA	-1.94
Fluency				
Supermarket				
Items Provided	16	2	2	19
Attained (≥20)	Unattained	Unattained	Unattained	Unattained
Letters FAS				
Letter Fluency FAS	15	3	4	22
Percentile FAS	<10th	<10th	<10th	<10th
z-score FAS	-2.23	-3.22	-3.63	-2.02
Fruits and Vegetables				
Semantic: Fruits	8	2	1	9
z-score Fruits	-1.30	-3.12	-3.78	-1.56
Semantic: Vegetables	8	2	1	11
z-score Vegetables	-1.33	-3.33	-4.03	-1.32
Animal Fluency[b]				
Animals Named	14	6	4	12
QCL				
Mean Score Overall (/5)	3.44	3.85	4.44	4.75
General Quality of Life Score (/5)	5	4	5	4
SAQOL-39				
Physical Mean (/5)	4.81	4.31	4.22	4.56
Communication Mean (/5)	3.14	R:3.57 F/E:1.64	2.29	4.43
Psychosocial Mean (/5)	4.13	3.03	3.38	4.09
Mean Score (/5)	4.23	R:3.65 F/E:3.31	3.53	4.35

[a]Analyzed according to Yorkston & Beukelman (1980)
[b]From the word fluency subsection of the WAB-R
NA = not assessed; R = Romanian; F/E = French/English

5.4.1a *Participant 1*

According to the WAB-R, P01's profile was consistent with a mild Anomic aphasia. He had appropriate non-verbal semantic abilities as indicated by the PPTT, but impaired phonological reading abilities for regular words and nonwords. For sentence repetition on the BDAE, P01 presented with impaired single word (real and nonword) repetition, but sentence repetition was within the norms. He had a low rate of speech, but he provided an adequate number of content units during the verbal Cookie Theft task. The written Cookie Theft task was also within the norms. P01's verbal fluency was impaired on all tasks. His word finding abilities were moderately affected for his age and he made frequent phonological errors on the BNT. P01 indicated that his current word-finding strategies involved using the word in another language, pointing and/or mimicking.

5.4.1b *Participant 2*

According to the WAB-R, P02's profile was consistent with a moderate Broca's aphasia, with a minor apraxic component (speech difficulties due to motor planning). According to the PPTT and the PALPA, P02 had difficulties with both non-verbal semantic tasks as well as regular word and nonword phonological reading tasks. During the BDAE's oral Cookie Theft task, P02 had a slow speaking rate and provided an inadequate number of content units. The written Cookie Theft task was not completed due to P02's significant difficulties with writing related to dominant hand hemiparesis. At baseline, P02 also presented with low average single word repetition and below average nonword and sentence repetition. He was not within the norms for any verbal fluency task. His word finding abilities were severely affected for his age and he made a mixture of phonological and semantic paraphasias on the BNT. Many non-responses were also present. When asked what strategies he uses when unable to find a word, he indicated that he would try another language, use facial expressions and give himself time. It should be noted that although he was proficient in English prior to his stroke, P02 would frequently attempt to use Romanian and/or French words instead of English words due to his comfort in these other languages.

5.4.1c *Participant 3*

According to the WAB-R, P03's profile was consistent with a severe Broca's aphasia, with a partial apraxic component. She had appropriate non-verbal semantic abilities, but impaired regular word and nonword phonological reading abilities as well as impaired repetition and verbal fluency abilities. The PALPA nonword reading section needed to be discontinued due to an inability to verbally produce the stimuli. P03's score was outside of the normal range for the oral and written Cookie Theft tasks on the BDAE. P03's word finding abilities were severely affected for her age on the BNT. On this task, she often did not provide a response. When she did respond verbally, she made a mixture of phonological (dominant) and semantic errors. P03's written word abilities were stronger than her spoken word abilities. When unable to say a word aloud, she was occasionally able to write it. When asked what strategies she uses when unable to find a word, she indicated that she would give herself time, gesture and try to write out the word.

5.4.1d *Participant 4*

According to the WAB-R, P04's profile was consistent with a mild Anomic aphasia. Although not below the cut-off score on the PPTT, P04's non-verbal semantic abilities were borderline. For phonological reading on the PALPA, P04 was impaired on regular word reading and on 5 and 6 letter nonword reading tasks. Her performance was within the normal range for nonword reading with 3 and 4 letter nonwords. Her repetition abilities were within the normal range, but her verbal fluency abilities were impaired. P04 was within the normal range for almost all components of the oral Cookie Theft task (with the exception of speaking efficiency) and was within the normal range for the written task. P04's word finding abilities were moderately affected for her age and she made more semantic errors than phonological errors on the BNT. P04 indicated that she would take her time and try to spell out difficult to find words.

5.4.2 Assessments Related to Interventions

At the beginning of both assessment sessions, participants were asked to name 300 colour images from the Bank of Standardized Stimuli (BOSS; Brodeur, Guérard, & Bouras, 2014), consisting of images from a variety of categories (animals, body parts, clothes, electronics, fruits, furniture, household items, instruments, other food, plants, tools and vegetables). For immediate scoring feasibility, the images were presented in a predetermined randomized order with a MATLAB_R2018b program, where the randomized order was different for each baseline session. During the task, participants had 3 seconds to name each image and a 2 second pause was provided between each item. Images were shown in the centre of a 21.5-inch iMac desktop computer.

When the information was made available in their methodologies, other studies most typically employed a 5 second (Best et al., 2002) to 30 second (Neumann, 2018) time window for probe word naming. The present study presented probes with a 3 second naming time frame so as to obtain a measure of an individual's immediate access to words, without providing time for the use of assistive strategies. Short probe word naming opportunities were included to target an individual's more immediate word retrieval improvement, whereas within-intervention naming improvements (where more time was provided) investigated an individual's naming under increasingly ideal circumstances. Considering contexts where naming is required, including during conversations (i.e., immediate retrieval) and retrieving words for grocery lists (i.e., where delayed retrieval is still appropriate), the investigation of gains for both contexts are warranted.

When scoring a participant's success on the BOSS naming task, sound distortions and pronunciation differences due to a participant's accent were scored as correct, whereas paraphasias and nonresponses were scored as incorrect. If a participant indicated that they were unable to name an item due to never having known its name, that item was eliminated from possible inclusion in treated and untreated word lists. Words for which common synonyms were often provided (e.g., "couch" for "sofa") were also excluded from being used in the word lists.

From a subset of sixty images that a participant failed to name on at least one of the two initial baseline naming attempts, twenty images were pseudo-randomly assigned to each of the three image groups (PCA targeted, SFA targeted and control lists), while controlling for category, word frequency (0 – 2.88), number of syllables (1–4) and naming performance during the baseline assessments (0/2–1/2). Measures of word frequency (log spoken frequency) were obtained from the CELEX database (Baayen, Piepenbrock, & Gullikers, 1995). Paired two-tailed t-tests were completed for pairwise comparisons of the three lists to establish that they were not statistically different in terms of word frequency and syllable length (all p values > 0.05).

Tables 4-7. Means and standard deviations per participant and per word list for the word frequency and syllable length psycholinguistic parameters

P01 – Psycholinguistic Parameters			
	SFA	PCA	Control
Word Frequency	0.32 (0.45)	0.33 (0.57)	0.36 (0.60)
Syllable Length	2.00 (0.92)	2.00 (0.86)	2.00 (0.92)

P02 – Psycholinguistic Parameters			
	SFA	PCA	Control
Word Frequency	0.85 (0.62)	0.83 (0.64)	0.84 (0.72)
Syllable Length	1.70 (0.86)	1.70 (0.73)	1.70 (0.92)

P03 – Psycholinguistic Parameters			
	SFA	PCA	Control
Word Frequency	0.92 (0.74)	0.94 (0.50)	0.88 (0.90)
Syllable Length	1.50 (0.76)	1.45 (0.60)	1.50 (0.69)

P04 – Psycholinguistic Parameters			
	SFA	PCA	Control
Word Frequency	0.47 (0.59)	0.48 (0.61)	0.49 (0.64)
Syllable Length	1.90 (0.64)	1.85 (0.75)	1.90 (0.79)

*Presented as: average(standard deviation)

After generating the three word lists, a third baseline was collected immediately preceding the first intervention session for this sixty-image subset. Naming success at all three baseline naming opportunities can be found in Table 8 below. For the first two baselines, while all 300 initial images were named, naming success is only reported for the sixty-image subset.

Table 8. Pre-intervention (baseline) experimental assessment, probe word naming results

	P01			P02			P03			P04		
	SFA	PCA	CTL	SFA	PCA	CTL	SFA	PCA	CTL	SFA	PCA	CTL
B 1	0	0	0	0	1	1	1	1	0	0	0	1
B 2	0	0	0	1	0	0	0	0	1	3	3	2
B 3	4	4	5	2	2	2	0	3	0	5	4	10

B1 = baseline 1; B2 = baseline 2; B3 = baseline 3; CLT = Control

While generating the lists, other factors were considered. For stronger participants, P01 and P04, an attempt was made to include items that they could not name on both baseline sessions. However, due to the need to balance the three lists, a limited number of items were available for P04. For this reason, P04's word lists each contained three words that were correctly named during one of the two baseline session. For more severe participants, P02 and P03, in order to maintain motivation, one word successfully named on one of the two baseline sessions was included per list.

5.5 SFA and PCA Interventions

5.5.1 Intervention Sessions

Colour images from the BOSS (Brodeur et al., 2014) were presented in the centre of a 21.5-inch iMac desktop computer within a MATLAB_R2018b program (Figure 2). Interventions were provided by the primary experimenter, who was trained on SFA and PCA protocols.

Following the baseline assessments, twelve intervention sessions lasting 60 minutes to 90 minutes each (depending on the speed that the participant completed the naming protocols for the entire word list) were completed individually with participants over the span of six weeks according to a within-subject alternating counterbalanced treatment design. Each week, participants attended one SFA session and one PCA session, where a different list of twenty images was targeted for SFA and PCA. Two participants were randomly assigned to begin with SFA therapy (P02 and P04) and two participants were assigned to begin with PCA therapy (P01 and P03). The more severe participants (P02 and P03) typically required 90 minutes, whereas the less severe participants (P01 and P04) typically required 60 minutes to complete the intervention sessions. As the interventions progressed, participants became slightly faster at the tasks.

Both SFA and PCA interventions were modelled off pre-established protocols (Boyle & Coelho, 1995; Leonard et al., 2008), where certain modifications to the original protocol were included based on their success in previous studies (Hashimoto, 2012; Gravier et al., 2018; Mehta & Isaki, 2016; DeLong et al., 2015). In both interventions, the image to be named was displayed in the middle of the screen and the associated features were presented around it (Figure 2). For SFA, six semantic features are typically included. However, many words are not suited for both the *use* and *action* features (e.g., there is no immediately identifiable *use* of a kangaroo), and many words do not provide distinct information for both the *use* and *action* features (e.g., the *use* and *action* of a blender are both blending/blends). Considering that previous studies have combined these features and observed similar gains to the standard protocol (Gravier et al., 2018; Mehta & Isaki, 2016), these two features were combined in the present study. Combining the *action* and *use* features also allowed for the same number of features to be present in both SFA and PCA (five each), resulting in a greater potential for comparison between the therapies. For PCA, the five phonological features typically included in previous studies were included in the present study.

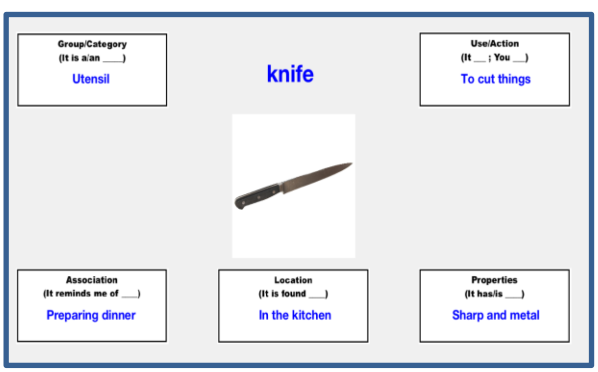
Since collecting information on the trajectory of within-intervention improvements (e.g., increased feature generation autonomy and increased naming success) to determine the presence of gains supplementary to probe word naming improvements was integral to the present study, both SFA and PCA protocols were presented through a MATLAB_R2018b program that retained the characteristics of the original hard-copy presentation method. During each session, either the twenty SFA or twenty PCA images were presented and treated in a random order. First, the image appeared alone in the center of the screen for 15 seconds. If the image was named correctly, the response was reinforced without re-naming the image (e.g., "yes, that's the word we were looking for"). Even if the participant correctly named the image, they were asked to provide answers for the features one at a time with guidance from the experimenter. Although inconsistently specified in previous studies, the order of SFA feature generation in this protocol was: *group*, *use/action*, *properties*, *location* and *association*. For PCA, features were generated in the following order: *first sound, first sound associate, last sound, number of syllables* and *rhyme*. It should be noted that the target image remained on the screen for the entire feature generation process, while the features were gradually displayed in their respective locations on the screen as they were generated.

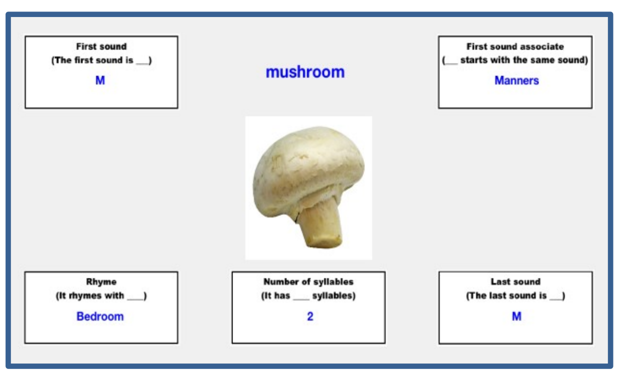
Following the initial attempt to name the item, participants were given approximately 10 seconds to self-generate an answer for each feature, where only one response was required per feature before proceeding to the subsequent feature in the chart. Participants were encouraged to generate different answers for each of the features (e.g., not using *construction* as an answer for both the *action* and *association* features for the word *hammer*). However, if a participant insisted that the repetition was their preferred response, it was accepted. If the participant did not provide an answer or provided an incorrect answer (e.g., [s] is the first sound in *bed*) to the feature in question, time was allocated for the experimenter to provide a forced choice option (verbally and in writing) between two words, a correct option and an option with dissimilar characteristics (e.g., forced choice options for the *first sound* feature for the word *bed* could be [b] versus [x]). The order of presentation for the correct and incorrect forced choice options was randomized.

