8-22-2014

Neurobehavioral Response to Increased Treatment Dosage in Chronic Aphasia

Jennifer L. Mozeiko
University of Connecticut - Storrs, jennifer.mozeiko@uconn.edu

Neurobehavioral Response to Increased Treatment Dosage in Chronic Aphasia

Jennifer Mozeiko

University of Connecticut, 2014

Purpose- This study investigated changes in oral-verbal expressive language and patterns of neural activation associated with improvements following a "double dose" of high intensity aphasia treatment in four participants with stroke-induced, chronic aphasia. Generalization of treatment to untrained materials and to discourse production was also analyzed as was the durability of the treatment effect.

Methods- Participants with aphasia (PWAs) were assessed using standardized measures, discourse tasks, a social validation measure, and neuroimaging at four time points to document behavioral and neural changes throughout each of two thirty-hour Treatment Periods of constraint induced language therapy (CILT; Pulvermüller et al., 2001). Assessments took place Pre-Treatment, Post-Treatment Period I (30 hours), Post-Treatment Period II (30 additional hours) and at Follow-up, eight weeks after treatment completion. Daily probes of trained and untrained materials were also administered. A slow event related, confrontational naming paradigm was employed using stimuli customized for each PWA based on pre-treatment testing. Region of Interest (ROI) analyses were conducted to assess changes in activation in three main

language areas associated with overt naming: bilateral inferior frontal, middle temporal and superior temporal gyri.

Results- Despite participant heterogeneity, behavioral results for each PWA indicated a positive response to treatment following Treatment Period I and also following Treatment Period II with medium to large effect sizes following both Treatment Periods compared to Pre-Treatment. The treatment effect extended to untrained stimuli and to discourse productivity or efficiency and was maintained eight weeks following treatment completion. Hemispheric laterality shifted over the course of treatment but direction of shift varied among participants, brain regions and between various time points.

Discussion- Neural and behavioral data tended to support the utility of a second treatment period although recovery patterns varied widely among individuals. These results are relevant for rehabilitation in chronic aphasia confirming that significant language gains continue well past the point of spontaneous recovery and that they can occur in a relatively short time period. Importantly, changes are not confined to a single treatment period suggesting that individuals with chronic aphasia may benefit from multiple, high intensity doses of treatment.

Neurobehavioral Response to Increased Treatment Dosage in Chronic Aphasia

Jennifer L. Mozeiko

B.A. University of Connecticut (1993)

M.A. University of Connecticut (2005)

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

at the

University of Connecticut

2014

APPROVAL PAGE

Doctor of Philosophy Dissertation

Neurobehavioral Response to Increased Treatment Dosage

Presented by

Jennifer L. Mozeiko, B.A., M.A.

Co-Major Advisor

_____

Carl A. Coelho, Ph.D.

Co-Major Advisor

_____

Emily B. Myers, Ph.D

Associate Advisor

_____

Tammie J. Spaulding, Ph.D.

University of Connecticut

2014

Acknowledgments

Over the past eight years, I have received support and encouragement from a great number of individuals. Thanks to my advisor, Carl Coelho, first and foremost. I knew early on that pursuing a PhD was a journey for which I would need to choose my guide wisely. Carl has been both a role model and mentor, and I am so grateful to have had his unwavering support. I found another mentor in Emily Myers whose help and advice was essential in navigating the fMRI design, analysis and interpretation in this thesis and who I have found to be inspiring on multiple levels. I would also like to thank Tammie Spaulding, my committee member, who offered me encouraging words, thoughtful criticism, and time and attention during very busy semesters.

Many thanks also: To Gabriela Depa and Francesca Rush, undergraduate research assistants and promising clinicians, who spent hours transcribing discourse samples and assisting with discourse analyses. To Leora Cherney, who provided valuable insights on the prospectus, shaping what would lead to an ultimately better thesis. To all of the participants and their families who drove from across the state to attend treatment sessions, scanning sessions and endless assessments. It was they who best motivated me to relentlessly push forward as they themselves have done and continue to do.

To my labmate and friend, Karen Lê, who has walked this road (and farm) with me and to all of the other doctoral students, Stephanie Del Tufo, Sayako Earle, Dana Arthur and the others who have lent their support and provided encouragement--thank you.

Thanks to those with essential roles in my village: To my sister, Kristin Mozeiko, for doing it first and coming out the other side. To Nancy Pufahl for being there when it counted. To Mindy Carpenter, Sabrina Bowersett, Meredith Silagy—three women who are so important in

my children's lives and in mine. And to my mom, Janice Fuller, for raising me to "Just Do It" long before Nike trademarked the phrase.

Finally, I am grateful to my husband, Jason Pufahl, for his absolute consistency, patience and support and to my children Owen and Hazel Pufahl for the persistent reminder that life trumps work.

Table of Contents

Chapter I

Introduction

There are several approaches to the treatment of aphasia based on competing paradigms. Each has its merits but few have consistently demonstrated lasting effects or generalization to functional communication, the ultimate goal of treatment. This may be due, in part, to their failure to generate the neural reorganization sufficient to maintain any behavioral change (Leon, Maher, & Gonzalez Rothi, 2011).

Though typical outpatient speech language therapy is administered at a "dose" of two-three hours per week, treatments most consistently flagged as effective in generating behavioral change—though not necessarily generalizability— are those delivered at doses greater than that (Bhogal, Teasell, & Speechley, 2003; Cherney, Patterson, Raymer, Frymark, & Schooling, 2010; Kelly, Brady, & Enderby, 2010; Robey, 1998; Teasell et al., 2009). Neuroimaging studies confirm corresponding neural reorganization after high dosage treatments (Breier, Maher, Novak, & Papanicolaou, 2006; Crosson et al., 2009; Fridriksson, Bonilha, Baker, Moser, & Rorden, 2010; Musso et al., 1999). Fewer studies also document generalizability and maintenance of gains but of those that do, it has been demonstrated that short term, high dosage therapy (20-30 hours over two weeks) can result in stable improvements (Barthel, Meinzer, Djundja, & Rockstroh, 2008; Maher et al., 2006). However, protracted dosage (ten hours over five weeks) has been shown to have a better maintenance effect than a more intensive (ten hours over two weeks) dosage for individuals with anomia (Sage, Snell, & Lambon Ralph, 2011). Thus, although there is strong evidence that higher dosages are effective in generating immediate behavioral change, further investigation is needed to determine the factors contributing to optimal dosage and to maintenance of treatment effect for various clinical populations.

Current studies that claim to administer "intensive treatment" vary widely in their definitions of the parameters that make up a "dose" of treatment (e.g., session duration and session frequency), further complicating its contribution to treatment efficacy. Assumptions about the meaning of intensity stem from various literature reviews of dosage. For example Bhogal et al., (2003) found a significant immediate treatment effect for therapy administered at 8.8 hours per week over 11 weeks whereas Robey (1998) made more general conclusions that a minimum of two hours of treatment per week were more effective than less. These reviews do not purport to define intensity, however, they are often referenced to justify use of the term "intensive" in any given treatment study. Just as a medication dosage is more than the number of pills to ingest (i.e., unit amount based on patient weight, number of days to take medication, number of times a day to take medication), an aphasia treatment dosage is more involved than the number of hours of treatment administration.

Cherney and colleagues (Cherney et al., 2010; Cherney, Patterson, & Raymer, 2011; Cherney, Patterson, Raymer, Frymark, & Schooling, 2008) completed a thorough investigation of the effect of intensive aphasia treatment by conducting a series of systematic reviews of treatment studies that sought to compare aphasia treatments using higher and lower intensity conditions. Levels of intensity varied between studies, as did treatment type and participant type (chronic and acute populations were both included). Results indicated that the question of dosage is not straightforward and that more is *not* necessarily better. Cherney (2012) pointed out that other variables in the therapeutic process, specifically treatment *type*, also affect treatment outcome, making it difficult to separate out the contribution of intensity.

Despite this, studies using intensive dosages have become more common allowing neuroimaging to become more widely used as another measure of treatment outcome and to

investigate the neural reorganization that takes place as a result of treatment. High dosage treatments performed over shorter time periods make neuroimaging more feasible for treatment research as there is less likelihood of participation drop out and a greater chance that changes are due to treatment versus other factors experienced in longer time periods. Neuroimaging of treatment remains in the investigative stages and requires data from cumulative studies in order to best answer current questions. One example is the question of whether regaining use of spared left perilesional tissue results in better language outcomes than recruitment of right hemisphere homologues. This issue is discussed in more depth in Chapter III, p. 61. At this time much of the data support regained use of left perilesional areas for strongest functional gain but there is also compelling evidence supporting the right (i.e, Hartwigsen et al., 2013; Richter, Miltner, & Straube, 2008). Though a matter of continued debate, these data are likely more complimentary than contradictory. Language is a cognitive process subserved by many related processes all of which contribute to effective function and therefore various routes of compensation may be possible following brain damage. (Refer to Saur & Hartwigsen, 2012 for an in depth review of the neurobiology of language recovery). It remains to be seen, then, whether the contribution of one hemisphere or one brain region is more essential than another to the language recovery process post stroke.

The primary objective of this dissertation study was to investigate the question of increased treatment dosage. Despite reported gains as a result of hours-long treatment sessions over months of time, U.S insurance companies do not yet cover intensive treatments for clinical use since evidence remains equivocal (Cherney, Patterson, & Raymer, 2011). Discrepancies in terminology have made it difficult to quantify gains and to assess whether there is such increased value in intensive language treatment. By using a treatment that has been researched using a

fairly standard dosage and one that has been replicated several times, this study aims to add to the mounting evidence that factors of neuroplasticity apply to language remediation in ways similar to the way they apply to motor remediation (e.g., Mark & Taub 2004; Nudo, Plautz, & Frost 2001; Kleim, 2004). Specifically, it is predicted that rigorously "exercising" damaged language networks will result in neural restitution or compensation in people with chronic aphasia as demonstrated by language gains that generalize to functional oral verbal expression and as well as corresponding increases in neural activation. It is predicted that an additional period of treatment will result in additional benefits.

Within this construct is a continued exploration of a treatment that has been under some scrutiny by aphasia researchers. The question remains whether Constraint Induced Language Therapy (CILT) is a viable means of remediation for those with a range of aphasia deficits, even for those with mild aphasia for whom reported gains have been most limited (Meinzer et al., 2008). An in depth discussion of candidacy for intensive treatment and for CILT is in Chapter III, pg. 63, however, is predicted that when participants with mild aphasia are provided sufficiently challenging material, CILT will be equally beneficial to this population.

A second main objective was to investigate changes in neural activation over four time points. The brain's response to a task is a series of physiologic changes in blood vessels within a specific region which may vary depending on the task. These include changes in blood flow and therefore in blood oxygenation levels allowing fMRI imaging to produce a map of neurons that are active and those that are not. Though increasing in number, there are few treatment studies that use neuroimaging to help assess change. Those that have tend to use pre- and post-treatment scans only; however, this does not necessarily provide a complete picture of what has occurred during treatment. "Post-treatment" scans have been reported to occur at time points

immediately post-treatment (Fridriksson, Richardson, Fillmore, & Cai, 2011; Meinzer et al., 2008) to eight months post-treatment (Menke et al., 2009). Given similar language performance outcomes, neural results would likely still be very different just due to the time post-treatment. More time post-treatment means more opportunity for other life experiences to impact neural change and for treatment response to decay. Follow-up scans showing a maintained or decayed response would also provide important data but only in when compared to changes that occurred immediately post treatment. In order to best assess the recovery process in response to treatment, neural data from multiple time points is desirable. In the current study, Functional Magnetic Resonance Imaging (fMRI) is performed at four time points in order to provide additional insight into the neural changes that occur during the recovery process in people with chronic aphasia. It is predicted that neural activation will increase during the period of treatment and then will decline for trained words as production becomes more automatic, suggesting that learning has been consolidated.

Exploration of these aims begins in Chapter II with a preliminary study (Study One) outlining outcomes of eight participants. Four participants received 30 hours of CILT over two weeks and four received the same amount over ten weeks. In Chapter III additional background information for the current study and the four research questions are presented. Chapters IV and V describe the methods and the results of the current study and Chapter VI is a discussion of the results within the framework of the research questions.

Chapter II

Study One

Introduction

Prior to initiation of the dissertation study, a first treatment study was designed to investigate the contribution of intensity to Constraint Induced Language Therapy (CILT; Pulvermüller et al., 2001). CILT, also known as Constraint Induced Aphasia Therapy (CIAT), or most recently as Intensive Language Action Therapy (ILAT; Difrancesco, Pulvermüller, & Mohr, 2012) is a treatment of aphasia based on a successful physical therapy protocol, Constraint Induced Movement Therapy (CIMT; Taub, Miller, & Novack, 1993). CIMT is used in physical therapy for limb weakness after stroke and is based on the philosophy of "learned non-use", the tendency to rely on the stronger limb thereby hindering rehabilitation of the affected limb (Taub, Uswatte, & Pidikiti, 1999; Taub, 2004). Studies have shown increased limb use and evidence of motor cortex reorganization (Taub et al., 1999) following CIMT which employs three key principles: 1) massed practice 2) restraint of the unaffected limb 3) forced use of the affected limb (Taub et al., 1999).

In 2001, Pulvermüller and colleagues applied these principles to language treatment for individuals with chronic aphasia. In CILT, compensatory non-verbal communication modalities (such as gesturing, writing or drawing) are restrained and participants are required to produce exclusively verbal requests and responses. The initial study (Pulvermüller et al., 2001) and several subsequent follow-up studies (e.g., Barthel, Meinzer, Djundja, & Rockstroh, 2008; Johnson et al., 2013; Kurland, Pulvermüller, Silva, Burke, & Andrianopoulos, 2012; Maher et al., 2006; Meinzer, Djundja, Barthel, Elbert, & Rockstroh, 2005; Rose, Attard, Mok, Lanyon, &

Foster, 2013; Sickert, Anders, Münte, & Sailer, 2014; Szaflarski et al., 2008) all showed significant improvement in the amount and quality of communication on a variety of outcome measures including standardized aphasia batteries, communication activity logs, and narrative discourse samples. The variables contributing to remediation, however, remain ambiguous. Restraining compensatory communication is a radical change for speech-language pathologists who have been trained to assist in the maximization of functional communication. Therefore, before adopting such a paradigm shift, it is prudent to determine the contribution of each CILT factor to the success of treatment.

As discussed, several studies cite the importance of intensity to a treatment regimen for those with chronic aphasia, however a systematic review of studies that controlled the treatment in order to compare intensive and non-intensive dosages found the results equivocal (Cherney, Patterson, Raymer, Frymark, & Schooling, 2010). Contributing to these results was a large randomized control study that demonstrated the rigors of an intensive program ineffective for those with acute aphasia (Bakheit, Shaw, Barrett, et al., 2007). In addition, some studies showed slight gains for a distributed plan of treatment (Ramsberger & Marie, 2007; Sage, Snell, & Lambon Ralph, 2011).

In the non-impaired population, it is distributed practice, also known as spaced repetition, that has shown to be more effective *except* in the learning of complex tasks (Donovan & Radosevich, 1999). However, the implication might be that language re-learning for an individual with chronic aphasia would be considered such a complex task, requiring intensive training to jump-start the cognitive-linguistic system.

Distributed practice has logistical advantages in aphasia rehabilitation in that it is the way treatment is currently scheduled in outpatient clinics, allowing a speech and language pathologist

to see multiple clients per day. Repeated opportunities (as opposed to mass practice) for learning and re-learning have potential advantages for the participant as well. By extending the treatment period, the PWA has multiple chances to learn and thus perhaps more opportunity for adaptive neural change to occur since recent literature suggests that sleep is an important factor in promoting learning-dependent synapse formation (Yang et al., 2014).

Cherney and colleagues (2008) provided a systematic review summarizing evidence for intensity of treatment and for CILT on language and functional outcome measures. Data suggested that performance on language outcome measures was generally better and maintained longer following CILT than on other intensively administered treatments. Importantly, there are few studies that have specifically controlled for intensity. Maher (2006) and Kurland and Pulvermüller (2012) each compared CILT to a group therapy encouraging multimodality communication, much like Promoting Aphasics' Communicative Effectiveness therapy (PACE; Davis, 2003) which promotes the use of all communicative modalities including gesture, drawing and writing. Improvements were noted in both groups but Kurland (2012) reported better naming performance and Maher (2006) reported better maintenance of gains following for those who received CILT. Most recently, Rose (2013) used Multi-Modal Aphasia Therapy (M-MAT), for which the goal is also verbal language production however, clinicians use multimodal cues to facilitate production. Again, there was a positive change in aphasia severity in both groups, and reported improvements in language production. Neither treatment was reported as having an advantage over the other. One study compared CILT to an individually tailored therapy (Barthel et al., 2008) with, again, comparable results. In summary, results have tended to favor CILT marginally but no study has yet found a clear advantage for it suggesting that intensity may be a main contributor to positive outcomes following this treatment.

The present study is a Phase II study (Robey, 2004) in which the treatment was controlled in order to analyze the contribution of intensity to CILT for individuals with chronic aphasia. CILT was delivered in what appears to have emerged as a standard dose for this treatment at a Total Intervention Duration of 30 hours over two weeks to two dyads. The same treatment was also administered in a more traditional dosage of 30 hours over ten weeks to two additional dyads. This latter dosage of three hours per week is more akin to what an individual might receive as an outpatient restricted by typical insurance coverage. Given the heterogeneity of the participants, variable response to treatment was anticipated, however, gains in productivity of discourse and on standardized tests were predicted for individuals who received the intensive CILT (CILT-I). It was hypothesized that those who received a more standard distribution of treatment (CILT-D) were less likely to demonstrate change on these measures when compared to their pre-treatment performance.

Methods

Participants.

Eight participants were recruited from an aphasia group based at the University of Connecticut Speech and Hearing Clinic. Inclusion criteria included: (a) a single left-hemisphere stroke (b) onset of at least one year prior to participation in the study, (c) premorbid right-handedness, as confirmed by a spouse or family member, (d) no reported history of other neurological or learning disorders (e) monolingual English speakers and (f) access to reliable transportation (see Table 1). All participants had adequate hearing and visual acuity, some with hearing aids and corrective lenses, to participate in the study. Individuals' communicative deficits varied widely and most demonstrated some degree of concomitant apraxia of speech (AOS). Differential diagnosis of AOS is difficult, particularly for those with more severe

aphasia deficits where symptoms of groping and variability of errors may be attributable to the aphasia (Duffy, 2012). AOS is generally thought to negatively impact aphasia treatment but participants with AOS have been included in previous CILT studies with positive results (e.g., Kurland et al., 2012; Kurland, Silva, Burke, & Iyer, 2011; Maher et al., 2006) and thus was not considered as criteria for exclusion.

Table 1.

Characteristics of Participants receiving Intensive (I) and Distributed (D) Treatment

ID	I1	I2	I3	I4	D1	D2	D3	D4
Age	26	53	67	72	63	47	51	77
Months Post Onset	67.2	18	134.4	42	96	13.2	21.6	13.2
Sex	M	F	M	F	M	M	F	M
Handedness	R	R	R	R	R	R	R	R
Hemiplegia	mild	none	severe	severe	none	severe	none	moderate
Education (years)	12	14	12	12	12	12	13	16
Previous Employment	McDonalds cook	FAA Technician	Mechanic	Homemaker	Paving and construction	Window installation	Hospital food service	Computer aided design
AOS	mild	severe	mod	mod	severe	mod	mild	mild
CADL	90	90	40	8	26	35	81	77
	95%	95%	25%	5%	50%	95%	95%	50%

R-CPM								
R- WAB AQ	67.7	24.8	32.3	27.4	28.9	50.1	84.2	73.6
				Selected WAB AQ Subtests				
Yes/No Questions	100.0%	90.0%	80.0%	75.0%	70.0%	95.0%	85.0%	90.0%
Auditory-Verbal Comp.	100.0%	81.7%	45.0%	11.7%	60.0%	95.0%	93.3%	98.3%
Word Fluency	30.0%	0.0%	10.0%	10.0%	0.0%	10.0%	60.0%	20.0%
Object Naming	88.3%	3.3%	25.0%	21.7%	11.7%	66.7%	100.0%	76.7%
WAB Classification	Broca's aphasia	Not Classifiable	Broca's aphasia	Global aphasia	Not Classifiable	Broca's aphasia	Anomia	Conduction aphasia

Note. ID- I-Intensive and D-Distributed; AOS-apraxia of speech; CADL- Communication Activities of Daily Living (Holland, Frattali, & Fromm, 1999) RCPM- Raven's Coloured Progressive Matrices(Raven, Court, & Raven, 1988); R-WAB AQ- Revised Western Aphasia Battery Aphasia Quotient (Kertesz, 2006)

While taking part in the study, from the time of baseline collection to follow-up testing four weeks post-treatment, individuals did not participate in any other form of language rehabilitation, including social aphasia groups. Informed consent was obtained from all participants in the study, which was approved by the University of Connecticut Institutional Review Board.

Design.

This study used a modified multiple baseline design across subjects (McReynolds & Kearns, 1983) in order to detect potential changes in discourse production, the primary outcome variable of interest, resulting from treatment on a case-by-case basis. In this way, it was possible to track potential generalization of treatment to connected speech across eight participants of varying aphasia severity. A multiple baseline design is the preferred method in aphasia treatment (Kiran et al., 2012; Thompson, 2006) and when performed across subjects, all participants should begin baseline testing concurrently with staggered treatment initiation for individuals and continued baselining for those yet to initiate treatment. CILT, however, was designed for small groups and is less conducive to the required staggering of baselines at the individual level. Since treatment for each dyad was conducted at different time periods (CILT-I dyads received treatment in July and August; CILT-D dyads received treatment from September to November), staggered and protracted baseline periods were possible at the small group level but would have required participants, seven of whom relied on caregivers for transportation, to make several additional trips to the Speech and Hearing Clinic. This was not financially or logistically feasible for most of them. Instead, a minimum of three baselines was taken for each individual at least 24 hours between baselines and no more than one month between baseline

points. Fewer baseline points may be considered a limitation however all of the participants in this study were at least one year post CVA and none were receiving alternate therapy. Therefore, it is likely than any change following baseline is a result of treatment. Replication of results was demonstrated across participants following a stable baseline.

Standardized measures of aphasia, cognition and functional communication were administered pre- and post- therapy and one month after the completion of treatment as additional measures of responsiveness to treatment.

Standardized Assessments.

The Western Aphasia Battery- Aphasia Quotient (WAB- AQ; Kertsz, 1982), the Communication Activities of Daily Living-2 (CADL-2; Holland, Frattali, & Fromm, 1999) and Raven's Coloured Progressive Matrices (R-CPM; Raven, Court, & Raven, 1998) were administered pre-treatment, immediately post-treatment and one month post-treatment. The WAB provides an AQ score yielding an estimation of aphasia severity and classification parameters. The test has good test retest reliability ($r = .88$, $p < 0.001$) and internal consistency ($r = .974$; Shewan & Kertesz, 1980) and a five point gain is thought to be clinically significant (Shewan & Donner, 1988). The object naming subtest of the WAB was used to gauge potential treatment generalization to untrained words.

The CADL provides a way to quantify the ability of someone with aphasia to communicate using their residual skills in day to day encounters. It also has good test-retest reliability ($r = 0.88$) and internal consistency ($r = 0.99$; Aten, Caligiuri, and Holland 1982). The R-CPM (reliability and consistency unavailable for PWAs) measures general cognitive abilities without requiring processing or production of oral verbal language. In addition, it has been used as a prognostic index. The Quick Assessment for Apraxia of Speech (validity and consistency

unavailable; Tanner & Culbertson, 1999) was completed pre-treatment only to help characterize the language deficits of the participants. When possible, AOS was distinguished from aphasia and classified as mild, moderate or severe using differential diagnosis guidelines recommended by Duffy (2005, p. 422).

Baseline Testing.

Three to five baseline probes testing discourse production were administered on different days, always at least 48 hours apart, but within a two week time period during the period of pre-treatment testing. Treatment began once stability was achieved for the efficiency measure of Correct Information Units (CIUs)/minute. Stability was defined as a lack of consistent increase or decrease in slope though day to day performance variation, not unusual for PWAs, was evident for several participants. It often took several baselines to establish consistency. CIUs/min were calculated for each picture description (see data analysis below, p. 20, for details on this discourse analysis) and averaged with the other two for each baseline point as well as for subsequent probes during and following treatment.

Baseline testing was always done first, prior to any other testing scheduled for that day. In order to control for potential learning effect, ten different Rockwell prints were used to stimulate language production throughout all baseline, treatment and post treatment probes. Three Rockwell prints were shown to each participant at each probe and for each they were prompted with "Can you tell me what is happening in this picture?" The next three prints from the ten were administered at the next baseline, keeping the same ten pictures in rotation for all subsequent baseline probes, treatment probes and follow-up probes. Ten pictures were chosen in order to decrease chance of a learning effect and since connected speech resulting from multiple

stimuli are said to be more representative of change due to treatment (Brookshire & Nicholas, 1994). No time limit was given for responses.

Intervention.

Dyads were created by matching people of comparable aphasia severity according to performance on all pre-treatment testing measures. Four participants (two male, two female) received intensive CILT (CILT-I) and four participants (three male, one female) received distributed CILT (CILT-D). The treatment itself was identical.

Traditional CILT according to the protocol initially described by Pulvermüller and colleagues (2001) and with further refinement from Maher et al. (2006) was administered to both groups. Since the completion of this study, CILT has been described in even more detail and gesture restrictions have been further clarified (Difrancesco et al., 2012). The activity central to treatment is, essentially, the well-known "Go Fish" game in which one participant asks another for a card that matches one of those he has been dealt. If the person has the requested card, it is surrendered; if not, the requestor must "go fish" or draw from the deck. The activity continues until one player is holding no remaining unmatched cards. The player with the most pairs wins.

There are several levels of task difficulty as outlined by Maher et al. (2006). Level One required a single word response given a deck of high frequency words. Level Two was the same but required introduction of the carrier phrase, "John, do you have the..." Level Three required use of an adjective, "Do you have the green pear," and Level Four required the use of two adjectives, "Do you have the sliced, green pear?" Criterion was reached when both participants in a dyad achieved fluidity or approximately 80% accuracy at a level. Since the same stimuli were used for Level One and Level Two, these two levels could be trained simultaneously by setting different production targets for each individual.

Participants who received CILT-I attended treatment for a Session Duration of three-hours, Session Frequency of five days per week, for a Total Number of 10 sessions over a Total Intervention Duration of two weeks. After the first 90 minutes, they received a 10 minute break to stretch and have a snack. Treatment then continued for an additional 90 minutes. Those who received CILT-D participated in one-hour sessions, three days a week, for ten weeks. No breaks were provided within the 1-hour sessions. The latter dosage would be considered the more traditional treatment schedule. Both groups received a total of 30 hours of treatment. Card sets were created to include nouns of high and low frequency occurrence and items of varying number and color.

Central to CILT is the employment of *forced use* of the verbal modality and *restraint* of all communication modalities except for oral verbal language. All participants were required to produce and respond to verbal communication regularly throughout the session. Each was clear that the "rules" of the game required no use of alternative communicative modalities such as writing or gesture. Vague gesticulations accompanying verbal productions were accepted but gesturing as a means of communication was discouraged as outlined in the clarification of CILT methods (Difrancesco et al., 2012) . Shaping was also a component of treatment requiring increasingly more challenging linguistic goals. For example, the single word, "brush" or even an approximation such as /brə/ was acceptable in the beginning but with each success new goals were created toward the goal of a full sentence consisting of a carrier phrase plus the requested item, "Jen, do you have the paint brush?" Participants were instructed on individual linguistic targets (word approximation, single word or introduction of the carrier phrase) prior to each session and the clinician provided cueing as necessary in order that a correct response is elicited and avoiding the production of errors (errorless learning). This took no more than a minute prior

to the initiation of treatment each day and was usually the same as the day before, in which case no further instruction was provided. The clinician, a licensed SLP and the author of this manuscript, participated in game play and modeled expected requests and responses for each participant. Cardholders were provided for individuals with hemiplegia who could not hold at least five cards fanned out and for any other participant who chose to use one.

Treatment Stimuli.

Treatment stimuli consisted of 120 full color stimulus items per level which were divided into four 30-card decks. Word frequency data were derived from the MRC psycholinguistic database (Coltheart, 1981). This relatively large number of stimuli relative to those from other studies is based on evidence that the goal of treatment is not word learning but rather neuroplastic brain remodeling as has been documented following intensive aphasia treatments (e.g., Schlaug, Marchina, and Norton 2009; Crosson et al. 2009; Meinzer et al. 2004). Greater numbers of stimuli have been demonstrated to result in increased word learning with equal durability than shorter lists of stimuli for both individuals with severe and those with mild naming impairments (Snell, Sage, & Ralph, 2010).

Treatment Probes.

Discourse probes identical to those administered at baseline were also administered after every six hours of treatment in order, resulting in five probes per participant. Participants were scheduled to arrive 30 minutes early in order to complete testing prior to that day's treatment session. Treatment probes were also administered during post-treatment follow-up sessions.

Data Analysis.

Results of the WAB-AQ, the CADL-2 and the RCPM along with changes in discourse performance were each analyzed and described to assess each individual's response to treatment.

All discourse elicitation and standardized assessment administration were digitally video-recorded. Discourse measures were then transcribed verbatim and analyzed for CIU count by the author, according to the procedure developed by Nicholas and Brookshire (1993). CIUs are words and intelligible paraphasias that are relevant to the picture being described. Words do not need to be used in a grammatically correct manner in order to be included in the CIU count but if they did not accurately describe the picture, they were not counted. For example, if the picture was of a boy falling off a stool and the participant said "She is falling off the chair," no credit would be given for "she" or for "chair." False starts, revisions and extraneous commentary such as "I don't know how to say it but," were also not included. CIUs provide a measure of productivity which is important for some participants. For others, efficiency of verbal production was more relevant. For this, CIUs per minute and the CIUs as a proportion of total word count (WC) were calculated. All three measures were calculated for each participant.

Ten percent of the transcripts were re-analyzed by the author and also her academic advisor for reliability of CIU counts. Both inter- and intra-rater reliability calculations were generated six months after initial counts were made. Point to point intra-rater agreement of 95.7% was performed by the author. Point to point inter-rater agreement between the author and her advisor was 91.3% and differences were resolved by discussion so that final agreement was 100%. The CIU/min is a calculated measure combining CIU count and time. Its reliability is affected by the reliability of the CIU count discussed above.

Effect sizes (ES) were calculated in order to avoid the Type I error that often occurs with visual inspection alone (Beeson & Robey, 2006). The d statistic was calculated as described by Busk and Serlin (1992, pp.197-198) by subtracting the mean of the baseline probes from the mean of the two final probe scores and dividing the result by the standard deviation of the

baseline scores. In the calculation of Total CIUs for participant I4, the first treatment probe was included in the baseline mean due to no baseline variability for this participant. Strength of effect benchmarks (large=10, moderate= 7, small=4) were based on the reports of Beeson and Robey (2006). Effect sizes are strongly influenced by baseline variability.

Results

Due to the heterogeneity of the participants, results are interpreted individually, each participant acting as his or her own control. I1, I2, I3 and I4 received 30 hours over two weeks and are described first. D1, D2, D3 and D4 received 30 hours of treatment over 10 weeks and are described next.

I1.

Treatment Performance.

Of the eight participants, I1 was the least motivated and often arrived late to treatment sessions. Despite this, I1 progressed through Level One by the end of the first week. He was producing full carrier phrases plus a high frequency word (Level Two) with only minimal cueing needed to initiate the carrier phrase by the end of the treatment duration.

Standardized Tests.

Performance on each of the standardized tests appear in Table 2. Pre-treatment, I2 scored near ceiling levels on the CADL demonstrating good use of residual language and functional communication. His scores on the RCPM were also high pre-treatment. WAB AQ scores were high in auditory comprehension subtests and lower for oral verbal language production. The selected subtests in Table 2 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, I2 made an 8.38 point change on the WAB AQ attributable to naming and also to word repetition. Smaller gains were also

observed on both the CADL and the RCPM. I1 did not return for follow-up testing, thus treatment maintenance was not assessed for this participant.

Table 2.

I1- Summary of Assessment Scores at each testing period

Assessment	Pre-treatment	Post-treatment	Pre-post-treatment change	Follow-up 1 month post-treatment
CADL	90.00%	96.00%	6.00%	N/A
RCPM	89.00%	92.00%	3.00%	N/A
WAB AQ	67.72%	76.10%	8.38%	N/A
WAB AQ Subtests				
Yes/ No Questions	100.00%	95.00%	-5.00%	N/A
Auditory Word Recognition	100.00%	100.00%	0.00%	N/A
Sequential Commands	90.00%	100.00%	10.00%	N/A
Object naming	88.33%	100.00%	11.67%	N/A
Fluency	30.00%	50.00%	20.00%	N/A

Note. I1 did not return for the one month post follow-up assessment.

Probes of Generalization to Connected Speech.

Visual inspection reveals a slight upward trend of CIUs as a proportion of total words from pre-treatment to one week post treatment (see Figure 1). Effect sizes were minimal or non existant on all measures. Of note, probes of generalization took place each morning at 9:00 am and I1 often reported being up late the night before with friends. This may have had an effect on performance on this measure.

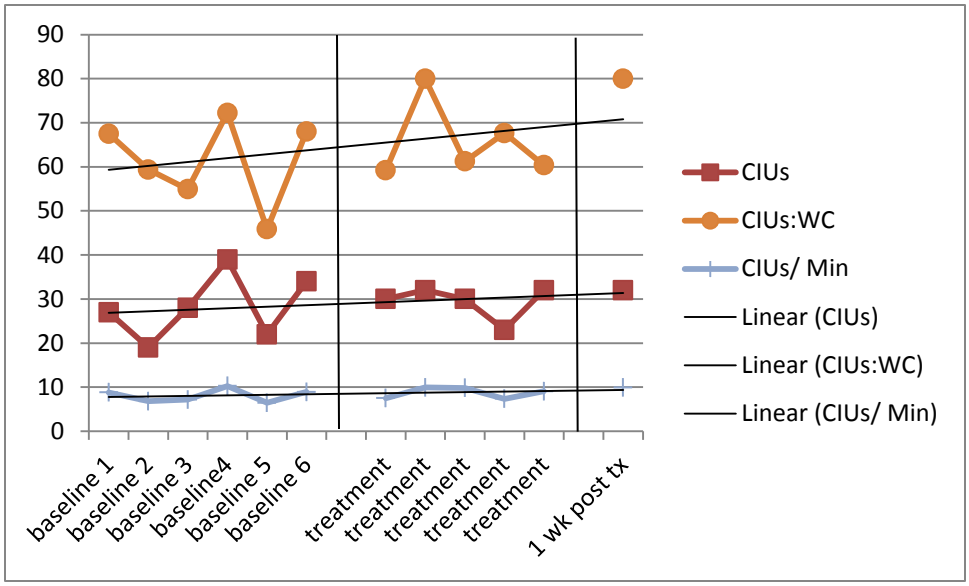


Figure 1-I1- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: 0.72, none; 1.5 none; 2.2, none.

I2

Treatment Performance.

I2 was highly motivated to improve verbal production. Due to severe AOS, this participant relied on writing to communicate making treatment sessions all the more challening for her. I2 was not able to produce or repeat a single word at the start of treatment. By the end she could name approximately 20 words but required cueing, including reminders of articulator placement. She never achieved criteria (80% accuracy) for Level One.

Standardized Tests.

Performance on each of the standardized tests appear in Table 3. Pre-treatment, I2 scored near ceiling levels on the CADL demonstrating good use of functional communication despite almost no oral verbal language. She effectively used writing and gesture to communicate. Her scores on the RCPM were also high pre-treatment, higher than any of the eight individuals

participating in the study. WAB AQ scores were moderately high in auditory comprehension subtests and very low for oral verbal language production, consistent with her severe AOS. The subtests shown in Table 3 are those for which the greatest change was observed among those who demonstrated change compared to the other subtests. Following thirty hours of treatment, I2 made a 7.8 point change on the WAB AQ attributable to auditory comprehension subtests. No gains were observed on either the CADL or RCPM immediately post treatment. Small gains were seen on the CADL and the WAB AQ at follow-up testing one month post treatment. Again, gains were most marked in auditory comprehension measures but there was a 6.7% gain in object naming at this time point as well.