A consensus document of forced choice options for SFA feature generation was pre-established among four individuals. The experimenter and three volunteers separately completed a table indicating their response for all five SFA features for the 300 baseline words. Since words often had multiple appropriate feature responses (e.g., *properties* of a crocodile could be *scales,* *sharp teeth,* etc.), all responses were compared to determine if one or many responses were provided with a higher frequency. If so, when needed, one of these responses (randomly chosen) would be the "correct" forced choice provided to participants. On rare occasions, when consistent responses were not obtained, one of the available options was chosen. For PCA, while the majority of the features only had one possible response (e.g., there are three syllables in *banana*), a document outlining 1-3 rhymes per word was completed from a rhyme generation engine. For words where only one common rhyme was available (e.g., *arctic* for *garlic*), that rhyme was the appropriate option provided to participants. When multiple common rhymes were available (e.g. *love, dove* or *shove* for *glove*), one of them was randomly chosen as the force choice option.

Once an appropriate answer was provided for all features in the chart, 20 seconds were allocated for the experimenter to review all of the features and for the participant to attempt to (re)name the image. If correctly named, their answer was reinforced with a repetition of their response (e.g., "yes, the answer is dog"). If the image was not named or was incorrectly named, the participant was asked to repeat the correct answer after the experimenter. The correct answer was then displayed with its image and feature responses on the screen (Figure 2) before the process began again with the next image. It was expected that as participants became increasingly familiar with the SFA and PCA protocols, experimenter prompting and support would gradually decrease.

Figure 2. Example completed SFA (left) and PCA (right) feature analysis charts

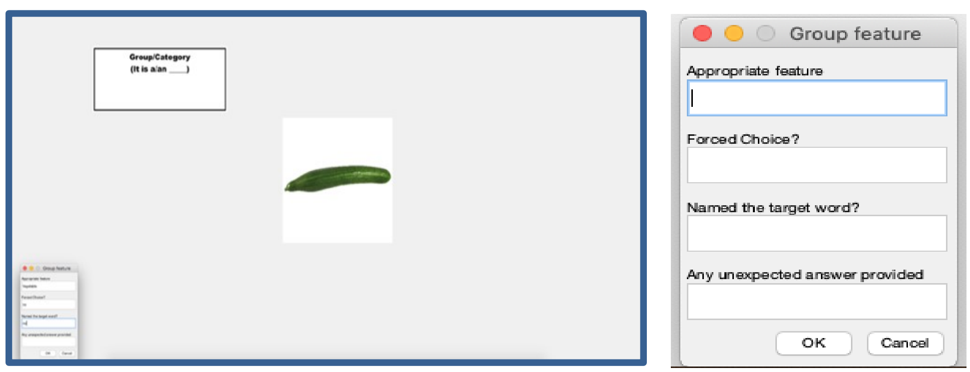




It should be noted that when a participant either named the image/feature or took the maximal amount of time to respond per step, a popup window (Figure 3) appeared asking the experimenter to indicate: the participant's answer, if this answer was correct (when naming the image at the start and at the end), if the answer was a forced choice (for the features), if the participant correctly named the image after identifying the feature and if the participant provided any other unexpected answers. While all of these responses were saved, only correct feature responses and the item's name following the final naming attempt were displayed on the screen.

Since the MATLAB program automatically provided a prompt to enter a response after the end of the initial 15 second period, it provided an estimate of the time taken to initially name each item (i.e., named the item prior to feature generation). Timing information was converted into a binary measure indicating if the participant successfully named the target word within the 15 seconds prior to the feature generation process ("yes" or "no").

Figure 3. Example display of the popup window generated during feature generation



5.5.2 Within-Intervention Probes

During the intervention phase of the study, the twenty targeted SFA and twenty targeted PCA words were probed every second week. The twenty-word control list was not probed during the intervention phase to reduce the occurrence of practice effects (DeLong et al., 2015). Participants were presented with SFA and PCA images in a randomized order according to a

MATLAB_R2018b program, where participants had 3 seconds to name each image and a 2 second pause between items. The order of presentation was different each time the images were probed.

5.5.3 Post-Intervention Assessments

During the week immediately following the final intervention session, two follow-up sessions of two hours each were completed to collect post-intervention assessment measures. All treated BOSS images (both SFA and PCA) as well as the untreated (control) list were probed. Investigations of carry-over gains present from the original images to presentations of dissimilar instances of the items (e.g., a green apple versus a red apple) and to instances of the item presented in a natural context (e.g., an apple alone versus an apple in a grocery store) were also conducted.

Volunteers searched for carry-over images on Google with "free to use or share" usage rights. Chosen images were in colour. Cartoons, line drawings and clip art were excluded. For the carry-over items depicting dissimilar instances of the items, volunteers selected images with white or grey backgrounds that presented the item in a different format than the original image (e.g., a metal cane instead of a wooden cane). For carry-over items depicting the item in a natural context, images were chosen when the target word was easily-recognizable (e.g., centered in the picture). A red arrow pointing to what should be named was additionally added to these images.

Specifically, the naming process at post-intervention was conducted in the same manner as for the intervention probes (randomized order with 3 seconds to name the image and a 2 second pause between items). Additionally, divided over the two sessions, all assessments completed during the baseline phase of the study, with the exception of the hearing screening, the PPTT, the PALPA and the BDAE were completed again with the participants to identify if changes occurred.

Although it could be argued that practice effects occurred on the probe word naming tasks at post-intervention and at maintenance since the same word lists (although presented as different pictures) were presented three times, once for the originally treated pictures and once for either set of carry-over items, each presentation was separated by multiple other tasks. The originally trained pictures were the only image set presented on the first post-intervention session. One carry-over image set was presented at the start of the second post-intervention session, whereas the other carry-over images were presented at the end of that second assessment session. This procedure was also applied during the maintenance assessments.

Specific terminology for the image sets will be used for the remainder of the thesis. Specifically, *Original Pictures* are the images that were used during baseline assessments, SFA and PCA interventions as well as probed at post-intervention and maintenance phases of the study. Two sets of carry-over images, *Different Pictures* and *Natural Context* conditions, were probed at post-intervention and maintenance phases of the study only. *Different Pictures* (DP) refer to image sets depicting different exemplars of the SFA, PCA and control images, whereas *Natural Context* (NC) refer to image sets depicting the SFA, PCA and control images in natural contexts (as opposed to having them depicted on a white background).

5.5.4 Maintenance Assessments

One-month following the post-intervention phase of the study, two maintenance sessions of two hours each were completed to identify if any changes experienced at post-intervention were maintained following a period of rest. The same measures were collected as for post-intervention. All standardized assessments completed at post-intervention were completed and all targeted (SFA and PCA), control and carry-over items (*Different Picture* and *Natural Context)* were probed.

5.5.5 Inter-Rater Reliability

Following the completion of data collection, a point-to-point inter-rater reliability task was completed by three volunteers for the post-intervention and maintenance phases of SFA, PCA and control item probe naming tasks. Due to unforeseen complications with the video recording device (i.e., it would turn off unexpectedly), as the only participant with video recordings at this phase of the experiment, inter-rater reliability judgements were also completed for P03's probe word naming during the third baseline. The volunteers were asked to indicate if the participant named the correct word ("1") or did not name the correct word ("0") within the 3 seconds allotted. These judgements were compared to those of the primary experimenter completing on-line judgements of success during the presentation of the probes. Inter-rater reliability judgements were completed for P02, P03 and P04. The task was not completed for P01 since video recordings of the probe sessions were not obtained (consent was not provided for video recording). A single percentage indicating inter-rater agreement is provided per participant, which combines all judgement comparisons for that participant's naming. Naming success agreement across the four raters was 85% for P02, 97% for P03 and 91% for P04. P02's inter-rater reliability percentage may be lower due to the low volume of his voice and due to difficulties in distinguishing his ambiguous pronunciations of certain cognate probe words (e.g., *tomato* versus *tomate*) on recordings.

6. Data Analysis

Visual representations of the data as well as statistical analyses, when possible, were completed to address each of the six hypotheses outlined. Visual representations were completed using Microsoft Excel, whereas statistical analyses were completed using SPSS (version 24). Nonparametric statistical analyses for two related samples (the McNemar Test) were completed. Depending on the hypothesis and the number of comparisons completed, the McNemar Test's threshold for significance was alpha = 0.05, alpha = 0.017 (0.05 divided by 3) or alpha = 0.025 (0.05 divided by 2). For the latter two cases, the alpha level was reduced when a Bonferroni correction for multiple comparisons was applied. Specific methods of data analyses (visual versus statistical), and when appropriate their respective alpha level, are outlined per hypothesis.

Effect sizes (ES) were calculated for the first three hypotheses according to the d statistic formula introduced by Beeson and Robey (2006). ES were used to outline the magnitude of change observed from baseline (baselines 1, 2 and 3) to post-intervention for the *Original Picture* sets (SFA treated, PCA treated and control) as well as for the carry-over images (*Different Picture* and *Natural Context*). ES were interpreted according to Beeson & Robey's (2006) study on lexical retrieval interventions. Specifically, the ranges used to interpret ES are: a small effect ranged from 4.0 to 6.9, a medium effect ranged from 7.0 to 10.0 and a large effect was 10.1 and above (Hashimoto, 2012; Beeson & Robey, 2006). ES less than 4.0 were considered to be negligible.

For the analyses of the WAB-R and BNT standardized assessments, the presence or absence of statistically significant changes from baseline were determined according to benchmarks provided by Gilmore, Dwyer, & Kiran (2018). To determine these benchmarks, Gilmore et al. (2018) completed a comprehensive literature search and subsequently extracted the data from relevant studies. Random-effects meta-analyses were completed to determine summary effect sizes following aphasia therapy (i.e., benchmarks for significant raw score point changes). Benchmarks were provided separately for within-group and between-group designs. In accordance with the current study's design, the benchmarks chosen for the WAB-R and BNT analyses corresponded with Gilmore et al.'s (2018) within-group design benchmarks. For the WAB-R aphasia quotient (AQ) score, the benchmark of a 5.03 point change was considered a significant change, whereas a 3.18 point change on the BNT was considered a significant change.

For analyses of the QCL and SAQOL-39 scales, statistical analyses were completed using a computer program provided by Crawford & Garthwaite (2006). The program uses sample test-retest data (means and standard deviations), the sample size and the correlation coefficient between the test-retest variable to generate a regression equation so as to compare an individual's degree of change on pre/post administrations of standardized assessments to predicted scores obtained from the sample data. Two-tailed significant t-tests were provided by the program and interpreted at alpha = 0.05 (Crawford & Garthwaite, 2006). Since, to the best of our knowledge, test-retest data with all required indices (e.g., standard deviation) were not available for the WAB-R or BNT, this method of determining statistical significance was not possible for these measures.

Data necessary to generate predicted scores for the QCL were obtained from Babbitt, Cherney, & Halper (2008). This source contained QCL-17 pre and post data for eight participants. For the SAQOL-39, data used to generate the predicted scores was obtained from Hilari, Lamping, Smith, Northcott, Lamb, & Marshall (2009). While the seventy-one participants were in the acute stage post-stroke, the test-retest data corresponded to three- and six-months post-stroke time points.

7. <u>Results</u>

7.1 Standard Outcomes: Probe Word Naming

The first two hypotheses refer to naming improvements from baseline during 3 second probe word naming for SFA and PCA treated word lists and the untreated (control) word lists. Tables outlining the number of words successfully named at post-intervention (Table 9) and maintenance (Table 10) phases of the study for the three naming conditions are provided. Graphical representations (Figure 4) illustrate the progression of naming success scores for the three baseline naming attempts, the three within-intervention probes as well as naming scores for the *Original Pictures* condition and carry-over items (*Different Pictures* and *Natural Context* conditions) at post-intervention and maintenance phases of the study.

Although probe word naming for the untreated list is reported separately from SFA and PCA treated lists, since statistical analyses pertain to the same "set" of three comparisons, it is relevant to apply a Bonferroni correction for multiple comparisons (alpha = 0.017; 0.05 divided by 3).

Results obtained from statistical analyses and associated ESs are discussed per hypothesis.

Table 9. Post-intervention phase experimental assessment, probe word naming results

	P01			P02			P03			P04		
	SFA	**PCA**	**CTL**	**SFA**	**PCA**	**CTL**	**SFA**	**PCA**	**CTL**	**SFA**	**PCA**	**CTL**
OP	20	19	9	9	7	1	3	2	3	8	10	9
DP	19	19	5	7	6	2	4	3	4	10	9	11
NC	17	18	9	7	8	2	5	5	3	11	10	11

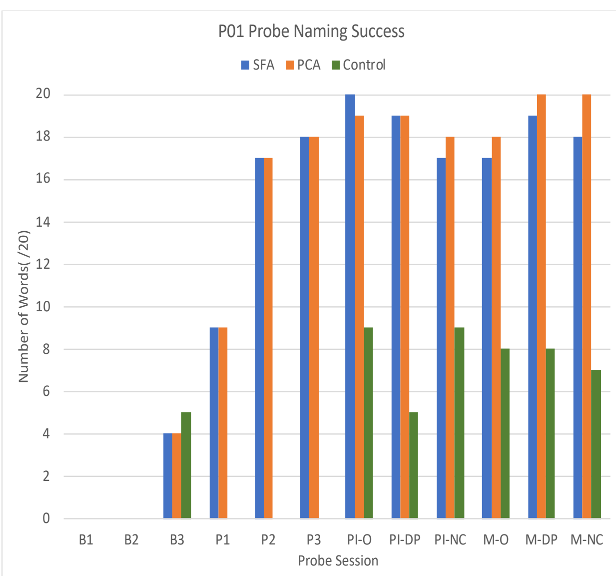
OP = *Original Pictures*; DP = *Different Pictures*; NC = *Natural Context*

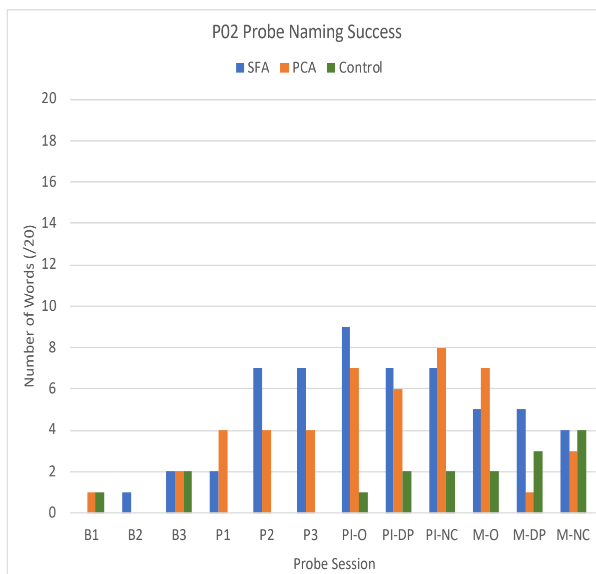
Table 10. Maintenance phase experimental assessment, probe word naming results

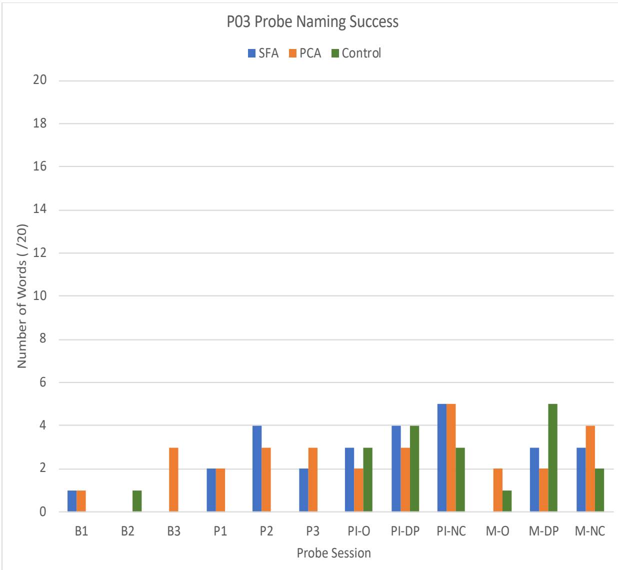
	P01			P02			P03			P04		
	SFA	**PCA**	**CTL**	**SFA**	**PCA**	**CTL**	**SFA**	**PCA**	**CTL**	**SFA**	**PCA**	**CTL**
OP	17	18	8	5	7	2	0	2	1	10	8	14
DP	19	20	8	5	1	3	3	2	5	8	6	10
NC	18	20	7	4	3	4	3	4	2	11	12	10

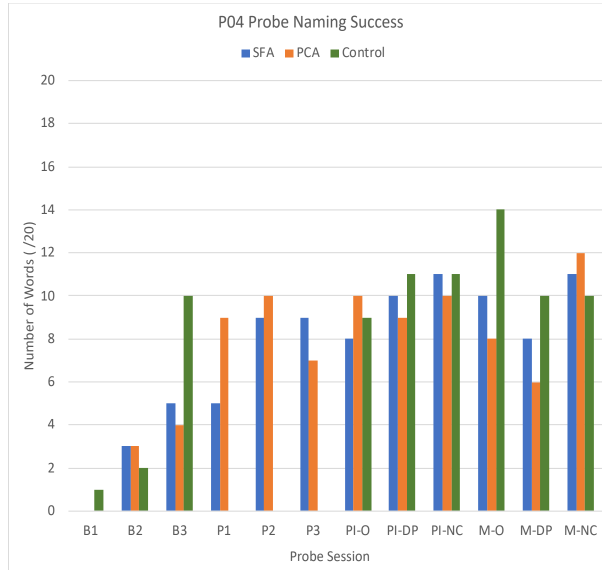
OP = *Original Pictures*; DP = *Different Pictures*; NC = *Natural Context*

Figure 4. Visual representation of each participant's naming success on SFA, PCA and control probe word naming across the entire study









Legend: B1 = Baseline 1; B2 = Baseline 2; B3 = Baseline 3; P1 = Probe 1 (within-intervention); P2 = Probe 2 (within-intervention); P3 = Probe 3 (within-intervention); PI = Post-Intervention; M = Maintenance; O = *Original Pictures*; DP = *Different Pictures*; NC = *Natural Context*.

7.1.1 Hypothesis 1: A participant's naming is expected to improve following the therapies, but gains may be different for SFA and PCA for a given participant, depending on factors such as an individual's baseline aphasia severity.

The McNemar Test (alpha = 0.017; trend: $p < 0.05$) was used to determine significant improvements in probe word naming for the *Original Pictures* condition, where naming results obtained at the third (last) baseline were separately compared for SFA and PCA to their corresponding naming success at post-intervention. At post-intervention, P01 significantly improved from baseline on both SFA ($p < 0.001$) and PCA ($p < 0.001$) probe word naming, where a medium ES was obtained for both SFA (ES = 8.08) and PCA (ES = 7.65) word lists. P02 experienced a trend of improvement for the SFA word list (p=0.039, medium ES = 8.00), but not for PCA (p=0.063). Alternatively, P04 experienced a trend of improvement for PCA (p=0.031, negligible ES = 3.69), but not for SFA (p=0.250). P03's performance did not approach significance for either SFA (p=0.250) or PCA (p=1.000) word lists.

Table 11. Effect sizes indicating the magnitude of change for the *Original Pictures* condition at post-intervention as compared to the baseline phase for SFA and PCA treated word lists

	P01	P02	P03	P04
SFA Word List	8.08*	8.00^{t}	4.60	2.12
PCA Word List	7.65*	6.00	0.44	3.69^{t}

*Significant improvement (p < 0.017); ^{t} Trend of improvement (p < 0.05)

Results regarding the maintenance of SFA and PCA naming gains for the *Original Pictures* condition were only completed for participants for whom a significant improvement or a trend of improvement was observed at post-intervention. The McNemar Test (alpha = 0.017) was used to determine which effects were maintained. Comparisons were made between baseline scores and scores at maintenance, where a significant result indicated the maintenance of effects and a nonsignificant result indicated that effects were not maintained. For P01, improvements on both SFA (p < 0.001, small ES = 6.78) and PCA (p < 0.001, medium ES = 7.22) were maintained. However, the trends of improvement obtained for P02 on the SFA list and for P04 on the PCA list were not maintained (baseline-maintenance comparison – P02: p=0.375; P04: p=0.219).

Table 12. Effect sizes indicating the magnitude of change for the *Original Pictures* condition at maintenance as compared to the baseline phase for SFA and PCA treated word lists

	P01	P02	P03	P04
SFA Word List	6.78*	4.00	-0.57	2.91
PCA Word List	7.22*	6.00	0.44	2.73

*Significant improvement (p < 0.017); ^{t} Trend of improvement (p < 0.05)

7.1.2 Hypothesis 2: Despite inconsistent results in the literature, limited generalization to the untreated list is expected. If generalization from the treated lists to the untreated list is observed, it is expected to be of a smaller magnitude (smaller effect size) than gains observed on treated lists.

The McNemar Test (alpha = 0.017; trend: p < 0.05) was used to compare the third (last) baseline results for the *Original Pictures* untreated (control) word list to corresponding results at post-intervention and maintenance. Significant naming improvements were not obtained (all p >

0.017) for any of the four participants for the untreated word list. No trends for statistical significance were observed (all $p > 0.05$).

Table 13. Effect sizes indicating the magnitude of change for the *Original Pictures* condition at post-intervention as compared to the baseline phase for the untreated word list

	P01	P02	P03	P04
Untreated Word List	2.54	0.00	4.60	0.95

*Significant improvement ($p < 0.017$); ^{t} Trend of improvement ($p < 0.05$)

It should be noted that P04 had an unstable baseline for the control word list. Naming success at the first and second baseline sessions was significantly different (alpha = 0.05) from naming at the third baseline (baseline 1 and 3: p=0.004; baseline 2 and 3: p=0.021). During the first two baseline sessions, P04 would attempt to spell difficult to access words and remain focused on an item, despite the presentation of new items. Instead of presenting all available items, the third baseline only probed words included in SFA treated, PCA treated and control word lists. It is possible that when the number of words decreased from 300 to 60 at the third baseline, the amount of cognitive resources required and the build-up of interference from previous items decreased, possibly allowing for a greater availability for image processing and successful naming.

The maintenance of effects from post-intervention were not investigated since the untreated word list did not show significant gains for any of the participants when naming success was compared from baseline to the post-intervention phase.

7.2 Expanded Outcomes: Within-Intervention Naming and Feature Generation

The following two hypotheses, hypotheses 3 and 4, investigated within-intervention gains across the progression of intervention from session 1 to session 6. While hypothesis 3 refers to a participant's increasing autonomy at self-generating SFA/PCA features, hypothesis 4 investigates the frequency of successfully naming a target item during the SFA and PCA interventions process.

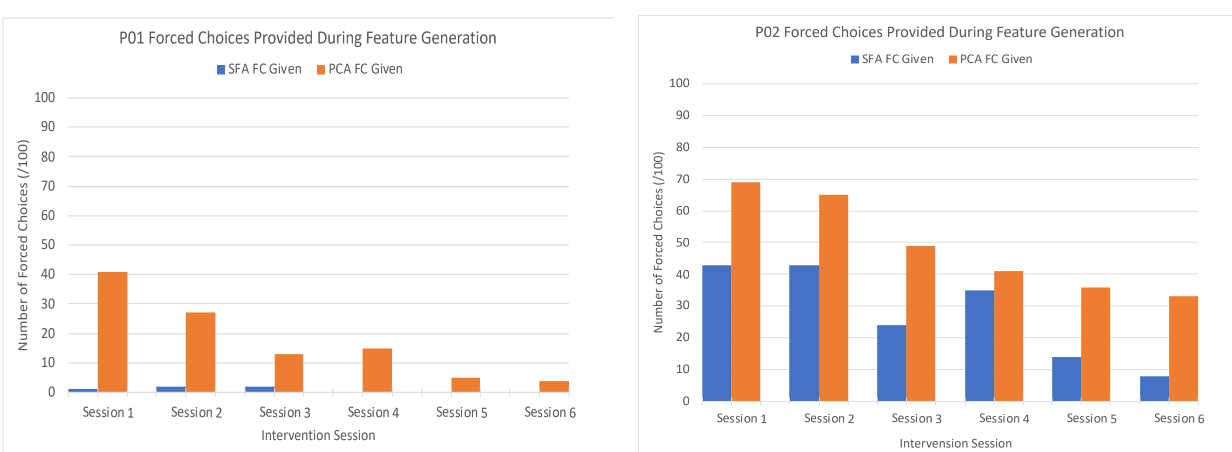
Results obtained from statistical analyses and associated ESs are discussed per hypothesis.

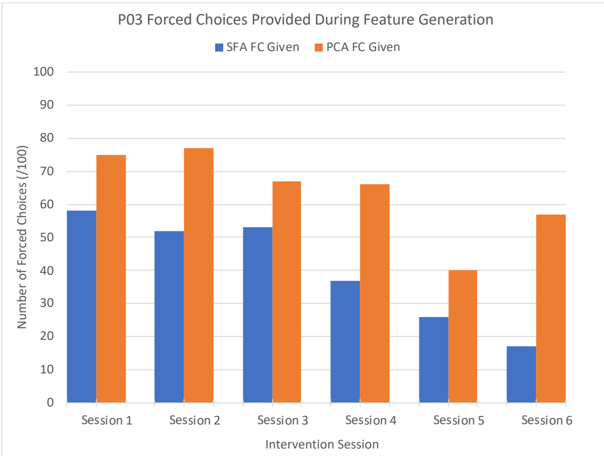
7.2.1 Hypothesis 3: The frequency of forced choice decisions required during the feature generation process is expected to decrease over the progression of intervention sessions, while a participant's autonomy for generating features will increase.

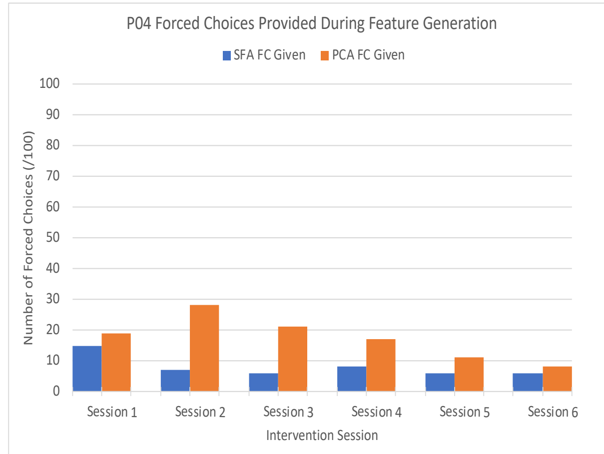
Visual representations indicating the progression of the number of forced choices required during SFA and PCA feature generation from session 1 to session 6 are represented graphically per participant (Figure 5). The maximal number of forced choice options possible per SFA or PCA session was one hundred (five forced choice options for each of the twenty words for either intervention). A breakdown of the number of forced choices provided per feature, per participant and per intervention, can be found in the Appendix (Figure I).

The McNemar Test (alpha = 0.05) compared the frequency a participant required any of the five forced choice options ("yes" versus "no") during feature generation at the first (session 1) and last intervention session (session 6) for SFA and PCA. A significant reduction in the number of forced choices required for SFA feature generation was found for P02 ($p=0.003$) and P03 ($p=0.016$), whereas a significant reduction in the number of forced choices required for PCA feature generation was found for P01 ($p<0.001$) and P04 ($p=0.021$). A small number of forced choices were consistently required for P01 and P04, even in the first SFA intervention session.

Figure 5. Forced choices required for SFA and PCA feature generation per session







Regarding the delivery of the forced choice options during intervention sessions, when an SFA or PCA feature could not be generated by the participant, a forced choice option between the correct answer and a dissimilar option was provided by the experimenter. However, multiple forced choice options were provided on the same paper, occasionally allowing P03 to search for an appropriate *category/group* feature on the page instead of generating one on her own. For this reason, P03's SFA forced choice measure should be interpreted with caution.

7.2.2 Hypothesis 4: The frequency of successfully naming a target item during SFA and PCA interventions will increase over the progression of therapy. Specific investigations regarding the frequency of initial naming (naming an item prior to feature generation) and the frequency of items never named within an intervention session were completed.

Separate visual representations indicating the progression of naming success during the SFA (Figure 6) and PCA (Figure 7) intervention sessions outline the proportion of items named successfully within the first 15 seconds of the SFA and PCA protocols (prior to feature generation), the proportion of words successfully named during the feature generation process (i.e., named successfully for the first time during the rest of the SFA or PCA protocol) and the proportion of words never named during that intervention session. These figures illustrate the distribution of these three proportions across all six intervention sessions for each of the four participants. The (numerical) proportion of items distributed across these three categories at session 1 and at session 6 are indicated in Tables 14 and 15 and directly map onto the data presented in Figures 6 and 7 respectively. For each session and each intervention, the maximal number of items that could be successfully named at any time was twenty words.