Table 3.

I2- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post-treatment change	Follow-up 1 month post-treatment
CADL	90.00%	90.00%	0.00%	96.00%
RCPM	97.00%	95.00%	-2.00%	97.00%
WAB AQ	24.80%	32.60%	7.80%	33.00%
WAB AQ Subtests				
Yes/ No Questions	90.00%	90.00%	0.00%	85.00%
Auditory Word Recognition	81.67%	86.67%	5.00%	95.00%
Sequential Commands	43.75%	45.00%	1.25%	61.25%
Object naming	3.33%	3.33%	0.00%	10.00%
Fluency	0.00%	0.00%	0.00%	0.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech

Visual inspection shows a slight increase in slope for productivity, the primary variable of interest for this participant and effect sizes for this measure were large (8.0) (see Figure 2). It should be noted that productivity was at its peak one week post treatment and this was not maintained at the one month post follow-up, unlike standardized measures. Proportion of CIUs to total words increased most significantly as repeated single words (this, this, this) were replaced with some content words. There was no effect size for the efficiency measure of CIUs/ minute as was expected for this participant.

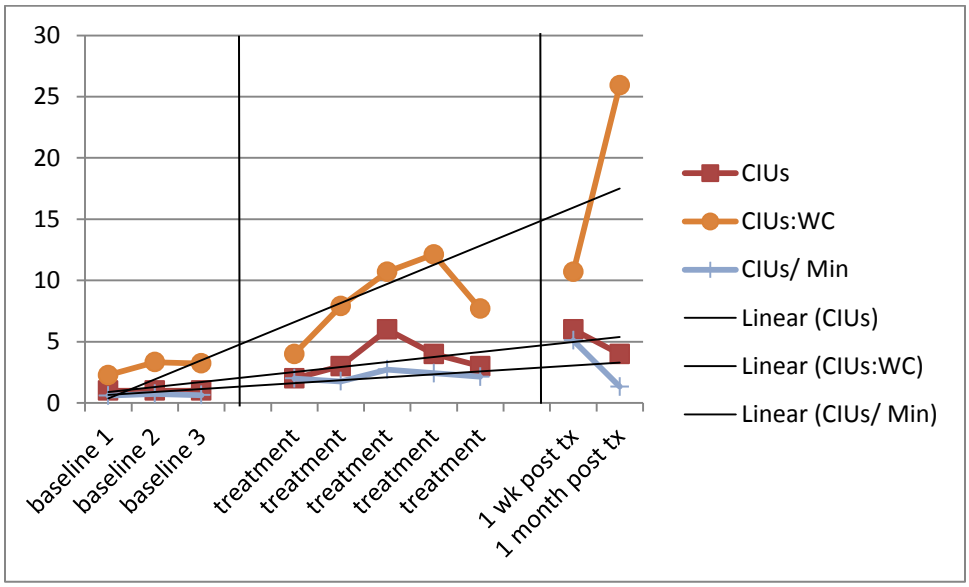


Figure 2. I2- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: 8.0, large; 26.3, large; 1.5, none

I3
Treatment Performance

I3 presented with moderate-severe AOS though in this case, AOS was more difficult to diagnose due to severity of aphasia in which all communicative modalities very severely impaired. This participant made greater gains in treatment performance than any other participant despite the fact that he was also the one furthest post stroke (> 11 years). He could not name one item prior

to treatment and by the end could name most of the trained items with a minimal visual or phonemic cue and could name 30 independently.

Standardized Tests

Performance on each of the standardized tests appears in Table 4. Pre-treatment, I3 scored in the 40th percentile on the CADL. Many errors were judged to be a result of auditory comprehension deficits. Scores on the RCPM were in the 60th percentile. This test requires no auditory comprehension component therefore all errors were due to difficulty in observing the patterns on this test. An initial WAB AQ score of 32 was comprised of auditory comprehension subtest scores that declined as complexity increased and generally low scores on oral verbal language production subtests. The selected subtests in Table 4 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, I3 made a 14 point change on the WAB AQ with gains in several areas but most pronounced on auditory comprehension subtests. A 25% gain was also observed on the CADL and 5% on the RCPM immediately post treatment. As with his Treatment Performance, I3's gains on standardized tests exceeded that of all other participants. All gains were maintained at follow-up testing one month post treatment and I3 demonstrated increased gains on object naming subtests.

Table 4.

I3- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post-treatment change	Follow-up 1 mo. Post-treatment
CADL	40.00%	65.00%	25.00%	57.00%
RCPM	62.16%	67.57%	5.41%	67.57%
WAB AQ	32.30%	46.00%	13.70%	47.70%
		WAB AQ Subtests		
Yes/ No Questions	80.00%	90.00%	10.00%	90.00%
Auditory Word Recognition	45.00%	70.00%	25.00%	70.00%
Sequential Commands	20.00%	40.00%	20.00%	46.25%
Object naming	25.00%	36.67%	11.67%	51.67%
Fluency	10.00%	5.00%	-5.00%	10.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech

Visual inspection shows a slight increase in slope for productivity, the primary variable of interest for this participant and effect size for this measure was small (3.0) (see Figure 3). Proportion of CIUs to total words increased most significantly as repeated single words (here, here, here) were replaced with some content words, though baseline variability for this measure was too great to yield any effect size. There was no effect size for the efficiency measure of CIUs/ minute as was expected for this participant.

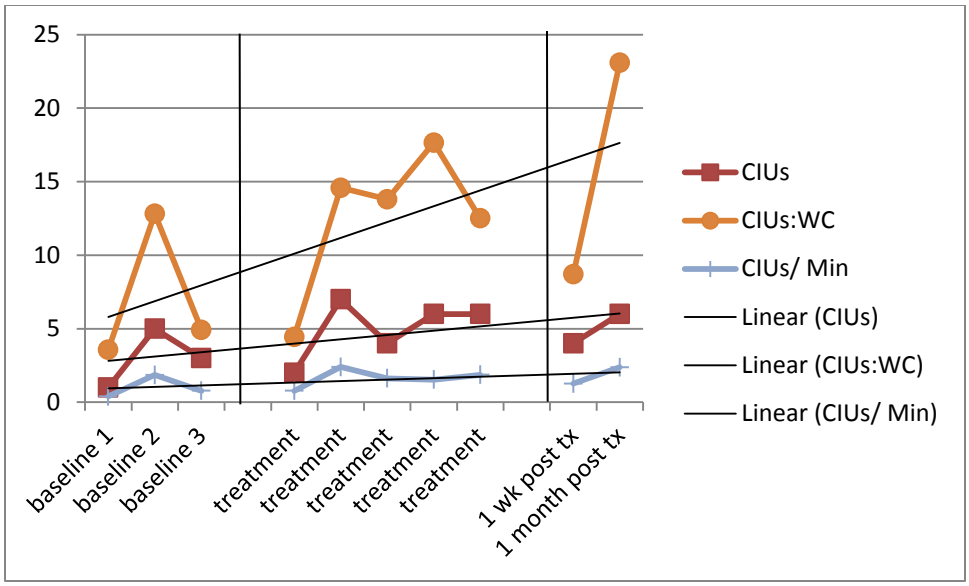


Figure 3. I3- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: 1.5, none; 3.2, small; 1.8, none.

I4

Treatment Performance

I4 presented with more severe aphasia deficits than any of the other participants in both expressive and receptive language. Though initially very motivated and upbeat, she struggled through the treatment sessions and became very frustrated by the end of the two weeks, having made very little progress. Like I3 (who was paired with I4 for treatment), she could not name one item prior to treatment but she did not demonstrate the same gains and all words had to be cued or repeated in order for production. I3 could name five words independently at the end of two weeks.

Standardized Tests

Performance on each of the standardized tests appears in Table 5. Pre-treatment, I4 scored 8% on the CADL as auditory deficits precluded understanding of most of what was presented in this test. Scores on the RCPM were in the 48.65%. This test requires no auditory

comprehension component therefore all errors were due to difficulty in observing the patterns on this test. An initial WAB AQ score of 27.4 was comprised of auditory comprehension subtest scores that declined as complexity increased and generally low scores on oral verbal language production subtests. The selected subtests in Table 5 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, I4 made a 3.3 point change on the WAB AQ which would not be considered clinically significant; however the 30% gain in auditory word recognition was notable. An 11% gain was also observed on the CADL and a decrease of 5.4 % was observed on the RCPM immediately post treatment. This loss was recovered and all gains were maintained at follow-up testing one month post treatment. Like I2 and I3, I4 also demonstrated increased gains on object naming subtests at this time point. *In the case of all three participants, the gains from pre-treatment to follow-up treatment exceeded those observed from pre-treatment to immediately follow-up treatment.*

Table 5.

I4- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post-treatment change	Follow-up 1 month post-treatment
CADL	8.00%	19.00%	11.00%	21.00%
RCPM	48.65%	43.24%	-5.41%	48.65%
WAB AQ	27.40%	30.70%	3.30%	32.10%
		WAB AQ Subtests		
Yes/ No Questions	75.00%	70.00%	-5.00%	65.00%
Auditory Word Recognition	11.67%	41.67%	30.00%	43.33%
Sequential Commands	8.75%	31.25%	22.50%	32.50%
Object naming	21.67%	23.33%	1.67%	30.00%
Fluency	10.00%	15.00%	5.00%	10.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech.

Visual inspection shows a consistent increase in slope for productivity, the primary variable of interest for this participant and effect size for this measure was large (9.1) (see Figure 4). Proportion of CIUs to total words increased most significantly as repeated single were replaced with some content words, though baseline variability for this measure was too great to yield any effect size. There was a small-moderate effect size (5.7) for the efficiency measure of CIUs/ minute. Productivity gains began to decay one-week post treatment and continued to show decline at one-month post treatment.

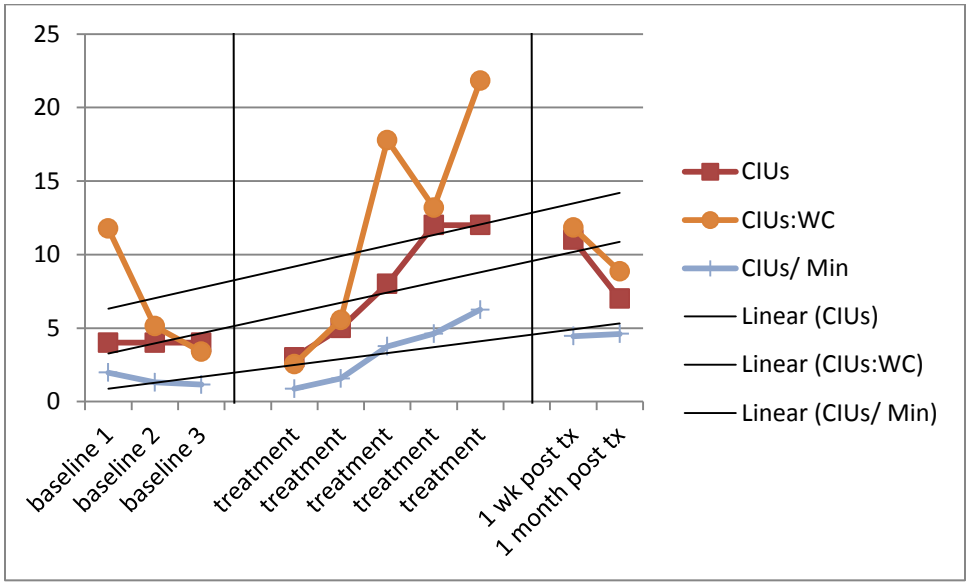


Figure 4. I4- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: 9.2, large; .81, none; 5.7, small-moderate.

D1

Treatment Performance.

D-I was also severely impaired in both expressive and receptive language. D1 participated willingly but demonstrated some complacency, seeming to have little expectation for

progress. Though errorless learning was emphasized in this treatment for all participants, D-I demonstrated consistent impulsivity and was unable to wait for cues before producing incorrect responses. The clinician instituted a hand signal to alert him when it was his turn to talk but this was only mildly effective. Perhaps as a result of this, D-I made little progress in treatment progressing from independent production of two words to eight by the end of the ten week treatment duration.

Standardized Tests

Performance on each of the standardized tests appears in Table 6. Pre-treatment, D1 scored 26% on the CADL, demonstrating good use of gesture to convey some answers though auditory comprehension was again a barrier for success on this test. Scores on the RCPM were in the 75.7%. This test requires no auditory comprehension component therefore all errors were due to difficulty in observing the patterns on this test. An initial WAB AQ score of 28.9 was comprised of auditory comprehension subtest scores that declined as complexity increased and generally low scores on oral verbal language production subtests. The selected subtests in Table 6 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, D1 made a 2.1% change on the WAB AQ which would not be considered clinically significant. A 10% and 20% increase were noted on fluency and on yes/no questions respectively. Greater than 10% decreases were noted on the RCPM and on auditory sequencing. Other measures tended to be nearly unchanged post treatment. Losses tended to be recovered at the one-month follow-up period with increases observed in fluency (5%) and object naming (10%). An additional 10% decline was observed in auditory word recognition at this time point.

Table 6.

D1- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post- treatment change	Follow-up 1 month post-treatment
CADL	26.00%	29.00%	3.00%	23.00%
RCPM	75.68%	64.86%	-10.81%	78.35%
WAB AQ	28.90%	31.00%	2.10%	34.70%
WAB AQ Subtests				
Yes/ No Questions	70.00%	90.00%	20.00%	75.00%
Auditory Word Recognition	60.00%	58.33%	-1.67%	48.33%
Sequential Commands	46.25%	23.75%	-22.50%	48.75%
Object naming	11.67%	15.00%	3.33%	25.00%
Fluency	0.00%	10.00%	10.00%	15.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech

Despite lack of progress during treatment, some mild improvement was noticed in this participant's productivity and efficiency of language as shown in Figure 5. Visual inspection shows a consistent increase in slope for all three measures of discourse. Gains in all three measures began to decay following treatment with continued drop off at one month post treatment, resulting in negligible effect sizes.

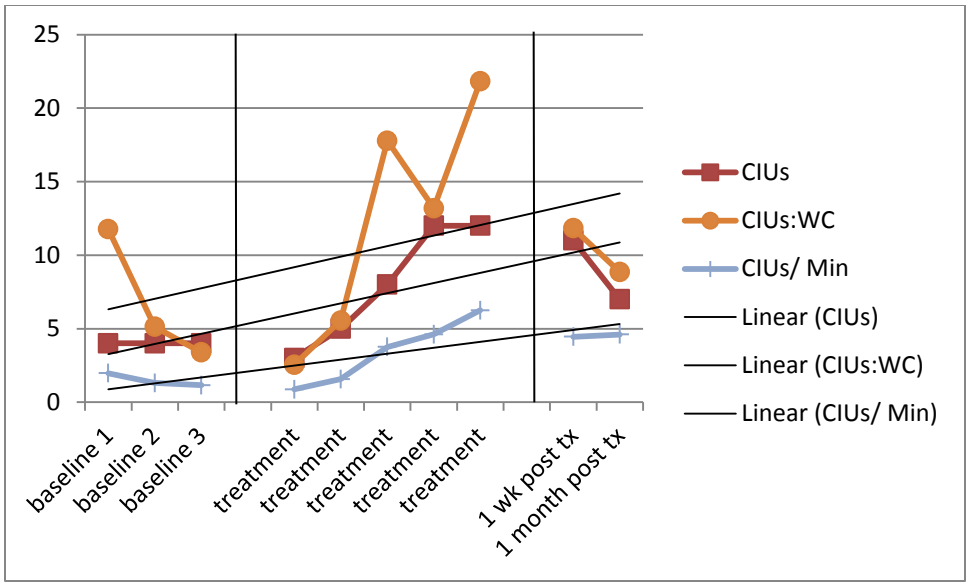


Figure 5. D1- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: .58, none; .97, none; 2.7, none.

D2

Treatment Performance.

D2 put forth maximal effort during all treatment sessions, was responsive to cueing and made slow but incremental progress throughout treatment. Though he remained at Level One, he was independently naming 22 words by the end of treatment compared to five words on day one of treatment.

Standardized Tests.

Performance on each of the standardized tests appears in Table 7. Pre-treatment, D2 scored 35% on the CADL, demonstrating general confusion with how to answer questions despite several attempts to model the expected response. This participant lived in a situation where he rarely interacted with other people since the time of his stroke suggesting a lack of exposure may have contributed to initial low score on this measure. Scores on the RCPM were relatively high at 89.2%. An initial WAB AQ score of 50.1 revealed generally intact oral verbal

comprehension for simple yes/no questions and at the word level and obvious breakdown at the sentence level. The selected subtests in Table 7 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, D2 made a 8.6% change on the WAB AQ which is considered clinically significant. A 10% and 16% increase were noted on fluency and object naming subtests, respectively. Of significance was a 43% gain on the CADL immediately post treatment with an additional 11% gain on this measure at one month post treatment. No other participant in the study demonstrated this large a gain on any test. Other measures tended to be nearly unchanged post treatment. Gains in object naming decreased one month post treatment but were increased compared to pre-treatment.

Table 7.

D2- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post- treatment change	Follow-up 1 month post-treatment
CADL	35.00%	78.00%	43.00%	89.00%
RCPM	89.19%	91.89%	2.70%	91.89%
WAB AQ	50.10%	58.70%	8.60%	61.60%
		WAB AQ Subtests		
Yes/ No Questions	95.00%	95.00%	0.00%	90.00%
Auditory Word Recognition	95.00%	95.00%	0.00%	93.33%
Sequential Commands	73.75%	66.25%	-7.50%	70.00%
Object naming	66.67%	83.33%	16.67%	78.33%
Fluency	10.00%	20.00%	10.00%	20.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech

D2 demonstrated improvement in productivity, the primary variable of interest for this participant, as shown in Figure 6. Visual inspection shows a consistent increase in slope for this measure that dips only slightly at one week and one month follow-up testing yielding an effect

size of 4.0 (small). Although a small effect size was calculated for words/minute (5.9), this is not supported by visual inspection and is likely increased due to such minute variability in baseline. Negligible effect size was seen for the efficiency measure of CIUs as a proportion of total words. Efficiency was not a variable of interest for this participant.

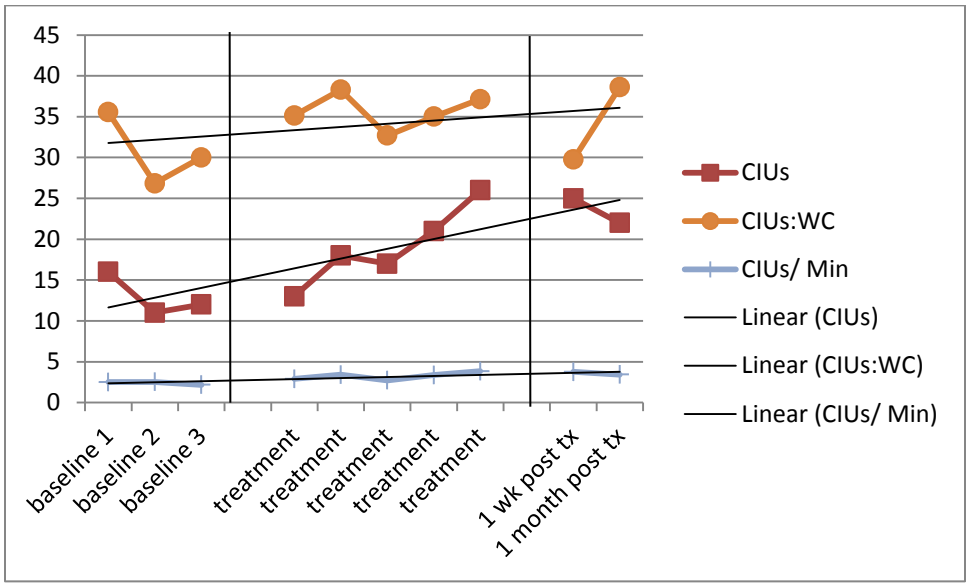


Figure 6. D2- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: 4.0, small; 2.4, none; 5.9, small-medium.

D3

Treatment Performance

Of all eight participants, D3 was the least impaired in expressive language. She began at Level Three (carrier phrase plus low frequency word) and quickly increased to Level Five (carrier phrase plus object requiring two adjectives. For example "Jen, pass me the sliced green pear" when there are also cards with two green pears, a single green pear and pears of other color and number within the same deck.) Though her expressive language exceeded the other participant in the group (D4), she found it more difficult to keep track of who was holding the card of interest and therefore rarely "won" a game in the first few weeks. Attention and memory

were not tested prior to treatment but appeared to be an area of deficit for this participant based on her game performance. She progressed throughout the treatment period but still won less than 25% of the time.

Standardized Tests

Performance on each of the standardized tests appears in Table 8. Pre-treatment, D3 scored 81% on the CADL, demonstrating general aptitude with using residual language for functional communication. Scores on the RCPM were relatively high at 89.2%. An initial WAB AQ score of 84.2 revealed generally mild deficits in both expressive and receptive language. The selected subtests in Table 8 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, D3 did make changes of greater than 10% on some subtests of the WAB but gains and declines were about equivalent resulting in little overall change. It is not clear why a 14% decline was also noted on the CADL. This change was largely reversed at the one month post treatment follow-up. A 5% decline was also noted on the WAB AQ which is worth highlighting since equal increases are considered significant. This change was characterized by a decline in the sequential commands subtest as well as lesser declines in score in fluency and in naming. Overall, D3's performance at one month post treatment is slightly less than was observed pre-treatment introducing the possibility of this treatment having had a potential negative effect on performance.

Table 8.

D3- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post- treatment change	Follow-up 1 month post-treatment
CADL	81.00%	67.00%	-14.00%	78.00%
RCPM	89.19%	81.08%	-8.11%	83.80%
WAB AQ	84.20%	83.90%	-0.30%	78.10%
		WAB AQ Subtests		
Yes/ No Questions	85.00%	95.00%	10.00%	90.00%
Auditory Word Recognition	93.33%	98.33%	5.00%	98.33%
Sequential Commands	83.75%	81.25%	-2.50%	60.00%
Object naming	100.00%	88.33%	-11.67%	85.00%
Fluency	60.00%	80.00%	20.00%	45.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech.

D3 demonstrated improvement in productivity (see Figure 7; small effect size of 3.5) but this was not a primary variable of interest for her since she was productive pre-treatment, just inefficient with discourse characterized by long pauses, many filler words and phrases such as "ah, um, you know, I don' t know" and circuitous language. The primary variable of interest for this participant was efficiency and D3 did show a small effect size for CIUs per minute. Producing more content words per minute means she was using fewer fillers and choosing more appropriate words-- a good outcome for D3, despite her performance on standardized tests. Negligible effect size was seen for the efficiency measure of CIUs as a proportion of total words.

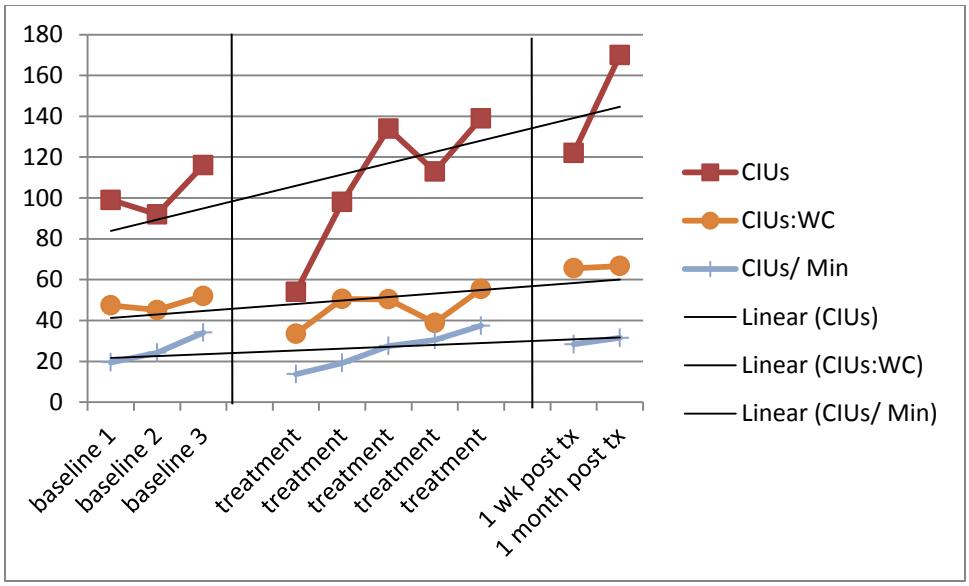


Figure 7. D3- Narative Discourse Probes- Productivity and Efficiency. Effect sizes for total CIUs, the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: 3.5, small; 5.09, small; .53, none.

D4

Treatment Performance.

It was not possible to assist D4 in order to achieve errorless production. Phonemic and semantic cueing, even repetition failed in assisting this participant. Given enough time, he could usually produce the phrase but like D2, he was not able to benefit from the errorless production aspired to with all participants. Despite this, he made clear progress within treatment sessions beginning at Level Three (carrier phrase plus low frequency word) and quickly increased to Level Five along with D3(carrier phrase plus object requiring two adjectives)

Standardized Tests.

Performance on each of the standardized tests appears in Table 9. Pre-treatment, D4 scored 77% on the CADL, demonstrating general aptitude with using residual language for functional communication. Score on the RCPM was 75.6%. An initial WAB AQ score of 73.6 revealed deficits understanding complex sentences (sequential commands) and on several

expressive language subtests. The selected subtests in Table 9 are those for which the greatest change was observed among those who demonstrated change. Following thirty hours of treatment, D3 did not demonstrate a change in overall AQ score but did increase by 10% on the sequential commands subtest. Like D3, he decreased by 12 points on the CADL post treatment; but unlike D3, this decrease was maintained at the one month post follow-up. Overall, D4's performance at one month post treatment was identical to that observed pre-treatment on the WAB AQ and the RCPM.

Table 9.

D4- Summary of Assessment Scores at each testing period.

Assessment	Pre-treatment	Post-treatment	Pre-post- treatment change	Follow-up 1 month post-treatment
CADL	77.00%	65.00%	-12.00%	62.00%
RCPM	75.68%	81.08%	5.41%	83.78%
WAB AQ	73.60%	74.70%	1.10%	73.00%
		WAB AQ Subtests		
Yes/ No Questions	90.00%	85.00%	-5.00%	85.00%
Auditory Word Recognition	98.33%	96.67%	-1.67%	98.33%
Sequential Commands	53.75%	65.00%	11.25%	62.50%
Object naming	76.67%	85.00%	8.33%	75.00%
Fluency	20.00%	15.00%	-5.00%	15.00%

Note. CADL- Communication Activities of Daily Living; RCPM- Raven's Coloured Progressive Matrices; WAB AQ- Western Aphasia Battery Aphasia Quotient

Probes of Generalization to Connected Speech

Two figures are used to depict performance on the three discourse outcome measures for D4 since his productivity was very high and required a different scale than that used for the efficiency measures (see Figures 8 and 9). D4 was overly productive, as he discovered a strategy by which extensive circumlocution often, eventually, helped to achieving the point or the word

he was working toward. Therefore, efficiency was the outcome measure of interest (CIUs/min) for this participant. Effect sizes were negligible, however.

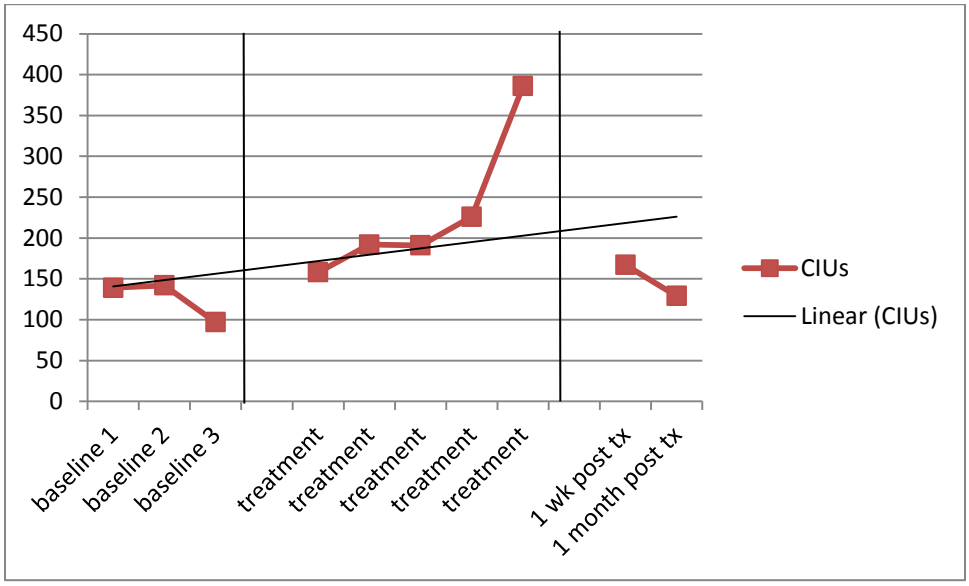


Figure 8. D4- Narative Discourse Probes- Productivity. Effect sizes for total CIUs was .87, none.

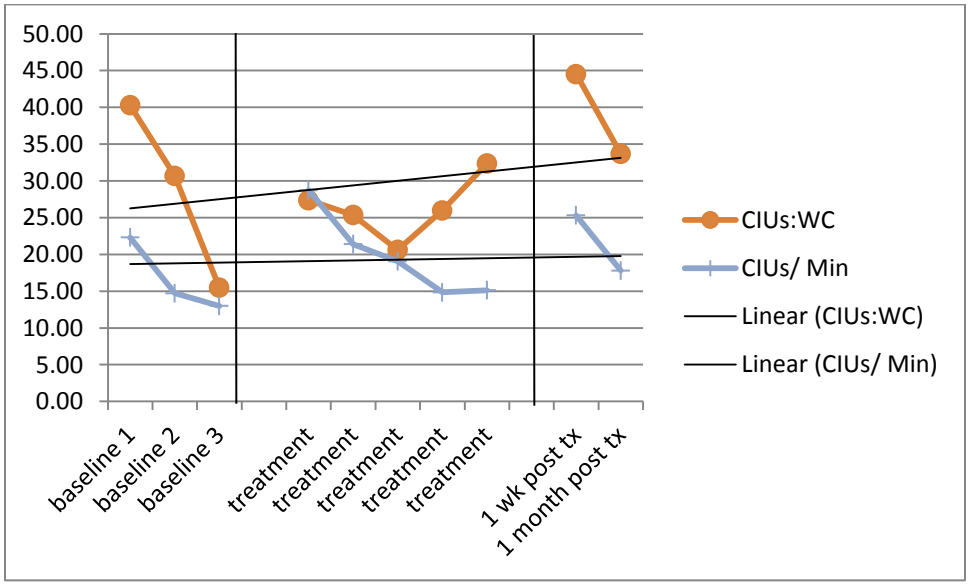


Figure 9. D4- Narative Discourse Probes- Efficiency. Effect sizes for the proportion of CIUs to total word count (CIUs:WC) and CIUs per min (CIUs/min) were as follows: .82, none; 2.32, small.

Summary of Results.

Seven of the eight participants attended all 30 hours of treatment, before-treatment session Treatment Probes, post-treatment testing and follow-up testing. Follow-up data were obtained at one month post- treatment. Participant I1 was often 20-30 minutes late for sessions and did not return for one-month follow-up testing. Most participants made gains on either one of the standardized measures or on the primary outcome variable of interest in this study—that is the generalization of treatment to either discourse productivity or efficiency, depending on pre-treatment discourse patterns (see Table 10).

Table 10

Summary table of performance on standardized batteries and discourse measures

ID	Aphasia severity	WAB-AQ	R-CPM	CADL	CIUs	CIUs/min	CIUs:WC
I1	mild-mod	↑	↑	↑	—	—	—
I2	mild-mod	↑	—	↑	↑	—	↑
I3	severe	↑	↑	↑	—	—	—
I4	severe	↑	—	↑	↑	↑	—
D1	severe	↑	—	—	—	—	—
D2	severe	↑	—	↑	↑	—	↑
D3	mild-mod	↓	↓	↓	↑	—	↑
D4	mild-mod	—	↑	↓	—	—	—

Note. Increases and decreases refer to a greater than 5% change using from pre-treatment to post-treatment or from pre-treatment to follow-up. The greatest positive value or the greatest negative value depending on whether scores tended to trend up or down.

Treatment performance.

All eight participants were fully engaged throughout all 30 hours of treatment and improved on trained materials to varying degrees. For those with more moderate to severe aphasia deficits (I3, I4, D1, D2) treatment progress was seen within Level One. For the participants with severe AOS (I2 and D1) very little treatment progress was observed. I2

advanced to Level Two with the addition of the carrier phrase, and D3 and D4 to Level Five, requiring two adjectives and a noun.

Standardized measures.

All four of the participants who received CILT-I and two who received CILT-D demonstrated a greater than five point gain on the WAB AQ, either at post-treatment testing or at follow-up testing. The object naming subtest of the WAB was examined for treatment generalization to untrained words. All but one participant (D3) had increased object naming scores in post-treatment tests. I2, I3, I4 and D1 demonstrated their largest gains on this subtest at the one-month follow-up. I1 did not return for follow-up testing. WAB AQ gains were maintained for five of the seven participants who returned for follow-up testing one-month post-treatment.

Three participants (I3, I4 and D2) demonstrated an increase of two standard deviations on the CADL-2. Two who did not show demonstrable gains were those whose pre-treatment scores were at or close to ceiling (I1 and I2) and two demonstrated decreases, one of which persisted at the one-month follow-up (D4). Follow-up data showed that gains were maintained on this measure and, for I4 and D2, continued to increase at the one-month follow-up.

Probes of generalization to connected speech

Productivity and efficiency of discourse were measured for all eight participants over time and both visual inspection and calculation of effect sizes were used to gauge responsiveness to treatment. Increased productivity indicated more words that were directly relevant to the pictures being described. Increased efficiency indicated increased self-monitoring resulting in fewer repeated words, false starts, irrelevant and filler words.

I2, I4 and D2 each demonstrated an effect of treatment in their target area of either productivity, efficiency or both. D3 demonstrated an effect on productivity but this was not the target for this participant whose speech was fluent and circuitous. Calculated effect sizes for participant I3 yielded numbers just below the benchmark for "small" effect due to too much baseline variability but there was a clear upward trend in productivity. Effects were generally maintained and in all cases final data points exceeded pre-treatment scores.

Discussion

This study was a preliminary investigation of individual responses to CILT delivered at two dosages aiming to assess response differences due to the contribution of intensity to this treatment. Although positive gains tended to be the case in both groups, results were somewhat stronger and maintained somewhat better for those who received the intensive dose in that all four who received CILT-I made gains on the WAB AQ (including a 15 point gain from I2 who was more than 11 years post stroke) but only two of the four who received CILT-D showed an equal response. Two of the four who received CILT-I and two of the four who received CILT-D made gains in their target area of narrative discourse improvement but visual inspection shows better maintenance at the one-month follow-up for those who received CILT-I. All gains noted were in line with previous studies that have demonstrated consistent patterns of improvement following CILT administered at this dosage (e.g., Kurland et al., 2012; Maher et al., 2006; Pulvermüller et al., 2001).