Figure 6. Position of naming success or lack thereof across SFA sessions

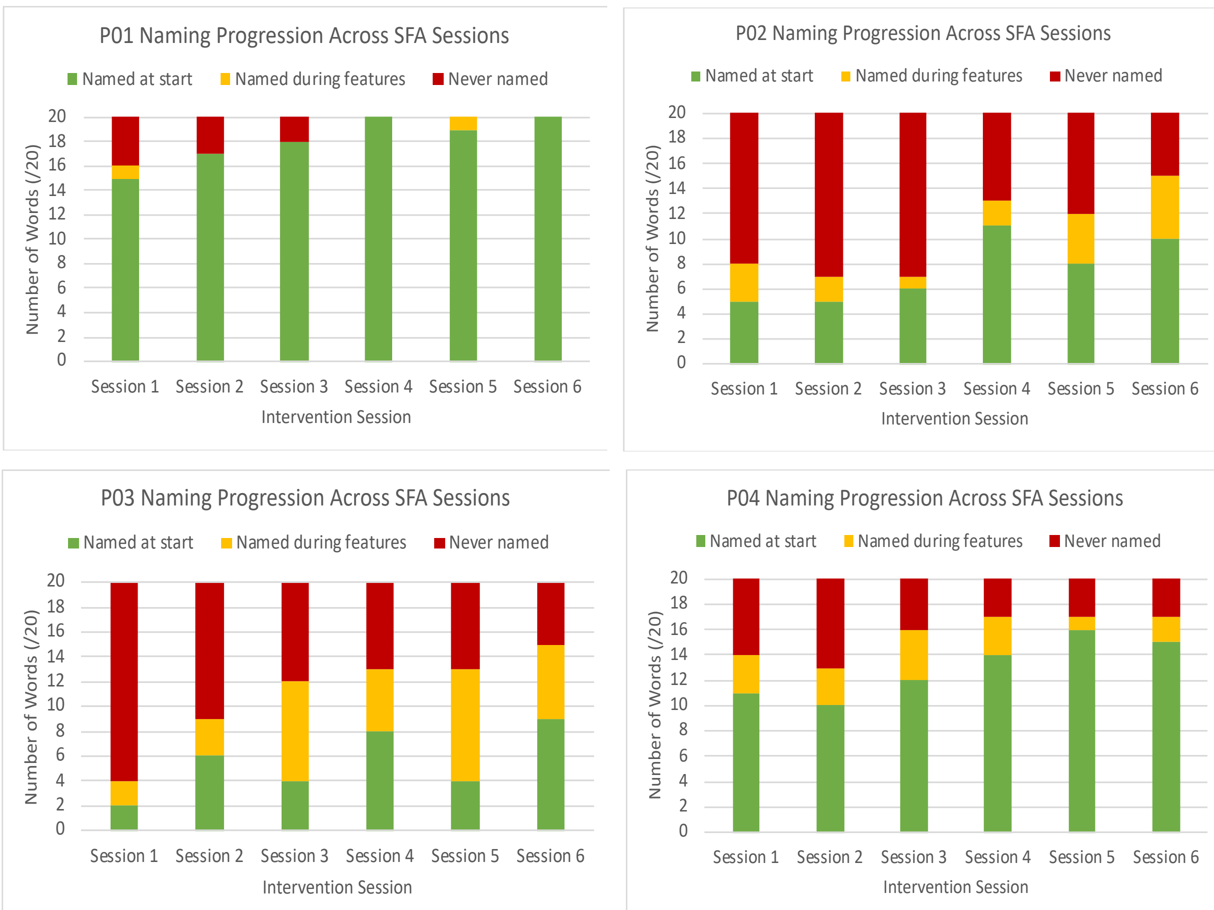


Table 14. Proportion of naming (within the first 15 seconds prior to feature generation process, during feature generation process and those never named) during SFA

	P01			P02			P03			P04		
Position	**S**	**FG**	**N**	**S**	**FG**	**N**	**S**	**FG**	**N**	**S**	**FG**	**N**
Session 1 (%)	75	5	20	25	15	60	10	10	80	55	15	30
Session 6 (%)	100	0	0	50	25	25	45	30	25	75	10	15

S = named within the first 15s at start; FG = named during feature generation; N = never named

Figure 7. Position of naming success or lack thereof across PCA sessions

Table 15. Proportion of naming (within the first 15 seconds prior to feature generation process, during feature generation process and those never named) during PCA

	P01			P02			P03			P04		
Position	**S**	**FG**	**N**	**S**	**FG**	**N**	**S**	**FG**	**N**	**S**	**FG**	**N**
Session 1 (%)	45	20	35	15	50	35	5	25	70	60	35	5
Session 6 (%)	100	0	0	65	15	20	25	55	20	65	35	0

S = named within the first 15s at start; FG = named during feature generation; N = never named

Timing information was converted into a binary measure indicating if the participant successfully named the item prior to the 15 second cut-off (“yes”), that is before feature generation, versus if the participant did not name the item within the time allotted (“no”). Items successfully named after this cut-off were considered items named during the feature generation process. Remaining items were considered items never named during that SFA or PCA intervention session.

To address significant changes in the proportion of items named within the first 15 seconds prior to the feature generation process and for items named during the feature generation process from session 1 to session 6, two McNemar Tests (alpha = 0.025; trend: $p < 0.05$) were completed. Considering these statistical analyses pertained to the same "set" of two comparison, a Bonferroni correction for multiple comparisons was appropriate (alpha = 0.025; 0.05 divided by 2).

The first set of analyses compared the proportion of items successfully named within the first 15 seconds of the intervention (prior to feature generation) at session 1 and at session 6. Analyses were completed separately for SFA and PCA. P01 experienced a trend of improvement for the number of items successfully named within this period for SFA ($p=0.031$) and a significant improvement for items named within this period for PCA ($p=0.001$). A significant increase was found for P02 for the PCA word list ($p=0.006$) but not the SFA list, whereas P03's naming during this period experienced a trend of improvement for SFA treated items ($p=0.039$), but not PCA treated items. No significant levels or trends of improvement were reached for either list for P04.

The second set of analyses compared the proportion of items never named at session 1 and at session 6 for SFA and PCA. P01 experienced a significant reduction in the proportion of words never named for PCA ($p=0.016$) but not for SFA. Significant reductions in the proportion of words never named were also obtained for both SFA ($p=0.007$) and PCA ($p=0.003$) for P03. No significant changes or trends were observed for either word list for P02 and P04. Thus, while visual representations appear to indicate that the proportion of items never named during SFA and/or PCA decreased across the progression of sessions for most participants, only P01 (PCA) and P03 (SFA and PCA) reached a statistically significant reduction in the number of items never named.

When informally comparing the progression of naming in SFA and PCA, P01 reached 100% naming success by session 6 for both SFA and PCA interventions, indicating that the interventions may have been of comparable effectiveness. For P02, although more words were named at the start of the PCA protocol by session 6 (statistically confirmed above) and despite dissimilar progressions of treatment for each intervention, both interventions appeared to be comparable by session 6 with regards to the total proportion of words successfully named at any instance during the intervention protocol (75% for SFA and 80% for PCA). While P03 was able to name more words within the first 15 seconds of the SFA protocol by session 6 (statistical trend of improvement), she successfully named a comparable total proportion of words during both SFA (75%) and PCA (80%) interventions. The visual progression of interventions across all sessions appears similar for SFA and PCA. Although a greater degree of change seemed to have occurred for P04 during SFA interventions, she was progressively able to name a larger proportion of words during PCA intervention. By session 6, she was able to name all words during PCA, but not during SFA interventions. However, during PCA, the proportion of words that required feature generation to be successfully named remained almost identical at sessions 1 and 6.

While the number of words never named decreased for all participants (albeit to a significant degree for only a select number of participants), it should be noted that some words remained “resistant” to SFA and/or PCA. For example, even after generating all the semantic features for “plunger”, P04 remained unable to access the word. For P03, a word that was consistently unnamed despite generating all PCA features was “mushroom”. These resistant words were present despite the fact that all three word lists were created while controlling for semantic category, word frequency, number of syllables and naming performance during the first two baseline assessments. Additionally, although some words appeared “resistant” to interventions, items included in interventions were taken from a variety of semantic categories (e.g., food, body parts, objects, etc.) and “resistant” words were not consistently associated with a specific category.

Analyses investigating gains induced by each of the five features during the two interventions were planned. However, each specific feature was associated with a very limited number of gains, even in the first treatment sessions. Bar charts showing the point at which a word was successfully named during SFA (Figure II) and during PCA (Figure III) can be found in the Appendix. It was expected that if more words were successfully named for the first time at a given feature, then that feature could be assumed to add more value to the intervention protocol. Possibly due to the high level of success shown by some participants when given more time to name pictures during interventions, and to the carry-over of naming gains from one session to the next, no clear patterns emerged at this time regarding the added value of any one feature to the naming process.

7.3 Generalization and Carry-Over

Hypotheses 5 and 6 refer to a participant's naming success on carry-over items (i.e., dissimilar depictions of the same item and items presented with natural context backgrounds) and improvements experienced on standardized assessments. These investigations provide information on alternative forms of generalization apart from response generalization. For hypothesis 5, graphical representations of the progression of naming success are given in Figure 4 above.

7.3.1 Hypothesis 5: Successful naming of carry-over items (other exemplars of the targeted items and targeted items depicted in natural contexts) is expected. However, overall accuracy is expected to be lower than for original items, as shown by smaller effect sizes when comparing baseline naming of original items to post-intervention naming of carry-over items, than when comparing baseline naming of original item to post-intervention naming of original items.

The McNemar Test (alpha = 0.017; trend: $p < 0.05$) was used to compare the naming success for carry-over items (*Different Picture* (DP) and *Natural Context* (NC) conditions) from the third (last) baseline to post-intervention and maintenance. P01 experienced significant carry-over of treatment gains for SFA (DP: $p < 0.001$, medium ES = 7.65; NC: $p < 0.001$, small ES = 6.78) and for PCA (DP: $p < 0.001$, medium ES = 7.65; NC: $p < 0.001$, medium ES = 7.22) for both carry-over conditions. P04 experienced a trend of improvement for the *Natural Context* condition carry-over items for SFA (p=0.031, negligible ES = 3.31) and PCA interventions (p=0.031, negligible ES = 3.69). No significant changes or trends were observed (all p values > 0.05) for the *Different Picture* condition carry-over items. No significant changes or trends were observed for either intervention or either set of carry-over conditions for P02 or P03 (all p values > 0.05). Significant improvements or trends of improvement were also not observed for either carry-over condition for the untreated word list (all p values > 0.05) for any of the four participants.

Table 16. Effect sizes indicating the magnitude of change at post-intervention for the carry-over conditions (*Different Picture* and *Natural Context*) when compared to baseline scores

	P01		P02		P03		P04	
	DP	**NC**	**DP**	**NC**	**DP**	**NC**	**DP**	**NC**
SFA	7.65*	6.78*	6.00	6.00	6.33	8.05	2.91	3.3^{t}
PCA	7.65*	7.22*	5.00	7.00	1.09	2.40	3.21	3.69^{t}
Untreated	1.15	2.54	1.00	1.00	6.33	4.60	1.35	1.35

DP = *Different Picture*; NC = *Natural Context;*
*Significant improvement ($p < 0.017$); ^{t} Trend of improvement ($p < 0.05$)

Results regarding the maintenance of SFA and PCA naming for carry-over conditions are only detailed for participants for whom a significant improvement or a trend of improvement was observed from baseline to post-intervention. Comparisons were made between scores at the third (last) baseline and scores for the *Different Picture* and *Natural Context* conditions at maintenance. A significant result on the McNemar Test (alpha = 0.017) indicated the maintenance of effects, whereas nonsignificant results indicated that the effects were not maintained. For P01, SFA *Different Picture* (p<0.001, medium ES = 7.65) and *Natural Context* (p<0.001, medium ES = 7.22) carry-over effects as well as PCA *Different Picture* (p<0.001, medium ES = 8.08) and *Natural*

Context (p<0.001, medium ES = 8.08) carry-over effects were maintained. P04's trend of improvement for the SFA *Natural Context* condition was not maintained (p=0.031, negligible ES = 3.31), whereas the trend of improvement for the PCA *Natural Context* condition was maintained (p=0.008, small ES = 4.65) four weeks post-intervention. For P04, the change from a trend of improvement at post-intervention to a significant improvement at maintenance for the *Natural Context* condition is assumed to have arisen due to the fluctuations in her naming performance throughout the study (see Figure 4). Additionally, while the statistical significance changed, the change is driven by a limited number of words (2 word increase from post-intervention to maintenance). Maintenance was not statistically assessed for the untreated word list for *Different Picture* and *Natural Context* conditions since statistically significant results or trends were not observed for the untreated word list for any of the four participants at post-intervention.

Table 17. Effect sizes indicating the magnitude of change at maintenance for the carry-over conditions (*Different Picture* and *Natural Context*) when compared to baseline scores

	P01		P02		P03		P04	
	DP	**NC**	**DP**	**NC**	**DP**	**NC**	**DP**	**NC**
SFA	7.65*	7.22*	4.00	3.00	4.60	4.60	2.12	3.31
PCA	8.08*	8.08*	0.00	2.00	0.44	1.75	1.76	4.65*
Untreated	2.19	1.84	2.00	3.00	8.05	2.88	1.15	1.15

DP = *Different Picture*; NC = *Natural Context*

*Significant improvement (p < 0.017); ^{t} Trend of improvement (p < 0.05)

7.3.2 Hypothesis 6: Considering the limited amount of change reported in the literature on measures of standardized assessments and connected speech, limited gains are expected on standardized assessments and on measures of quality of communication and quality of life.

As indicated in the methodology section, some of the standardized assessments were completed at all three phases of the study, namely baseline, post-intervention and maintenance. The following section outlines changes that arose for any of the four participants on assessments that were completed again at post-intervention (Table 18) and at maintenance (Table 19).

Table 18. Post-intervention assessment results

Post-Intervention Assessments	P01	P02	P03	P04
WAB-R				
Part 1				
Spontaneous Speech (/20)	18	12	12	18
Auditory Verbal Comp. (/10)	9.55	8.6	7.25	10
Repetition (/10)	7.4	6	2.8	9.4
Naming & Word Finding (/10)	8.8	4.1	4.3	7.9
Aphasia Quotient (/100)	87.5	61.4	52.7	90.6
Aphasic? (<93.8)	yes	yes	yes	yes
Classification	Anomic	Broca's	Broca's	Anomic
Severity	Mild	Moderate	Moderate	Mild
BNT				
Scores				
Spontaneously Given Correct Resp.	49	13	4	27
Stimulus Cues Provided	10	44	42	17
Correct Resp. Following Stimulus Cue	0	2	0	0
Phonemic Cues Provided	11	45	56	33
Correct Resp. Following Phonemic cue	2	7	18	11
Total Number Correct	49	15	4	27
z-score	-0.93	-8.33	-12.8	-9.93
Paraphasias				
Phonological	21	6	12	1
Verbal	8	14	7	9
Neologistic	1	1	0	5
Multi-Word	0	0	0	0
Perceptual	0	0	0	0
Non-responses	0	23	27	5
Fluency				
Supermarket				
Items Provided	17	4	1	21
Attained (≥20)	Unattained	Unattained	Unattained	Attained
Letters FAS				
Letter Fluency FAS	16	2	4	17
Percentile FAS	<10th	<10th	<10th	<10th
z-score FAS	-2.15	-3.31	-3.63	-2.47
Fruits and Vegetables				
Semantic: Fruits	13	2	2	12
z-score Fruits	0.21	-3.12	-3.50	-0.72
Semantic: Vegetables	6	2	2	8
z-score Vegetables	-2.00	-3.33	-3.76	-2.14
Animal Fluency[a]				
Animals Named	16	2	4	10
QCL				
Mean Score Overall (/5)	3.82	4.35	4.75	4.70
General Quality of Life Score (/5)	4	4	5	5
SAQOL-39				
Physical Mean (/5)	5.00	4.69	4.47	4.88
Communication Mean (/5)	2.86	R:4.57 F/E:3.29	3.07	5.00
Psychosocial Mean (/5)	4.50	3.84	3.44	4.43
Mean Score (/5)	4.41	R:4.32 F/E:4.09	3.79	4.72