The prediction that positive outcomes would not be seen with CILT-D did not bear out, however, as positive gains were observed for some in this group as well. For one (D2), outcomes equaled or exceeded those who received CILT-I. D2's most notable change was on the CADL with gains on this measure above those demonstrated by any other participant. It is

possible that the distributed treatment promoted growth in functional communication areas targeted in this measure. The greater number of sessions meant more times interacting with others in the waiting room and perhaps increasing the amount of social exposure was important for this participant who reported being confined to his home except for his trips to participate in the treatment study. This benefit of distributed treatment programs should not be overlooked when assessing dosage. This would appear to be more critical for those immediately post stroke when visits to their speech pathologist may be their only form of social interaction, depending on premorbid social proclivity as well as the family's ability to support participation in other social endeavors.

D3 also demonstrated positive outcome. Unlike D2, this did not extend to all measures but she did improve and maintain effect of treatment in discourse efficiency. Although greater response was expected for those who received CILT-I, a *negative* response was not anticipated for those who received CILT-D and yet declines on the CADL were observed for D3 and D4. It is not clear how treatment could be responsible for this decline which was reversed at follow-up testing for D3 but not for D4.

The two participants with severe AOS (I2 and D1) made minimal progress within treatment sessions as well as and on most outcome measures of oral verbal production, but I2 (with participant with the more severe AOS of the two) did demonstrate change in other areas where D1 did not. I2 also demonstrated large effect sizes in productivity. Large effect size was seen due to a very stable baseline and a small number of new spontaneous word productions. Whether this progress translates to functional change is questionable and one might speculate that, given equal time and effort, this same participant may have been better served by a

treatment targeting her AOS such as Sound Production Treatment (Wambaugh, Martinez, McNeil, & Rogers, 1999).

Other participants who demonstrated relatively small increases in verbal production when assessing for functional value (I3, I4, D2) tended to make larger increases on oral verbal comprehension measures. These increases are thought to bestow greater benefit in contributing to enjoyment of activities of daily living as well as in contributing to potential future improvement if better comprehension precedes better production. Substantive increases on comprehension subtests have been reported in previous studies using CILT (e.g., Breier et al., 2009; Szaflarski et al., 2008). Though CILT is targeted at verbal production, successful game play requires careful attention to what other participants are producing. Auditory comprehension may warrant closer scrutiny in future examination of CILT application.

Seven of the eight participants explicitly indicated that they enjoyed the treatment. It is likely that camaraderie with other people with aphasia contributed to enjoyment, perhaps above the satisfaction of working hard to achieve a goal. The encouragement and support of other group members should be evaluated more closely in their contribution to treatment gains and to other post-stroke life achievements such as regaining a driver's license, transitioning to independent living or becoming involved again in previously enjoyed social activities.

Three of the four participants who received CILT-I and one of the four who received CILT-D increased naming by scores greater than what was seen immediately post treatment at the one-month post-treatment time. Continued increases once treatment has ceased have been observed in previous studies following intensive regimens such as CILT (e.g., Johnson et al., 2013; Szaflarski et al., 2008) and MIT (e.g., Schlaug et al., 2009) and are one indication that benefits of intensive treatments extend beyond the treatment itself. Although D2 participated in

CILT-D, he appears to have received equal benefit suggesting that optimal dose may vary depending on the individual; however, this is difficult to assess without knowing his response to CILT-I. Some studies have attempted to deal with inter-participant variability by exposing each individual to two successive treatments types using a cross-over single subject design (Kurland et al., 2012; Rose et al., 2013), however, this method is also limited in that order of treatment is thought to play a significant role in treatment response. Intensive therapies that are thought to provide a system boost to those with chronic aphasia. If so, it follows that the greatest gains would be observed after any treatment provided first. Perhaps extending the period between treatment administrations would allow this very useful cross-over design to be more illustrative in terms of treatment comparison.

Future Studies.

It is believed that the natural and dynamic nature of the CILT contributes to recovery. Treatment comparison studies include those that have involved card games (Barthel et al., 2008) or encourage verbal production given multimodal cueing (Rose et al., 2013). When administered intensively these treatments resulted in outcomes comparable to the results of CILT groups. Perhaps it is not the treatment at all, but simply the intensity of stimulation that is priming the system for change. If this is the case, the dosage must also be refined. Three hours per day for ten days has shown positive results but this is exhausting for the participant and possibly more than is needed. However, it is possible that it is exactly this pushing of limitations that is necessary to instantiate lasting change and thus even more might result in an even more positive outcome. On a related note, if a participant's performance does plateau after 10, 20 or 30 hours we do not yet know whether it is reasonable to assume the system has been "maxed out" or if a break followed by the resumption of treatment may result in further improvements.

As more success is observed with individuals and small groups using CI therapy principals, the gaps in our knowledge of the treatment of aphasia deficits become more apparent. Larger participant groups are necessary in order to best examine the duration and length of therapy in varying intensities. The ultimate goal is to make best recommendations for optimizing treatment hours given individual insurance limitations.

Chapter III

Background

Study One would benefit from replication with a greater number and, ideally, more homogeneous participants, however, preliminary data do suggest a contribution of treatment intensity to successful gains seen following CILT. Some participants also demonstrated gains following CILT-D, the distributed version of CILT, pointing to other factors contributing to success observed following this treatment.

The current study seeks to increase the benefits seen following CILT and other intensive treatments by increasing the dosage even more while programmatically assessing for language change and neural activation using multiple outcome measures. Variables important to an in-depth understanding of the chosen treatment and assessment are discussed below.

Intensive Language Treatment

Lack of consensus as to the efficacy of intensive language treatment was discussed previously (see Introduction to Study One, pp.8-10). Studies using intensive doses have tended to demonstrate positive language gains; however, when the treatment is controlled and lower doses are compared, the results don't always favor intensity. It is possible that the differences in intensity in those studies (e.g., Ramsberger & Marie, 2007; Snell et al., 2010) were not different enough to impact treatment effect for their participants. Also, as previously discussed, definition of what constitutes "intensive" treatment varies widely. For the purposes of this current study and in order that other studies may be compared appropriately, dosage parameters proposed by Warren et al. (2007) and modified by Chreney (2012) are used. These include Session Duration, Session Frequency, Intervention Duration and Number of Sessions. These parameters are often

not defined in aphasia treatment studies, including those purporting to provide high intensity treatment and may contribute to the equivocal results reported to date.

Extended, months-long Intervention Durations, consisting of daily Session Frequency and several hours-long Session Durations have been administered (Basso, 2001; Code, Torney, Gildea-Howardine, & Willmes, 2010; Mackenzie, 1991; Poeck, Huber, & Willmes, 1989) provide support that *unambiguously* high dosages of treatment have positive immediate post-treatment outcomes reflected in standardized language test scores and qualitative analysis. Recently, high dosage treatments of much shorter Intervention and Session Durations such as Constraint Induced Language Therapy (CILT; Pulvermüller et al., 2001)—and several variations thereof (Barthel, Meinzer, Djundja, & Rockstroh, 2008; Faroqi-Shah & Virion, 2009; Goral & Kempler, 2009; Maher et al., 2006; Mozeiko, Myers, & Coelho, 2011; Szaflarski et al., 2008)—have also shown positive effects and report maintenance of gains with fair consistency. CILT and others that utilize a "mass practice" approach provide high Session Frequency and long Session Durations over a relatively short Intervention Duration (one-three weeks). The results are often lasting language gains that in some instances continue to increase for two months after the completion of treatment (Barthel et al., 2008; Breier, Maher, Novak, & Papanicolaou, 2006; Maher et al., 2006; Meinzer, Djundja, Barthel, Elbert, & Rockstroh, 2005; Mozeiko, Myers, & Coelho, 2011).

Constraint Induced Language Therapy

CILT, described in detail in the previous chapter, was used again for the current study due to a) positive language gains for a range of aphasia types b) fair consistency within the literature in the dosage in which it tends to be administered allowing for comparison among

treatment studies c) logistic feasibility and d) its contribution to experimental control in so far as the participants within each dyad experience identical treatment conditions.

This treatment was designed to make use of two important principles of experience-dependent neural plasticity including Intensity of Treatment and Use It to Improve It. These are two of several fundamental experience-dependent training principles discussed by Raymer and colleagues (2008) in a narrative review that includes a summarization of basic science evidence relevant to aphasia treatment research. Intensity of Practice refers to the animal and motor literature which tends to result in neural changes as a result of increased repetition. Intensity of Practice is also reviewed as it relates to the learning literature showing potential benefit for more complex material. Use It to Improve It is based on the notion that the potential rehabilitation of oral verbal language is negatively impacted by the non-use of that modality of communication. In other words, forced use of the oral verbal modality, even when very impaired, will improve potential rehabilitation. For a more thorough description of experience-dependent training principles that influence neuroplasticity, the reader is directed to read Raymer and colleagues' (2008), Translational Research in Aphasia: From Neuroscience to Neurorehabilitation.

CILT is not unlike other evidence-based aphasia treatment programs that make use of principles guiding adaptive neuroplasticity. Several others also focus exclusively on improving the oral language modality. Limiting the use of other language modalities, such as written or gestured language, is less explicit in other protocols; however, there is a clear expectation of oral verbal language production when that is the modality being trained. For example, in Response Elaboration Therapy (RET; Kearns, 1985) no one is instructing the participant to avoid gesture or written communication, but verbal production is the expectation. CILT is also described as making use of an approach that involves shaping, scaffolding and reinforcement but these are

commonly used techniques in the treatment of aphasia, and within the field of Speech Language Therapy (SLT).

What makes CILT most attractive for study is that in each of the several studies in which it has been replicated, similar and well-defined dosage parameters have been reported (see Table 11). This makes CILT an ideal treatment option for which to begin an investigation of dosage parameters and potential change in effect when these parameters are manipulated.

Table 11

Dosage parameters reported in studies that have used CILT.

Study	Session Duration- *Hours*	Session Frequency- *Sessions/ Week*	Number of Sessions	Intervention Duration- *Weeks*	Total Intervention Dosage- *Hours*
		Chronic Aphasia			
Pulvermüller et al., 2001	3	5	10	2	30
Meinzer et al., 2004	3	5	10	2	30
Meinzer et al., 2005	3	5	10	2	30
Pulvermüller, Hauk, Zohsel, Neininger, & Mohr, 2005	3	5	10	2	30
Kempler, Goral, & Tison, 2006	1.25	5	20	4	25
Maher et al., 2006	3	4	8	2	24
Meinzer et al., 2006	3	5	10	2	30
Breier, Maher, Schmadeke, Hasan, & Papanicolaou, 2007	3	4	12	3	36
Meinzer, Elbert, Djundja, Taub, & Rockstroh, 2007	3	5	10	2	30
Barthel et al., 2008	3	5	10	2	30
Szaflarski et al., 2008	3 to 4	5	5	1	18
Meinzer et al., 2008	3	5	10	2	30
Richter, Miltner, & Straube, 2008	3	5	10	2	30
Szaflarski et al., 2008	3 to 4	5	5	1	15-20

Berthier et al., 2009	3	5	5	2	30
Breier et al., 2009	3	4	12	3	36
Goral & Kempler, 2009	1.25	5	20	4	25
Faroqi-Shah & Virion, 2009	3	4	8	2	24
Kirmess & Maher, 2010	1.5 to 3	5	10	2	15-20
Kurland, Baldwin, & Tauer, 2010	3 .	5	10	2	30
Kempler & Goral, 2011	NR	NR	NR	4	30
Kurland, Silva, Burke, & Iyer, 2011	NR	NR	NR	NR	NR
Breier, Juranek, & Papanicolaou, 2011	3	4	12	3	36
Mozeiko et al., 2011	3 .	5	10	2	30
Kurland, Pulvermüller, Silva, Burke, & Andrianopoulos, 2012	3 .	5	10	2	30
Johnson et al., 2013	3	5	15	3	45
Rose, Attard, Mok, Lanyon, & Foster, 2013	3.25	4	8	2	32
Acute and Subacute Aphasia					
Bakheit, Shaw, Carrington, & Griffiths, 2007	1	5	60	12	60
Kirmess & Maher, 2010	3	5	10	2	30
Sickert, Anders, Münte, & Sailer, 2014	2	5	15	3	30

Note. Studies are listed in order of publication within the population treated (chronic or acute and subacute). Dosage parameters (session frequency, etc.) used according to definitions provided by Warren (2007) and Cherney (2012). NR-not reported

Increasing Treatment Duration

The work discussed in the previous chapter contributes to the evidence that high session frequency and long session length are important elements of CILT. It has not yet been tested whether there is value in also increasing the Total Treatment Duration beyond two weeks. Programs in which a combination of therapies are administered—group, computer and individual—have resulted in positive changes on standardized language batteries for treatment durations spanning longer than one month. In addition, the effects of two different high dosage treatments on a single individual have been compared in recent studies (Kempler & Goral, 2011; Kurland et al., 2012; Rose et al., 2013). In each, two different treatments were administered in equal time blocks one after the other in order to compare effects. Often a continuation of positive effect is observed though the effect tends to be attenuated compared to that seen following the first treatment. Since results demonstrate that gains continue after a first dose, it is believed that a double treatment of a single treatment type will show enhanced positive effects.

Although the main benefit of a double dose of treatment is predicted to be an increase in gains on trained materials, there are other potential advantages. One is to determine whether a maximum treatment effect is reached during this period. The maximum treatment effect refers to each participant's maximum potential performance on trained items, given a particular treatment. For one person, the maximum treatment effect may be production of 50 novel adjective noun combinations with consistent improvement over time; for another 10 new word productions a quarter of the way through treatment may constitute maximum effect. Various patient populations will likely respond differently and within different time frames to various treatments. Ideally, a double dose would result in continued language gains for the entirety of treatment but it is also possible that performance will plateau for long periods for those with chronic aphasia.

In either case, observation of peak performance is more likely given two additional weeks of training. Information about time of maximum treatment can help to determine optimal dosages in the future.

Another potential benefit of a double dose of treatment is better maintenance or durability of the newly learned behavior due to additional practice. Overlearning refers to the concept that practice extending beyond the point of mastery can lead to task automaticity. The effect of hundreds of task repetitions to achieve automaticity has been well documented in the motor literature (Kleim & Jones, 2008). Overlearning in the aphasia literature is not as well-documented. An fMRI study of a participant with aphasia with alexia demonstrated a clear left hemisphere (LH) shift in laterality when comparing mastery of reading word items (95%) to a two week overlearning period of these same words (Kurland et al., 2007). Left laterality—greater left than right hemisphere (RH) activation— is typical for language function in the noninjured brain. An increase in left laterality indicates either a decrease in activity in the RH, increased activity in the LH or both. This study suggests that neural change is, potentially, a better indicator than behavioral data as to when treatment has ceased to be effective. If true, it would be unwise to discharge patients from service when behavioral changes have plateaued, as is the current practice. Overlearning has positive implications for generalizability to both untrained materials as well as to settings outside the clinic as these behaviors are more likely to be those that are easiest to perform without assistance. In the case of word learning, overlearned nouns are likely to be attempted outside the clinic; therefore they will be reinforced in more natural settings and thus more likely to be assimilated into the language repertoire.

Generalization of Target Behavior

Despite the fact that generalization has been deemed the "gold standard in treatment research" (Thompson & Shapiro, 2007, p. 37) its appearance in aphasia treatment studies remains inconsistent. Generalization refers to a) a transfer of skills to environments outside the clinic setting and b) improvements in behaviors not targeted during treatment. Home practice that reinforces skills learned during treatment are considered one of the most important protocol components for inducing both behavioral and neuroplastic changes after Constraint Induced Motor Therapy (CIMT; Morris & Bickel, 2011). CILT is modeled after CIMT yet current studies using CILT rarely describe any transition of practice to home leaving the possibility of generalization to chance. In one CILT study, Meinzer and colleagues (2005) did incorporate a home practice regimen (CIAT[1] plus) and reported greater maintenance of gains and greater family and self-reported communicative effectiveness compared to CILT without the home practice regimen at a six month post-treatment follow-up. Their home component consisted of daily communication practice with a family member and an individually defined interaction with a non-family member (e.g., asking for a loaf of bread at the bakery). Due to differences in participants' support systems and home environments, this was not incorporated into the current study. Therefore, any generalization observed could be attributed to how the participant responded to the treatment alone.

Although CILT is considered to be a treatment of verbal expression, there is evidence of transfer to untrained verbal comprehension after treatment. In studies that reported standardized battery subtests pre and post CILT (e.g.,Breier et al., 2006; Mozeiko et al., 2011; Szaflarski et

[1] CILT is also referred to as Constraint Induced Aphasia Therapy (CIAT)

al., 2008), gains in spoken language comprehension were noted. Careful documentation of these changes could benefit future treatment of verbal comprehension in chronic aphasia.

Candidacy for High Dosage Aphasia Therapy

Candidacy for any treatment must be considered but particularly for one with the challenging clinical logistics and rigor of a high dosage protocol. At this time, much of the CILT research to date has included participants of varying severity. Meinzer and colleagues (2007) analyzed data from their various CILT studies and demonstrated that 38 of 44 people made improvements on standardized tests and that results were not correlated with aphasia chronicity or with age; they *were* correlated with initial severity of aphasia. The authors attributed their findings to the learned non-use hypothesis positing that those who have withdrawn from verbal communication the most are the most likely to benefit from CILT. It is also possible, however, that the materials were designed for those with more moderate aphasia symptoms and therefore not sufficiently challenging to those with milder aphasia types.

Once a more mildly affected participant has been sufficiently challenged, changes still may not be evident if standardized tests are used as the only measurement of change. More sensitive measures of change are also required to better assess treatment efficacy for various clinical subtypes. The standardized tests often used as the sole measure of change in previous studies of CILT are not sensitive enough to capture changes for participants already performing near ceiling. Discourse production and comprehension tasks are examples of more sensitive methods of documenting functional change for all, but *particularly* for those with mild aphasia symptoms. In probes of word learning, response time would provide a more sensitive measure increasing the value of response accuracy scores. It is possible that during overlearning, when

other behavioral measures look to have plateaued, that response times will continue to decrease and will correlate with continued neural change.

Neuroimaging and High Dosage Aphasia Therapy

In the past decade neuroimaging has been used increasingly to document changes resulting from aphasia therapy regimens, providing valuable information on brain reorganization after stroke. High dosage, short duration therapies are particularly attractive to researchers since neural change can be documented over relatively short time spans. At this time several high dosage aphasia treatment studies utilizing neuroimaging have demonstrated that perilesional activation tends to be associated with gains in language production (Breier et al., 2009; Fridriksson, Bonilha, Baker, Moser, & Rorden, 2010a; Kurland et al., 2012). However, the reorganized language system also makes use of other undamaged LH areas and RH lesion homologue areas (Crosson et al., 2009; Musso et al., 1999; Richter et al., 2008). These differences are likely not contradictory, but rather attributable to variance in lesion sites as well as to premorbid heterogeneity of language organization (Saur et al., 2010) as well as to the degree of cerebral perfusion in the lesioned left hemisphere (Fridriksson et al., 2011). Although the best language outcomes have tended to be associated with a shift back to left hemisphere laterality, lesion extent or perilesional hypoperfusion may render adjacent tissue unavailable for recruitment. In these cases, the right hemisphere must be utilized, albeit less efficiently.

Another important variable that influences the site and degree of physiologic change is the type of language treatment. For example, Crosson and colleagues (2009) initiated oral naming with complex left hand movements as a way of guiding the recovery process rightward. Musso and colleagues (1996) trained oral comprehension, a premorbidly bilateral activity, and demonstrated the role of the RH in recovery. In contrast, studies that report increased LH

activation tend to be those that required premorbidly LH demanding tasks such as oral naming (e.g., Fridriksson et al., 2011).

Response to various types and dosages of aphasia treatment can also provide insight into the potential for identifying regions of activation that are most predictive of change. Ideally, baseline neuroimaging will someday be used to guide decisions regarding aphasia treatment. Many neuroimaging studies report changes in activation pre- to post-treatment around the inferior frontal gyrus (IFG) (Fridriksson et al., 2010a; Meinzer et al., 2006), where Broca's area resides, but change has also been observed in subcortical areas such as the precuneus and posterior thalamus (Fridriksson et al., 2007). Menke, et al.(2009) reported data from a follow-up scan performed eight months post a high dosage anomia treatment providing evidence that neural recovery is still in flux immediately following treatment. Their findings indicated that immediate post-treatment gains were correlated with brain regions responsible for memory, attention and integration of information but that maintained gains were correlated with more traditional language regions. As in the above discussion on the contributions of the right and LH, post-treatment outcome data are also likely complimentary. If so, it will be apparent when studies are replicated with multiple scans tracking neural activation before, after and several months following cessation of treatment. It will take data from multiple longitudinal studies to begin generalizing findings to the larger population and to determine whether there are specific brain regions that are critical for recovery when using a specific treatment type.

Absence of behavioral gains following treatment can provide equally valuable information via neuroimaging. It is conceivable that a specific lesion site may interfere with successful treatment. For example, Fridriksson and colleagues (2010a) determined that participants with damage to posterior brain regions in the middle temporal and occipital lobe

were less suitable candidates for the cueing treatment approach used in their study. At this time there are no studies that report lesion differences in those who benefitted from CILT and those who did not. Studies of functional imaging following CILT are emerging (e.g, Breier et al., 2006; Meinzer et al., 2008; Richter et al., 2008) but more data are required before generalizations can be made about candidacy and predicting response to treatment. Richter and colleagues (2008) with the largest fMRI imaging CILT study to date (n=16) make a compelling argument that inefficiency in language processing such as that seen in increased RH activation prior to treatment is most susceptible to change following treatment. However, Breier and colleagues (2009) present evidence suggesting that those with greater initial RH activation were those less likely to *maintain* gains observed immediately post CILT. These studies used reading comprehension, silent word stem completion (Richter et al., 2008) and aural comprehension tasks (Breier et al., 2006) during neuroimaging. It would appear that these tasks test the treatment's impact on these skills, and the brain areas recruited for these skills, rather than the impact on verbal production—the targeted language modality. Those groups that used an overt naming paradigm (Kurland et al., 2012; Meinzer et al., 2008) following CILT found that perilesional activation was associated with behavioral gain. Results from these three studies of one treatment type demonstrate that neuroimaging task and stimuli variability bears on conclusions drawn about activation and lateralization. They also demonstrate that the effects of CILT (possibly the effect of *any* high dosage treatment) are observable in various brain regions, depending on the task. This suggests the treatment had effects reaching beyond spoken expressive language.

Difficulty of task and stimuli is yet another variable worth considering during fMRI paradigm design. Word recognition has been shown to be influenced by stimuli characteristics

such as word frequency and phonological neighborhood density—the number of phonologically similar words to the stimulus item (Prabhakaran, Blumstein, Myers, Hutchison, & Britton, 2006). Prabhakaran and colleagues' (2006) findings demonstrated an increase in the blood oxygen level-dependent (BOLD) response for words that were more difficult to access such as low frequency and high density words. There is some evidence that word production is similarly influenced (Frank, Tanenhaus, Aslin, & Salverda, 2007).

In current aphasia treatment studies, the difficulty level of stimuli is often not a point of focus. The tendency is to compare change pre and post-treatment between subjects and to correlate this with lesion sites for each individual. These studies tend to use task items in the scanner that are thought to be sufficiently challenging to all participants and change post-treatment would indicate that this is true. It is unlikely, however, that the stimuli are equally challenging for all PWAs. By introducing varying difficulty levels, and customizing the level of difficulty to the PWA, we have the opportunity to better contrast the effect of learning within each individual.

It seems feasible that with enough training, previously difficult material could eventually elicit the same BOLD response as easier stimuli. Activation during language tasks that are *easy* for a specific individual could theoretically provide information about relatively well-functioning brain areas that we aim to stimulate with similar but more challenging tasks. It may be that varying difficulty levels, however, will call upon different neural mechanisms since easy items, or those that do not require training, are served by a functional part of the brain versus trained items which must make use of re-activated areas.

By using neuroimaging in conjunction with controlled treatments, researchers have the opportunity to further elucidate the factors leading to optimal dosage, the neural mechanisms

mediating the recovery process, and the timing of recovery. This process should yield data that will aid in the determination of optimal treatment programs based on individual differences.

Summary of Problem

There is compelling evidence that increasing the duration of high dosage treatments will yield positive effects in language behavior that are both durable and generalizable to functional verbal language (Bhogal et al., 2003; Cherney et al., 2008; Robey, 1998). Positive reports appear to be particularly consistent after utilization of the mass practice schedule approximating 30 hours over two weeks (Barthel et al., 2008; Meinzer et al., 2005). Nonetheless, the most recent review of treatment studies that compare differing dosages, reported no clear differences between intensive and nonintensive treatments across studies (Cherney, Patterson, & Raymer, 2011). It seems likely that equivocal results reported for those with chronic aphasia are due, at least in part, to a lack of definition in what constitutes intensity. In order to determine optimal dosage parameters, each parameter must be manipulated within various treatment protocols for various patient populations. The current study focuses on the Intervention Duration by doubling the dosage of a single, mass-practice treatment with a no-treatment period inserted in the middle to optimize any potential consolidation effect. Manipulation of the other dosage parameters and, potentially, of a no-treatment period will be the basis of future studies.

Neuroimaging of aphasia treatment is in the frontier stages and preliminary questions have yet to be definitively answered. Whether perilesional recruitment is necessary for positive outcome and whether it is predictive of success in the treatment of verbal production are questions still in debate. Due to the heterogeneity of this population and of the human brain in general, longitudinal imaging in conjunction with monitoring of behavioral changes offers the best opportunity to examine the recovery process over time.

Current Study

The current study investigated the effects of increasing the treatment duration of an already high-dosage treatment. The protocol included the administration of CILT in two blocks separated by a five-week no treatment period. The break was established in response to the increased treatment response reported in some studies one month following the completion of CILT (Johnson et al., 2013; Maher et al., 2006; Mozeiko et al., 2011; Szaflarski et al., 2008). This made it more likely that any increases following Treatment Period II were indeed due to treatment and not to continued changes as a result of Treatment Period I. Each treatment block took place in three hour daily sessions over a period of two weeks. Participants with aphasia (PWA) achieved at least 80% accuracy at each of six increasingly difficult treatment levels in order to move on to the next level of treatment. The starting level depended upon the results of probes administered pre-treatment.

PWAs also underwent functional imaging while performing a naming task with a word list customized for each participant prior to the first scanning session (refer to section II.E for information on details and source of stimuli). Two categories of words were used: 1) Difficult words or those the participant is unable to produce[2] and 2) Easy words, those that are consistently produced correctly. Word frequency data were noted but did not influence the Easy/ Difficult categories which were determined by individual participant performance during naming (see Stimuli section, below, for detail). Scans will be performed at four time points: Pre-treatment, Post-Treatment Period I , Post-Treatment Period II and at a two-month Follow-up.

[2] "Production difficulty" refers to word finding deficits. PWAs with concomitant, severe motor speech disorders will be ineligible for participation in the study.

Therefore, each participant served as his or her own control. Changes in lateralization and activation patterns over time were examined in relation to stimulus type.

Research Questions and Predictions

This study will address the following research questions related to behavioral and neural activation changes over time measured by treatment probes, generalization probes, standardized tests, qualitative assessment and fMRI.

Research Question 1: Dosage effect. What is the effect on response accuracy and response time of CILT for trained material after one and two treatment doses (a dose=30 hours over two weeks)?

Predictions following Treatment Period I (30 hours of treatment):

P_{1a}- There will be an increase in response accuracy for trained materials. Repeated exposure and practice naming the same stimuli is predicted to result in improvement in accuracy of this material. No increase in accuracy for trained materials would indicate a lack of treatment responsiveness.

P_{1b}- There will be a decrease in response time on materials for which criterion was reached (80% accuracy). As competency increases with naming, it is predicted that response time will decrease. On materials for which criterion is not reached, decreases in response time are predicted to be less or may not decrease at all as competency precedes speed. Considering the speed of word production necessary in typical, connected speech, it is supposed that increased rate may be necessary for generalization to discourse. Speed of naming in aphasia rehabilitation is understudied in the aphasia literature. Conroy, Sage, and Ralph (2009) did find a relationship between naming accuracy and connected speech though *not* between speed of naming and connected speech; however, they suggest that speed warrants further research.

P_{1c}- There will be an increase in activation for the production of Hard Trained words, corresponding with improvements in naming following Treatment Period I (Scan One vs. Scan Two). This prediction is based on recent neuroimaging literature as it relates to aphasia treatment in which perilesional activation increases, possibly as a result of reperfusion to the area or anatomic remodeling, tend to be associated with a successful language response to treatment (Dorothee Saur & Hartwigsen, 2012). As processing becomes more efficient within an area, it is also possible that activation will be more attenuated as has been shown to be the case in neurotypical word learning (Meltzer, Postman-Caucheteux, McArdle, & Braun, 2009). This was Richter's (2008) rationale for his group's findings of decreased RH activity following aphasia treatment with increased performance and no change in LH activity. Individual differences are anticipated due to a range of lesion severity. Those with less extensive damage are predicted to have the greater leftward shift. This is based on the supposition that less damage means there is more healthy tissue available to recruit in the LH.

Predictions following Treatment Period II (60 hours of treatment):

P_{1d}- There will be an additional increase in response accuracy for trained materials. Again, more practice is predicted to increase participant accuracy which is the most direct measure of the efficacy of this treatment. No additional increases in accuracy would indicate a lack of responsiveness to a second treatment dose.

P_{1e}- There will be an additional decrease in response time for trained materials. Increased repetition is thought to increase rate as well as accuracy. No decrease in response time would not indicate lack of direct treatment response but may have implications for durability or generalizability of response (Conroy, Sage, & Ralph 2009)

P_{1f}- There will be an additional increase in activation for Hard, Trained words after Treatment Period II (Scan Two vs. Scan Three) corresponding with improvements in naming. If this prediction bears out, it would likely mean that reperfusion or anatomic remodeling has occurred (Dorothee Saur & Hartwigsen, 2012). If the null hypothesis is true and additional activation is not observed when naming improvements continue, it may be a result of increased efficiency of current neural networks (Meltzer et al., 2009).

Research Question 2: Generalization of effect. What is the effect on response accuracy and response time of CILT for untrained material after one and two treatment doses?

P_{2a}-There will be an increase in response accuracy and a decrease in response times for untrained materials; an increase on standardized test scores; an increase in functional communication and an increase in productivity and/ or efficiency after each Treatment Period. If the result of treatment is neural remodeling or increased neural efficiency, as predicted, improvements should extend to items not explicitly trained as observed for some participants following one dose of CILT (e.g., Mozeiko, Myers, & Coelho 2011; Maher et al. 2006; Szaflarski et al. 2008). A second dose of CILT is predicted to be observed for those who did not demonstrate generalization after one dose. If the prediction does not bear out and generalization does not occur for any participant, a double dose may not be warranted since generalization is the ultimate goal of treatment.

P_{2b}- There will be an increase in activation after each Treatment Period (Scan One vs. Scan Two and Scan Two vs. Scan Three) for the production of Hard, Untrained words, corresponding with improvements in behavioral naming of all Trained and Untrained words. No difference in activation will be observed for those who do not improve on Trained

words. Again, if neural remodeling is the result of training and performance improves for Untrained exemplars, it is expected that and increase in activation will be observed for each successive scan. If increased efficiency of functioning networks results, attenuation of activation would be expected.

Research Question 3: Durability of effect. Will treatment effects be maintained at follow-up assessment eight weeks after treatment completion?

P_{3a}- Gains in response accuracy for trained material and untrained materials, decreases in response times for trained and untrained materials, gains on standardized test scores and in discourse productivity and efficiency will all be maintained. Any neural change that occurs within Treatment Period I will be reinforced during Treatment Period II providing the best possible opportunity to maintain that change. If maintenance does not occur, it may be concluded that a) additional treatment periods are needed in order to maintain gain b) that the treatment was not the optimal treatment for the participant and other treatment protocols should be explored.

P_{3b}- There will be either no change or a decrease in activation for all conditions, two months post-treatment (Scan Three vs. Scan Four) corresponding with maintained behavioral accuracy in the scanner. If the participant is no longer receiving intensive treatment, it is unlikely that activational changes will persist. It is possible, however, that behavioral gains will lead to increased use of the oral-verbal modality leading to more practice and thus, further change.

Research Question 4: Neural activation. Will laterality changes be observed over the course of treatment?

P_{4a}- There will be an overall shift from greater RH activation to greater LH activation for those with the most extensive lesions (Scan One vs. Scan Four) and a general increase in LH activation for those with the most spared LH tissue. For all, LH activation is expected to increase over time for those who experience positive language gains. The LH is the dominant hemisphere for language for all participants and activation in the perilesional area is most consistently associated with the best recoveries (e.g., Fridriksson et al. 2011; Meinzer and Breitenstein 2008). It is possible that gains will be realized without a shift of laterality, or even with a rightward shift, since activation increases in the RH have been associated with positive behavioral changes, as well (e.g., Hartwigsen et al., 2013; Richter, Miltner, & Straube, 2008)

Chapter IV

Methods

Participants

Six participants were recruited from the University of Connecticut aphasia group and from other local aphasia groups based on interest in the study and willingness to commit time for all assessment, treatment periods and scanning. Participants were compensated for their time in the scanner and for transportation to and from all treatment and scanning sessions. All participants provided informed consent prior to initiation of the study. The study had previously been approved by the University of Connecticut's Institutional Review Board (IRB). Inclusion and exclusion criteria included: single, left hemisphere (LH) stroke at least 12 months prior to the study; monolingual, native English speaker; right handed; no reported history of psychiatric illness or acute, unstable medical conditions; ability to name at least two items on the Boston Naming Test (BNT); normal or corrected hearing and visual; understanding of study and ability to provide informed consent; and toleration of and candidacy for 3 Tesla (3T) scanning. See summary of screening procedures in Table 12.

Following several months of recruitment, it was necessary to include two participants who did not meet all criterion. One participant was left-handed; another acquired aphasia as a result of an anoxic event. Inclusion of these two individuals was necessary so the study could be initiated in a more timely fashion. Participants were divided into two groups of three based on severity of aphasia as determined by pre-treatment testing. One triad was comprised of individuals with mild-moderate aphasia and the second, individuals with moderate-severe aphasia. These triads were necessary as the treatment was group based. It should be noted that

for the purposes of this dissertation *only the findings from the four participants who met all recruitment criteria will be discussed*.

Table 12

Screening tests prior to study enrollment

Test	Description/Purpose	Inclusionary/ Exclusionary Criteria
Vision Screening	Snellen chart; question color blindness	Pass- 20/30 at 2.3 feet with or without corrected vision and answer of "no" for color blindness. Fail- refer to optometrist prior to enrollment; exclude if "yes" to color blindness
Hearing Screening	Portable audiometer. Test both ears at 500, 1000, 2000, and 4,000Hz	Pass-35 dB Fail- refer to audiologist prior to enrollment
Boston Naming Test (BNT)	To assess word finding ability	Pass- $\geq$ 2/60 Fail- < 2/60
Institute of Living fMRI screening form	To assess candidacy for 3T fMRI. Requirements are more stringent for 3T than for some lower powered clinical scans.	Pass- all questions must be satisfactory to fMRI technicians in order to be enrolled in study. Documentation must be provided to confirm that any implants are safe for scanning. Fail- exclude from study
Assessing apraxia of speech (Duffy, 2005, p. 95)	To assess motor planning/programming	Descriptive only

Demographic data for each of the four participants appear in Table 13. There were three male participants and one female ranging in age from 47 to 79 years. All participants attained at least a high school level education and three were employed prior to their stroke; one had retired.

Participants were between 31 and 58 months post onset. Two participants, M1 and M2 were classified as fluent anomic. Error patterns and affected brain regions differed between these two participants but both demonstrated generally good grammaticality, occasional paraphasias, good repetition, relatively intact auditory comprehension and fairly good reading and writing skills. The more severe participants did not fit neatly into a proscribed aphasia classification. One participant, S1, had a moderate-severe nonfluent aphasia. Although he scored in the 90th percentile for accuracy on the apraxia subtest of the Western Aphasia Battery (WAB), he did demonstrate speech characteristics consistent with apraxia of speech including distorted sound substitution errors and multiple unsuccessful attempts to correct errors for spontaneous language (Duffy, 2005, p. 95). The fourth participant, S2, had moderate fluent aphasia characterized by normal prosody and strings of grammatically appropriate jargon interspersed with actual content words and with both phonemic and neologistic paraphasias. Language samples from each participant are provided in Table 14.