R = Romanian; F/E = French/English

[a]From the word fluency subsection of the WAB-R

Table 19. Maintenance assessment results

Maintenance Assessments	P01	P02	P03	P04
WAB-R				
Part 1				
Spontaneous Speech (/20)	18	12	12	19
Auditory Verbal Comp. (/10)	9.87	8.45	6.9	10
Repetition (/10)	7.5	6.4	3.6	9.2
Naming & Word Finding (/10)	9.4	3.8	3.9	8.3
Aphasia Quotient (/100)	89.5	61.3	52.8	93
Aphasic? (<93.8)	yes	yes	yes	yes
Classification	Anomic	Broca's	Broca's	Anomic
Severity	Mild	Moderate	Moderate	Mild
BNT				
Scores				
Spontaneously Given Correct Resp.	49	11	4	24
Stimulus Cues Provided	9	32	19	11
Correct Resp. Following Stimulus Cue	0	0	0	0
Phonemic Cues Provided	11	49	56	35
Correct Resp. Following Phonemic cue	2	12	20	14
Total Number Correct	49	11	4	24
z-score	-0.93	-9.20	-12.8	-10.93
Paraphasias				
Phonological	19	13	8	12
Verbal	3	7	3	7
Neologistic	0	0	1	0
Multi-Word	0	0	0	0
Perceptual	0	0	0	0
Non-responses	2	21	27	3
Fluency				
Supermarket				
Items Provided	17	2	4	19
Attained (≥20)	Unattained	Unattained	Unattained	Unattained
Letters FAS				
Letter Fluency FAS	20	1	2	20
Percentile FAS	<10th	<10th	<10th	<10th
z-score FAS	-1.82	-3.39	-3.81	-2.21
Fruits and Vegetables				
Semantic: Fruits	9	1	2	9
z-score Fruits	-1.00	-3.42	-3.50	-1.56
Semantic: Vegetables	5	0	1	10
z-score Vegetables	-2.33	-4.00	-4.03	-1.59
Animal Fluency[a]				
Animals Named	18	3	3	13
QCL				
Mean Score Overall (/5)	3.53	3.38	4.88	4.88
General Quality of Life Score (/5)	4	4	5	5
SAQOL-39				
Physical Mean (/5)	4.88	4.13	4.47	4.94
Communication Mean (/5)	2.71	R:3.64 F/E:2.21	3.79	4.71
Psychosocial Mean (/5)	4.81	3.84	3.44	4.56
Mean Score (/5)	4.46	R:3.92 F/E:3.67	3.92	4.74

R = Romanian; F/E = French/English

[a]From the word fluency subsection of the WAB-R

Although statistical analyses could not be completed for the degree of change on specific sections of the WAB-R, Gilmore et al. (2018) identified that for within-group designs, a 5.03 point change on the WAB-R aphasia quotient (AQ) could be considered a benchmark for a significant change on that measure. Using this benchmark, both P02 (5.3 point change) and P03 (10.1 point change) experienced a significant change on their AQ score from baseline to post-intervention, whereas the other two participants did not. Considering the AQ change from baseline to maintenance remained within the same benchmark window, the changes from baseline can be considered to be maintained. Additionally, P03's AQ increased sufficiently to change the severity classification of her aphasia. At baseline, she was classified with a severe Broca's aphasia, while at post-intervention and at maintenance, she was reclassified with a moderate Broca's aphasia.

For the BNT, the McNemar Test (alpha = 0.05) was first completed to compare the frequency of words named correctly and the frequency of words named incorrectly at baseline and at post-intervention. Significant naming improvements were not obtained for any of the four participants (all p values > 0.05). However, a trend of improvement was obtained for P01 (p=0.063). Since statistically significant results were not obtained from baseline to post-intervention, investigations of maintenance were not completed. As a complement to the analysis completed with the McNemar Test, results were interpreted in accordance with significant change benchmarks reported in the literature. Gilmore et al. (2018) identified that a change of 3.18 points on the BNT could be considered a significant change on this measure. Using this benchmark, P01 experienced a significant change (6 point change) from baseline to post-intervention. This change was maintained since he achieved the same score of 49/60 at post-intervention and at maintenance. Changes on the BNT from baseline to post-intervention did not reach the benchmark for any of the other participants. Measures of maintenance were not investigated for these other participants.

When examining changes from baseline on verbal fluency measures, the only participant to successfully attain the cut-off of twenty words on the supermarket task at post-intervention was P04. However, the cut-off was not re-attained (i.e., not maintained) four weeks later. With regards to the other verbal fluency measures for which a set cut-off was not included, considering the limited degree of change from baseline (i.e., a maximal change of five words for only a limited sample of participants and fluency tasks) as well as the fact that the literature has reported the existence of practice effects during test-retest performances for verbal fluency tests in healthy adults (Strauss, Sherman, & Spreen, 2006), statistical analyses were not completed. In particular, considering that "relatively large changes in performance are required to conclude that real decline or improvement has occurred as opposed to being due to the effects of practice and random measurement error" (Strauss et al., 2006, p.515), test-retest changes were not deemed appropriate for the FAS letter, fruit, vegetable or animal fluency tasks.

As an informal measure of generalization, all verbal fluency tasks, namely the supermarket fluency, FAS letter fluency, fruits, vegetables and the animal verbal fluency subtest in the WAB-R, were examined to determine if any of the items from the SFA treated, PCA treated and/or untreated word lists were exclusively mentioned at post-intervention and/or maintenance. Any words mentioned at baseline were removed from the respective totals mentioned at post-intervention and maintenance (Table 20). For P03, although only spoken words were considered in the fluency score totals for the assessments, both spoken and written mentions of treated and untreated items are included in the table. Although to very limited degrees, all participants experienced at least some amount of carry-over of treated and/or untreated items to the verbal fluency tasks at post-intervention and at maintenance. It should be noted that treated and untreated words were never mentioned during the letter (phonemic) fluency task for any participant.

Table 20. Total number of SFA, PCA and untreated (control) words exclusively mentioned during verbal fluency tasks at post-intervention and at maintenance

	P01		P02		P03		P04	
	P-I	M	P-I	M	P-I	M	P-I	M
SFA	1	1	0	0	1s, 3w	1s, 4w	2	1
PCA	3	0	1	1	1s, 1w	1s, 1w	0	0
Untreated	0	0	0	0	0	1s, 3w	1	1

P-I = post-intervention; M = maintenance; s = spoken; w = written

With regards to the measures of quality of communication (QCL) and quality of life (SAQOL-39), statistical analyses were completed using a computer program provided by Crawford & Garthwaite (2006).

For the QCL, significant changes were not obtained (all p values > 0.05) for any of the four participants when baseline scores were compared to those reassessed at post-intervention. For the SAQOL-39, baseline to post-intervention score comparisons were completed for the SAQOL-39's total score and for each of the three subcategories, namely the physical, psychosocial and communication categories. When significant changes were observed from this comparison, maintenance was investigated by comparing scores obtained at baseline to scores collected at the maintenance phase. Outputs were interpreted as maintained if significant effects were obtained from the baseline to maintenance score comparisons. P01 reported to be more affected by his communication at post-intervention (p=0.021) and even more so at maintenance (p=0.005) since his scores progressively worsened. Although suspected to result from alternative factors than the interventions due to score fluctuations, P02's physical score significantly increased at post-intervention (p=0.045), but then significantly decreased at maintenance (p<0.001). For P02, considering his differential degrees of comfort with Romanian as opposed to French and English, he reported these languages as two separate scores for the communication subcategory. Thus, analyses for the communication subcategory and the SAQOL-39's total score were completed separately for P02's Romanian versus English/French scores. Results indicated a significant increase on the communication subcategory for both Romanian ($p=0.020$) and French/English ($p=0.005$) as well as for the total SAQOL-39 score when Romanian ($p=0.007$) or French/English ($p=0.002$) scores were used. Changes were not maintained four weeks post-intervention ($p > 0.05$).

At post-intervention and maintenance phases of the study, participants were once again asked to indicate strategies they use to find their words. Although the effect of the treatment is difficult to ascertain, certain influences from the SFA (e.g., indicating what an item is used for) and PCA (e.g., counting syllables) protocols can be observed. As compared to his baseline responses, at post-intervention, P01 indicated that he often describes a word by drawing it and explaining it, pointing and gesturing as well as taking his time when trying to find a difficult word. At maintenance, he added that he will sometimes think of the number of syllables in the word. At both post-intervention and at maintenance, P02 mentioned that he uses translation dictionaries, takes his time and thinks of his memories of the word. At both phases, P03 indicated that she tries to write a difficult-to-name word, takes her time and thinks of the word's sounds. At both post-intervention and maintenance, P04 indicated that she describes the item (e.g., what it looks like or what it is used for), takes her time and tries to find the first letter.

8. Discussion

This study described the effect of two popular treatments that use the generation of word-related features to improve naming, namely Semantic Feature Analysis (SFA) and Phonological Components Analysis (PCA). Six weeks of concurrent SFA and PCA word-finding interventions were completed with four individuals with chronic post-stroke aphasia. Improvements were investigated by comparing baseline naming accuracy to probe naming accuracy at post-intervention as well as by tracking the progression of within-intervention gains. Progress accomplished during feature generation and a PWAs' reliance on the therapist for the generation of features during SFA and PCA interventions were measured. The generalization of treated items to untreated items, the generalization to carry-over measures (items depicted in a dissimilar format and items depicted in a natural context) and the generalization to standardized assessments (e.g., the WAB-R) were also investigated. Overall, participants experienced differential patterns and degrees of improvement. One participant experienced significant gains on all probe word naming tasks for treated items and generalization to carry-over measures as well as improvements on a subset of within-intervention objectives (e.g., reduction in forced choices required for PCA) and standardized measures (e.g., the BNT). Other participants experienced more limited effects, showing improvements on only within intervention objectives and/or on certain standardized assessments.

The discussion will largely follow the order of the hypotheses presented above. Specifically, improvements on probe word naming will first be discussed for SFA and PCA treated word lists, followed by a discussion on the improvements for the untreated word list (i.e. naming generalization from the treated lists to the untreated list). Within-intervention gains as well as comparisons between these gains and probe word naming will then be provided. Subsequent discussions will refer to carry-over effects (items depicted in a dissimilar format and items depicted in a natural context), changes on standardized assessments as well as a discussions on different participant profiles across all six hypotheses. Lastly, PCA's mechanism of action and the disparity between statistical and clinical significance will be discussed.

8.1 Hypothesis 1: Improvements on SFA and PCA Treated Word Probes

As hypothesized, a certain number of participants experienced significant improvements or trends of improvement following SFA and/or PCA interventions. Although only one participant experienced statistically significant gains following the interventions in the present study, results are generally in line with previous research, where differential degrees of gains have been reported per study and per participant (Boyle, 2010). DeLong et al. (2015) reported improvements for 4/5 participants following SFA intervention, while Wambaugh et al. (2013) indicated that 8/9 participants improved on SFA treated lists containing items from animate and inanimate categories. For PCA, Leonard et al. (2008) reported significant improvements in 7/10 participants. As a specific example where both SFA and PCA were compared within the same individuals, in van Hees et al. (2013), 7/8 participants improved following PCA and 4/8 improved following SFA. Similarly to P01, some participants experienced improvement on both interventions. Some participants experienced significant results on only one intervention in a similar manner to the trends of improvement observed for P02 and P04. Similar to P03's profile, one participant in van Hees et al. (2013) did not experience significant improvements following either intervention.

Although interventions were not allocated on the basis of an individual's impairments or remaining strengths since support for both methods of treatment allocation can be found in the literature (van Hees et al., 2013; Neumann, 2018), certain results obtained in the current study suggest a relationship between a participant's baseline profile and treated probe word naming success. P01 presented with appropriate semantic abilities, but impaired phonological abilities at baseline. Since he improved on both SFA and PCA word lists, his profile is in accordance with both methods of treatment allocation. P02 presented with impaired semantic and phonological abilities at baseline. He experienced a trend of improvement following SFA intervention but no improvement following PCA. Thus, his profile does not favour one model over the other. Since P03 did not experience any significant improvement or trends of improvement for either intervention, despite adequate semantic, but impaired phonological abilities at baseline, her profile also does not favour one model over the other. At baseline, P04 presented with borderline semantic abilities and slightly impaired phonological abilities (i.e., impaired regular word reading and 5 and 6 letter nonword reading, but adequate 3 and 4 letter nonword reading and repetition abilities). Considering that she experienced a trend of improvement on the PCA word list, her profile is most in accordance with the impairment-based method of allocation. While support can be provided for either treatment allocation method, results are most in accordance with predicting an intervention's success based on other factors such as aphasia severity, since allocation on the basis of impairments or remaining strengths was not definitively predictive of naming success.

Many factors, such as an individual's aphasia severity (Efstratiadou et al., 2018), have been reported to predict the degree of improvements observed in intervention studies. This was a factor that appeared to predict intervention success in the present study as P01, a participant presenting with mild Anomic aphasia, experienced the most success on probe word naming, whereas P03, a participant presenting with severe Broca's aphasia, did not experience any significant gains on probe word naming. Thus, globally and in coherence with the literature on treatment efficacy (Efstratiadou et al., 2018; Breitenstein et al., 2017), gains for treated lists tended to vary with aphasia severity, where the milder the aphasia severity, the greater the gains experienced by the participant. Another factor that may inadvertently influence the degree of success following SFA and/or PCA interventions is the total number of sessions completed. Gravier et al. (2018) indicated that the number of intervention sessions tended towards predicting naming improvements for treated items but was not a statistically significant predictor of improvement. However, these specific aspects of the procedure are inconsistently reported across studies, making it difficult to conclusively appraise their influence (Neumann, 2018; Efstratiadou et al., 2018).

Intensity and dosage also count among potential factors in treatment success (Breitenstein et al., 2017; Brady et al., 2016). Gains have been observed following high-intensity speech and language services as compared to deferred interventions (Breitenstein et al., 2017). Although the definition of "high-intensity" treatment is unclear (Hinckley & Carr, 2005), the current study was completed at a lower intensity (60 minute sessions twice per week) than many other SFA and PCA interventions. Both Boyle & Coelho (1995) and Leonard et al. (2008) completed 60 minute sessions three times per week, Neumann (2018) completed two-hour sessions two or three times weekly and van Hees et al. (2013) completed 45 to 90 minute sessions three times per week. However, when compared to the intensity of S-LP services often provided in clinic (assumed at 60 minute sessions once per week), the current study was completed at a higher intensity. Thus, probe word naming gains may have been slightly reduced in the present study since it was completed at a lower intensity than other studies. Importantly, since interventions are often completed at a lower intensity in clinical settings than in research settings, expectations of success in the clinic should be adjusted in accordance with the therapy's intensity.

Another factor that may influence the degree of success could be the amount of time provided to name each word during probe word naming. The majority of SFA and PCA studies, with the exception of van Hees et al. (2014) that utilized a probe naming time of 3 seconds, provided participants with approximately 5 seconds to 15 seconds to name each probe word (Wambaugh et al., 2013; Leonard et al., 2008; Best et al., 2002). Shorter versus longer probe naming durations may coincide with the restorative versus compensatory approaches, where naming gains on shorter probes may only be experienced if SFA and PCA do in fact restore elements of the impaired processing network. If longer probe naming durations were allowed, more time would be available for individuals to use compensatory approaches for naming. Thus, the extent of lexical access recovery may not be what is targeted during longer probe naming durations. As a study that included a shorter probe naming duration of 3 seconds, the lack of time to use compensatory strategies (i.e., naming success relied entirely on the extent of lexical retrieval recovery) may have contributed to the slightly lower proportion of participants who experienced significant gains (e.g., van Hees et al., 2014). It should be noted however, that for participants P02 and P03 that presented with mild apraxia of speech, the degree of lexical restoration may not have been completely isolated during probe word naming. Their difficulties with this task may have also been a result of motor planning efforts in generating target words (i.e., as opposed to only having difficulties with the immediate access of words). To account for both approaches and to obtain a measure of each participant's success when provided with more time, within session gains were also examined. As will be discussed in section 8.3, more participants reached significant levels of naming success during intervention sessions, as compared to probe word naming.

8.2 Hypothesis 2: Generalization of Treated Items to Untreated Items

The generalization of SFA and/or PCA treated items to the untreated item list was not obtained for any of the four participants in the present study.

Although a certain amount of generalization to untreated items was hypothesized, a pattern of inconsistent response generalization can be found in the literature (DeLong et al., 2015; Boyle & Coelho, 1995; Wambaugh et al., 2013; van Hees et al., 2013; Neumann, 2018). In Boyle's (2010) review on SFA, all three possibilities of generalization to untreated items were reported, namely that some studies observed improvements, some did not observe any effects and some observed variable effects in terms of the magnitude and/or the extent of generalization across participants. For example, van Hees et al. (2013) did not observe any generalization to untreated items for any of their eight participants, while Neumann (2018) indicated a weak to moderate generalization to untrained items for all four of their participants.

An explanation for the inconsistent response generalization results could be the influence of practice effects. DeLong et al. (2015) reported that generalization to untreated items could be explained, at least in part, by the amount of exposure (practice effects) an individual had to the items in the untreated word list. Although the repetition frequency of probe word naming for untreated word lists has become more restricted with the dissemination of knowledge regarding practice effects, studies such as Leonard et al. (2008) still probed untreated items every three sessions (i.e., every week according to their intervention intensity). While probed slightly less frequently, van Hees et al. (2013) still probed all three word lists (PCA treated, SFA treated and untreated) every fourth intervention session. van Hees et al. (2013) did not observe any response generalization for any of their participants, whereas Leonard et al. (2008) observed generalization to untreated items for 3/7 participants. Since Leonard et al. (2008) probed untreated items slightly more frequently, the frequency of probe repetition may in fact play a role in the presence and degree of generalization to untreated items. Thus, considering the untreated word list was not probed at all during the treatment phase in the present study, the absence of potential practice effects may have contributed to the lack of generalization to untreated items.

8.3 Hypotheses 3 & 4: Within-Intervention Improvements

A participant's need for the experimenter to provide them with forced choices during the feature generation process was investigated. Forced choices were provided during the present study since this factor of dose form (the task of the intervention session) was determined by Gravier et al. (2018) to be a contributing factor for the presence and degree of naming success for both treated and untreated items. Gravier et al. (2018) noted that the average number of features generated by the participant (i.e., those not directly provided by the clinician without the client's involvement), predicted the degree of naming success. It should be noted however, that this conclusion regarding a client's feature generation only applies to SFA, as PCA was not used in that study. In the present study, significant reductions in the need for forced choices during SFA were found for two participants (P02 and P03). Similarly, a significant reduction in the need for forced choices during PCA was found for two participants (P01 and P04). It should be noted that for P01 and P04, even at the first SFA session, only a small number of forced choices were ever required, lessening the potential of observing significant differences.

Considering Gravier et al.'s (2018) investigations were only completed for SFA, it could be argued that requiring a participant to self-generate phonological features and/or to choose an appropriate phonological feature from a forced choice set may not be as appropriate as the completion of the same task for semantic features. Since phonology is largely arbitrary and since PWA have been reported to have difficulty self-generating some phonological information such as a word's first sound (Lambon Ralph et al., 2000), the generation of phonological features without first being provided with the word may occasionally act as a "guessing game". This will be further discussed when addressing PCA's mechanism of action in section 8.7. Despite these concerns, statistically significant reductions in the need for phonological forced choices were experienced in the present study. Thus, although possibly difficult at the start, a PWA's ability to accurately self-generate phonological features by the final intervention session improved. Globally, as the protocols were practiced, a participant's facility with SFA and PCA increased, as did their autonomy for the self-generation of features. However, it is still uncertain if this autonomy, which could translate into a compensatory strategy, would generalize to naming novel words or if the gradual autonomy only occurred due to specifically practicing the task with those treated words.

Within-intervention analyses were completed in addition to probe word naming analyses to investigate if discrepancies existed between these two naming conditions. For example, P03 did not experience any significant or trends of improvements on probe word naming, but she experienced significant reductions in the words never named during both interventions.

Since only 3 seconds were provided for probe word naming, probe naming can be considered a measure of immediate word retrieval. This measure corresponds with the restorative framework, where SFA and PCA were thought to restore impaired networks following the stroke (Boyle & Coelho, 1995). Contrarily, with the addition of more time and feature generation, within-intervention gains correspond more closely with the compensatory framework (DeLong et al., 2015; Wambaugh et al., 2013). Although direct comparisons to probe word naming were not possible due to the differential conditions in which the data was collected, the within-intervention results indicated that even when improvements were not obtained during probe word naming (i.e., immediate recall), gains could still be obtained in more ideal circumstances. While different, both avenues of investigation and intervention are warranted since rapid word retrieval is necessary for communication effectiveness during some forms of interactions, whereas delayed word retrieval while using compensatory strategies examines the maximal degree of naming gains possible.

Furthermore, although some features were occasionally difficult to generate, the process of feature generation appears to be a valuable contribution to naming success. Specifically, some words that were not initially retrieved in the first 15 seconds of the naming protocols were successfully named (or added) during feature generation. However, considering Hashimoto (2012) only included three SFA (subset targeted: *group/category, properties* and *association*) and three PCA (subset targeted: *rhyme, first sound* and *number of syllables*) features and still observed significant naming improvements during probes (10 seconds provided for probe word naming), questions emerge regarding the scope of feature generation necessary during SFA and PCA protocols. Specifically, the inclusion of all five features may not be necessary to obtain naming improvements. Despite this supposition, no patterns clearly emerged in the current study to favour a single or a subset of the SFA or PCA features for naming success. In other words, no one feature suddenly increased the number of words successfully named during that feature's generation for either SFA or PCA. Thus, while feature generation appears to assist within-intervention naming even for individuals for which probe word naming did not significantly improve, SFA and PCA protocols could become more efficient through the reduction of the number of features (i.e., less time required to target each word can lead to targeting a greater number of words per session). However, while heeding the nature, order and salience of the features, further research would be necessary to determine which features to maintain and which to eliminate.

8.4 Hypothesis 5: Carry-Over to Other Pictures

For one participant, P01, gains significantly carried-over to both the *Different Picture* and *Natural Context* conditions for both SFA and PCA interventions, while another participant, P04, experienced a trend of improvement for the *Natural Context* condition for SFA and PCA items. All gains were maintained four weeks post-intervention for P01 and P04.

While many previous SFA and PCA studies have investigated generalization to untreated items and generalization to other measures such as standardized assessments and connected speech (Neumann, 2018; Coelho et al., 2000; Peach & Reuter, 2010; Rider et al., 2008; van Hees et al., 2013), generalization to carry-over items of this nature has only been completed to a limited extent. Using different semantic and phonological treatments, Howard et al. (1985) probed participants with two different sets of the same images, a treated set and a set to investigate naming generalization to different images of the same item (similar to the *Different Picture* set used in the present study). With a correlation of 0.77 between the two sets, that study determined that improvements were not picture specific. Howard et al.'s (1985) results are in line with the current study, where P01, who experienced significant naming improvements on SFA and PCA's *Original Picture* condition, experienced comparable gains when the items were presented as different exemplars (*Different Picture* condition). Extending beyond Howard et al. (1985), gains in the current study were not limited to isolated objects presented on a white background. Gains experienced for the *Original Pictures* condition presented in isolation were similar to those experienced for objects even when presented in context (*Natural Context* condition). These gains are coherent with Hashimoto and Frome's (2011) SFA study, where their participant named a comparable subset of items when they were presented all together in a naturalistic context (similar to the present study's *Natural Context* condition) and when items were presented during therapy.

Although the rationale is not explicitly stated in naming treatment studies in aphasia, the standard presentation format for images to be named during word-finding interventions are images presented in isolation (i.e., on white backgrounds). Since the investigation of and intervention for word-finding difficulties for PWA is often completed through the use of confrontational naming tasks designed to represent and elicit a single, specific word (Conroy, Sage, & Lambon Ralph, 2009), it may have been thought that images presented in natural contexts (i.e., not in isolation on a white background) would result in difficulties deciding what single word to elicit (e.g., a bowl of candies could elicit *candy* or *snack)*. Additionally, when naming has been examined in healthy individuals, faster reaction times have been observed when items were presented with low visual complexity (Meteyard & Bose, 2018). Furthermore, considering the tradition of presenting items to be named as black-and-white line drawings (Snodgrass & Vanderwart, 1980) so as to remove distractor and non-essential features to the recognition of a word's label (Brodeur et al., 2010), presenting items other than in isolation may have been viewed as unnecessary.

However, the addition of context such as colour to these line drawings has also been found to significantly increase naming accuracy and speed (Rossion & Pourtois, 2004). Furthermore, Cuetos et al. (2002) showed that the more details provided in a drawing, the more detailed its semantic representation, and the easier it is for patients to name the item. Thus, since the addition of at least a certain number of contextual cues assists in the naming process and may even lead to requiring fewer additional cues (e.g. phonological cues), it is possible that providing items to be named with contextual background information could also be increasingly assistive for naming tasks. This modification would need to be thoroughly investigated, however. For example, Hoffman et al.'s (2015) study on semantic dementia noted that there can be a cost to training multiple representations of items as these labels may become over-generalized. Considering PWA present with different training objectives and strategies as compared to individuals with semantic dementia, where individuals with semantic dementia have to relearn lost vocabulary (Hoffman et al., 2015), whereas individuals with aphasia know the vocabulary, but have difficulty accessing it (Efstradiadou et al., 2018), the costs to training multiple representations of items may not be as impactful for interventions with PWA. While this impact of overgeneralization may be less relevant for PWA, training individuals with aphasia with concomitant cognitive deficits on multiple representations of items could pose alternative difficulties. Prior research studies have indicated that cognitive impairments, such as executive function deficits, have been associated with reduced treatment success (Dignam et al., 2017). Reduced abilities in domains such as cognitive flexibility and inhibition may make tasks that "introdu[ce] greater variation into the learning experience" (Hoffman et al., 2015, p. 240) increasingly difficult. Therefore, while there may still be a risk to training too many representations of the same image (Hoffman et al., 2015), utilizing this approach to further the generalization of naming gains could be investigated and/or eventually recommended with PWA without or with minimal cognitive impairments.

8.5 Hypothesis 6: Carry-Over to and Changes on Standardized Assessments

Although limited changes on standardized assessments occurred, some significant and noteworthy changes were still found. The difference between P02 and P03's WAB-R AQ score from baseline to post-intervention reached Gilmore et al.'s (2018) benchmark for a significant change on this measure. P03's improvement also resulted in a change in severity rating from severe to moderate aphasia. On the BNT, P01 reached Gilmore et al.'s (2018) benchmark for a significant change. Additionally, while to different degrees per participant and to a limited degree overall, words from all three lists (SFA treated, PCA treated and untreated) were exclusively mentioned by participants at post-intervention and/or at maintenance when asked to name as many words as they could think of in a minute during the supermarket task, the animals subtest of the WAB-R as well as during the fruit and vegetable fluency tasks. Although a very limited total number of words, since these totals were controlled for baseline naming (items also mentioned at baseline were removed), they indicate at least some degree of additional generalization to different tests that do not involve pictures.

With regards to quality of life, although not maintained, P02 reported significant ameliorations on the SAQOL-39 total score and communication subcategory. Interestingly, these increases were found for both his Romanian and French/English scores. Conversely, P01, the participant who experienced the most significant naming improvements, reported a significantly worsened score on the communication subcategory of the SAQOL-39 at post-intervention and again at maintenance. It is possible that although he improved during interventions, P01 became increasingly aware of his communication difficulties. This observation of a decline rather than an improvement following aphasia intervention has been considered a possibility in other research. For example, when examining quality of life outcomes post-aphasia intervention, Hoover, Caplan, Waters, & Carney (2017) used two-tailed tests during statistical analyses in case declines on assessments, including measures of quality of life, occurred. Thus, measures of quality of life and communication should be monitored during SFA and PCA interventions.

Within the clinical context, to address both ameliorations and declines on measures of quality of life, realistic goal setting and counseling should be completed with clients. While completing an intervention to reduce a client's communication deficits is important, it is also imperative to work with the client and their family to determine goals that are meaningful, but also realistic for them to target. Considering achieving normal naming abilities may not be realistic within the chronic stage of aphasia, being faced with an inability to reach this goal may even reduce an individual's self-perception and quality of life. Instead, counseling on realistic goals and the development of adaptive strategies, as opposed to focusing on removing the deficits themselves, could assist in managing an individual's expectations. In addition, the act of finding alternative communication strategies could further improve their communicative success and quality of life.

8.6 Observations Related to Different Participant Profiles

When comparing the results obtained during probes, those obtained within the intervention sessions themselves and any changes observed on standardized assessment measures, it becomes clear that participants experienced differential patterns and degrees of improvement. Every participant experienced at least some degree of improvement on at least one measure, whether it was on probe word naming, on increasing feature generation autonomy and/or for increased naming success during the intervention sessions themselves.

Although she experienced some improvements, P04's gains were limited, especially considering the mild severity of her aphasia which has been reported in the literature to be a predictor of intervention success (Efstratiadou et al., 2018; Breitenstein et al., 2017). It should also be noted that her probe word naming success fluctuated across the entire study. While definitive explanations for small gains and probe word naming fluctuations cannot be provided, a potential influencing factor could have been related to difficulties in executive functioning (EF; Dignam, Copland, O'Brien, Burfein, Khan, & Rodriguez, 2017). This factor may be considered for example, since P04 frequently self-reported difficulties with her short-term memory (e.g., she frequently needed to audio record or write down information to remember it). Although cognitive testing was not completed in the present study and is difficult to complete with PWA due to their difficulties with receptive and expressive language, research indicates that "non-linguistic cognitive deficits, including EF impairments, commonly co-occur with aphasia and importantly, may influence language profiles and outcomes" (Murray, 2017, p. 721). Future studies could consider including cognitive assessments within the baseline assessment battery. Adding these assessments could continue to inform if and to what degree EF and other cognitive abilities are predictive of success on commonly used intervention tools such as SFA and PCA.