Table 13

Participant Characteristics

	M1	M2	S1	S2
Age	54	47	56	79
MPO	57.96	57	51	30.96
Sex	M	M	M	F
Handedness	R	R	R	R
Hemiplegia	mod-severe	none	mod	mild arm monoplegia
Occupation				

	Owner, steel fabrication company	Treasury project manager	Mechanical engineer	Insurance company purchasing office
Education	16	16+	16	12
BNT	91.67%	76.67%	5.00%	5.00%
RCPM	89.19%	86.49%	94.59%	48.65%
R-WAB AQ	95	87.6	38.5	51.7
Spontaneous Speech	95%	95%	35%	65%
Auditory-Verbal Comp.	100%	91%	85%	71%
Word Fluency	65%	50%	10%	10%
Object Naming	93%	97%	40%	60%
Reading	100%	96%	46%	44%
Writing	80%	88%	25%	52%
Language Production	Fluent anomic. Slow, deliberate, often circumlocutory speech.	Fluent anomic. Slow, effortful speech marked by hesitations, incorrect word choice, multiple self- corrections. Very functional. Gets message across	Severe nonfluent. Few words, often repeated. Uses stereotypies and overlearned phrases.	Moderate fluent. Long sentences with normal sounding prosody but little content. Lacking awareness of language deficits.

Note. MPO-months post onset; BNT-Boston Naming Test (Kaplan, Goodglass, Weintraub, & others, 1983); RCPM- Raven's Coloured Progressive Matrices(Raven et al., 1988); R-WAB AQ-Revised Western Aphasia Battery Aphasia Quotient (Kertesz, 2006).

Table 14

Example of a Rockwell Picture and Participant Responses

Stimuli	Prompt from Clinician
	Tell me what is happening in this picture.

Participant	Participant Response
M1	They are in a restaurant uh, but… it's… the table is set for four people and th-they have the mother and the son and the two fully grown man, men. Um, I don't know if they are all together but the mother and young dau- young son, are um engaged in a prayer, um , before their meal and, uh, the two men are, uh, having a cup of coffee and a cigarette.
M2	Ok, looks like they are, um I'm gonna say it's uh, a woman with their boys, um in a restaurant. The woman is.. she is.. the woman is saying grace for, um before her meal. The two b- the, the older boys are waiting for her to finish her, finish the, finish grace. The little boy seems to also waiting to go the women. I mean the little boy is bowed, bowing himself with grace, for grace, with grace, on grace…with the woman.
S1	Oh yeah, the other ones are zerty bezert, and the other ones, all, all of them, was zerty bezert, boom boom boom, and they're what!? And then zerty bezert. I'm sorry I can't speak.

S2	Okay, this this little little gal was over here and she was a this was a monner over here. And and he was this this barrow he was on her madderer, he was her madderer. Over here. He was just madder over here to tell her this little bon. He probably… I don't know if she was, maybe, maybe having a little out of it or if she was telling him or whatever it is and she's telling her this iter tella her.

Note. M1 and M2 are the two participants with mild aphasic deficits. S1 and S2 are the two more severely affected.

All four participants had large lesions in which the middle cerebral artery was implicated, according to radiologists' reports. M2 and S1 had lesions extending throughout much of the frontotemporal regions and subcortically. M1 and S2's lesions were more temporoparietal and also extended to subcortical regions. See Figure 10 and Table 15 for detailed lesion locations and volume sizes for each participant.

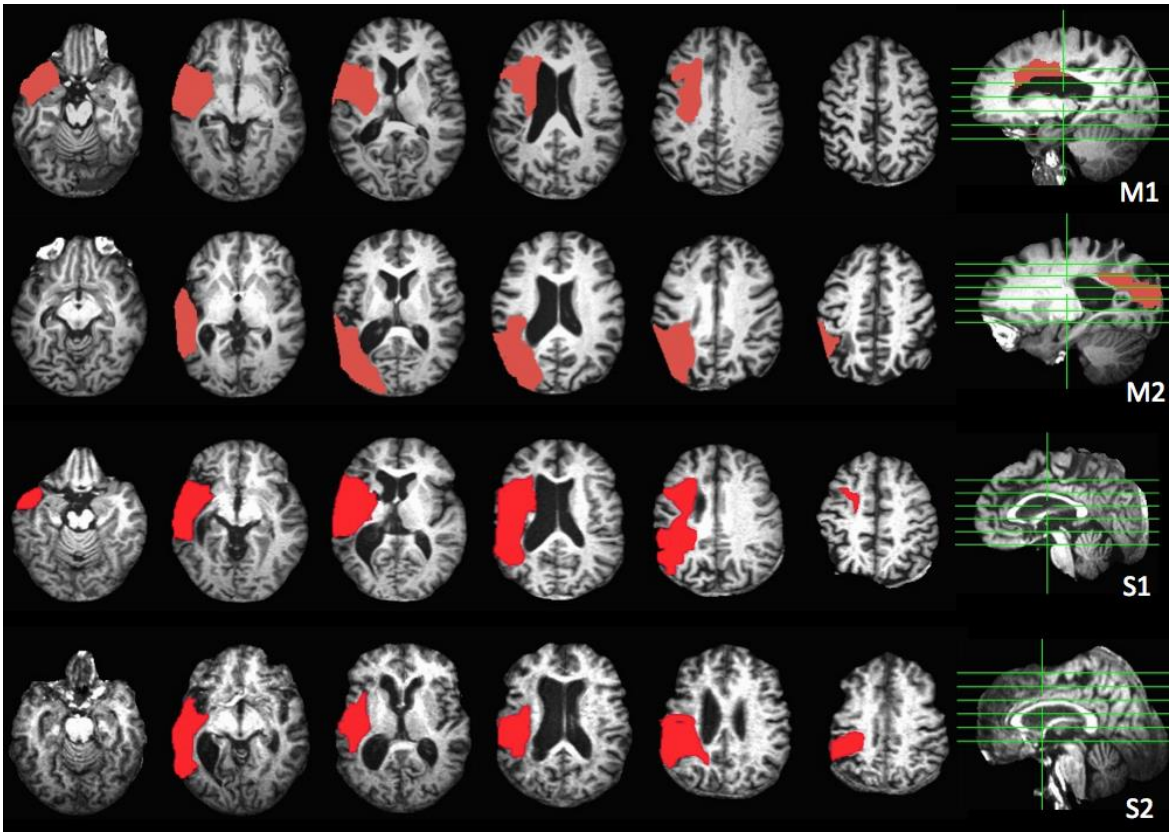


Figure 10. Magnetic resonance images indicating the extent of lesions in individual participants. Red indicates lesioned areas which were manually traced using Analysis of Functional Neuroimages (AFNI; Cox, 1996).

Table 15

Participant Lesion characteristics

	M1	M2	S1	S2
Lesion Volume	75,475 mm^3	64,869 mm^3	99,671 mm^3	67,250 mm^3
Brain areas implicated	13.3% Superior Temporal Gyrus	23.2% Middle Temporal Gyrus	22.1% Superior Temporal Gyrus	20.1% Superior Temporal Gyrus
	11.1% Inferior Frontal Gyrus	21.0% Superior Temporal Gyrus	14.0% Middle Temporal Gyrus	14.6% Inferior Parietal Lobule
	10.9% Insula	13.2% Middle Occipital Gyrus	11.9% Insula	12.6% Insula
	5.7% Precentral Gyrus	9.4% Inferior Parietal Lobule	6.3 % Inferior Frontal Gyrus	9.3 % Post-central Gyrus
	5.3 % Middle Temporal Gyrus	7.5 % Supra-marginal Gyrus	5.6 % Precentral Gyrus	8.7 % Middle Temporal Gyrus
	3.9 % Middle Frontal Gyrus	3.9 % Cuneus	4.9 % Inferior Parietal Lobule	6.7 % Supra-marginal Gyrus
	3.7 % Lentiform Nucleus	3.1 % Angular Gyrus	2.6 % Supra-marginal Gyrus	4.6 % Precentral Gyrus
	2.1 % Cingulate Gyrus	2.4 % Superior Occipital Gyrus,	2.3 % Post-central Gyrus	1.9 % Transverse Temporal Gyrus
	1.6 % Claustrum	1.0 % Precuneus	1.8 % Lentiform Nucleus	
			1.5 % Middle Frontal Gyrus	
			1.3 % Transverse Temporal Gyrus	

Note. % indicates the amount of the participant's lesion that overlaps with that left hemisphere brain region.

Experimental Design

A modified multiple baseline design across participants was used in conjunction with a multiple probe technique to evaluate the effects of treatment (Thompson, 2006). Prior to

treatment, probe testing was completed for each level in the treatment hierarchy to determine starting level of treatment (levels of treatment are discussed in Treatment Stimuli, section 2.5.1). Baseline probes were then conducted a minimum of four times per participant on the level at which they placed.

All participants also received baseline probes of productivity and efficiency of discourse production. Productivity refers to the quantity of relevant information, most impaired in those with more severe aphasia. Productivity is less problematic for those with mild aphasia but may be slower or characterized by excessive fillers, repetitions, false starts and mazes. Efficiency measures allow us to better quantify production within a finite time period or within the sample itself. Discourse was not explicitly trained but was probed consistently throughout the treatment to assess for generalization.

Baseline probes conducted during varying time periods from within a two week time span to a within a six month span per participant and were taken four to six times in order to serve as experimental control. Baselines were always conducted a minimum of 48 hours apart. Extension and variation of the baseline phases across participants confirmed baseline stability and increased the likelihood that changes observed during the treatment phase were, indeed, due to treatment (Thompson, 2006). In a typical multiple baseline design across participants, the treatment start times also tend to be staggered, but this was not possible given the small-group treatment design. The multiple probes allow for investigation of performance on increasingly difficult linguistic targets over the duration of the treatment period.

Once baseline testing was complete, participants received their first baseline fMRI scan and completed pre-treatment testing (for a timeline of treatment and assessment, see Figure 11). Treatment Period I began the following week with a triad of participants seen together for a three

hour session, every day, for two weeks. Participants each received a post-treatment fMRI scan and two post-treatment testing sessions the following week. They then received no treatment for the next four weeks.

Following the no-treatment period, treatment and generalization probes were administered and then Training Period II was conducted for another two weeks, followed by a third fMRI scan and two days of post-treatment testing. The fourth and final fMRI scan took place eight weeks after the completion of Treatment Period II followed by two more days of follow-up testing. The eight week follow-up period was decided based on studies using CILT that consistently demonstrate maintenance of gains four weeks post treatment (Johnson et al., 2013; Maher et al., 2006; Mozeiko et al., 2011; Szaflarski et al., 2008). Maintenance at time points long after treatment is more suggestive of permanent change but fewer studies report maintenance data beyond a month's time. Eight weeks is not long enough to determine whether changes will be longstanding but may be more informative than a four-week follow-up, without much increased risked of participant attrition. Tasks from each of these phases of treatment will be described in detail in subsequent sections.

Treatment probes identical to those used in baseline testing were administered prior to each training session starting after the first six hours (two days) of CILT. Probes for generalization to discourse were administered starting after the first nine hours of treatment and every six hours thereafter. The treatment probes consisted of ten stimulus items included in the previous day's treatment session—Trained items, ten equivalent Untrained stimulus items, ten from the previously trained level to track maintenance, and also ten stimulus items from the subsequent level, the latter served as a means of baselining until items were included in the trained sets. On alternate days, probes of generalization to discourse were administered. See

Table 16 for a treatment and assessment schedule. Probes and other assessment measures are described in more detail in the next section. All treatment and assessment took place at the UConn Speech and Hearing Clinic in Storrs, CT. All neuroimaging sessions took place at the Institute of Living in Hartford, CT.

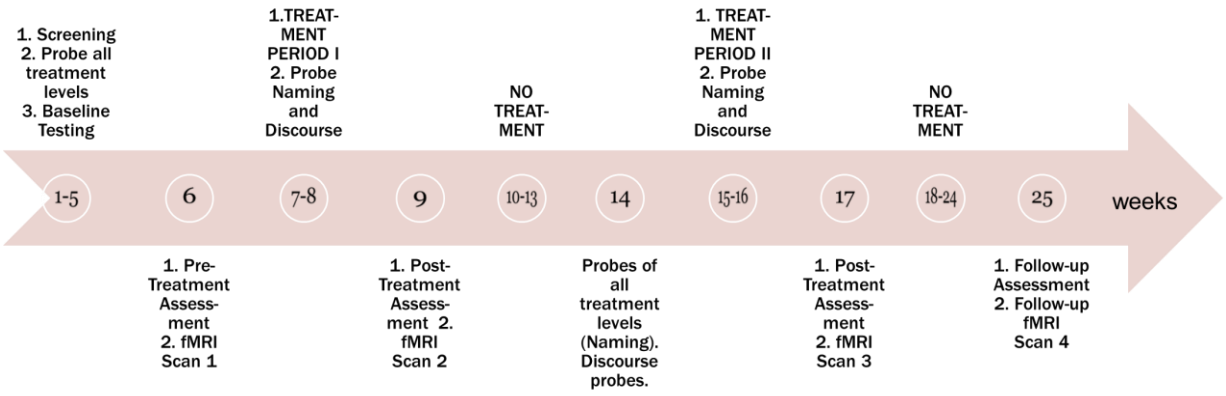


Figure 11. Timeline of behavioral assessments, neuroimaging and treatments.

Table 16

Treatment and Assessment Schedule

Treatment Phase	Task	Time task is initiated	Time to complete task
Pre-Treatment	Baseline starting level and probe subsequent levels.	Baseline collection was staggered 2 weeks-6 months before treatment; 3-5 baseline points taken. Probes taken during week prior to treatment.	3-5 days
	Language Testing	One week pre-treatment	2-3 days
	FMRI Scan One	Four days pre-treatment	1 hour
Treatment Period I	Treatment Period I	Monday- 9:30 AM	30 hours over two weeks, daily except weekends

	Probes of treated and untrained material	SEVERE- after every 3 hours of treatment, starting day 2 MILD- after every 6 hours of treatment, starting day 3	SEVERE-20 min; prior to start of treatment MILD- 30 min; prior to start of tx
	Probe of generalization to discourse	SEVERE-after every 3 hours of treatment, starting day 2 MILD- after every six hours of treatment, starting day 4	SEVERE-5-10 min; prior to start of treatment MILD- 15 min; prior to start of tx.
Post-Treatment Period I	Language Testing	3 days post-tx 1	1-2 days
	All Probes	3 days post-tx 1	30 min.
	FMRI Scan Two	6 days post-tx 1	1 hour
Treatment Period II	Treatment Period II	Monday- 9:30 AM	30 hours over two weeks, daily except weekends
	Probes of trained and untrained material	SEVERE- after every 3 hours of treatment, starting day 2 MILD- after every 6 hours of treatment, starting day 3	SEVERE-20 min; prior to start of treatment MILD- 30 min; prior to start of tx
	Probe of generalization to discourse	SEVERE-after every 3 hours of treatment, starting day 2 MILD- after every six hours of treatment, starting day 4	SEVERE-5-10 min; prior to start of treatment MILD- 15 min; prior to start of tx.
Post-Treatment Period II	Language Testing	3 days post-tx 2	2-3 days
	Probes	3 days post-tx 1	30 min.
	FMRI Scan Three	6 days post-tx 1	1 hour
Follow-Up	Language Testing	eight weeks post-treatment Period II	2-3 days
	Probes	eight weeks post-treatment Period II	30 min.
	FMRI Scan Four	eight weeks post-treatment Period II	1 hour

Standardized Tests

All standardized tests were administered four times- pre-, post-Treatment Period I, post-Treatment Period II and as a follow-up to treatment, eight weeks following completion. These included the Revised Western Aphasia Battery (R-WAB; Kertesz, 2006), which yields an Aphasia Quotient (AQ) used as a measure of severity. The WAB is a widely used, validated, standardized aphasia assessment (Bakheit, Shaw, Barrett, et al., 2007; Kiran, Sandberg, & Sebastian, 2011; Thompson, den Ouden, Bonakdarpour, Garibaldi, & Parrish, 2010). Shewan and Kertesz (1980) report good test-retest reliability ($r = .88$, $p < .001$) and internal consistency ($r = .974$) for the WAB. Administration of the WAB takes approximately one hour. The full WAB, which includes an assessment of reading and writing, was administered three times, pre- and post-Treatment Period II and eight weeks following treatment as it was in these subtests that changes were anticipated. Only those subtests comprising the AQ score were administered post-Treatment Period I. A five-point increase on the AQ tends to be used as the benchmark indicating clinical significance (Shewan & Kertesz, 1980) though results of Rasch analysis suggested a variable standard error of measurement (SEM) according to aphasia severity (> 2 points for AQs 30-70 ranging to > 6 points for AQs < 20 and > 90) (Hula, Donovan, Kendall, & Gonzalez-Rothi, 2010). In order to most easily compare results with that of other studies, the five-point benchmark is used in this study.

The BNT (Kaplan et al., 1983) is a 60-item confrontational naming test included as an additional measure of untrained spontaneous naming. No SEM has been reported for the BNT for people with aphasia but Flanagan and Jackson (1997) reported an SEM of 1.02 for non brain injured individuals. The Computerized Revised Token Test (McNeil et al., 2008) was used as a more sensitive measure of auditory comprehension. No SEM has been reported on this battery

but test-retest reliability on subtests ranged from .79-.91 and .85 overall. Reports from previous studies using CILT indicate that much of the change on pre- to post- treatment WAB AQ scores are attributed to changes in receptive language (Johnson et al., 2013; Mozeiko et al., 2011; Szaflarski et al., 2008) When not specified, a two standard deviation change on normed tests or 20% change on non-normed tests is considered clinically significant (Robey, Schultz, Crawford, & Sinner, 1999).

The participant and spouse or family member were also asked to complete the Communicative Effectiveness Index (CETI; Lomas et al., 1989) prior to and following each treatment period (see Appendix A) in order to assess for potential functional changes outside of the clinic. The CETI is a rating scale completed by the clinician with both the participant and family members to assess for functional communication and ability to interact effectively with other people. The CETI is said to have strong psychometric properties including good test-retest reliability (r = .94), SEM (5.87), inter-rater reliability(r = .73) and good construct validity (r = .89) with the WAB (Lomas et al., 1989).

Baseline Measures

Prior to treatment initiation, several tasks were conducted.

Determination of starting treatment level. The level at which to begin treatment for each individual (refer to Table 17) was determined by presenting ten stimuli per treatment level via E-prime. The expected response type was modeled by the clinician prior to presenting each of the eight levels and responses were scored. Any level resulting in a score of less than 80% of all possible points was considered appropriate for inclusion in training. Triads were formed according to individual performance. Despite some variation in ability, triads were matched with

all participants initially placing at the same entry level (Level One for those with severe aphasia and Level Four for those with mild aphasia).

Responses to trained targets. Once the level of treatment was established, baseline assessment of the starting level and subsequent level of treatment took place a minimum of three times prior to commencement of treatment. For the two participants with mild aphasia, the starting point was at Level Four, thus expected production was carrier phrase + adjective + object (e.g., "Do you have four plums?") and for the two with more severe aphasia, it was at the first level with production of a high frequency word (e.g., "ball"). Different stimulus items of equivalent difficulty were administered at each baseline session to avoid potential learning effect.

Two of the baseline measures took place during the week prior to treatment. Two to three additional baseline measures were taken between four months and two weeks prior to treatment. Baselines of starting level were stable prior to treatment.

Narrative discourse. Three of twelve Norman Rockwell pictures were presented with the request to "tell me what is happening in this picture." Story descriptions were transcribed and scored for correct information units (CIUs) according to guidelines by Nicholas and Brookshire (Nicholas & Brookshire, 1993). CIUs, a measure of discourse productivity, were also used to calculate discourse efficiency measures CIUs per minute and also the proportion of CIUs to total words. Both efficiency measures were used in order to contrast whether a potential increase in efficiency was due to increase in speed, word selection or both.

Treatment Protocol

Constraint Induced Language Therapy (CILT) was administered over two treatment periods for a total of 60 hours of treatment. Each period consisted of 30 hours of treatment over

two weeks with a five week break in between. Treatment began every morning at 9:30 AM following daily treatment probes. CILT uses an interactive game approach, following the rules of the well-known card game "Go Fish." Each participant was offered a card holder and dealt 5-7 cards, depending on familiarity with the deck, and was then instructed to request matching cards from other players. Participants were asked to respond as completely as possible, as modeled by the clinician who participated in all games. The clinician also modeled requests, cued responses when necessary and reminded participants to use the verbal modality of communication only. Gesture and written communication were disallowed as a substitution for the verbal modality. The "constraint" aspect of this treatment has been a point of debate in the literature considering evidence of a facilitatory nature of gesture in speech production. Individuals with aphasia have been found to produce more spontaneous gestures in comparison with controls during confrontational naming or spontaneous conversation (Carlomagno, Pandolfi, Marini, Di Iasi, & Cristilli, 2005; Hadar, Burstein, Krauss, & Soroker, 1998; Rose, Douglas, & Matyas, 2002). A recent pilot study by Jenkins et al. (2014) showed decreased complexity of narrative discourse when gesture was restrained compared to unrestrained gesture during storytelling.

Difrancesco and Pulvermuller (2012) provided a detailed explanation of "constraint" as it was originally conceptualized for this language therapy, dispelling the notion that the use of hands was forbidden. Nonspecific gestures or hand movements accompanying verbal language were permitted, in keeping with these guidelines. Each participant was actively involved and both produced and responded to requests for the full three hours. They were provided a ten-minute break after the first ninety minutes.

Fading semantic and phonemic cues were provided to promote errorless learning (Maher et al., 2002). Once mastery of the targeted materials was achieved by all participants in the group, a new deck requiring a more difficult response was introduced. A treatment hierarchy was designed with the protocol described by Maher and colleagues (2006) in mind; however, additional levels were created to a) challenge the least impaired participants and b) prepare for increased progress through the hierarchy anticipated with the addition of a second treatment period.

Levels are as follows and also summarized in Table 17. Level One consists of a high frequency word deck in which requesting the object by name is the goal (e.g., "cat"). Level Two is a low frequency word deck in which requesting the object by name is required (e.g., "anchor"). Level Three is comprised of mixed frequency objects requiring the carrier phrase (e.g., "Jen, do you have the anchor." Level Four uses a mixed frequency object deck requiring a single adjective to differentiate between nouns (e.g., frying pan vs. dust pan). Level Five also uses mixed frequency objects but additional stimuli are includes so that the request must incorporate multiple adjectives in order to differentiate between cards (e.g., red frying pan vs. black frying pan). In Level Six, another mixed frequency word deck is used, this time requiring production of two objects and a preposition (e.g., "The cat is on the chair"). Level Seven is a mixed frequency word deck requiring the production of at least two objects, two adjectives and one preposition (e.g. "The black cat is on the pink chair"). Level Eight uses a deck comprised of complex pictures and requires the production of a complete descriptive sentence (e.g. "Two girls are sleeping in the canoe while a boy fishes"). There is no one proscribed sentence per picture but the description must be adequate such that another participant recognizes the stimulus item in question.

Starting level was based on individual performance during baseline testing and individuals were grouped according to level. Points were scored based on production (see Table 17) and a minimum of 90% of all possible points per level were necessary before progressing to the next level, with one exception. If, after one week (15 hours of treatment), Level One (high frequency objects) was not achieved, Level Two (low frequency objects) was introduced and trained simultaneously in order to ensure exposure to stimuli necessary in the neuroimaging task.

Table 17.

Treatment hierarchy and scoring

Level	Expected Production	Maximum Points/ turn
1	High frequency object (1)	1
2	Low frequency object (1)	1
3	Mixed frequency object (1)+ carrier phrase (1)	2
4	Mixed frequency objects (1) + adjective (1) + carrier phrase (0)*	2
5	Mixed frequency objects (1) + 2 adjectives (2) +carrier phrase (0).	3
6	Two mixed frequency objects (0)** + preposition (1)	1
7	Mixed frequency objects (2) + adjective (2) + preposition (1)	5
8	One sentence picture description. CIU:WC	100

Note. CIUs refer to Correct Information Units, described in Treatment Stimuli section. WC refers to word count.
*Once the carrier phrase was mastered it was expected in all future productions and not awarded additional points.
**The preposition was the focus of training at this level. Accurate preposition must be produced and used appropriately given the ordering of the other words in the sentence.

After Treatment Period I participants received one week of post-treatment testing and scanning followed by no treatment for four additional weeks. Performance on probes one week and five weeks post-treatment was assessed for maintenance or change and to help determine the starting level of Treatment Period II.

The risk of including a period of no-treatment was stopping before a participant achieves neural change. In contrast, continuing with this rigor for longer than two weeks posed a greater

risk of exhausting participants, deteriorating motivation and possibly losing participation entirely. The primary determination for the five week no-treatment period, however, was driven by current findings reported post-CILT. Investigators have reported that language gains are maintained and, for some participants, even continue to increase up to a month post-treatment (Johnson et al., 2013; Maher et al., 2006; Mozeiko et al., 2011; Szaflarski et al., 2008), providing evidence that that neural changes continue post- intensive treatment.

Two treatment studies that have compared CILT to other treatment types have used a cross-over treatment design by which one treatment was used for approximately 30 hours over two weeks and then a second treatment type was used for an additional 30 hours (Kurland et al., 2012; Rose et al., 2013) . The study by Rose and colleagues (2013) incorporated a one week break between treatment periods; Kurland et al., (2012) did not report a break. In both studies, greater gains were observed after the first treatment administration, regardless of type. Fewer gains followed the second treatment. The smaller gains reported for standardized tests from these studies following the second treatment are no greater than changes seen in follow-up reports 4-weeks post CILT. Given that changes following a course of intensive treatment have been observed *without* any treatment, it is difficult to know how much gain should be attributed to the second treatment and how much is actually the result of neural change brought about by the first treatment. Avoidance of this confound motivated the five-week break.

Treatment stimuli

Treatment stimuli consisted of 120 full color stimulus items *per level* which were divided into four 30-card decks. An additional 120 items *per level* were created that were never included in the training process and seen only ten at a time during the treatment probe sessions. Included within the Trained and Untrained stimuli were customized words which were considered either

Hard or Easy for each participant during pre-treatment testing. The process of determining these word lists is described in the subsequent section on fMRI Stimuli.

This number of stimuli items used was far greater than tend to be reported in aphasia treatments. However, considering the accumulating evidence that intensive language therapies are directly responsible for neuroplastic brain remodeling in those with chronic aphasia (Crosson et al., 2009; Meinzer et al., 2004; Schlaug, Marchina, & Norton, 2009), confining training to a small set of word lists would be counterproductive. Neural changes appear to be a result of re-activating primary language areas surrounding the lesion site or by strengthening secondary language areas, such as the homologue to the lesion in the right hemisphere. Therefore, activation of these areas using a large number of sufficiently challenging materials and not memorization of a set word list was the primary goal of treatment. Greater numbers of stimuli have been demonstrated to result in increased word learning with equal durability than shorter lists of stimuli for individuals with severe and those with mild naming impairments (Snell et al., 2010).

Treatment probes

Probes of trained materials took place after every six hours of treatment to assess progress in the participants with mild aphasia and after every three hours for those with more severe aphasia. The latter participants received double the probes given to the mild group because a) the time necessary to administer probes was much less for the participants with severe aphasia who did not progress as quickly and therefore didn't require the large number of probes given to those with mild aphasia and b) more day to day variability in naming performance in these participants warranted a greater number of data points. Probes were always administered

prior to treatment initiation, delivered via E-prime and identical to the delivery that occurred during baseline testing.

Probes of Untrained materials from the same level and on the subsequent level were administered at the same time to assess for stimulus generalization. Probes of levels previously achieved were also administered in order to track maintenance of gains over time. Response accuracy was scored for all participants. For those with mild aphasia, response times were also recorded as a measure of efficiency.

Participants were rarely exposed to repeated materials during probes of Untrained materials and would only see an item a maximum of twice throughout the treatment study period. This was done in order to reduce the chance of improvement due to word exposure. Word frequency data were derived from the MRC psycholinguistic database (Coltheart, 1981). This was relevant for training only for S1 and S2 who were at initial treatment levels requiring a separation of high and low frequency types.

Stimuli were presented at conversational levels as increased levels have not shown to influence auditory processing in individuals with aphasia (McNeil, Darley, Olsen, & Rose, 1983). E-prime 2.0 software (Schneider, Eschman, Zuccolotto, & Guide, 2002) was used to present naming probes in order to calculate response time and to facilitate participants' familiarity with the task used in the fMRI conditions.

Response times were calculated during the review of video files. To do this a beep was presented coinciding with the presentation of the stimulus item and that time point was subtracted from the time at which there was a *correct* initiation of the word or phrase production to yield a "response time". Accuracy of production was calculated according to the scale outlined in Table 17. Points earned were divided by the total possible points in order to

determine percent accuracy at each level. Level Eight responses (one-sentence picture description) were transcribed and scored for correct information units (CIUs) according to guidelines by Nicholas and Brookshire (Nicholas & Brookshire, 1993). These were used to calculate CIUs/minute in order to measure efficiency of oral verbal production.

Effect sizes.

Effect sizes for performance on trained stimuli were calculated using Busk and Serlin's (1992) variation on Cohen's d statistic advised by Beeson and Robey (2006) in order to avoid the Type I error that may occur with visual inspection alone. This is done by subtracting the mean of the baseline probes from the mean of the two final probe scores and dividing the result by the standard deviation of the baseline scores (Beeson & Robey, 2006). For studies of naming, the benchmarks recommended by Robey and Beeson (2005) are 4.0, 7.0 and 10.1 corresponding with small, medium and large effects. These benchmarks were based on single subject studies of lexical retrieval. It should be noted, for comparison purposes, that recent aphasia treatment studies such as those by Rose et al. (2013) and Thompson et al. (2010) use benchmarks of 2.6, 3.9, and 5.8 from Robey et al. (1999) which are based on single subject aphasia treatment studies but not specifically those of lexical retrieval. The former, more conservative benchmarks, were those used in the current investigation.

Discourse probes

Probes for generalization to narrative discourse were administered, before treatment sessions, on alternate days starting nine hours post-treatment and then every six hours after that. Three of twelve Norman Rockwell pictures were presented with the request, "tell me what is happening in this picture." Participants did not receive feedback on the quality of their descriptions and were given no time limit. Each of the twelve Rockwell pictures were seen no

more than four times during all phases of the study (over a span of four months) and never more often than once per week.

All picture descriptions were videotaped, transcribed verbatim and analyzed for CIUs which measure discourse productivity, were also used to calculate two discourse efficiency measures 1) CIUs per minute and 2) the proportion of CIUs to total words. Effect sizes were calculated for productivity and efficiency for each participant.

Treatment data analysis

Results of the study were examined on an individual basis, in keeping with single-subject experimental design conventions. There is not general agreement on the best means of analysis for research of single subject experimental design. Visual inspection, trend lines, binomial tests, analysis of variance, the C-statistic, standardized effect sizes, and "clinical significance"—often defined as a two standard deviation change on standardized tests or by 20% on nonstandardized measures— have each been used to describe the effects of aphasia treatment. Each has strengths but none are without limitations (Robey et al., 1999). The limitations of those that are sensitive to autocorrelation outweigh the benefit (Robey et al., 1999) and even those deemed necessary, such as visual inspection, have questionable validity on their own.

As such, responsiveness to treatment was examined based on a combination of measures with each assessment taking place a minimum of four times. All measures were calculated for each participant and then discussed qualitatively as they related to treatment response after one and two treatment periods. Outcome variables investigated include:

- Change in performance on standardized tests.
- Slopes and effect sizes for trained materials.
- Slopes and effect sizes for untrained materials and for generalization to discourse.
- Neural activation as it corresponded with behavioral change.
- Response times on treatment probes for participants with mild aphasia only.

The intensive nature of this treatment was expected to stimulate neural activation. By inducing the use of hundreds of words, including those that the PWA was likely to avoid, inactive or dysfunctional system processes were thought to become re-engaged and thus treatment was expected to result in gains beyond the trained stimuli to untrained stimuli and also connected speech. Since increased, more efficient verbal language production is the ultimate goal of treatment and not simply mastery of a set word list, the generalization measure of connected speech also serves as a main outcome measure. Point to point intra- and inter-reliability was performed by the author and by trained research assistants for CIU analysis and was found to be 97.2 and 94.3 respectively.

Neuroimaging

In order to investigate the neural activity corresponding with potential language changes over time, fMRI scans were conducted. Structural images were acquired prior to each imaging session. Blood-oxygen-level-dependent (BOLD) fMRI measures were acquired using a Spoiled Gradient Recalled Echo (SPGR) during a confrontational naming task in which PWAs were asked to name a set of "Easy" and "Hard" words as determined prior to the first scan (for more detail, see Neuroimaging Protocol section below). One set of 60 "Trained" words was included in the training set and one set of 60 "Untrained" words was never trained. In order to examine changes in activation over time, scans took place:

1) pre-Treatment
2) post-Treatment Period I
3) post-Treatment Period II
4) two months post-Treatment Period II.

Verbal responses were digitally recorded and transcribed and then accuracy and response times were documented. Changes in lateralization and activation patterns over time were examined in relation to stimulus type, i.e., Easy vs. Hard, for each Trained and Untrained set.

The particular contrasts of interest were those that could demonstrate an effect of time, an effect of training and an effect of difficulty. Anatomical regions of interest (ROIs) were then further analyzed in order to qualitatively compare activation across the same individuals.

FMRI stimuli

Participant responses to three presentations of 384 items from the Treatment Stimuli (described in Treatment Stimuli, above) were used to generate the stimulus items to be used in the scanner. For each participant a unique set of 120 stimuli were generated comprised of 60 "Easy" and 60 "Hard" words. For the participants with mild-moderate aphasia the Easy words were those produced accurately in three of three trials and Hard words were those that were not produced or produced inaccurately in at least one of three trials. For participants with more severe aphasia, Easy words were those that were either independently produced or stimulable given a semantic or phonemic cue for three out of three trials. Hard words those they could not produce, even given cueing, in all three trials. Half of the fMRI stimuli, 30 Hard and 30 Easy, were included in the stimuli to be trained and half were never trained. Hard and Easy lists for each individual were not comprised of all low frequency or all high frequency words but more high frequency words did tend to be included in the easier lists as shown in Table 18.

Table 18

Means and standard deviations for customized words lists

	M1 Hard	M1 Easy	M2 Hard	M2 Easy	S1 Hard	S1 Easy	S2 Hard	S2 Easy
mean	30.63	37.14	23.90	39.88	29.88	48.97	38.10	45.93
SD	70.44	65.50	38.17	89.26	118.12	87.31	66.53	91.21

Note. Means and standard deviations are based on a word's written frequency of occurrence as given in the norms of Francis and Kucera (1982). The frequency range is 0-69,971 (for "the").

A control condition as used in Kurland and colleagues' (2012) recent fMRI study was also used. Controls consisted of 30 empty stimuli requiring visual processing and a verbal response but requiring no semantic or phonologic processing. These comparison items consisted of meaningless lines and squiggles that required a "pass" or "no" response. A total of 150 stimuli (30 Trained Easy, 30 Trained Hard, 30 Untrained Easy, 30 Untrained Hard, 30 Controls) were divided into six runs. Each run included an equal percentage of Trained and Untrained Easy, Hard and control stimuli presented at random (example stimuli shown in Table 19).

Table 19

Example stimuli for one participant (SR4)

Stimulus Type	Example 1/ word-KF-Freq	Example 2/ word-KF-Freq	Example 3/ word-KF-Freq
Tr-Easy	 Carrot-1	 Zebra-1	 Sandwich-10
UT-Easy	 Box-70	 Fork-49	 Hammer-9
Tr-Hard	 Ant- 6	 Wagon- 55	 Grapes-3
UT-Hard	 Bed-127	 Belt- 29	 Pumpkin- 2

Note. Tr-Trained; UT-Untrained; KF-freq- refers to a word's written frequency of occurrence as given in the norms of Francis and Kucera (1982). The range of frequency in the file is 0-69,971 ("the"). These examples demonstrate that high frequency words were not always those that were easiest for a participant.

FMRI procedure

Images were acquired on a Siemens 3 T Siemens scanner. High-resolution three-dimensional T1-weighted anatomical images (MPRAGE) were acquired for co-registration. Each of the participants underwent four scanning sessions of approximately one hour each. The first took place pre-treatment (Scan One); the second post Treatment Period I (Scan Two); the third post Treatment Period II (Scan Three); and the Follow-up (Scan Four) took place eight weeks post treatment completion (Scan Four).

Each scan session began by aligning the participant's head to the magnetic field center. Participants were then asked to hold as still as possible during acquisition of a 1 mm^3 voxel magnetization prepared rapid acquisition gradient-echo (MPRAGE) anatomical (repetition time = 1900 msec, echo time = 4.15 msec, inversion time = 1100 msec, 1 mm^3 isotropic voxel size, 256 X 256 matrix). Participants were reminded of the procedure and the MPRAGE was immediately followed by functional acquisition during an overt naming task. Six runs of functional acquisition were acquired in ascending, interleaved order using gradient echo-planar imaging sequences (SPGR) (29 slices, 4mm thick, 3.44mm x 3.44mm axial in-plane voxel resolution, gap=0, 220 mm^3 field of view, flip angle = 70 º, TR = 6,000 ms, time acquisition= 1,500 ms; TE=27 ms). Each functional run was preceded by two additional TRs during which no data were recorded in order to allow for stabilization of longitudinal magnetization. Each of the six runs took approximately five minutes between which communication with the participant took place via intercom in order to relay instructions and to ensure patient comfort. A temporal sparse sampling design (Hall et al., 1999) was used so that there was a quiet scanner-off period for naming in order to best assess production and to avoid movement artifacts. Stimuli were presented for 4.5 seconds each followed by presentation of a crosshairs for 7.5 seconds.

Scanning took place during the first and last 1.5 seconds of crosshair presentation with a 4.5 second silent period in between (see Figure 12).

Stimulus types were randomized within each run and delivered using E-prime software and overt responses were digitally recorded. Responses were analyzed for accuracy naming latency. All participants wore noise attenuating headphones in the scanner. Stimuli were projected on a screen located behind the scanner, visible via a mirror angled above the participant's head.

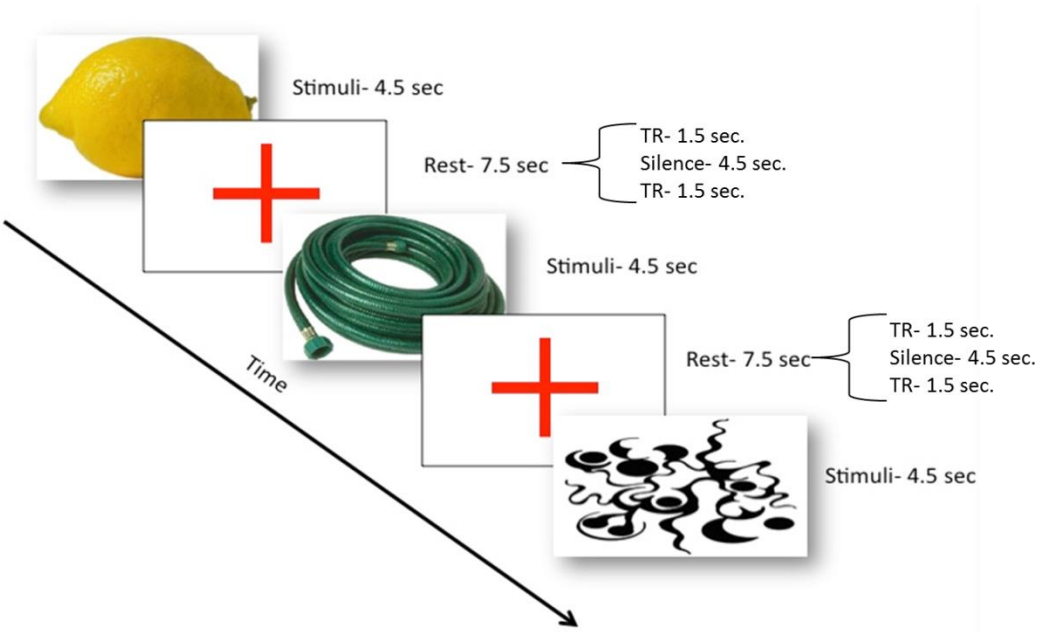


Figure 12. Sample of stimuli presentation in the scanner. Participant names stimulus or says, "pass" within 4.5 seconds, rests for 7.5 seconds during which time two volumes are collected (1.5 second each and 4.5 seconds apart), names the next stimulus item, rests again for scanning, etc., for a total of 25 items per run. The stimulus item with the black squiggles served as the control trial.

Scanner task practice

Participants were instructed to name the stimulus item as quickly as possible or to say, "pass" or "no" if they were unable to name the picture within 4.5 seconds. They were also to

say “pass” or “no” when they saw the control stimuli. Participants were asked to stay as still as possible in the scanner and to name the word they saw as quickly as possible using a loud, clear voice. They were to be silent at the appearance of the red crosshairs which appeared after each stimuli and control indicating that volumes were being acquired. Refer to Figure 12 for an example of stimuli presentation and timing.

All practice items were delivered via E-prime in order to simulate the actual scanner experience. Practice sessions took place at the UConn clinic on a laptop computer using unused stimuli two to three days before each scan and again 30 minutes prior to scanning at the scanning site. In addition, a reminder to “say the name as quickly as you can or say ‘pass’” was given immediately prior to task initiation. Participants repeated the task until they reported feeling comfortable with the protocol.

FMRI data analyses

The fMRI data of the four participants were analyzed individually to avoid problems associated with small sample sizes and grouping data from patients with heterogeneous lesions (Kiran et al., 2012). The goal of analysis was to identify voxels that changed in signal correlating with the timing of stimulus presentation throughout the experiment.

Preprocessing. Functional MRI data were processed using Analysis of Functional Neuroimages (AFNI; Cox, 1996). Functional data sets were corrected for slice acquisition time so it is as if the data were all acquired at the beginning of each TR. Runs were concatenated and motion-corrected using a six-parameter rigid-body transform (Cox & Jesmanowicz, 1999). Functional MRI volumes with movement in any direction greater than 5 mm were discarded. Functional data sets were co-registered to the structural images, resampled to 3 mm^3 voxels and transformed to Talairach and Tournoux space. Smoothing of the functional data was performed

by averaging data points with their neighbors in order to optimize signal. This was performed with a 4-mm Gaussian kernel, and converted to percentage-signal-change units. Random voxel artifacts outside the brain and lesioned areas were masked out.

Lesions were traced manually by the author, slice by slice, using AFNI software. Lesions were defined as areas of hypointensity as compared to homologous contralateral tissue. Each consists of a black core (isointense to cerebral spinal fluid (CSF)) and tends to be surrounded by necrotic tissue (hypointense, but not isointense to CSF). These areas were defined in order to calculate lesion volume, to determine what percentage of the lesion overlapped with various brain structures and to create lesion masks used to analyze activation perilesionally.

Behavioral Analyses. Accuracy of responses and naming latency were analyzed for each individual as they corresponded with changes in neural activation in regions of interest (ROIs) including the Inferior Frontal Gyrus (IFG), the Superior Temporal Gyrus (STG) and the Middle Temporal Gyrus (MTG) as these are three areas comprise an important network in lexical tasks in the LH for non-brain injured individuals (Schuhmann, Schiller, Goebel, & Sack, 2012). Naming accuracy data were analyzed using a within subject 1 x 4 ANOVA with the factors of accuracy and time (Scan One, Scan Two, Scan Three, Scan Four) to test for a main effect of time.

Functional Data Analysis. Functional data were analyzed by concatenating the runs from all four scan sessions and modeling each condition (TE, UTE, TH, UTH, control) as five separate regressors, one for each scan session, for a total of 25 unique regressors. General linear tests were used to examine patterns of activation for various conditions. In order to determine effects of change due to treatment the following linear contrasts were examined for each participant, serving as his or her own control. Error responses were included in the analysis.

- pre- vs. post-Treatment Period I (Scan One vs. Scan Two)
- pre- vs. post- Treatment Period II (Scan Two vs. Scan Three)
- pre- vs. post- both treatment periods (Scan One vs. Scan Three)
- pre- vs. 8-weeks-post- both treatment periods (Scan One vs. Scan Four)
- post-Treatment Period II vs. Follow-up eight-weeks-post treatment completion (Scan Three vs. Scan Four)
- Hard vs. Easy stimuli across all scans
- Trained vs. Untrained stimuli across all scans

To assess patterns of activation, each participant's preprocessed functional data were submitted to a regression analysis. We used a stereotypic hemodynamic response time for each individual to convolve with the start time of each trial. Best practices dictate that unique hemodynamic responses (HDRs) be created for individuals following stroke since they may be delayed and can influence detection of fMRI activation in some brain areas (Bonakdarpour, Parrish, & Thompson, 2007; Kiran et al., 2012). Given the timing of this protocol with only two scans following each stimulus response, unique HDRs could not be estimated. The slow event timing used in this protocol ensures no overlap between conditions and therefore we are looking at the same time point following each stimulus presentation. If the hypothesized HDR is a poor match to the actual HDR, results would be an underestimate of actual activation response. Importantly, because each participant serves as his or her own control, alack of fit between HDRs should be consistent for any individual.

The hemodynamic response model was entered into a regression analysis. The regression analysis returned a by-voxel fit coefficient which indicated the degree of coherence between BOLD fluctuation in that voxel and the predicted response generated by the stimuli. In other words, at each voxel, a model derived from the timing of the stimulus presentations is fit to the data to determine if there was an effect at that location. These by-voxel fit coefficients were used to run the contrasts stated above. Data were analyzed qualitatively at an individual level to describe changes in activity over time.

Due to the sheer number of tests being performed (at every voxel in the brain), correction for multiple comparisons is necessary to protect against false positives or Type I error inflation. In typical statistical tests, a Bonferonni correction would be used, however in neuroimaging this is not reasonable and would result in false negative or Type II error. The Monte Carlo simulation was used to determine how big a contiguous cluster of voxels, each one significant at an uncorrected threshold of $p = 0.05$, has to be in order to be significant at a threshold p that is corrected for multiple comparisons. See Appendix B for results of Monte Carlo simulation for each participant. P values of .025 and voxel clusters of 28 were used to demonstrate Trained vs. Untrained contrasts for all participants. Activation in the other contrasts was so pronounced in the two mild participants that a 0.001 p value was used with cluster size of 20 (more stringent than the corrected threshold of 6) in order to best demonstrate areas of significance. In the participants with more severe aphasia, activation was slightly less robust and a p value of 0.01 was used with a cluster size of 20. Note that all thresholds met significance at $p = 0.05$.

Region of Interest Analysis.

Activation was analyzed within the ROIs described above allowing for qualitative comparison of main effects across the same participants for each scan period. ROIs were defined on the basis of pre-set anatomical masks provided by AFNI (Lancaster et al., 2000). Time-series graphs were generated for each participant by calculating the mean of all activated voxels within each ROI.

Intact functional connectivity in the left STG has been associated with better language comprehension recovery (Warren, et al., 2009). These areas are less well studied following stroke though they are often referenced when discussion peri- versus contralesional recovery patterns. For example, the right IFG is thought to play a more important role in language

processing in those with larger lesions but has been associated with more naming errors (Fridriksson, Bonilha, Baker, Moser, & Rorden, 2010b). Lesions in the left MTG were associated with poor treatment outcome following a cueing treatment approach (Fridriksson et al., 2010b). The right MTG is thought to be a key region for recovery of semantic processing especially when the task demand is increased (Sebastian & Kiran, 2011). The STG and the MTG were major areas of infarct for all four of the participants and the IFG was an area of impact for M1 and S1 (refer to Table 15 for lesion characteristics for each participant).

The mean percent signal change averaging across each hemisphere of the brain was calculated, and laterality was calculated used the following formula (left-right)/(absolute value of left+ absolute value of right) (Seghier, 2008); 1.0 represented completely left lateralized activity and -1.0 represented completely right lateralized activity. Patterns of lateralization over the course of treatment were assessed by plotting these over time. Shift or failure to shift in lateralization was compared to naming performance. While whole brain lateralization provides a gross measure of whether the hemispheric activation has changed in regard to processing this particular language task, it is possible that large rightward shifts of some brain regions may be obscured by equal leftward shifts in other regions and vice versa. For this reason, laterality was also calculated for each ROI.

Chapter V

Results

In order to analyze change over time, each participant's results are described individually (for the two mild, M1, M2 and two severe participants S1, S2). Results are reported for each of the dependent measures including standardized language measures, percent accuracy at each level of treatment for trained and untrained materials, response time for trained and untrained materials for those with mild aphasia, productivity and efficiency of connected speech.

Findings from neuroimaging follow the behavioral results for each participant. These include contrasts for the three main conditions: task difficulty (easy vs. hard words), effect of training (trained vs. untrained words) and time (scan1 vs. Scan Three). Change in percent signal in language areas including the inferior frontal gyrus, the superior frontal gyrus and the medial temporal gyrus over the four scans will also be reported.

M1.

Standardized assessment.

M1 participated in all 60 hours of treatment and attended all baseline, assessment and follow-up sessions as scheduled. He presented with a pre-treatment WAB-AQ score of 95 of 100; 92% accuracy on the BNT and 91% accuracy on the C-RTT shown in Table 20. M1's oral expressive language production was slow and characterized by frequent circumlocutions (e.g., "book carrying device" for backpack). Considering that this participant was already performing near ceiling levels for the standardized tests, gains considered clinically significant (e.g., 20% change or five points on the WAB-AQ) were not anticipated. He did, however make a five point increase on the WAB-AQ with a final score of 97 immediately following the second treatment period. This change is attributed to increases on the fluency and object naming subtests. The

change on the BNT, from 92%-97%, was not surprising given that verbal production was targeted. However, gains on the writing subtests of the WAB, from 80%-98%, and on the Raven's Coloured Progressive Matrices (RCPM), 89% -97%, cannot be similarly explained. All other changes were either negligible or the sensitivity of the measures hit ceiling.

Table 20

M1- Summary of Assessment Scores at each testing period.

Assessment	Pre-tx	Post-tx 1	Post tx 2	Follow-up
BNT	92%	93%	94%	97%
CRTT	91.33%		94.00%	91.73%
WAB AQ	95%	97.6%	99.6%	97.8%
WAB CQ	95.2%		98.7%	98.6%
WAB LQJ	93.5%		98%	98.5%
Subtests from the Western Aphasia Battery				
spontaneous speech	95.00%	95.00%	100.00%	97.50%
auditory verbal comprehension	100.00%	100.00%	100.00%	100.00%
repetition	96.00%	100.00%	100.00%	100.00%
naming and word finding	89.00%	98.00%	98.00%	94.00%
object naming	93.30%	100.00%	100.00%	100.00%
word fluency	65.00%	90.00%	90.00%	70.00%
sentence completion	100.00%	100.00%	100.00%	100.00%
responsive speech	100.00%	100.00%	100.00%	100.00%
reading score	100.00%		100.00%	100.00%
writing score	80.00%		95.00%	98.00%
apraxia score	98.30%		100.00%	100.00%
constructional, visuospatial and calculation score	95.00%		94.00%	99.00%
RCPM	89.19%		83.78%	97.30%

Note. All scores shown as percent of the maximum score. BNT-Boston Naming Test; CRTT-Computerized Revised Token Test; WAB AQ-Western Aphasia Battery Aphasia Quotient; WAB CQ-Western Aphasia Battery Cortical Quotient; WAB LQ-Western Aphasia Battery Language Quotient; RCPM-Raven's Coloured Progressive Matrices

Probes of trained and untrained material.

Accuracy and response times were recorded for probes of trained and untrained stimuli from levels four through eight. Results are summarized in multiple baseline formats representing

percent accuracy and response time (in seconds). See Figures 13 and 14. These figures depict performance on trained and untrained materials over six phases: pre-treatment, Treatment Period I, no-treatment, Treatment Period II, immediate post-treatment and eight weeks post-treatment. M1 achieved criteria on each level prior to the other participants in his group.

Accuracy of untrained materials increased to nearly the same extent as trained materials at all levels. Much of the treatment time occurred in Levels Five (Treatment Period I) and Eight (Treatment Period II) based on criterion achievement for all group members. M1 demonstrated large effect sizes for both trained and untrained materials at both levels (see Table 26 for effect sizes for all participants). Level Five was completed with effect sizes of 24.3 (large) for trained and 25.6 (large) for untrained materials. Level Eight was completed with effect sizes of 14 (large) for trained and 12.3 (large) for untrained stimuli. Maintenance data for Level Five was collected six times and over a period of 15 weeks. For Level Eight there were only two follow-up data points with which to calculate effect size.

M1's response times tended to increase prior to achieving criteria and then to steadily decrease through the maintenance periods. Improvements (decreases) in response times at some treatment levels were not maintained. Response times for untrained materials did not improve (decrease) to the same extent as the trained materials.

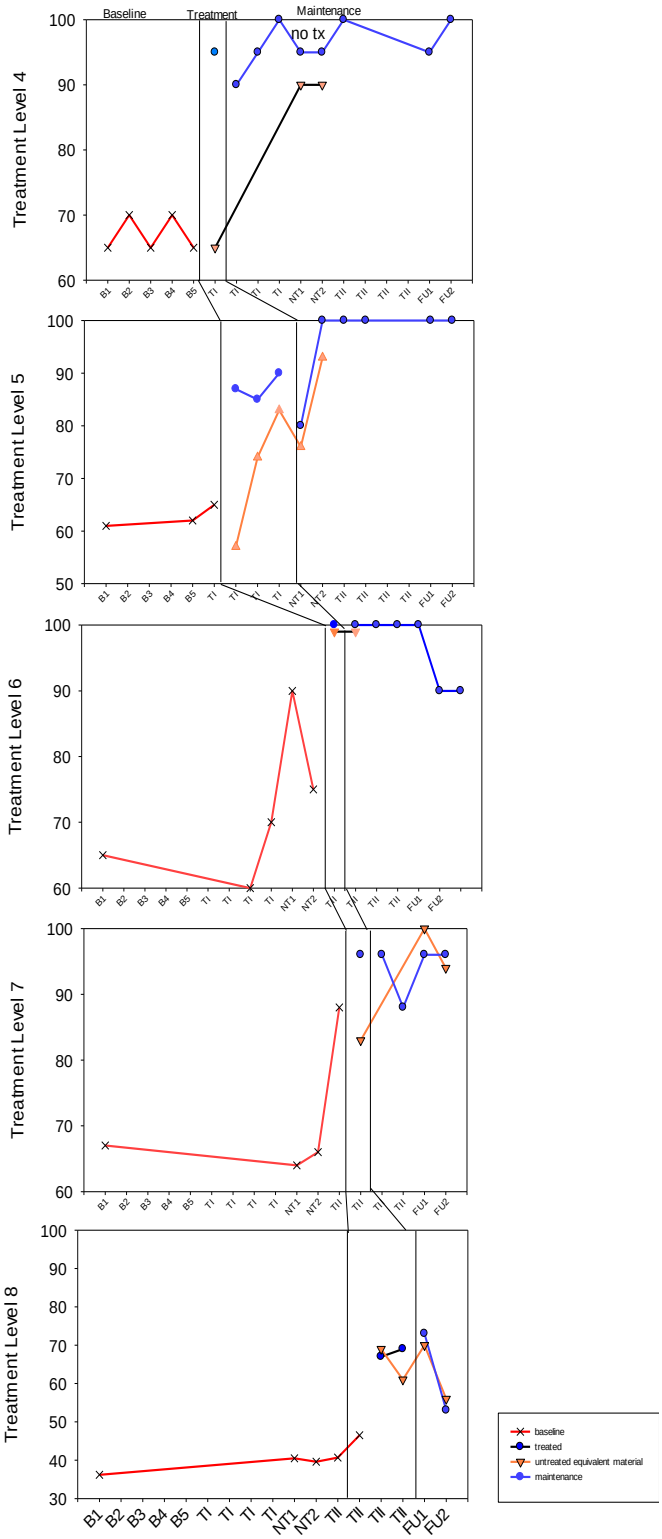


Figure 13. M1-Percent Accuracy for Treatment Levels 4-8. B-Baseline; TI-Treatment Period I; NT-no treatment period; TII-Treatment Period II; FU1-immediate post treatment; FU2- eight weeks post treatment

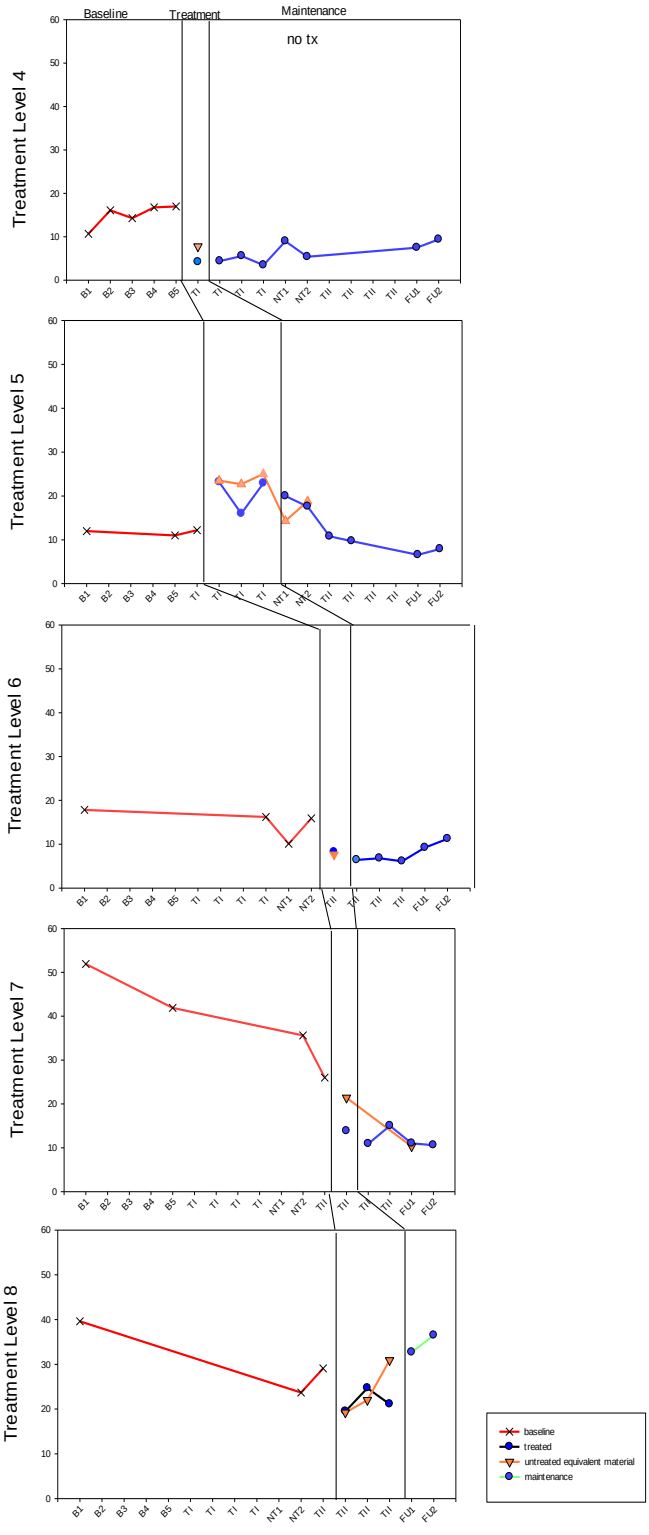
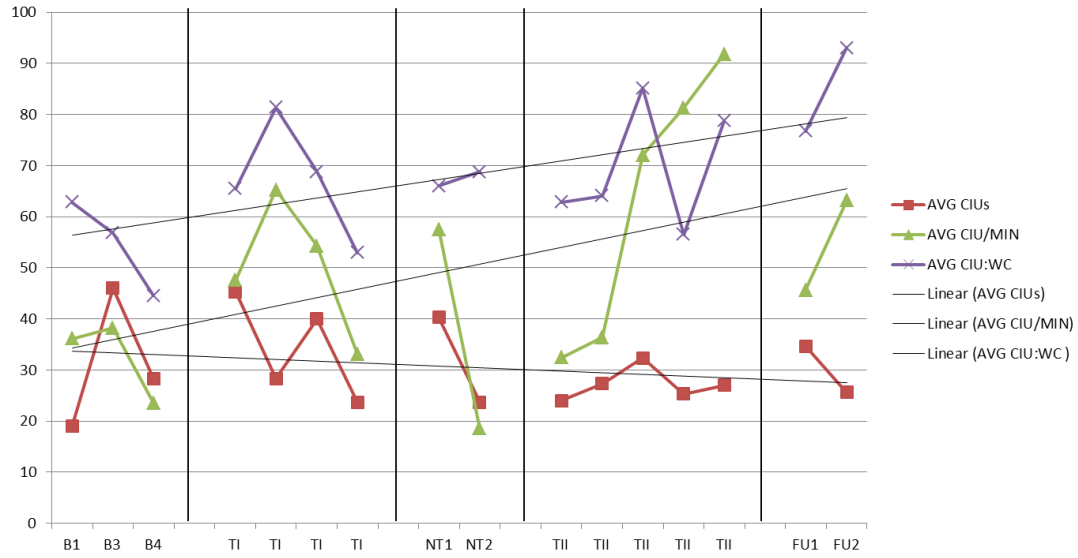


Figure 14. M1-Response Time (sec) for Treatment Levels 4-8. B-Baseline; TI-Treatment Period I; NT-no treatment period; TII-Treatment Period II; FU1-immediate post treatment; FU2- eight weeks post treatment

Generalization probes for narrative discourse.

Narrative discourse was sampled throughout the course of treatment. Three discourse measures, one index of productivity and two of efficiency of language production served as the generalization probes. Productivity was measured as the number of Correct Information Units (CIUs). For efficiency, the number of CIUs per minute and the proportion of CIUs to total word count were calculated. High variability was seen with both efficiency measures, though somewhat less for the proportion of CIUs per total word count as seen in Figure 15. Effect sizes were calculated for all measures. No change was seen in productivity, as anticipated for this participant. *Efficiency* of production was an outcome variable of interest for M1 since sheer productivity was not an area of deficit for this participant. Effect size for discourse efficiency using the proportion of CIUs to total word count was negligible following the first treatment period (1.8) and small (3.2) following the second. Great variability in response times resulted in minimal effect sizes for CIUs per minute but visual inspection reveals a greater slope for this measure than for any other. A steep, rising slope in CIUs per minute within Treatment Period II corresponds with treatment using the most complex stimuli and that this participant found most challenging. The steep increase in CIUs per minute was not maintained in follow-up testing but the proportion of CIUs to total words was maintained.



66 *Figure 15.* M1- Narative Discourse Probes. B1-4-baseline probes, Probes 1-8-treatment probes; NT1-immediate post first treatment period; NT2- immediate pre second treatment period; FU1-immediate post second treatment period; FU2- eight weeks post treatment.

Overt naming in the scanner.

Modest increases in naming accuracy (3%-10%) were observed for M1 between Pre-treatment (Scan One) and Post-Treatment I (Scan Two) for three of the four conditions. A 3% decrease was reported for the untrained easy condition. This may have been a reflection of performance variability seen in many people with aphasia. Additional increases were observed between Post-Treatment I (Scan Two) and Post-Treatment II (Scan Three) for all conditions, as shown in Figure 16. There was a 10% increase in accuracy on Trained Hard words between the first and second scan, reflecting effect of training of specific words. There was a 27% increase between the second and third scan, a 30% increase from the first to third scan on Untrained Hard Words. It is possible that having seen these words in the two previous scans impacted word learning but if so, accuracy gains might have been better distributed between each of the four scans. As it stands, the majority of the gain for this condition happened following the training

that most challenged this participant. Gains on all stimulus types were maintained eight weeks post treatment except for the Untrained Hard words which declined from post-Treatment Period II results by 6% but were 23% increased from pre-treatment levels.

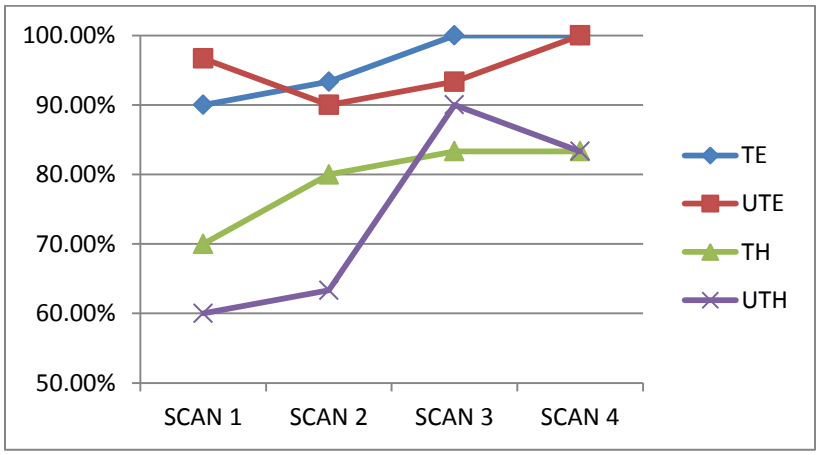


Figure 16. M1- Percent Accuracy for Scanner Naming. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Response times were consistent across scans for the Trained, Untrained Easy and Control conditions but decreased for the Trained and Untrained Hard conditions as seen in Figure 17. Response times were stable and slightly faster at the eight week follow-up scan.

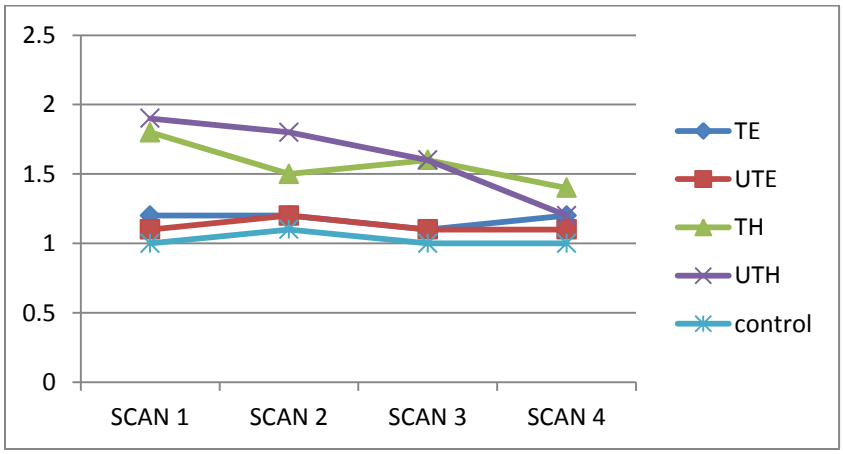


Figure 17. M1-Response times for scanner naming. Y axis denotes response times in seconds. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard; Control-"pass" response. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Contrasts.

Several linear contrasts were specified in order to examine differences for all conditions between each time period and for each condition across all time periods. Statistical maps were corrected for multiple comparisons by including only clusters that were significant at a corrected statistical threshold of $p < 0.001$ as determined by Monte Carlo simulations. Robust differences were seen for most contrasts after this correction and so we modified the threshold to only show clusters exceeding 20 voxels, though the Monte Carlo indicated that six was statistically significant. Less activation was seen for the trained vs. untrained condition so a threshold of $p < 0.025$ was used along with the 28 voxel clusters as determined by Monte Carlo simulation.

The three main contrasts of interest 1) Trained vs. Untrained words 2) Hard vs. Easy words and 3) Pre-Treatment (Scan One) vs. Post-Treatment (Scan Three) are displayed in Figure 18. Ten large clusters showed significantly greater activation for Untrained words relative to Trained words. Largest clusters were observed in the left thalamus, the right inferior occipital gyrus and right middle occipital gyrus and the left inferior occipital gyrus. Hard words showed significantly greater activation than easy words in 28 brain regions. Largest clusters were noted in the RIFG, right thalamus, left superior frontal gyrus and right medial frontal gyrus. Scan Three showed significantly greater activation than Scan One in 14 brain regions. Again, largest clusters were seen in right middle frontal gyrus, RSTG, bilateral cingulate gyrus, and left middle frontal gyrus. Mean activation peaks in all significant brain regions are shown for all three contrasts in Table C1, Appendix C.

The other contrasts were those of timing (scan1 vs. Scan Two, Scan One vs. Scan Three, Scan One vs. Scan Four, etc.) and best analyzed in relation to each other. A description of this analysis is described in the next section, Anatomical Regions of Interest. For all contrasts,

changes in activation were widespread throughout the brain as anticipated with large voxel clusters in regions associated with language functions including bilateral IFG and bilateral STG (see Table C1 in Appendix C and Figure 18).

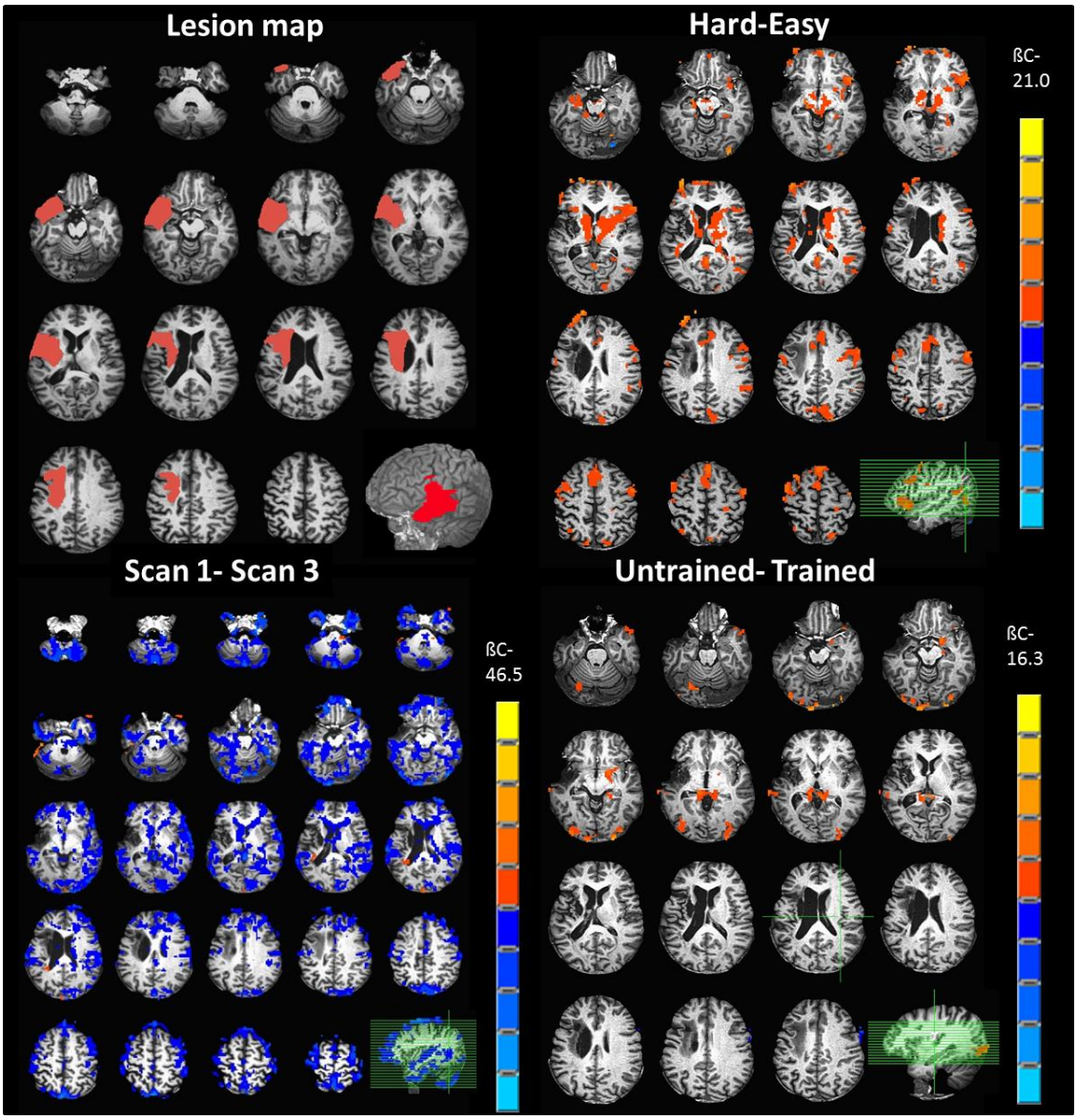


Figure 18. Lesion map and three contrasts for participant M1. ßC- beta coefficient for each contrast represents the intensity limits. Lesion map (top left) and three contrasts of interest: Hard-Easy (upper right) $p < .001$, 20 voxels; Scan One- Scan Three (lower left)- $p < .001$, 20 voxels. Untrained-Trained (lower right)- $p < .025$, 28 voxels. Blue represents more activation for the right side of the equation relative to the left as in Scan One- Scan Three. Red represents more activation for the left side of the equation relative to the right as in Hard-Easy.