Conversely, P01's profile of comprehensive gains across SFA, PCA and carry-over items as well as improvements on standardized measures suggest the sizable improvements still possible at the chronic stage of aphasia. Thus, this study supports the recommendation that interventions are still warranted past the acute stage of aphasia (Moss & Nicholas, 2006) and that while adaptive or communication-focused service delivery models are assistive (Holland, Fromm, Forbes, & MacWhinney, 2017), impairment-based interventions can still be effective. Furthermore, recent studies have begun to revise the assumptions regarding brain plasticity, the ability for the brain to change and adapt, in that even an adult's brain maintains at least some ability to adapt. Without direct intervention apart from attending community/aphasia centers, Holland et al. (2017) indicated that significant improvements on the WAB-R could be seen on test-retest data within the chronic stage of aphasia. It was not apparent that demographic factors were at play. Thus, considering at least some degree of brain plasticity is retained in adulthood, even for those with aphasia, language-based recovery is still possible even years into the chronic stage (Hartwigsen & Saur, In Press).

For other individuals, such as P02 and P03, the 3 second probe word naming task may have been too challenging. However, when provided with additional time and support as during the intervention sessions themselves, they were able to experience benefits on other measures. For example, additional gains included increased naming speed (i.e., more words named earlier in the SFA or PCA protocols) and/or improved facility with the naming protocol (i.e., reduced reliance on the experimenter for feature generation). Thus, while improvements on short duration probe word naming (i.e., immediate retrieval as a consequence of network restoration) are an important outcome measure, the investigation of within-intervention measures can assist in identifying alternative avenues of an intervention's success during ideal circumstances. While less assistive during conversations since speech and word retrieval often require an element of speed, the use of compensatory strategies to assist in naming can still be assistive for a PWA's daily life (e.g., asking someone who works at a grocery store where to find certain items).

8.7 PCA Mechanism of Action

While PCA has been found to significantly improve naming difficulties for some PWA (Leonard et al., 2008; van Hees et al., 2013), Lambon Ralph et al. (2000) reported that individuals with anomia post-TBI had significant difficulty generating the first letter or first sound of target words. For this reason, it was suggested that this strategy may not be as effective for aphasia as it is for tip-of-the-tongue word retrieval difficulties experienced by individuals without neurological disorders. However, even though PWA have difficulty self-generating certain phonemic cues, when provided with the first sound of a target word, studies have demonstrated that they became more effective at naming the item (Lambon Ralph et al., 2000).

While being provided with phonological cues has been found to increase naming success (Lambon Ralph et al., 2000), clients may initially be "guessing" what answer to provide for a given feature since the PCA protocol indicates that clinicians should first allow clients the opportunity to self-generate features before providing them with an appropriate answer. In addition, considering the difficulties observed by PWA with regards to the self-generation of phonemic features, the inclusion of forced choices as recommended by Gravier et al. (2018) for SFA may need to be formally investigated in more detail for PCA. The "guessing game" of trying to find an appropriate feature may only be intensified if forced choices for feature generation are required. During the present study, this presence of additional guessing was occasionally seen with participants, where even when provided with a forced choice during feature generation, participants were still unsure which option was the correct answer for that feature.

In addition to the guidelines outlined above, the PCA protocol (as taken directly from SFA) indicates that clinicians should only provide a PWA with the difficult-to-name word following the generation of the entire feature analysis chart. Without having a similar anchor (i.e., the full word) for PCA, as having the picture anchor the semantic information for SFA, the generation of certain phonological features may be too difficult and/or unproductive for retrieving the target word. For instance, without providing semantic information related to the word's conceptual representation, the generation of certain features such as the *first sound associate* and *rhyme* could potentially lead individuals further away from the target. For example, P02 almost exclusively provided people's names for the *first sound associate* feature (e.g., *Barbara* as the *first sound associate* for *button*). While technically not incorrect since it begins with the same first sound as the target word, generating features such as these may create unrelated and/or odd semantic associations that could delay or further impact an individual's retrieval of the target word. A certain amount of semantic interference may even be generated if semantically related, but incorrect words beginning with the first sound are provided (e.g. *balloon* as the *first sound associate* for *ball*).

While some feature generation improvements occurred during the PCA protocol and significant naming gains could be observed for certain participants, it is hypothesized that targeting a word's phonological information on its own may not always be satisfactory or efficient for naming success (Lambon Ralph et al., 2000). Considering participants often required forced choices far into the feature generation process (e.g., even at the last intervention session, many participants still required frequent forced choices for the *rhyme* feature), the accumulation of phonological information was not always sufficient for the successful completion of the protocol without the experimenter's aid. Instead, the addition of supporting semantic information may be necessary to fully specify the target word form (Lambon Ralph et al., 2000). Phonological cues may even, only be of additional value when it is necessary to reduce the abundance of potential semantic interpretations (e.g., an underspecified round object could a *ball, orange, sun,* etc.) for images with low visual complexity (Meteyard & Bose, 2018). Thus, the combination of both semantic and phonological cues may be most assistive for naming success (Pease & Goodglass, 1978) and naming efficiency for individuals with aphasia.

8.8 Clinical versus Statistical Significance

Although significant naming improvements or trends of improvement as well as significant generalisation and carry-over effects were obtained for select participants following six weekly sessions of SFA and PCA interventions, the question of clinical significance should be raised. The same twenty SFA and twenty PCA words were practiced for an hour each week. After a total of six hours targeting those words per intervention, not all participants experienced significant gains, effect sizes ranged from negligible to medium and some words consistently remained difficult-to-name even with the generation of features. Furthermore, the generalization of effects from targeted words to untargeted words was not present in the current study and has been reported to be inconsistent in the literature. Even when effects had the potential to translate to a person's daily life in the form of quality of life improvements and/or the ability to name improved words in different contexts, it is possible to wonder if these mixed results justify the amount of effort required to complete the interventions. As asked by Armijo-Olivo (2018), are the beneficial effects large enough for the interventions to be considered clinically significant?

A solution to increase the clinical significance of SFA and PCA would be to target more words concurrently. However, when needing to progress through the entire SFA or PCA protocol for each item, especially at the beginning of therapy, there would not have been enough time in a session to target an additional number of words. These time restraints were particularly present for the individuals with a more severe aphasia. To reduce concerns related to timing, Mehta & Isaki (2016) began their SFA intervention by only targeting twelve words. However, once a participant could successfully name an item over four consecutive sessions and gained familiarity with the intervention, a new item was added. Six additional words were added for one of their participants, whereas fifteen additional words were added for another. Since a change was not mentioned, it is assumed that the original 60 minute session duration was maintained across all sessions despite words being added (Mehta & Isaki, 2016). To further enhance the possibility of targeting more items simultaneously, the number of features could also be reduced, as previously explained.

While this study was unable to determine which features (if any) were most valuable to feature generation, it did suggest that feature generation was assistive for naming in that additional words were named during feature generation that could not be named prior to these cues within the intervention sessions. It is possible that with additional participants of various severities and types of aphasia as well as with the inclusion of more items to be treated in SFA and PCA word lists, patterns regarding which features provide additional value to naming (i.e., features for which a larger number of items are named immediately following its generation) could emerge. Alternatively, to address this question of how many or which features to retain, larger-scale studies could be completed where a different number and combinations of cues could be provided to participants, for which the resulting gains across lists could then be compared. Furthermore, considering PWA often present with a combination of semantic and phonological impairments, the combination of SFA and PCA interventions (i.e., creating a single intervention that contains both semantic and phonological features) could be additionally assistive (Hashimoto, 2012).

Considering PWA in the chronic stage post-stroke infrequently receive speech and language services and when obtained are only provided for a limited duration of time (Wissel, Olver, & Stibrant Sunnerhagen, 2013), the choice of interventions provided should be carefully considered. Even if SFA and PCA can produce statistically significant benefits, practicing adaptive strategies such as the use of alternative means of communicating (e.g., gesturing, writing, drawing, etc.) may be more beneficial in the chronic stage. However, if word finding was still determined to be a priority, then it is recommended that treated words be chosen with care, where more frequently required, or individually important words could be more clinically relevant and justifiable of the amount of training required for improvement.

8.9 Limitations

Various limitations were present in the current study, regarding the methodology, the study design and in the execution of the interventions. Limitations were also present with regards to the available standardized assessments.

An important limitation in the present study was its descriptive as opposed to experimental nature. As a case series, the naming improvements experienced by certain participants cannot be definitively attributed to the treatment effects. While naming improvements and generalization measures were obtained, conclusive interpretations of these results cannot be provided considering the differential nature of these improvements across participants and the lack of stable baselines (see discussion below). Despite these limitations, the inclusion of more participants at a later time could allow for patterns across participants to be discerned, which could in turn lead to the strengthening and/or revision of the current results.

While generating the three word lists (SFA treated, PCA treated and control), the inclusion of items was decided based on results following the first two baseline sessions. Although it was assumed that naming success would be comparable at the third baseline session, creating the word lists with only two baseline scores resulted in an unstable baseline for P04's control word list. Additionally, the differences in naming success from the first two baselines to the third baseline may have been influenced by the modification to the protocol. For the first two baselines, participants were asked to name all 300 images, whereas during the third baseline session, participants were only asked to name the sixty-word subset included in the three word lists (SFA treated, PCA treated and control). Time allowed to name the stimuli remained the same. This modification may have resulted in different degrees of difficulty (e.g., a greater amount of cognitive resources without breaks may have been present during the first two baselines), making naming at the third baseline easier. In the future, to assure greater stability, word lists should only be generated following the naming of all 300 words in all three baselines.

Another limitation involved the choice of words included for multilingual participants. Although word lists controlled for category, word frequency, number of syllables and naming performance during the first two baseline assessments, a subset of the words included were English-French cognates. The presence of cognates (e.g., tomato-tomate, button-bouton) made interventions and the scoring of probe word naming increasingly difficult, in particular for P02 for whom French sometimes appeared easier to access than English. During interventions and probe word naming, P02 would occasionally respond in French first and then would sometimes become stuck on the French word and be unable to generate the closely related English version.

With regards to the standardized assessments, despite difficulties assessing executive functioning in aphasia and considering the inability to definitively explain P04's inconsistent profile, the addition of baseline cognitive testing may be of value in future research. Furthermore, for assessments such as the WAB-R and the BNT, the same concern discussed above regarding cognates was present. For P02, an added difficulty arose since both English-French and English-Romanian cognates were present in the BNT. During post-intervention and maintenance phases, P02 began attempting to name the words in his order of language facility, first in Romanian, followed by either French or English. The use of cognates as a naming strategy for multilingual participants has been reported in the literature (Roberts & Doucet, 2011). The standardized assessments also posed another difficulty in that a small subset of items in the word lists (e.g., hammer) were also included in certain assessments such as the WAB-R and the BNT. Although only a small subset of the words included in the SFA and PCA lists, the additional naming attempts for these duplicated words may have resulted in limited practice effects, secondary to the effects of the naming interventions.

An additional limitation may have been present with regards to the short, 3 second period provided for participants to name the probe words. While a short duration of time was included to attempt to examine the degree of restoration to the lexical network (i.e., where additional time may have allowed participants to use compensatory strategies), this measure of restoration may not have allowed an accurate representation of gains for the participants presenting with mild apraxia of speech. Thus, P02 and P03's lack of significant probe word naming gains may have been at least partially due to difficulties in physically generating the word (i.e., due to motor planning difficulties) as opposed to solely impairments in accessing the lexical items.

8.10 Future Directions

Considering that a large proportion of previous studies have mainly focused on naming gains during probe word naming as well as the generalization of treated items to untreated items, it is recommended that future research continue to investigate within-intervention methodologies and associated naming gains. Specifically, research could be completed to further investigate the value of each feature on naming by comparing groups of individuals receiving interventions that include a different subset of SFA or PCA features. Additional studies could also compare SFA and PCA to other word finding interventions to continue to investigate the value of feature generation and if their generation is a key element in anomia remediation. Another avenue of intervention could be to combine SFA and PCA features within a single protocol to determine if combining the semantic and phonological cues further contributes to naming success.

Future studies could also examine carry-over and generalization of naming gains to other contexts, such as the generalization of gains to standardized measures of daily life functioning and communication. Furthermore, considering the gains observed when images were presented in a natural context, SFA and PCA intervention gains could be compared between images actually targeted in natural context conditions as compared to the traditional method of presenting images in isolation. These investigations could elucidate if context is of a greater assistance or detriment to naming success at the chronic stage of post-stroke aphasia.

9. Conclusion

This study has provided novel information regarding additional avenues of improvements following SFA and PCA interventions within the chronic stage of post-stroke aphasia, namely with regards to the carry-over of effects to different situations and the presence of within-intervention gains. Within-intervention gains represent progress achieved in more ideal naming circumstances, where additional time and the use of compensatory strategies can allow for greater naming gains than the immediate recall required during short probe word naming contexts. The consideration of within-intervention gains may expand our understanding of S-LP treatment efficacy. Although investigations regarding which feature(s) provided the most value for naming success were not possible at this time, this study began defining the within-session process and progress of feature generation and naming success and supported the need to further study those aspects. Future studies should investigate the effect of combining both SFA and PCA cues into a single structured therapy so as to determine if naming success is augmented when both semantic and phonological information is provided. Investigations regarding which cues should be maintained for maximal probe word naming and within-intervention success are also recommended.

References

Acevedo, A., Loewenstein, D. A., Barker, W. W., Harwood, D. G., Luis, C., Bravo, M., . . . Duara, R. (2000). Category fluency test: Normative data for English- and Spanish-speaking elderly. *Journal of the International Neuropsychological Society, 6*, 760-769.

Armijo-Olivo, S. (2018). The importance of determining the clinical significance of research results in physical therapy clinical research. *The Brazilian Journal of Physical Therapy, 22(3)*, 175-176.

Baayen, R. H., Piepenbrock, R., & Gullikers, L. (1995). *The CELEX lexical database [webcelex].* Philidelphia: University of Pennsylvania, Linguistic Data Consortium.

Babbitt, E., Cherney, L., & Halper, A. (2008, August). *Measuring Communication Confidence in Persons with Aphasia.* Paper presented at the Clinical Aphasiology Conference, Jackson Hole, Wyoming. Abstract retrieved from http://aphasiology.pitt.edu/1906/

Beeson, P. M., & Robey, R. R. (2006). Evaluating Single-Subject Treatment Research: Lessons Learned from the Aphasia Literature. *Neuropsychology Review*, *16*(4), 161–169.

Best, W., Herbert, R., Hickin, J., Osborne, F., & Howard, D. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology*, *16*(1-2), 151-168.