Anatomical regions of interest.

Within each of the contrasts, large clusters were observed in the IFG, MTG and STG, the three language areas of interest. Preliminary analysis was done looking at these regions using anatomical ROIs. This allowed for qualitative comparison of activation across time for one individual. For M1, the IFG, STG and MTG all showed bilateral increases in the BOLD response. In the IFG and STG maximum changes were most evident following Treatment Period II (Scan Three). In the MTG, this change was most pronounced following Treatment Period I (Scan Two). Activation increased in both hemispheres but more so in the right than left in all ROIs. In the IFG, the mean percent signal increase between Scan Two and 3 for all conditions was 0.17 and 0.10 for the right and left hemispheres respectively. In the STG the mean increase for all conditions between Scan Two and Scan Three was 0.26 on the right and 0.10 on the left. In the MTG it was 0.26 on the right and 0.22 on the left.

The pattern of activity in bilateral IFG best reflected differences in the varying difficulty levels of stimuli presented (see Figure 19). The greatest to least activation was observed in this order: Trained Hard, Untrained Hard, Untrained Easy, Trained Easy and finally the control condition. At Pre-Treatment (Scan One) this differentiation was much less evident. The control condition elicited the least activation but the other four conditions were nearly equivalent. After Treatment Period I, more difference was seen between the Easy and Hard conditions and by Scan Three there was a clear difference between each of the conditions in what would be the expected order of difficulty. At the Follow-up (Scan Four), there was a decrease in activation in the right IFG for all conditions (mean change in percent activation = -0.07) relative to Post-Treatment Period II (Scan Three). Activation was nearly identical for the four conditions of interest and they were only slightly increased over the control condition. Since response accuracy was

maintained, it is possible that less effort was required resulting in less neural activation. The left IFG showed, however, continued greater activation for the Untrained Hard condition and for the two Easy conditions at this final time point and a decrease for the Trained Hard condition.

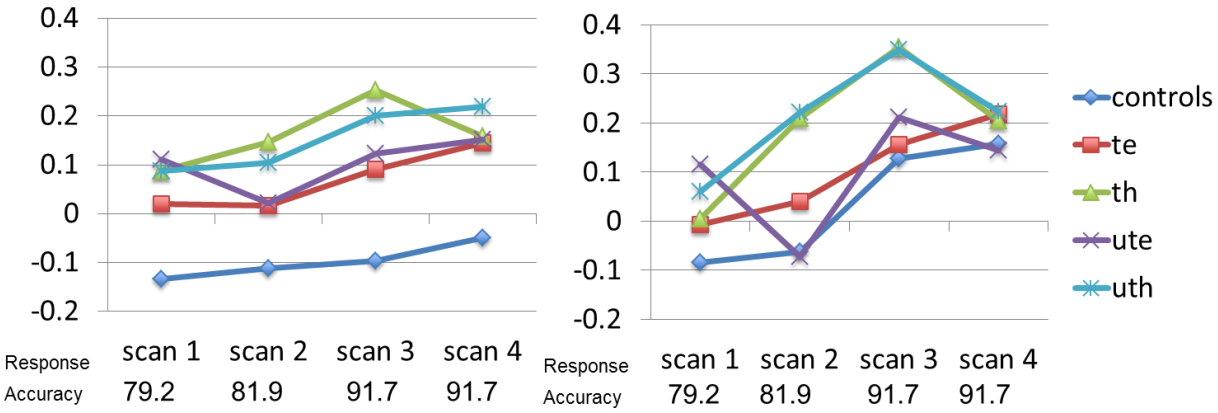


Figure 19. M1-Percent Signal Change in the Left (on left) and right (on right) Inferior Frontal Gyrus. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Bilateral STG also differentiated fairly well by condition though in this case the control condition elicited more activation than did the trained easy condition for every scan excepting the first (see Figure 20). As in the IFG, the distinction between all conditions was most pronounced at Scan Three and the least at Scan Four, in particular for the Hard vs. Easy condition, but also for the Trained vs. Untrained condition. There was a 0.25 percent signal change for the Trained Hard condition in the right STG from Pre-Treatment (Scan One) to Post-Treatment Period I (Scan Two) and an additional 0.40 percent signal change from Post-Treatment Period I (Scan Two) to Post-Treatment Period II (Scan Three). This total change of 0.61 was greater than the change seen for any other condition for all three ROIs. At Follow-up (Scan Four), activation in the right decreased by 0.43 percent indicating potentially less effort for this condition though activation was still increased more from pre-treatment levels (Scan One

to Scan Four change = .18 percent signal change) compared to the left STG (Scan One to Scan Four change = 0.19 percent signal change). Mean percent activation change for all conditions reflected this pattern with mean changes of 0.19 and 0.14 for differences from Scan One to Scan Four in the right and left STG respectively. The left STG accounted for 13.3% of M1's lesion which perhaps accounts for the reduced increase in activation as compared to the right STG.

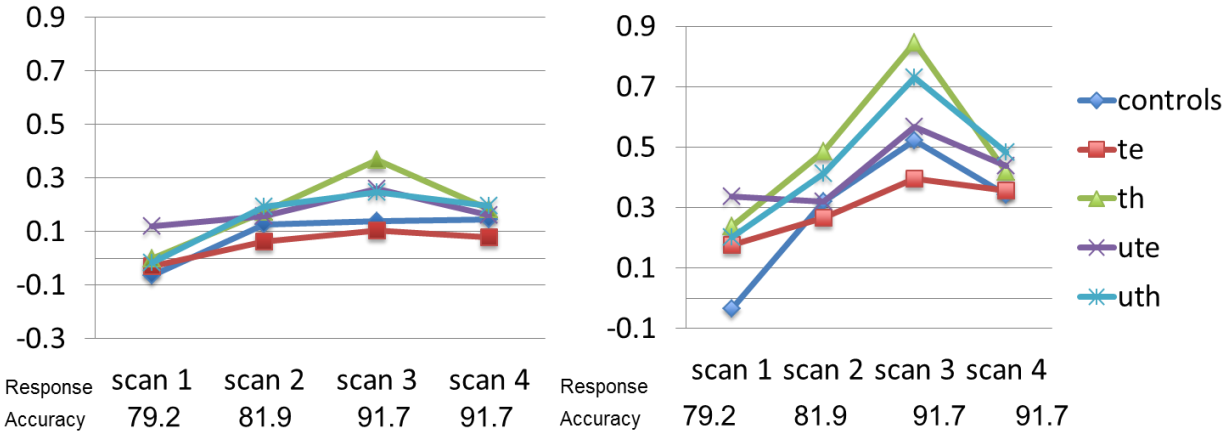


Figure 20. M1- Percent Signal Change in Left (on left) and right (on right) Superior Temporal Gyrus. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Unlike in the previous two ROIs, the largest increases in activation in the bilateral MTG were observed following Treatment Period I (Scan Two; see Figure 21). At Scan Three, following Treatment Period II, continued, smaller increases were observed for the Hard conditions, both Trained and Untrained but not for Easy conditions. Activation decreased or plateaued at Follow-up (Scan Four). Like the IFG, the right MTG seems to effectively differentiate between conditions in order of difficulty or effort level for the participant at the third scan and following Treatment Period II.

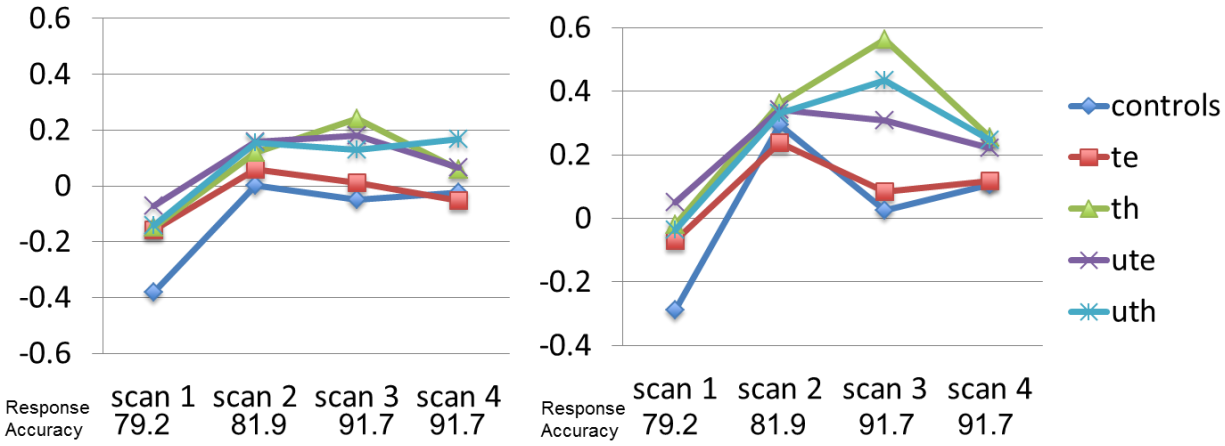


Figure 21. M1- Percent Signal Change in Left (on left) and right (on right) Middle Temporal Gyrus. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Laterality.

The strength of activation in the right and left hemispheres were calculated based on mean percent signal change across all conditions. In the first two scans, left hemisphere activation exceeded the right. Increased activation shifted rightward by the third scan and remained so for the follow-up scan eight weeks post-treatment (see Figure 22).

Laterality for each region of interest was plotted against the whole brain in order to see relative change in each area. For M1, the three ROIs do not account for the observed whole-brain shift rightward. The IFG does make a subtle shift rightward but the STG actually has reduced RH activation following Treatment Period I and the MTG shows a relatively stable RH preference.

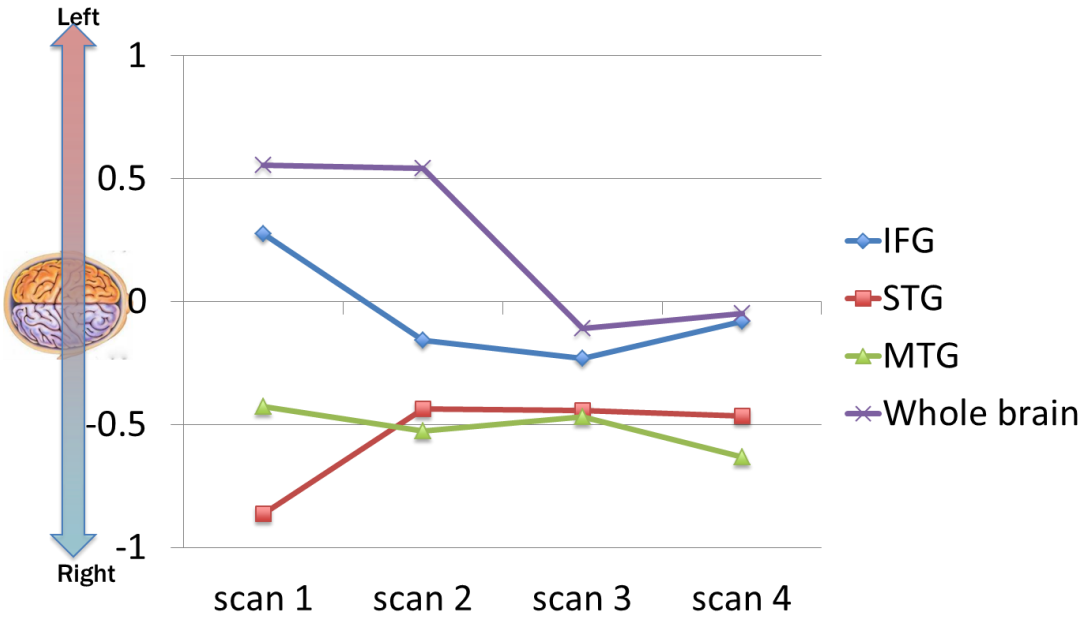


Figure 22. Mean percent signal change across all conditions was used to calculate laterality in the Inferior Frontal Gyrus (IFG), Superior Temporal Gyrus (STG), Middle Temporal Gyrus (MTG) and in the Whole Brain. Mean percent signal change (MPSC) of the left hemisphere (LH)- MPSC of the right hemisphere (RH)/ MPSC LH+MPSC RH. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

M2

Standardized assessment

M2 participated in all 60 hours of treatment and attended all baseline, assessment and follow-up sessions as scheduled. He presented with a pre-treatment WAB-AQ score of 88 of 100; 77% accuracy on the BNT and 72% accuracy on the C-RTT shown in Table 21. This participant's oral expressive language production was effortful, characterized by frequent self-revision, self-talk ("slow down!") and infrequent phonemic paraphasias. M2 made modest increases on standardized tests after Treatment Period I including an increase of 8.3 percentage points on the BNT and a change of 5.8 points on the WAB-AQ. Lesser increases followed the second treatment period with additional increase of 1.7% and 2% on the BNT and WAB AQ

respectively. Changes on the WAB AQ were attributed to increases on the fluency, object naming subtests and also in repetition. The change on the BNT, from 78%-90% mirrored the change seen on object naming subtests of the WAB. Like the previous participant, M2 also increased on the writing subtests of the WAB, from 88%-98% with the maximum change observed during at the eight week follow-up testing, and on the Raven's Coloured Progressive Matrices (RCPM), 89% -97%. Maximum change for this measure occurred following Treatment Period II. Auditory comprehension on the WAB AQ was relatively stable and a 9% maximum increase was observed on the C-RTT, a more sensitive measure of auditory comprehension.

Table 21

M2- Summary of Assessment Scores at each testing period

Assessment	Pre-tx	Post-tx 1	Post tx 2	Follow-up
BNT	76.67%	85.00%	86.67%	90.00%
CRTT	72.00%	n/a	83.33%	81.33%
WAB AQ	88%	93.80%	95.80%	93.90%
WAB CQ	90.30%	n/a	94.50%	94.70%
WAB LQJ	87.60%	n/a	94%	95.00%
		WAB Subtests		
spontaneous speech	95.00%	95.00%	100.00%	97.50%
auditory verbal comprehension	91.00%	96.00%	96.00%	85.50%
Repetition	69.00%	86.00%	87.00%	90.00%
naming and word finding	88.00%	97.00%	96.00%	94.00%
object naming	96.67%	100.00%	100.00%	100.00%
word fluency	50.00%	85.00%	80.00%	70.00%
sentence completion	100.00%	100.00%	100.00%	100.00%
responsive speech	100.00%	100.00%	100.00%	100.00%
reading score	96.00%	n/a	96.00%	100.00%

writing score	88.50%	n/a	86.00%	98.00%
apraxia score	98.30%	n/a	100.00%	100.00%
constructional, visuospatial and calculation score	89.00%	n/a	89.00%	95.00%
RCPM	86.49%	n/a	94.59%	89.19%

Note. All scores shown as percent of the maximum score. BNT-Boston Naming Test; CRTT-Computerized Revised Token Test; WAB AQ-Western Aphasia Battery Aphasia Quotient; WAB CQ-Western Aphasia Battery Cortical Quotient; WAB LQ-Western Aphasia Battery Language Quotient; RCPM-Raven's Coloured Progressive Matrices

Probes of trained and untrained material

Accuracy and response times were recorded for probes of trained and untrained stimuli from levels four through eight. Results are summarized in multiple baseline formats representing percent accuracy and response time (in seconds). See Figures 23 and 24. These figures depict performance on trained and untrained materials over six phases: pre-treatment, Treatment Period I, no-treatment, Treatment Period II, immediate post-treatment and eight weeks post-treatment.

Accuracy of untrained materials increased to nearly the same extent as trained materials at all levels. There was a predictable increase in performance for each level prior to commencement of training with a much sharper spike once training began. Increases on untrained and yet-to-be trained material indicate successful generalization.

Much of the treatment time occurred in Levels five (Treatment Period I) and eight (Treatment Period II) based on criterion achievement for all group members. M2 demonstrated large effect sizes for both trained and untrained materials at Level Five and medium effect sizes at Level Eight (see Table 26). Level Five was completed with effect sizes of 13.9 (large) for trained and 9.8 (large) for untrained materials. Maintenance data for Level Five was collected six times and over a period of 15 weeks. Effect sizes for Level Eight were also noted for the trained

and untrained materials and were 7.4 (medium) and 6.5 (medium) respectively. For Level Eight there were only two follow-up data points with which to calculate effect size.

M2's response times were either stable or decreased over time. At Level Four, treatment marked the period of decreased response time which was maintained. At Level Five, there was an initial increase in response time compared to baseline, as accuracy improved and then a steady decrease in response time through the maintenance period. At Level Six, there was a decrease in response time at treatment initiation which was generally maintained. At Levels Seven and Eight, decreases in response times were seen in the baseline phase and continued through treatment. Decreases were maintained for Level Seven but not for Level Eight.

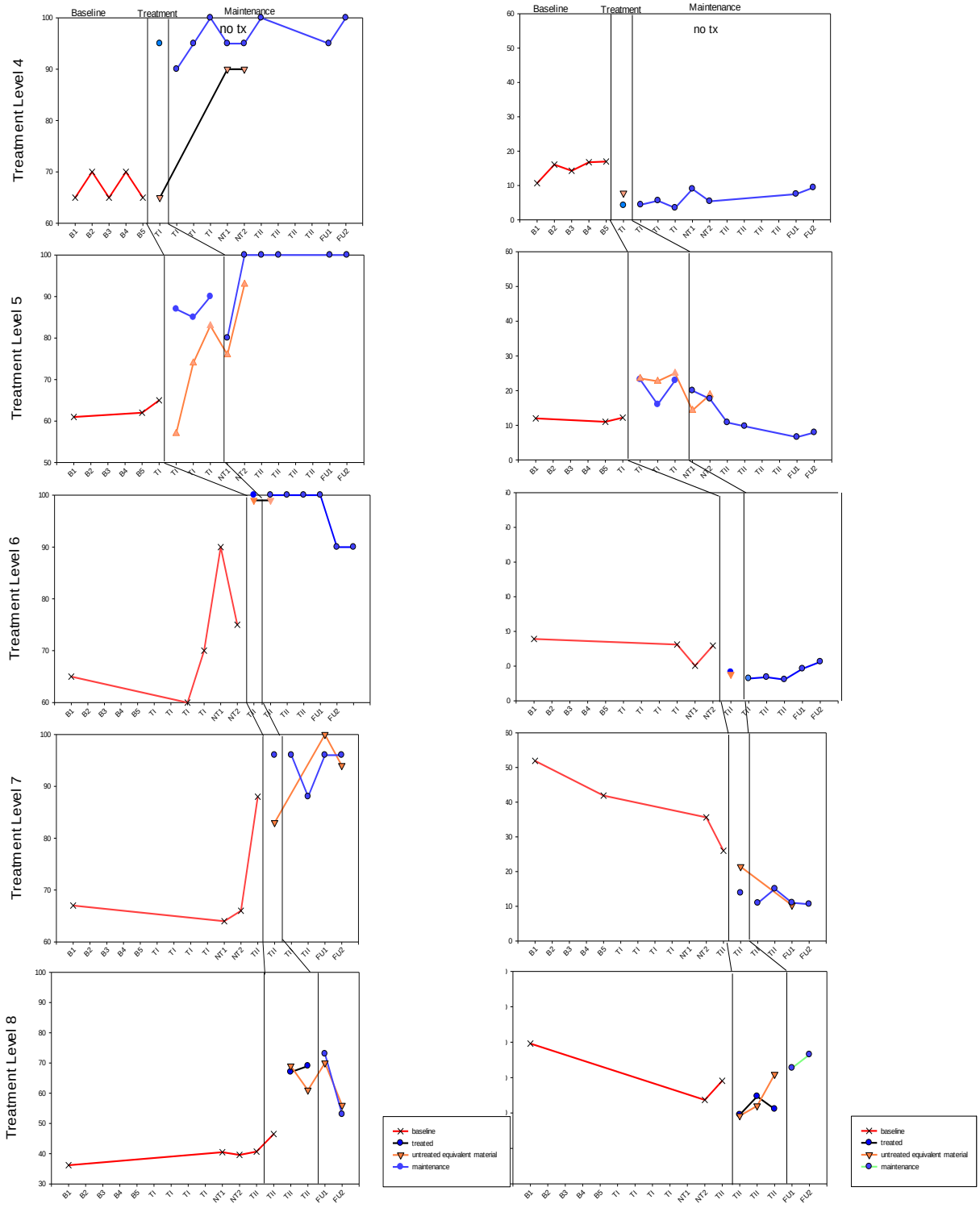


Figure 23. M2-Percent Accuracy on Treatment Levels 4-8. B-Baseline; TI-Treatment Period I; NT-no treatment period; TII-Treatment Period II; FU1-immediate post treatment; FU2- 8

Figure 24. M2-Response Time (sec) on Treatment Levels 4-8. B-Baseline; TI-Treatment Period I; NT-no treatment period; TII-Treatment Period II; FU1-immediate post treatment; FU2- eight weeks post treatment

Generalization probes for narrative discourse

As with M1, M2 demonstrated high variability in both efficiency measures as shown in Figures 25 and 26. Also like M1, no change was observed or expected in productivity and efficiency of oral verbal production was the outcome variable of interest. For the other three participants these measures are shown in one figure but the number of CIU's produced by M2 skewed the scale such that it was difficult to see the trend in the other measures. Therefore, separate figures for efficiency and productivity are shown. Effect size for discourse efficiency using the proportion of CIUs to total word count was negligible following the first treatment period (1.8) and medium (5.8) following the second. This participant had a much more stable baseline for CIUS per minute yielding a medium negative effect size (4.4) after Treatment Period I and a very large effect size (31.4) following Treatment Period II. Again, using visual inspection to inform these results, in this case it is clear that there was little change throughout and following Treatment Period I and a moderate increase in slope within and following Treatment Period II.

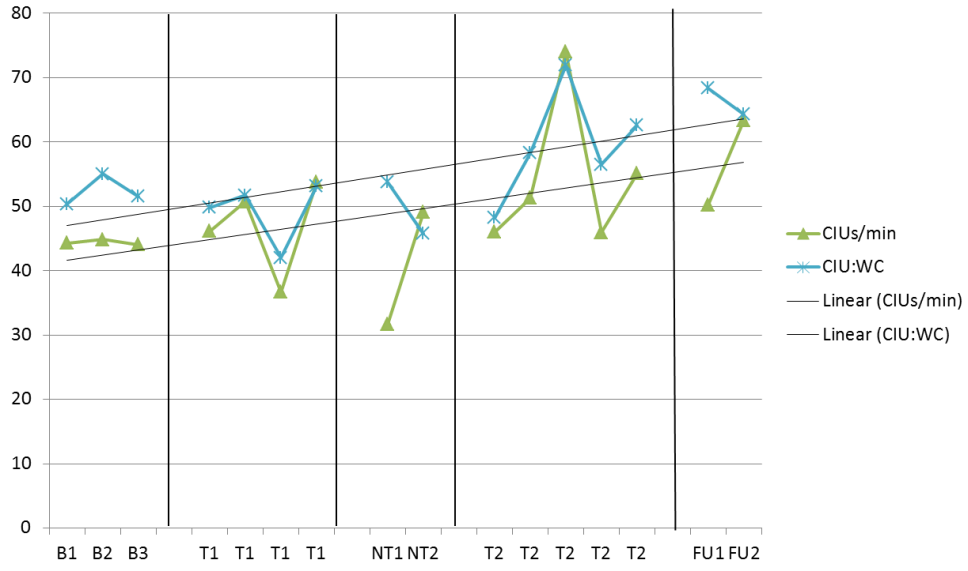


Figure 25. M2- Narative Discourse Probes- Efficiency. B1-4-baseline probes, Probes 1-8-treatment probes; NT1-immediate post first treatment period; NT2- immediate pre second treatment period; FU1-immediate post second treatment period; FU2- eight weeks post treatment. CIUs- Correct Information Units

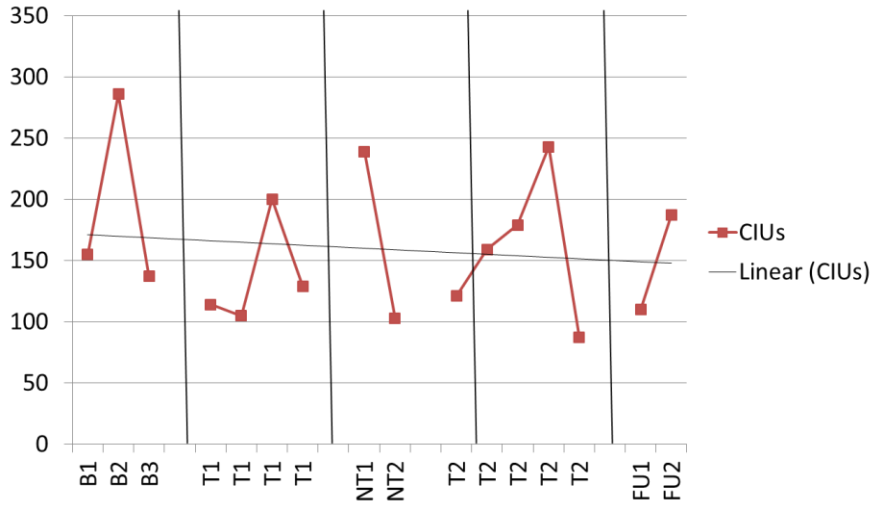


Figure 26. M2-Narrative Discourse Probe-Productivity. B1-4-baseline probes, Probes 1-8 treatment probes; NT1-immediate post first treatment period; NT2- immediate pre second treatment period; FU1-immediate post second treatment period; FU2- eight weeks post treatment.

Overt naming in the scanner.

M2 made a 7% increase in naming accuracy between Pre-Treatment (Scan One) and Post-Treatment I (Scan Two) for Trained Hard words as shown in Figure 27. No change was noted for Trained Easy and Untrained Hard words and there was a 7% decrease for the Untrained Easy condition. A two word improvement or decline (7%) would be attributable to the day to day variability seen in this participant's performance therefore it would appear as if Treatment Period I did not impact naming accuracy for this participant. Increases in performance were more pronounced from Post-Treatment I (Scan Two) to Post-Treatment II (Scan Three) with a 16% increase on Trained Hard materials and a 13% increase on Untrained Hard materials. A 7% increase and 3% decrease on Untrained Easy and Trained Easy materials, respectively, would again be attributed to day to day variability. Trained and Untrained Hard materials continued to increase at the fourth follow-up scan for a total increase of 30% and 20%, respectively. As with M1, it is possible that having seen these words in previous scans impacted word learning. Increases could also be result of successful training for Trained Hard words and generalization for Untrained Hard words. Accuracy on Trained and Untrained Easy words were maintained throughout all treatment periods and through follow-up with one to two word differences seen from session to session.

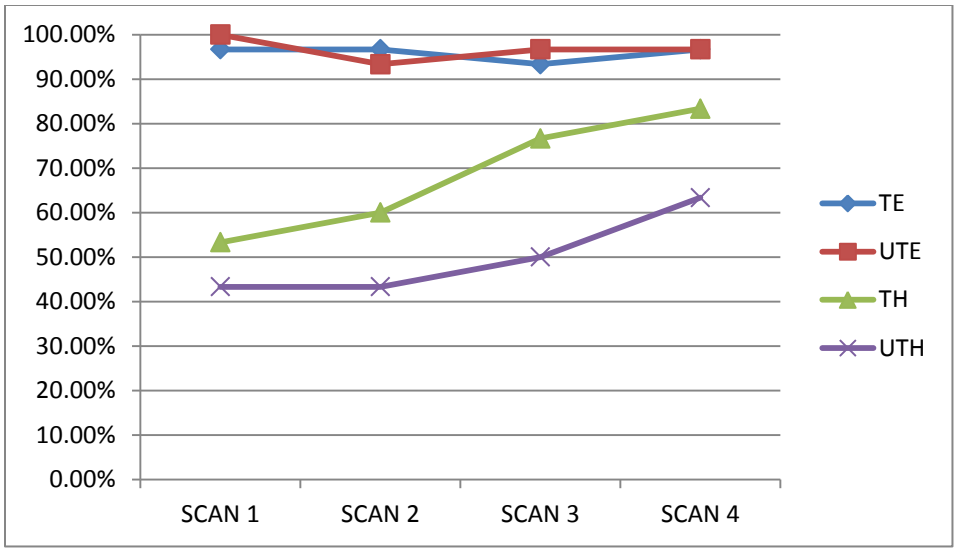


Figure 27. M2-Percent Accuracy for Scanner Naming. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Response times did not vary much across scans for the Trained, Untrained Easy and Control conditions with fastest responses recorded for the "pass" in the control condition, as expected. A decreased was observed for Trained Hard materials after Treatment Period I, demonstrating the effect of practice and an increase was seen for Untrained Hard materials as seen in Figure 28. Response times for three conditions was slightly increased at the follow-up scan, eight weeks post treatment, but was stable for the Trained Hard condition where the greatest gains in accuracy was seen.

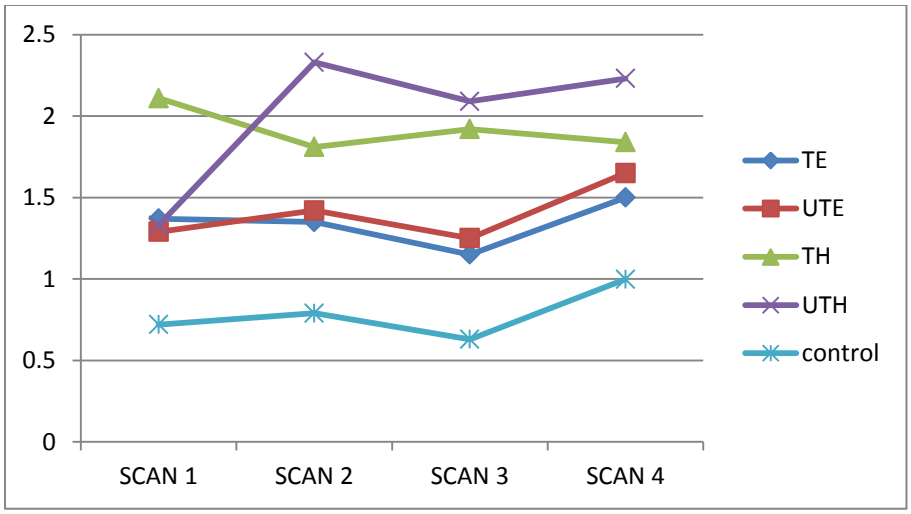


Figure 28. M2-Response times for scanner naming. Y axis denotes response times in seconds. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard; Control-"pass" response. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Functional neuroimaging data.

Several linear contrasts were specified in order to examine differences for all conditions between each time period and for each condition across all time periods. Statistical maps were corrected for multiple comparisons by including only clusters that were significant at a corrected statistical threshold of $p < .001$ as determined by Monte Carlo simulations. As with M1, robust differences were seen for most contrasts after this correction and so we modified the threshold to only show clusters exceeding 20 voxels, though the Monte Carlo indicated that six was statistically significant. Less activation was seen for the trained vs. untrained condition so a threshold of $p < .025$ was used along with 28 voxel clusters determined by Monte Carlo simulation.

The three main contrasts of interest a) Untrained vs. Untrained words b) Hard vs. Easy words and c) Pre-Treatment (Scan One) vs. Post-Treatment (Scan Three) are displayed in Figure 29. Eight large clusters showed significantly greater activation for Untrained words relative to

Trained words. Largest clusters were observed in the right superior frontal gyrus, the right precuneus and the left IFG. Hard words showed significantly greater activation than easy words in 25 brain regions. Largest clusters were noted in the bilateral anterior cingulate, the left lentiform nucleus, and bilateral subcallosal gyri. Scan Three showed significantly greater activation than Scan One in 47 brain regions. Again, largest clusters were seen in bilateral cuneus, right, precental gyrus and left IFG. Mean activation peaks in all significant brain regions are shown for all three contrasts in Table C2, Appendix C.

The other contrasts were those of timing (Scan1 vs. Scan Two, Scan One vs. Scan Three, Scan One vs. Scan Four, etc.) and were best analyzed in relation to each other. A description of this analysis is described in the next section, Anatomical Regions of Interest. For all contrasts, changes in activation were widespread throughout the brain, as anticipated, with large voxel clusters in regions associated with language functions including bilateral inferior frontal gyrus and bilateral superior temporal gyrus.

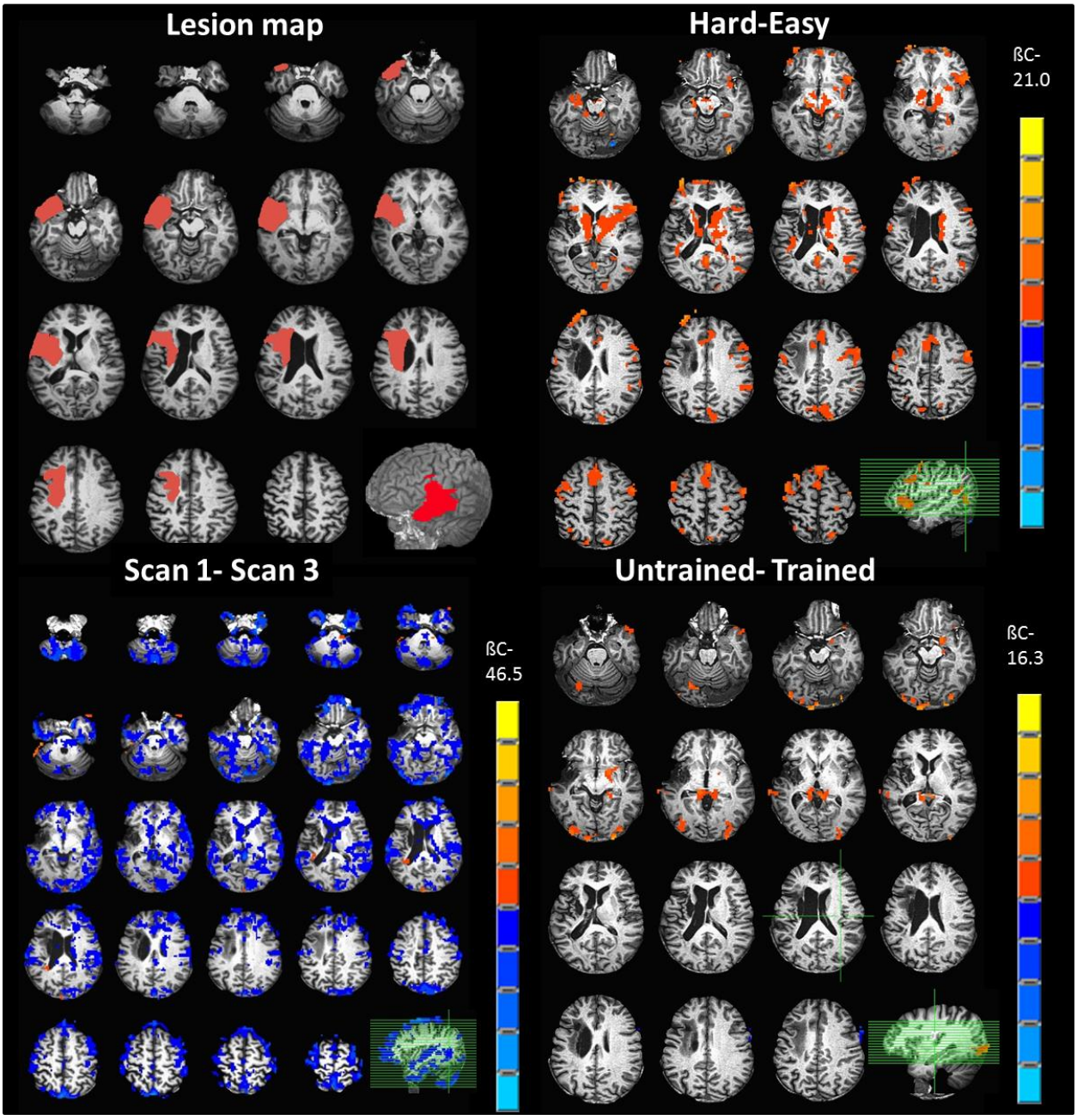


Figure 29. Lesion map and three contrasts for participant M2. ßC- beta coefficient for each contrast represents the intensity limits. Lesion map (top left) and three contrasts of interest: Hard-Easy (upper right) $p < .001$, 20 voxels; Scan One- Scan Three (lower left)- $p < .001$, 20 voxels. Untrained-Trained (lower right)- $p < .025$, 28 voxels. Blue represents more activation for the right side of the equation relative to the left as in Scan One- Scan Three. Red represents more activation for the left side of the equation relative to the right as in Hard-Easy.

Anatomical regions of interest.

Within each of the contrasts, large clusters were observed in the three predetermined language area ROIs—the IFG, STG and MTG. For M2, activation increased in both hemispheres but more consistently in RH regions compared to right and particularly following Treatment Period I (Scan Two) relative to Treatment Period II (Scan Three). In the right IFG, the mean percent signal decreased by -.06 after Treatment Period I, increased by .15 after Treatment Period II and decreased by .05 at Follow-up (Scan Four) as shown in see Figure 30. This is consistent with the activation pattern seen in the IFG for M1 and may reflect the effect of the increased challenge presented in the second treatment sessions. The left IFG increased consistently for the Untrained conditions with a total increase of .47 percent activation following Treatment Period II (Scan Three) and then decreased by 0.13 at Follow-up (Scan Four).

The pattern of activity in bilateral IFG best reflected differences in the varying levels of difficulty presented for M2 though not as consistently or well-differentiated as for M1. As with M1, the greatest to least activation was observed in this order following Treatment Period II (Scan Three): Trained Hard, Untrained Hard, Untrained Easy, Trained Easy and finally the control condition. At all other scan timepoints, Trained Easy appears at varying levels including the maximum activation compared to other conditions, at Scan Two. At Follow-up (Scan Four), there was a decrease in activation in the right IFG for all conditions (mean change in percent activation = -0.05) relative to Treatment Period II (Scan Three). Unlike M1, who showed nearly identical BOLD response for the four conditions of interest at Follow-up (Scan Four), M2 continued to show differentiation. He also continued to show gains in accuracy at this time point with a mean increase of 7.5%, unlike M2 who maintained previous gains. The left IFG showed, however, continued greater activation for the Untrained Hard condition at this final time point

and an overall slight increase in activation for all conditions relative to Scan Three (mean change in percent activation = .02 change).

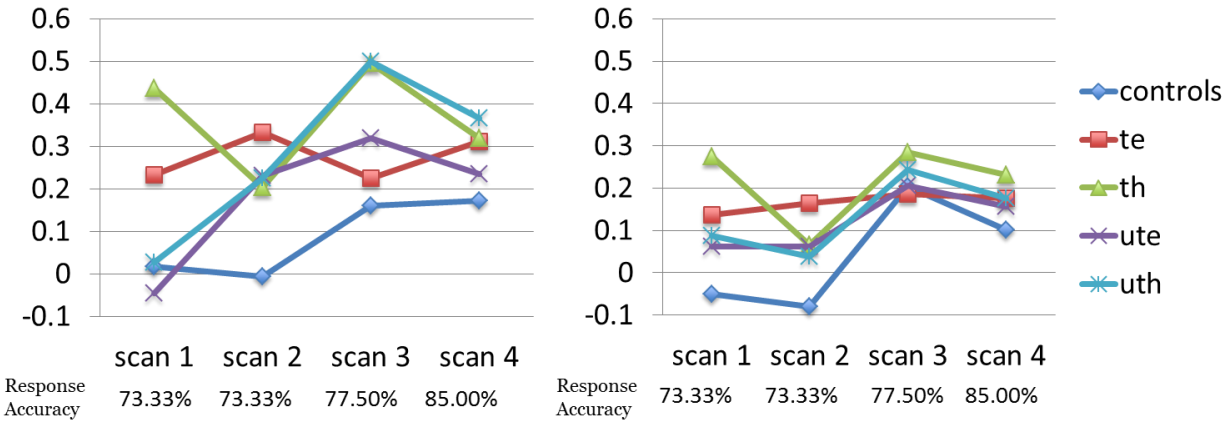


Figure 30. M2- Percent Signal Change in the Left IFG (on left) and right IFG (on right). Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Right and left STG also increased over successive scanning periods however there was not the same differentiation across conditions in this ROI and increases in the control condition matched increases in other conditions (see Figure 31). In the STG the mean increase for all conditions was 0.26 on the right and 0.05 on the left. The right and left MTG, showed the same pattern with a slight dip in activation following Treatment Period I and then an increase for all conditions that continued through to increase at Follow-up (Scan Four) as shown in Figure 32. This pattern was stronger in the right with a total mean change of 0.24 percent activation from Pre-Treatment (Scan One) compared to 0.07 in the left.

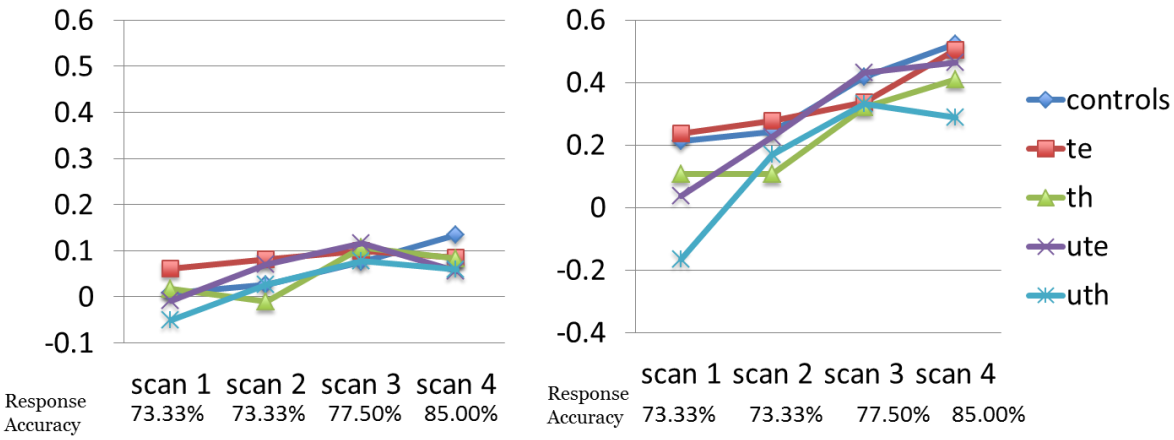


Figure 31. M2- Percent Signal Change in right (on right) and left (on left) STG. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Bilateral MTG did not differentiate across conditions for M2 however it may be the region that best corresponds with naming performance for this participant. Like the IFG, the bilateral MTG shows an initial dip in activation following Treatment Period I but like the STG the MTG also continued to show increase in activation at follow-up. There was no mean change in accuracy following Treatment Period I, a 4.2% change following Treatment Period II corresponding with a 0.084 percent increase in activation and an additional 7.5% increase corresponding with an additional 0.094 percent increase in activation at the Follow-up period.

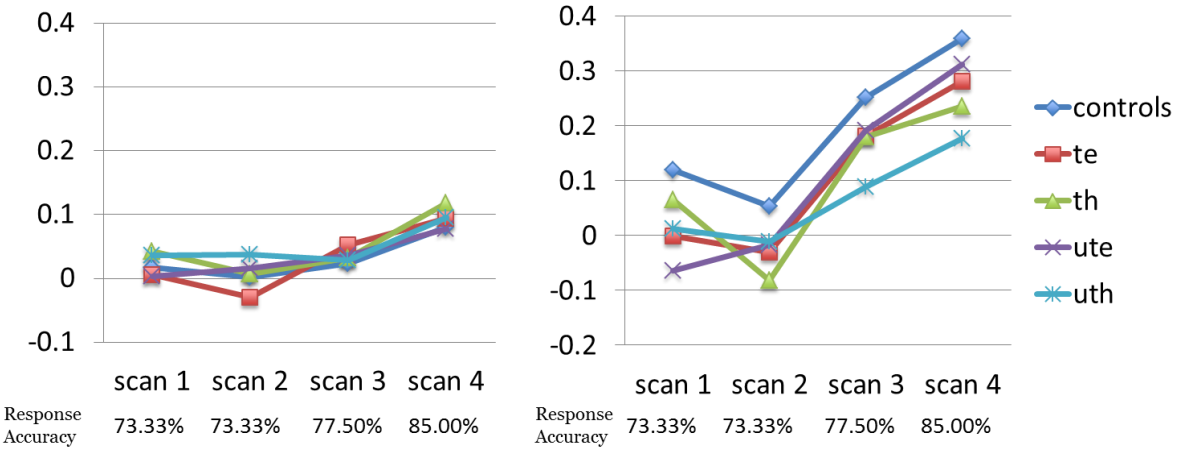


Figure 32. M2- Percent Signal Change in right (on right) and left (on left) MTG. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Laterality.

As with M1, the laterality index was used to plot changes in the relative strength of activation in the entire right vs. the left hemisphere. Laterality in the three ROIs was also calculated and is plotted against the whole brain as reference (see Figure 33). As with M1, the contributions of the three ROIs to the whole brain laterality are not obvious. Whole brain activation is fairly balanced between the right and left hemispheres at baseline. There is a strong rightward shift following Treatment Period I (Scan Two) and then a nearly equally strong shift leftward, back nearly to baseline which is maintained at Follow-up (Scan Four).

The MTG also starts fairly balanced between right and left hemispheres at Pre-treatment (Scan One) and shifts strongly to the left following Treatment Period I (Scan Two) and then even more strongly to the right following Treatment Period II (Scan Three). The IFG and STG show more consistent trajectories with less LH IFG activation over each successive scan and

slightly less RH activation STG over time. Laterality appears to be fairly stable for the whole brain and for all ROIs at the Follow-up (Scan Four).

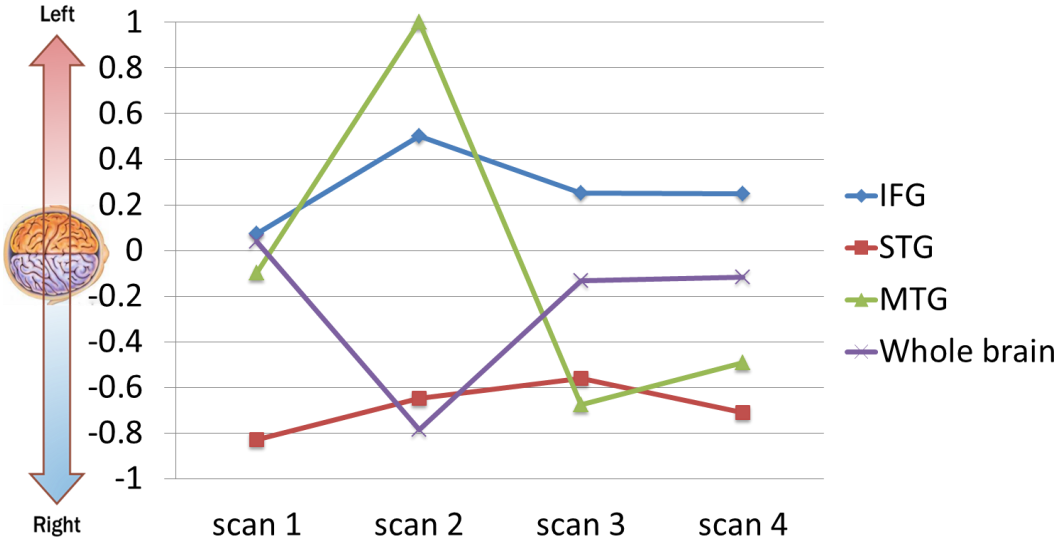


Figure 33. M2-Mean percent signal change across all conditions was used to calculate laterality. Mean percent signal change (MPSC) of the left hemisphere (LH)- MPSC of the right hemisphere (RH)/ (|MPSC LH|+|MPSC RH). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

S1

Standardized assessment.

S1 participated in 57 of the 60 hours of treatment and attended all baseline, assessment and follow-up sessions as scheduled. He presented with a pre-treatment WAB-AQ score of 38.5 of 100; 3.3% accuracy on the BNT and 76.7% accuracy on the C-RTT shown in Table 22. This participant's nonfluent oral expressive language production was characterized by several overlearned phrases, for example "I'm sorry," "That's not fair," "I don't know what you're talking about," and "No problem." He also had an entrenched stereotypy, "zerty bezert" which he used frequently in substitution for actual content words.

S1 made substantive increases on standardized tests after Treatment Period I including an increase of 6.6 percentage points on the BNT and a change of 14 points on the WAB-AQ. After Treatment II, there was an additional 11.7% change on the BNT but very little additional change on the WAB AQ . S1 demonstrated slight increases in auditory verbal comprehension as demonstrated on the WAB subtest as well as on the CRTT.

As with the previous participants, changes on the WAB AQ were seen primarily in object naming, word finding and fluency. For S1, other gains were also seen in sentence completion, word fluency and repetition. S1 demonstrated a 15% increase on writing subtests of the WAB, consistent with gains observed by all other three participants.

Table 22

S1- Summary of Assessment Scores at each testing period

Assessment	Pre-tx	Post-tx 1	Post tx 2	Follow-up
BNT	3.33%	10.00%	21.67%	18.33%
CRTT	76.67%	n/a	83.33%	81.33%
WAB AQ	38.50%	52.50%	52.90%	52.30%
WAB CQ	53.55%	n/a	62.80%	63.15%
WAB LQJ (100)	43.50%	n/a	54.65%	54.80%
Subtests from the Western Aphasia Battery				
spontaneous speech	25.00%	35.00%	35.00%	35.00%
auditory verbal comprehension	77.50%	85.50%	85.50%	81.50%
repetition	41.00%	65.00%	69.00%	50.00%
naming and word finding	24.00%	42.00%	40.00%	40.00%
object naming	28.33%	50.00%	40.00%	50.00%
word fluency	0.00%	0.00%	5.00%	15.00%
sentence completion	70.00%	80.00%	90.00%	90.00%
responsive speech	0.00%	40.00%	60.00%	60.00%
reading score	60.00%	n/a	58.00%	65.00%
writing score	22.50%	n/a	40.00%	37.50%

apraxia score	90.00%	n/a	90.00%	90.00%
constructional, visuospatial and calculation score	93.00%	n/a	90.00%	96.00%
RCPM	94.59%	na/	94.59%	89.19%

Note. All scores shown as percent of the maximum score. BNT-Boston Naming Test; CRTT-Computerized Revised Token Test; WAB AQ-Western Aphasia Battery Aphasia Quotient; WAB CQ-Western Aphasia Battery Cortical Quotient; WAB LQ-Western Aphasia Battery Language Quotient; RCPM-Raven's Coloured Progressive Matrices

Probes of trained and untrained material.

Accuracy was recorded for probes of trained and untrained stimuli from Levels one and two and results are summarized in multiple baseline format. See Figure 34. Level One was trained for all 60 hours of the study. Highest performance at this level was 80% accuracy on one occasion during Treatment Period II. Level Two was initiated and trained in conjunction with Level One after one week. Maximum performance at this level was 40% accuracy. These figures depict performance on trained and untrained materials over six phases: pre-treatment, Treatment Period I, no-treatment, Treatment Period II, immediate post-treatment and eight weeks post-treatment.

Accuracy of untrained materials increased in comparison to baseline but to a lesser extent than that of trained materials. Accuracy on Level Two stimuli (low frequency words) also increased *prior* to initiation of training indicating some level of generalization from treatment of high frequency words. Once training began, gains continued for trained words but appeared to plateau for untrained words.

S1 demonstrated moderate effect size of 6.4 for trained and minimal effect size (1.9) for untrained Level One materials at after Treatment Period I. He showed and medium-large effect sizes of 8.2 for trained and a small effect size of 3.8 following Treatment Period II.

Cues were provided after an unsuccessful naming attempt in order to reduce frustration and to allow the participant success with the trained task. After a week, it became clear that cueing was becoming more and more useful for S1. The tracking of cues only began when it was clear he was progressing with cues even when he did not appear to be progressing with spontaneous naming. By the end of Treatment Period II, a single initial phonemic cue resulted in 100% accuracy for trained words and 80% accuracy for untrained words. This was increased from 20% for both when documentation of cueing began on week two of Treatment Period I.

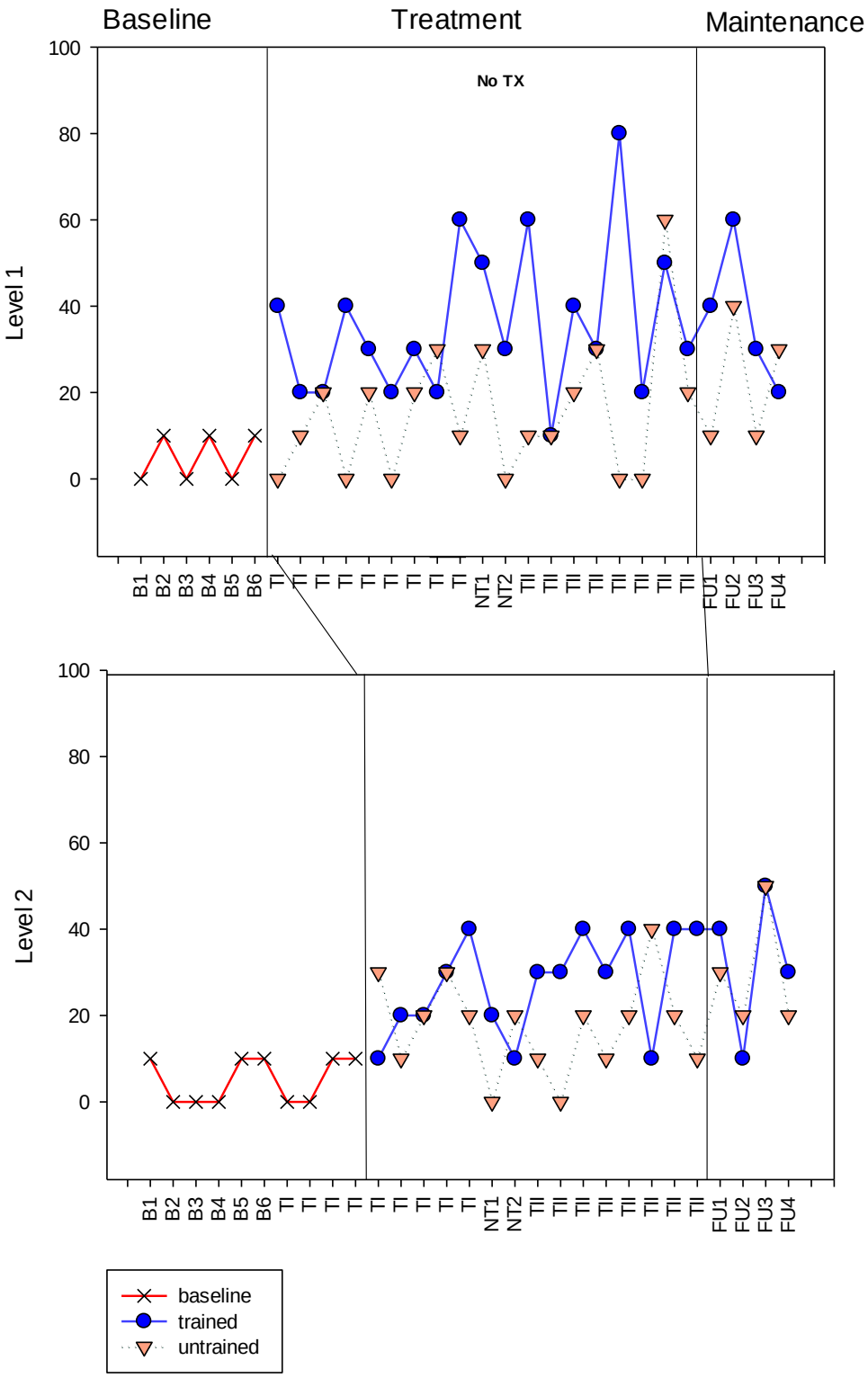


Figure 34. S1-Percent Accuracy on Treatment Levels 1-2. B-Baseline; TI-Treatment Period I; NT-no treatment period; TII-Treatment Period II; FU1-immediate post treatment; FU2- eight weeks post treatment

Generalization probes for narrative discourse.

Productivity was the main dependent variable for S1. Efficiency measures have been included to provide consistency between participants but since the production of informational content was so compromised in S1, productivity was the outcome variable of interest. Large effect size of 11.9 was calculated for productivity after Treatment Period I and moderate effect size of 6.3 following Treatment Period II. Massive variability in this measure for S1 was attributed to his use of over-learned phrases (see Figure 35). He was often able to use his limited repertoire in such a way that they were appropriate to the picture and could be counted as CIUs. In addition, there was an increase in spontaneous language in both semantic paraphasias (e.g., scissors for car jack) as well as emergence of accurate and appropriate novel words including: determined, hurried, pumping, dropping and praying. S1 made noticeably less use of his stereotypy "zerty bezert" though its use was never directly addressed in treatment. During the baseline period, he averaged 22.4 zerty bezerts per minute. In probes following Treatment Period I, he averaged 2.5 per minute. Immediately post treatment probes averaged .8 per minute and zero stereotypy usage was recorded at follow-up though sporadic use was still noted in conversation.

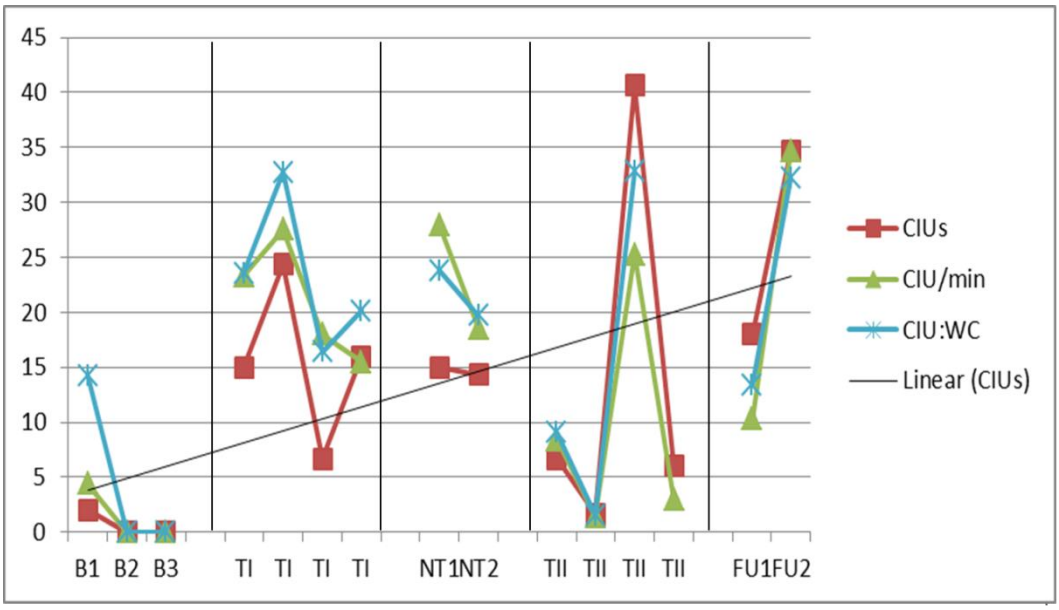


Figure 35. S1- Narative Discourse Probes- Efficiency and Productivity.B1-4-baseline probes, Probes 1-8-treatment probes; NT1-immediate post first treatment period; NT2- immediate pre second treatment period; FU1-immediate post second treatment period; FU2- eight weeks post treatment. CIUs- Correct Information Units

Overt naming in the scanner.

S1 made a 10% increase in naming accuracy between Pre-Treatment (Scan One) and Post-Treatment I (Scan Two) for Trained Easy and Untrained Easy and for Trained Hard words as shown in Figure 36. No change was noted for Untrained Hard words at this time or at any subsequent scan. A decline of 3.3% was observed at Post-Treatment II (Scan Three) for Trained Easy words and there was no change in any other condition. At Follow-up (Scan Four) , 6.7% increases were noted for Trained Easy and Untrained Easy words. A 3.3% increase was noted for Trained Hard words. As with those with mild aphasia, one to two word variability (3.3% to 6.7%) from scan to scan was anticipated and is not attributed to treatment performance.

During treatment naming probes, S1's greatest improvements were observed in his ability to be cued for a word. Uncued spontaneous naming did not improve much over the course of treatment and this was reflected in the scanner performance. Overall, there was an 4.3% increase

from baseline following Treatment Period I, no additional increase following Treatment Period II, and an additional 4% gain at Follow-up (Scan Four). Increases from Pre-Treatment (Scan One) to Follow-up (Scan Four) on Trained and Untrained Easy words may reflect an effect of treatment.

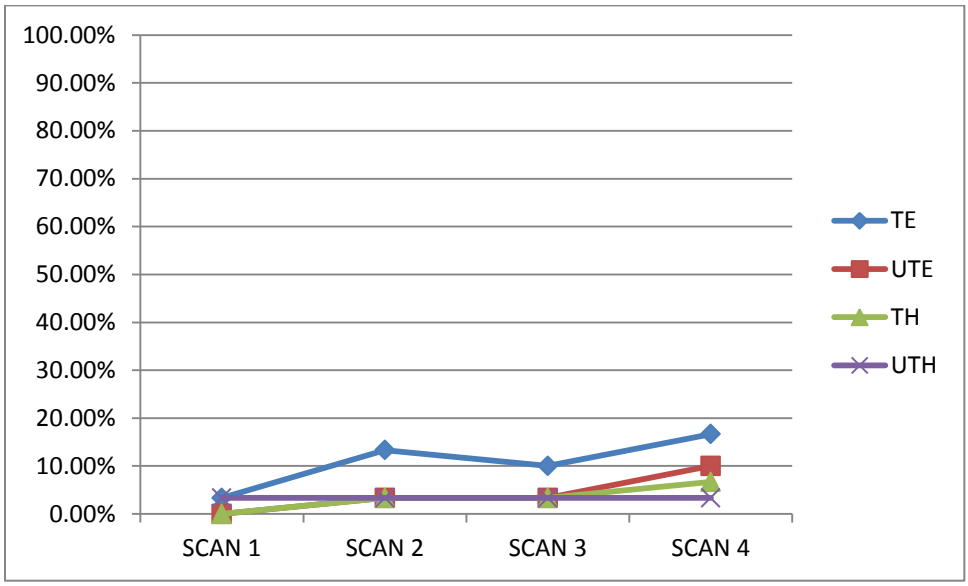


Figure 36. - Percent Accuracy For Scanner Naming. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

S1 consistently produced a "pass" response for the control stimuli but of the 100 other stimuli requiring spontaneous naming, he was only able to produce accurate responses for between four and ten stimuli per scan session. When reviewing response time data (Figure 37) it should be recalled that response times were based on times for accurate productions and there were only a maximum of ten spontaneous per session.

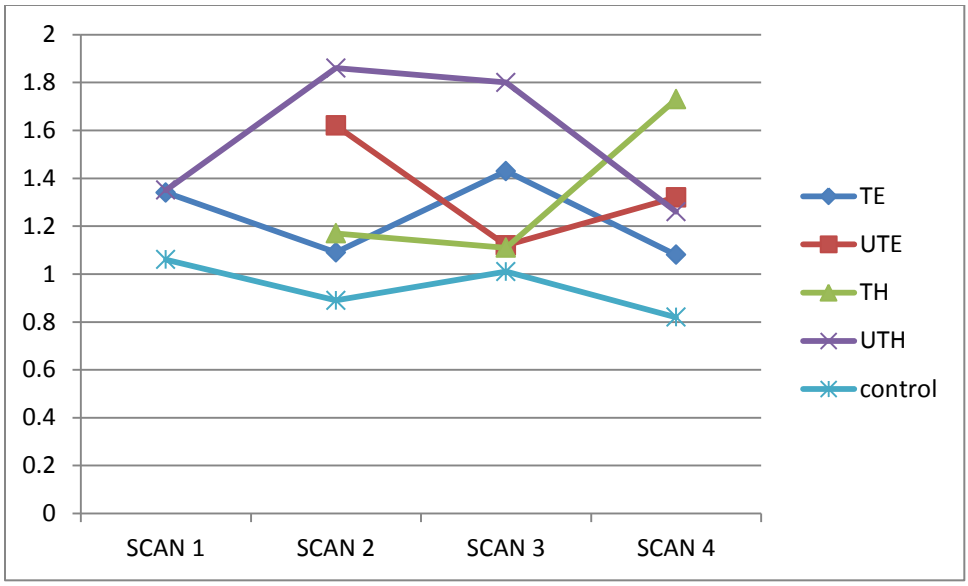


Figure 37. S1-Response times for scanner naming. Y axis denotes response times in seconds. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard; Control-"pass" response. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Functional neuroimaging data.

Several linear contrasts were specified in order to examine differences for all conditions between each time period and for each condition across all time periods. Statistical maps were corrected for multiple comparisons by including only clusters that were significant at a corrected statistical threshold as determined by Monte Carlo simulations. The most stringent threshold that still showed contrast differences were included. For the more severe participants these thresholds were less stringent than those applied to the participants with mild aphasia. As with M1 and M2, the most robust differences were seen for the Pre-Treatment - Post Treatment II (Scan One- Scan Three) contrast therefore this was corrected at $p < 0.001$, with cluster of six voxels or more. The Hard- Easy contrast was corrected at $p < 0.01$, with a threshold cluster size of 15 or more voxels. The Untrained- Easy contrast was corrected at $p < 0.025$, with a threshold cluster size of 28 or more voxels.

The three main contrasts of interest 1) Untrained vs. Untrained words 2) Hard vs. Easy words and 3) Pre-Treatment (Scan One) vs. Post-Treatment (Scan Three) alongside a montage of the lesioned regions are displayed in Figure 38. Untrained words showed significantly greater activation than trained words in eight brain areas. Largest clusters were observed in the right hemisphere in the superior frontal gyrus, the medial frontal gyrus and the middle frontal gyrus. Hard words showed significantly greater activation than easy words in nine brain regions. Largest clusters were noted in the bilateral cuneus and bilateral lingual gyri. Scan Three showed significantly greater activation than Scan One in 43 brain regions. In this contrast, largest clusters were seen in bilateral anterior cingulate, bilateral caudate, the left precentral gyrus. Mean activation peaks in all significant brain regions are shown for all three contrasts in Table C3, Appendix C.

The other contrasts were those of timing (Scan1 vs. Scan Two, Scan One vs. Scan Three, Scan One vs. Scan Four, etc.) and best analyzed in relation to each other. A description of this analysis is described in the next section, Anatomical Regions of Interest. For all contrasts, including the three described above, changes in activation were widespread throughout the brain, as with the participants with mild aphasia but with fewer of the largest clusters in regions associated with language functions. This is consistent with S1's overall poor naming performance in the scanner.

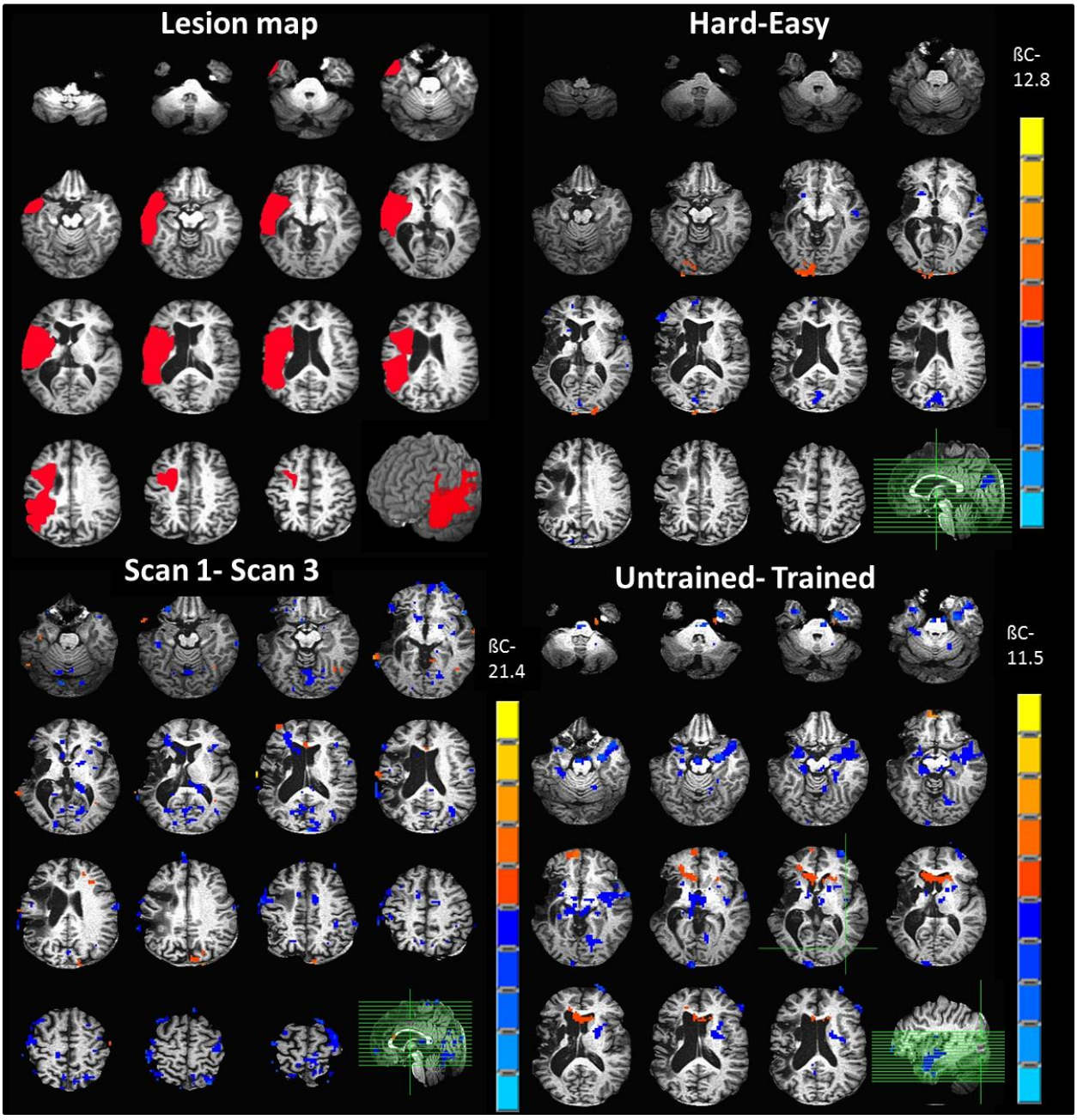


Figure 38. Lesion map and three contrasts for participant S1. ßC- beta coefficient for each contrast represents the intensity limits. Lesion map (top left) and three contrasts of interest: Hard-Easy (upper right) $p < 0.01$, 15 voxels; Scan One- Scan Three (lower left)- $p < 0.001$, 20 voxels. Untrained-Trained (lower right)- $p < 0.025$, 28 voxels. Blue represents more activation for the right side of the equation relative to the left. Red represents more activation for the left side of the equation relative to the right.