Blankestijn-Wilmsen, J., Damen, I., Voorbraak-Timmerman, V., Hurkmans, J., Brouwer de Koning, J., Pross, P., & Jonkers, R. (2017). The effect of static versus dynamic depictions of actions in verb and sentence production in aphasia. *Aphasiology*, *31(10),* 1166-1182.

Bose, A. (2013). Phonological naming therapy in jargon aphasia with neologisms: Effects on picture naming and neologisms. *International Journal of Language & Communication Disorders, 48,* 582–595.

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

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation*, *17*(6), 411-422.

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

Breitenstein, C., et al. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomized, open-label, blinded-endpoint, controlled trial in a health-care setting. *Lancet, 389(10078),* 1528-1538.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One, 5(5)*, e10773.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS One, 9*, e106953.

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

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

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

Crawford, J. R., & Garthwaite, P. H. (2006). Comparing patients' predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology, 20(3)*, 259-271.

Cuetos, F., Aguado, G., Izura, C., & Ellis, A. W. (2002). Aphasic naming in Spanish: Predictors and errors. *Brain and Language, 82,* 344-365.

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(4)*, 801–838.

DeLong, C., Nessler, C., Wright, S., & Wambaugh, J. (2015). Semantic feature analysis: further examination of outcomes. *American Journal of Speech-Language Pathology*, *24*(*4*), S864-S879.

Dignam, J., Copland, D., O'Briend, K., Burfein, P., Khan, A., & Rodriguez, A. D. (2017). Influences of cognitive ability on therapy outcomes for anomia in adults with chronic poststoke aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR, 60(2),* 406-421.

Efstratiadou, E. A., Papathanasiou, I., Holland, R., Archonti, A., & Hilari, K. (2018). A Systematic Review of Semantic Feature Analysis Therapy Studies for Aphasia. *Journal of Speech, Language & Hearing Research*, *61*(5), 1261-1278.

Gilmore, N., Dwyer, M., & Kiran, S. (2018). Benchmarks of significant change after aphasia rehabilitation. *Archives of Physical Medicine and Rehabilitation*. Advance online publication. doi: https://doi.org/10.1016/j.apmr.2018.08.177

Goldberg, D., & Williams, P. (1991). *A user's guide to the General Health Questionnaire*. Windsor: NFER-Nelson.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination-Third Edition (BDAE-3)*. Austin: ProEd.

Gravier, M. L., Dickey, M. W., Hula, W. D., Evans, W. S., Owens, R. L., Winans-Mitrik, R. L., & Doyle, P. J. (2018). What Matters in Semantic Feature Analysis: Practice-Related Predictors of Treatment Response in Aphasia. *American Journal of Speech-Language Pathology*, *27*(1S), 438-453.

Hartwigsen, G., & Saur, D. (In Press). Neuroimaging of stroke recovery from aphasia - Insights into plasticity of the human language network. *Neuroimage*. doi:10.1016/j.neuroimage.2017.11.056

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

Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *Journal of Communication Disorders*, *44(4),* 459-469.

Heart&Stroke. (2017). Different Strokes: Recovery Triumphs and challenges at any age. 2017 Stroke Report.

Hilari, K., Byng, S., Lamping, D. L., & Smith, S. C. (2003). Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. *Stroke, 34*(*8*), 1944-1950.

Hilari, K., Lamping, D. L., Smith, S. C., Northcott, S., Lamb, A., & Marshall, J. (2009). Psychometric properties of the Stroke and Aphasia Quality of Life Scale (SAQOL-39) in a generic stroke population. *Clinical Rehabilitation, 23*, 544-557.

Hinckley, J., & Carr, T. (2005). Comparing the outcomes of intensive and non-intensive context-based aphasia treatment. *Aphasiology, 19(10-11)*, 965-974.

Hoffman, P., Clarke, N, Jones, R. W., & Noonan, Krist, A. (2015). Vocabulary relearning in semantic dementia: Positive and negative consequences of increasing variability in the learning experience. *Neuropsychologia, 76,* 240-253.

Holland, A., Fromm, D., Forbes, M., & MacWhinney, B. (2017). Long-term recovery in stroke accompanied by aphasia: A reconsideration. *Aphasiology, 31(2),* 152-165.

Hoover, E. L., Caplan, D. N., Waters, G. S., & Carney, A. (2017). Communication and quality of life outcomes from an interprofessional intensive, comprehensive, aphasia program (ICAP). *Topics in Stroke Rehabilitation, 24(2),* 82-90.

Howard, D., Patterson, K. E., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985). The treatment of word retrieval deficits in aphasia: A comparison of two therapy methods. *Brain*, *108*, 817–829.

Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees Test: A test for semantic access from words and pictures.* Bury St. Edmunds: Thames Valley Test Company.

Jurica, P. J., Leitten, C. L., & Mattis, S. (1988). *DRS-2 Dementia Rating Scale-2: Professional Manual.* Lutz, FL: Psychological Assessment Resources.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lea & Febiger.

Kasselimis, D. S., Simos, P. G., Peppas, C., Evdokimidis, I., & Potagas, C. (2017). The unbridged gap between clinical diagnosis and contemporary research on aphasia: A short discussion on the validity and clinical utility of taxonomic categories. *Brain & Language, 164,* 63-67.

Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology, 10(2),* 159-180.

Kertesz, A. (2006). *Western Aphasia Battery-Revised*. Austin: Pro-Ed.

Kromrey, J. D. & Foster-Johnson, L. (1996). Determining the Efficacy of Intervention: The Use of Effect Sizes for Data Analysis in Single-Subject Research. *The Journal of Experimental Education, 65(1)*, 73-93.

Lambon Ralph, M. A., Sage, K., & Roberts, J. (2000). Classical anomia : a neuropsychological perspective on speech production. *Neuropsychologia, 38,* 186-202.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

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(4)*, 254-267.

Massaro, M. E., & Tompkins, C. A. (1992). 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.

Meinzer, M., Djundja, D., Barthel, G., Elbert, T., & Rockstroh, B. (2005). Long-Term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke, 36(7),* 1462-1466.

Meteyard, L. & Bose, A. (2018). What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. *Journal of Speech, Language, and Hearing Research, 61,* 658-674.

Moss, A., & Nicholas, M. (2006). Language rehabilitation in chronic aphasia and time postonset: A review of single-subject data. *Stroke, 37(12)*, 3043-3051.

Murray, L. (2017). Focusing attention on executive functioning in aphasia. *Aphasiology, 31(7),* 721-724.

Neumann, Y. (2018). A case series comparison of semantically focused vs. phonologically focused cued naming treatment in aphasia. *Clinical Linguistics & Phonetics*, *32*(*1)*, 1-27.

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

Nourbakhsh, M. R. & Ottenbacher, K. J. (1994). The Statistical Analysis of Single-Subject Data: A Comparative Examination. *Physical Therapy, 74(8)*, 768-776.

Paul-Brown, D., Frattali, C. M., Holland, A. L., Thompson, C. K., & Caperton, C. J. (2003). *Quality of communication life scales*. Rockville, MD: American Speech-Language-Hearing Association.

Peach, R. K., & Reuter, K. A. (2010). A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. *Aphasiology, 24*, 971–990.

Pease, D. M., & Goodglass, H. (1978). The effects of cueing on picture naming in aphasia. *Cortex, 14,* 178-189.

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

Roberts, P. M. & Doucet, N. (2011). Performance of French-speaking Quebec adults on the Boston Naming Test. *Canadian Journal of Speech-Language Pathology and Audiology, 35(3),* 254-267.

Robey, R. R., Schultz, M. C., Crawford, A. B., & Sinner, C. A. (1999). Single-subject clinical-outcome research: Designs, data, effect sizes, and analyses. *Aphasiology*, 13*(6),* 445-473.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception, 33(2),* 217-236.

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.

Sakamoto, R., Higuchi, A., Tsuda, K., Tanimoto, T., & Kami, M. (2017). Intensive speech and language therapy after stroke. *Lancet, 390,* 228.

Salmon, J. P., Matheson, H. E., & McMullen, P. A. (2014). Photographs of manipulable objects are named more quickly than the same objects depicted as line-drawings: Evidence that photographs engage embodiment more than line-drawings. *Frontiers in Psychology, 5,* 1-6.

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6(2),* 174–215.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14(2)*, 167-177.

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. (2014). Neural activity associated with semantic versus phonological anomia treatments in aphasia. *Brain & Language, 129,* 47-47.

Wallace, S. J., et al. (2018). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke, 14(2),* 180-185.

Wambaugh, J. L., Linebaugh, C. W., Doyle, P. J., Martinez, A. L., Kalinyak-Fliszar, M., & Spencer, K. A. (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., 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*(2), S334-S369.

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

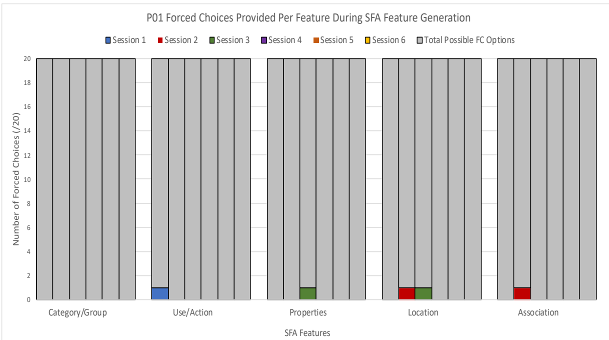
Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop "fishing"? A review of generalization following aphasia intervention. *Aphasiology, 29(11),* 1240-1264.

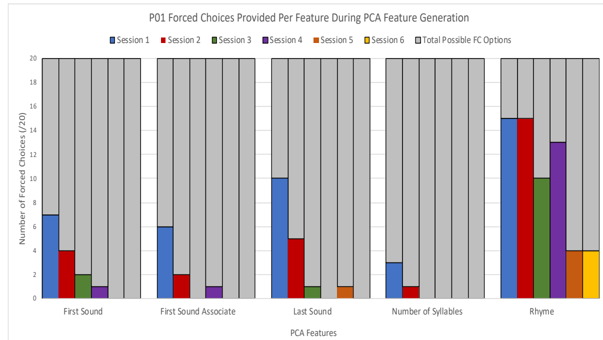
Wissel, J., Olver, J., & Stibrant Sunnerhagen, K. (2013). Navigating the poststroke continuum of care. *Journal of Stroke and Cerebrovascular Diseases, 22(1)*, 1-8.

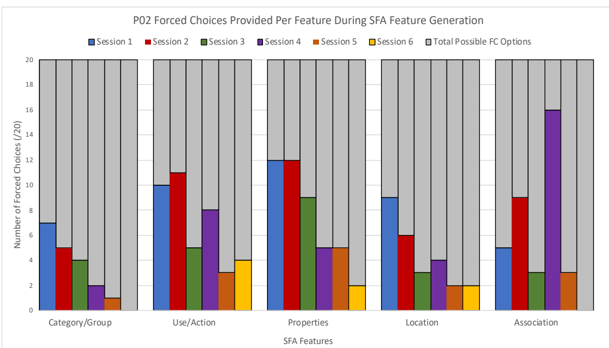
Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders, 45(1),* 27-36.

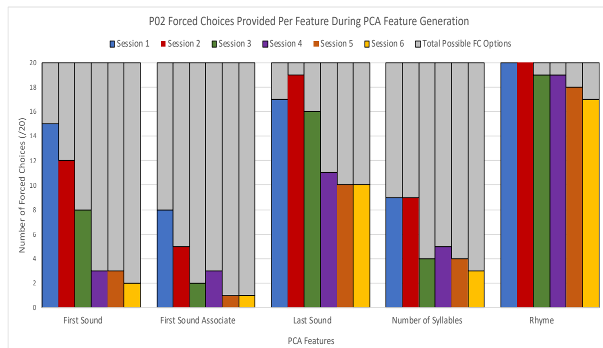
Appendix

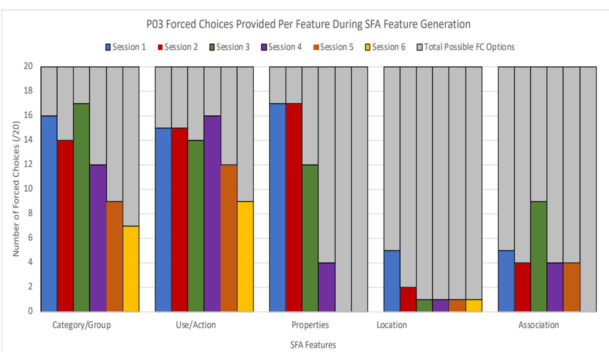
Figure I. Distribution of the number of forced choices required per feature from session 1 to session 6 for SFA and PCA interventions

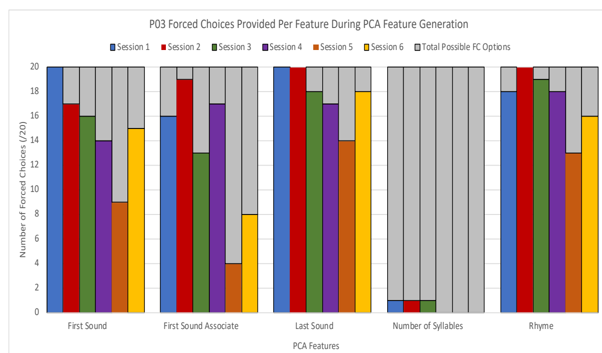


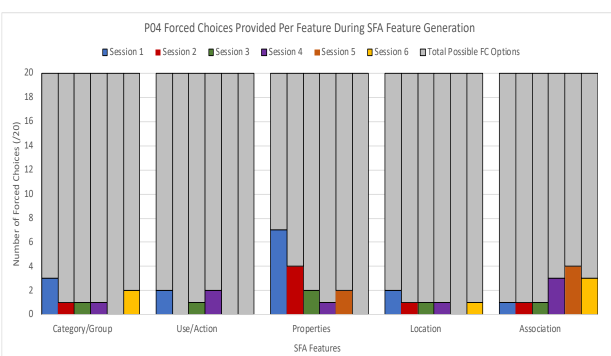












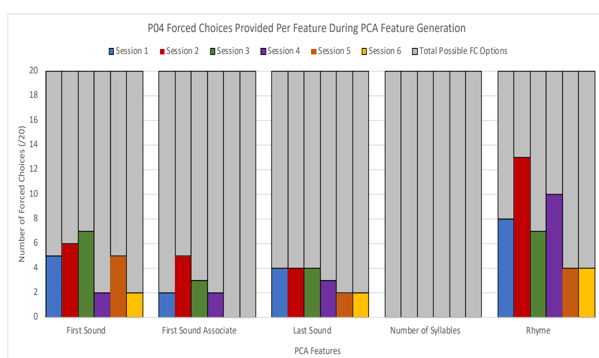


Figure II. The point at which a word was named (first 15s, at a feature, at end) during SFA

Figure III. The point at which a word was named (first 15s, at a feature, at end) during PCA