Anatomical regions of interest.

Although not the largest, there were statistically significant clusters observed in three predetermined language area ROIs—the IFG, STG and MTG-- within each of the contrasts run for S1. For S1, activation increased in both hemispheres but more consistently in RH regions compared to left. In some regions more increase was observed following Treatment Period I (Scan Two) and in some regions more following Treatment Period II (Scan Three). In the right IFG, the greatest increase in activation followed Treatment Period I with percent signal increase of 0.07, followed by a decline of -0.05 after Treatment Period II and of -0.007 at Follow-up (Scan Four) as shown in Figure 39. This differs from the right IFG pattern seen in M1 and M2 who received new, challenging stimuli in Treatment Period II. Since S1 did not meet criteria for the next level of training, no new challenge was introduced following Treatment Period I. The left IFG increased activation more consistently across scans for the Trained Easy condition. Both Untrained conditions initially resulted in a slight decrease in activation and then increased at Scan Three and Scan Four. Trained Hard decreased and then increased. Mean increases across all scans was negligible, the greatest, 0.04 percent activation, occurred at Follow-up (Scan Four).

If the right IFG is successful in differentiating level of difficulty, as appeared to be the case with M1 and M2, it shows that the Trained Easy condition is it is the most effortful for S1. Untrained Easy is second, Trained Hard third, Untrained Hard fourth and the Control again effectively served as the baseline with very little activation compared to all other conditions. Though different, this pattern makes sense for this participant. The Trained and Untrained Easy words are those he has had at least mild success producing in the past and S1 must use great effort to achieve this again. The Hard conditions here are not equivalent to the "hard condition"

seen by M1 and M2. For S1 the Untrained Hard and even the Trained Hard are potentially too challenging such that the participant may "pass" knowing that the word cannot be produced within the 4.5 second timeframe.

Unlike M1, who showed nearly identical BOLD response for the four conditions of interest at Scan Four, S1 like M2 continued to show the clear differentiation in the right IFG that became most apparent at Scan Three and persisted through Scan Four.

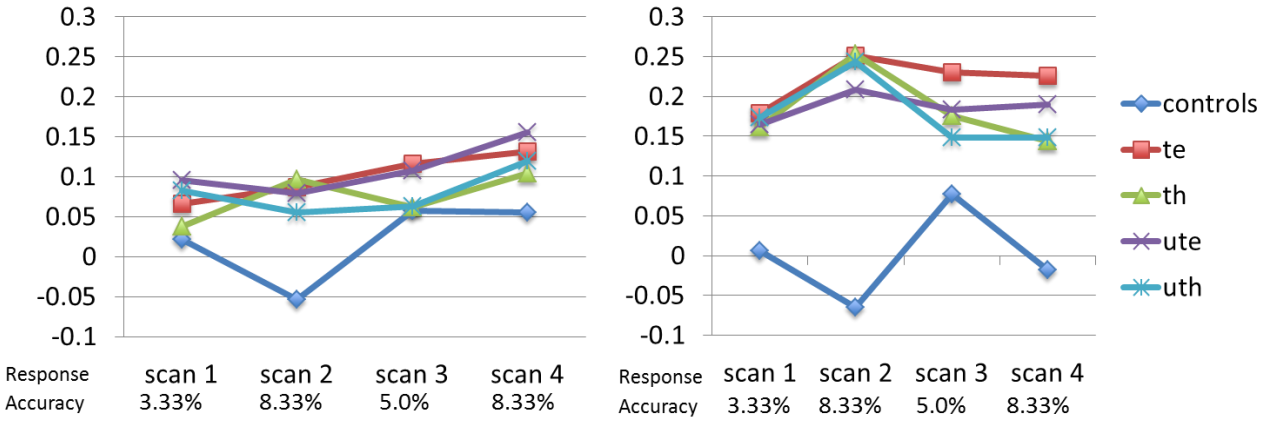


Figure 39. S1- Percent Signal Change in the Left IFG (on left) and right IFG (on right Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II-Scan Three; Follow-up eight weeks post-Scan Four.

Right and left STG showed opposite activation effects over time for S1, though the strength of changes in the right STG was much greater (see Figure 40). The right increased at Scan Two, following Treatment Period I and the left decreased. The right decreased at Scan Three, following Treatment Period II and the left increased. In the STG the mean percent activation change for all conditions from Pre-Treatment (Scan One) to Post Treatment Period I (Scan Two) was 0.11 on the right and -0.0008 on the left. From Post-Treatment Period I (Scan Two) to Post-Treatment Period II (Scan Three), the mean percent change was -0.19 on the right and 0.002 on the left. At follow-up the changes were 0.03 on the right and -0.0009 on the left.

As with all four participants, S1's left STG was a main site of lesion. 22% of his lesion overlapped with this region and may account for the decreased activation on left. It should be noted that although small, activation in this area is still differentiating between the Hard and Easy conditions and responding appropriately to stimuli.

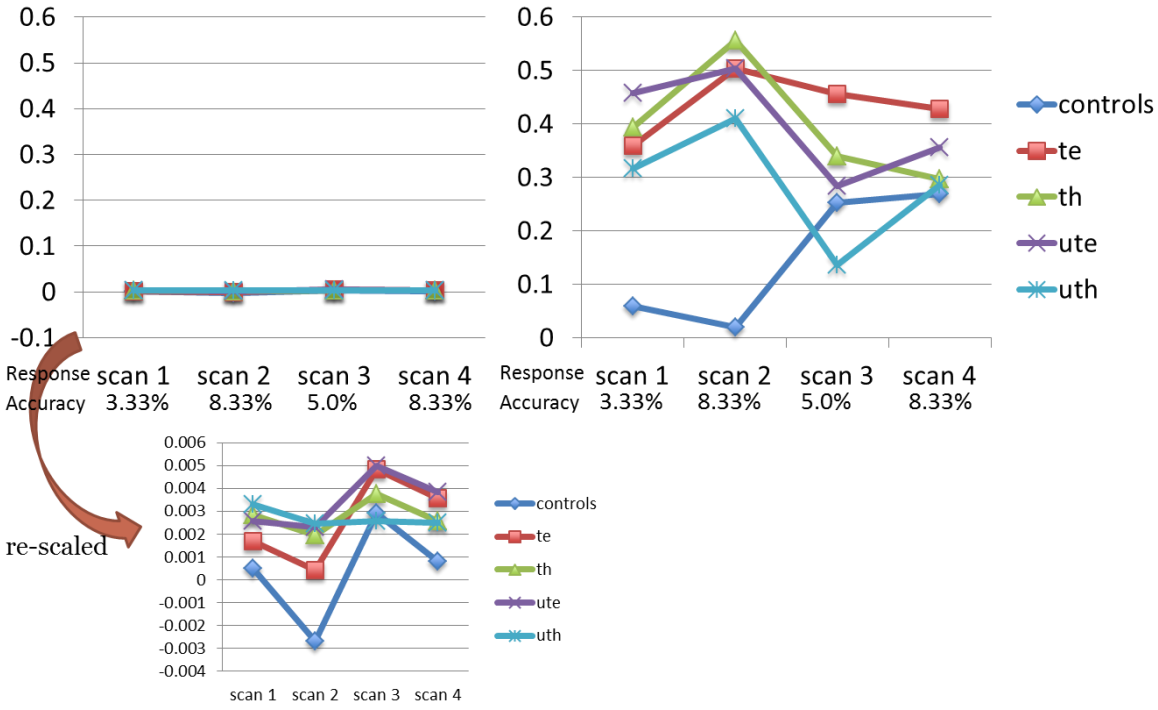


Figure 40. M2- Percent Signal Change in right (on right) and left (on left) STG with re-scaled left below in order to visualize changes. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

The right MTG clearly differentiated the Trained Easy condition, especially at Scans 3 and 4 when S1 was the most stimulable for trained items as shown in Figure 41. There is an increase in percent activation change bilaterally, across all scans for this condition, though it is more pronounced on the right. Mean change in activation across all conditions was at its maximum at Scan Two, following Treatment Period I with a .04 percent increase in the right STG. For the Trained Easy condition alone the increase was 0.12 percent activation change.

Changes in the left STG were negligible though the conditions were still differentiated fairly well at Scan Three in this hemisphere.

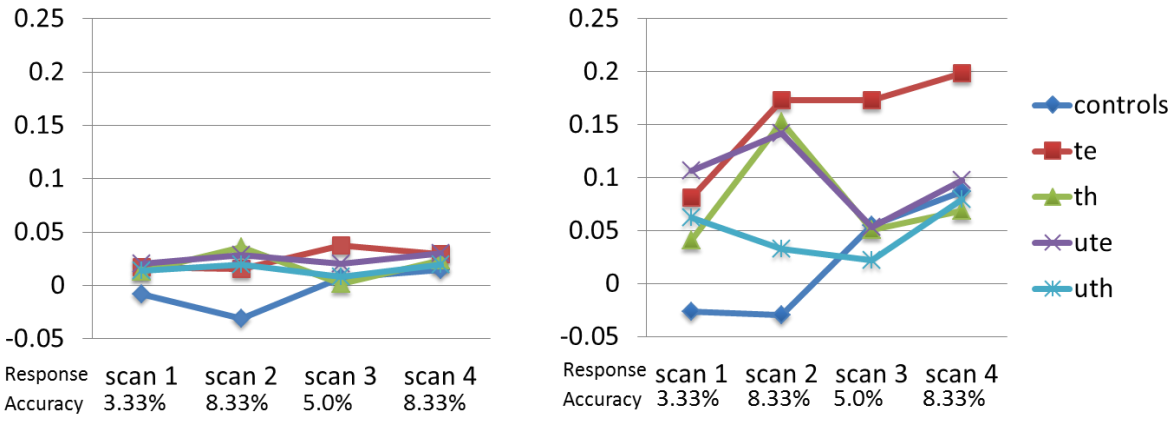


Figure 41. S1- Percent Signal Change in Right (on right) and Left (on left) Middle Temporal Gyrus. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Laterality.

The laterality index was used to plot changes in the relative strength of activation in the entire right vs. the left hemisphere. Laterality in the three ROIs were also calculated and are plotted against the whole brain as reference (see Figure 42). As with the two participants with mild aphasia, the contributions of the three ROIs to the whole brain laterality are not obvious. Whole brain activation starts with a slight left bias and then shifts strongly right after Treatment Period I. After Treatment Period II it shifts back leftward though this time with a slight right bias. The ROIs, in contrast all are more active in the RH to start and all increase in activity following Treatment Period I. After Treatment Period II, there is a subtle shift back to baseline laterality with additional RH deactivation in the IFG and what appears to be stabilization in the STG and MTG at Follow-up (Scan Four).

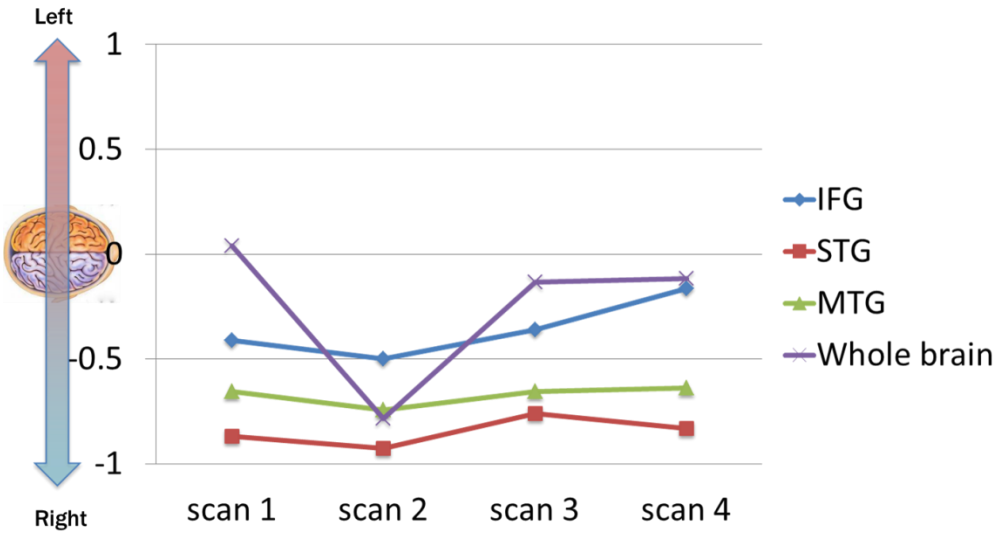


Figure 42. S1-Mean percent signal change across all conditions was used to calculate laterality. Mean percent signal change (MPSC) of the left hemisphere (LH)- MPSC of the right hemisphere (RH)/ (|MPSC LH|+|MPSC RH). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

S2

Standardized assessment.

S2 participated in all 60 hours of treatment and attended all baseline, assessment and follow-up sessions as scheduled. She presented with a pre-treatment WAB-AQ score of 51.7 of 100; 5% accuracy on the BNT and 76% accuracy on the C-RTT shown in Table 23. This participant's oral expressive language production was characterized by long, grammatically well-formed sentences, of which approximately 10-20% of the content words were meaningful.

S2 made large increases on standardized tests following Treatment Period I including an increase of 10 percentage points on the BNT and 12.3 points on the WAB-AQ. After Treatment II, results were mixed. The positive trajectory continued on the BNT and on the naming subtests of the WAB including naming and word finding, objects naming, and word fluency. Declines, however, were observed for subtests of sentence completion, repetition and auditory verbal

comprehension. At follow-up, most declines had reversed back to the initial increases seen after Post Treatment I and increases were maintained for object naming and for the naming and word finding subtests but not for fluency. S2 also showed a decline on the CRTT that continued at follow-up. The participant reported not having slept well the night before CRTT administration and her daughter added that S2 was "struggling today." It is not clear whether her performance on the CRTT was impacted as a result. It should be noted that auditory verbal comprehension scores also declined post Treatment Period II but recovered at follow-up testing, performed one day prior to CRTT administration.

Changes on the WAB AQ were attributed to all subtests except repetition and primarily to increases in word finding. The change on the BNT, from 5%-20% mirrored the change seen on object naming subtests of the WAB. S2 demonstrated a 12% increase on writing subtests of the WAB, consistent with gains observed by the other three participants but also a 35% increase on the reading subtests. She also improved on the RCPM (48.7% -70.3%) with the maximum change occurring after Treatment Period II. Maximum change for this measure occurred following Treatment Period II.

Table 23

S2- Summary of Assessment Scores at each testing period.

Assessment	Pre-tx	Post-tx 1	Post tx 2	Follow-up
BNT	5.00%	15.00%	20.00%	20.00%
CRTT	76.00%	n/a	70.00%	60.67%
WAB AQ	51.70%	64.00%	62.50%	64.40%
WAB CQ	60.60%	n/a	69.60%	70.60%
WAB LQJ (100)	52.60%	n/a	65.50%	68.60%
	Subtests from the Western Aphasia Battery			
spontaneous speech	65.00%	75.00%	75.00%	70.00%

auditory verbal comprehension	71.00%	92.00%	77.50%	83.00%
repetition	50.00%	40.00%	33.00%	39.00%
naming and word finding	28.00%	48.00%	52.00%	60.00%
object naming	33.33%	43.33%	58.33%	63.33%
word fluency	5.00%	20.00%	30.00%	15.00%
sentence completion	30.00%	90.00%	50.00%	100.00%
responsive speech	40.00%	90.00%	60.00%	90.00%
reading score	44.00%	n/a	65.50%	79.50%
writing score	55.00%	n/a	67.00%	61.00%
apraxia score	90.00%	n/a	91.60%	90.00%
constructional, visuospatial and calculation score	68.00%	n/a	82.00%	77.00%
RCPM	48.65%	n/a	70.27%	64.86%

Note. All scores shown as percent of the maximum score. BNT-Boston Naming Test; CRTT-Computerized Revised Token Test; WAB AQ-Western Aphasia Battery Aphasia Quotient; WAB CQ-Western Aphasia Battery Cortical Quotient; WAB LQ-Western Aphasia Battery Language Quotient; RCPM-Raven's Coloured Progressive Matrices

Probes of trained and untrained material.

Accuracy was recorded for daily probes of trained and untrained stimuli for Levels one and two. Level One was trained for all 60 hours. Highest performance at this level was 70% accuracy. Level Two was initiated and trained in conjunction with Level One after one week. Maximum performance at this level was also 70% accuracy. Results are summarized in multiple baseline format representing percent accuracy (see Figure 43). This figure depicts performance on trained and untrained materials over six phases: pre-treatment, Treatment Period I, no-treatment, Treatment Period II, immediate post-treatment and eight weeks post-treatment.

Accuracy of untrained materials increased in comparison to baseline but to a lesser extent than that of trained materials. Accuracy on Level Two stimuli (low frequency words) also

increased *prior* to initiation of training indicating some level of generalization from treatment of high frequency words. Once training began, gains continued for trained words but appeared to plateau for untrained words.

S2 demonstrated minimal effect size of 2.6 for both trained and untrained Level One materials at after Treatment Period I and medium effect sizes for both (7.8 and 6 respectively). Effect sizes on Level Two materials was also minimal (2.4 for both trained and untrained) following the Treatment Period I and did not increase after Treatment Period II (1.8 for trained and .84 for untrained.).

Cueing was not as effective for S2 as it was for S1 but was also tracked starting after 15 hours of treatment. During Treatment Period I, a phonemic cue added a 10% increase in accuracy to naming trials. In Treatment Period II, a phonemic cue began to be more useful and resulted in a 20-40% increase in accuracy per trial.

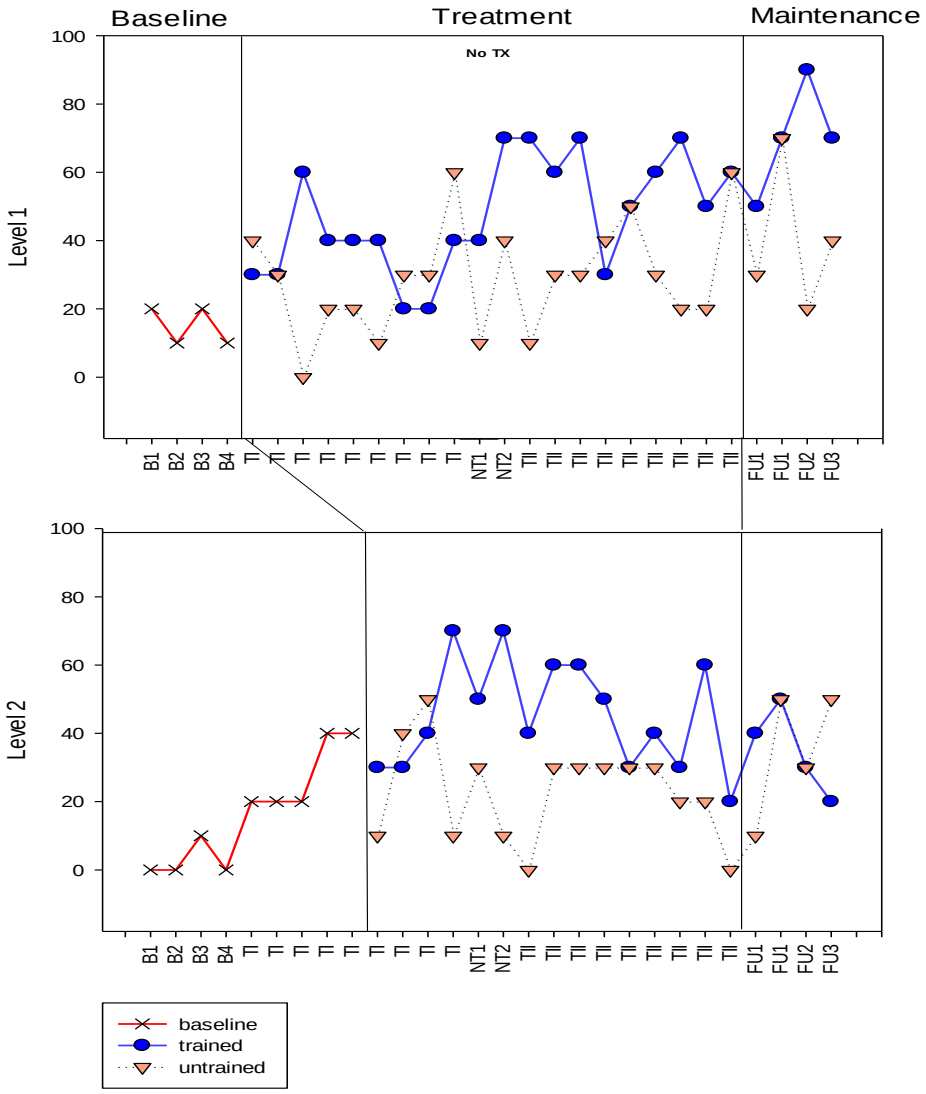


Figure 43. S2-Percent Accuracy on Treatment Levels 1-2. B-Baseline; TI-Treatment Period I; NT-no treatment period; TII-Treatment Period II; FU1-immediate post treatment; FU2- eight weeks post treatment.

Generalization probes for narrative discourse

For S2, like S1, productivity was the main dependent variable. Efficiency measures have been included to provide consistency between participants but since it was informational content that was so compromised in these participants, productivity was the outcome variable of interest. S2 was present for all discourse probes but due to a technical problem, the final Treatment Period II probe was not recorded and thus this data point could not be included.

S2's productivity increased as observed in Figure 44 with a large effect size of 10.3 observed after Treatment Period I 15.9 calculated after both treatment periods.

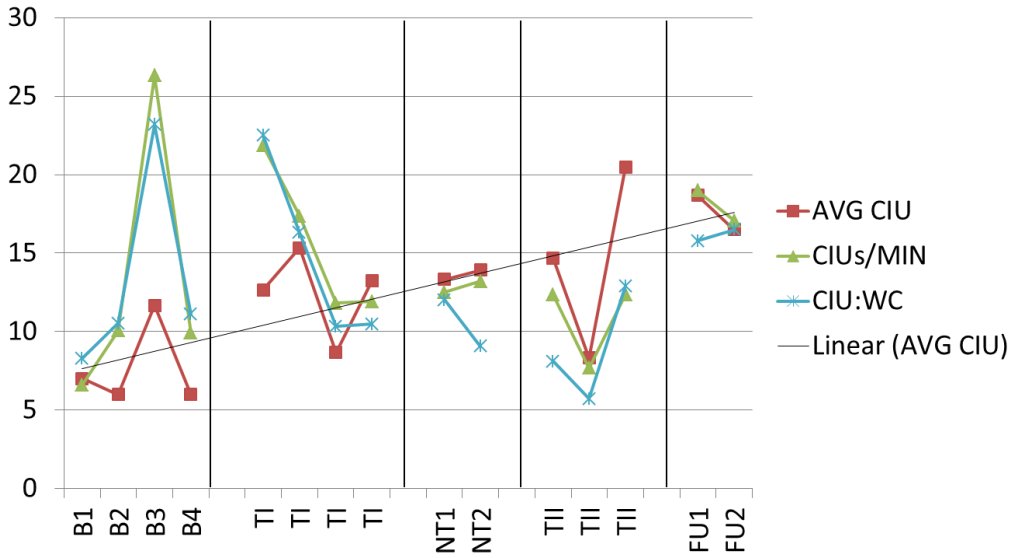


Figure 44. S2- Narative Discourse Probes- Efficiency and Productivity. B1-4-baseline probes, Probes 1-8-treatment probes; NT1-immediate post first treatment period; NT2- immediate pre second treatment period; FU1-immediate post second treatment period; FU2- eight weeks post treatment. CIUs- Correct Information Units.

Overt naming in the scanner.

S2 demonstrated a mean increase of 15.8% in spontaneous naming in the scanner across all conditions after Treatment Period I (Scan Two). Trained Hard increased by 13.3%, Untrained Easy and Untrained Hard both increased by 16.7% and the Trained Easy increased by 33.33%

(see Figure 45). Following Treatment Period II (Scan Three), there were additional increases for the Easy condition only. Trained Easy words increased an additional 3.3% and Untrained Easy increased by an additional 23.3%. There was no change in Trained Hard words and a 3.3% decline in Untrained Hard words. At Follow-up (Scan Four), gains tended to be maintained or within one to two words (3.3-6.6% change) which is attributed to day to day variability.

Performance in the scanner reflected performance during daily naming probes with one exception. As with her out-of-scanner spontaneous naming, S2 demonstrated larger effects of treatment following Treatment Period I than those following Treatment Period II. However, S2's in-scanner performance on Untrained Easy words exceeded her performance out of the scanner with accuracy increases exceeding that of the Trained Easy condition following Treatment Period II, Scan Three. Looking back at performance on daily probes did show that S2's best performance on Untrained items in each Treatment Period was on the final day of the final week of each treatment. Since scan periods followed closely after Treatment Period completion, this could be seen as consistent. Overall, S2 demonstrated a 15.8% increase from baseline following Treatment Period I and a 22.5% increase from baseline following Treatment Period II. Gains were maintained with a final mean accuracy of 21.7% more than baseline.

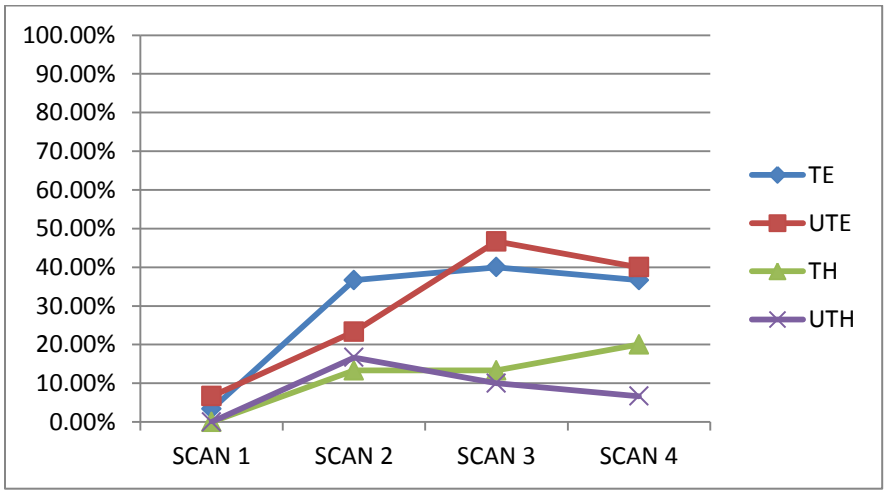


Figure 45. S1- Percent Accuracy For Scanner Naming. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

S2's response times tended to increase over time as accuracy increased (see Figure 46). Accuracy plateaued at Follow-up (Scan Four) however response times continued to increase.

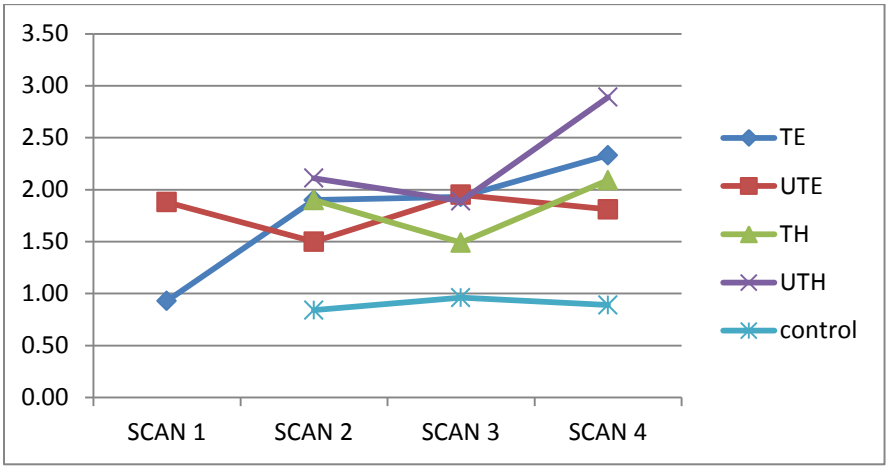


Figure 46. S2-Response times for scanner naming. Y axis denotes response times in seconds. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard; Control-"pass" response. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Functional neuroimaging data.

Several linear contrasts were specified in order to examine differences for all conditions between each time period and for each condition across all time periods. Statistical maps were corrected for multiple comparisons by including only clusters that were significant at a corrected statistical threshold as determined by Monte Carlo simulations. The most stringent threshold that still showed contrast differences were included. For the more severe participants these thresholds were less stringent than those applied to the participants with mild aphasia. S2 showed robust changes for Post-Treatment I- Post Treatment II (Scan Two- Scan Three) but less so for Post-Treatment II relative to Pre-Treatment (Scan One- Scan Three). This participant was extremely anxious about participating in the scanning aspect of this study. The other three

participants were much more comfortable. After having participated once, she was fine for all subsequent scans. It is possible that the increased activation observed in Scan One relative to Scan Two reflects a fear response which has been observed to activate a common affective circuit including regions such as the occipital-temporal lobe, the prefrontal cortex and thalamus (Stark et al., 2003).

Aside from the general increased activation at this first scan, relative to Scan Two, neural response tended to follow the pattern observed with the other participants. A corrected threshold of $p < 0.01$; cluster size > 15 voxels was used for the Pre-Treatment (Scan One)- Post-Treatment II (Scan Two) contrast. The Hard- Easy contrast was also corrected at $p < 0.01$, with a threshold cluster size of 15 or more voxels. The Untrained- Easy contrast was corrected at $p < 0.025$, with a threshold cluster size of 28 or more voxels.

The three main contrasts of interest 1) Untrained vs. Untrained words 2) Hard vs. Easy words and 3) Pre-Treatment (Scan One) vs. Post-Treatment (Scan Three) alongside a montage of the lesioned regions are displayed in Figure 47. Untrained words showed significantly greater activation than trained words in eight brain areas. Largest clusters included the bilateral cingulate and bilateral medial frontal gyri. Hard words showed significantly greater activation than easy words in 29 brain regions. The largest cluster incorporated the bilateral superior frontal gyrus and the right medial and middle frontal gyri. Scan Three showed significantly greater activation than Scan One in 46 brain regions. In this contrast, the largest cluster extended over the bilateral cuneus, left middle occipital gyrus, left middle temporal gyrus and left precuneus and left cerebellum. Mean activation peaks in all significant brain regions are shown for all three contrasts in Table C4, Appendix C.

The other contrasts were those of timing (Scan One vs. Scan Two, Scan One vs. Scan Three, Scan One vs. Scan Four, etc.) and best analyzed in relation to each other. A description of this analysis is described in the next section, Anatomical Regions of Interest. For all contrasts, including the three described above, changes in activation were widespread throughout the brain. Large clusters were associated with language ROIs and also in several other brain areas.

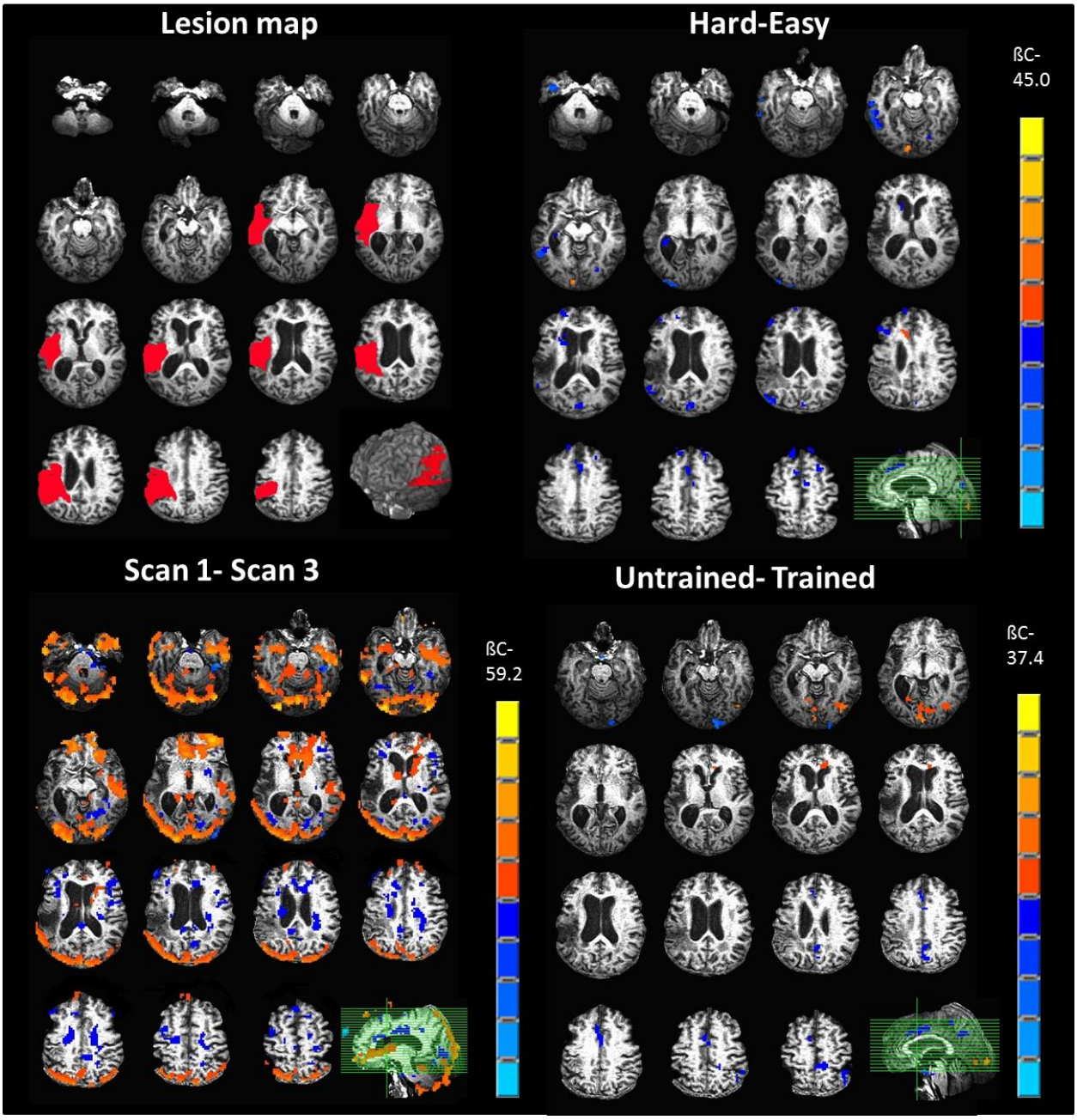


Figure 47. Lesion map and three contrasts for participant S2. ßC- beta coefficient for each contrast represents the intensity limits. Lesion map (top left) and three contrasts of interest: Hard-Easy (upper right) $p < 0.05$, 28 voxels; Scan One- Scan Three (lower left)- $p < 0.001$, 20 voxels. Untrained-Trained (lower right)- $p < 0.025$, 28 voxels. Blue represents more activation for the right side of the equation relative to the left. Red represents more activation for the left side of the equation relative to the right.

Anatomical regions of interest.

As with all other participants, statistically significant clusters were activated in the three predetermined language area ROIs—the IFG, STG and MTG-- within each of the contrasts run for S2. For S2, activation changes were observed in both hemispheres but more consistently in RH regions compared to left. For this participant, the control response of "pass" or "no" required more training. At Pre-Treatment (Scan One) the control response was usually produced incorrectly but a consistent "no" was produced in all subsequent scans. Consistent production continued to be more effortful for S2 and this is evident in each of the ROIs and in the case of this participant, does not serve as well as a baseline measure by which to compare the other conditions.

In the right IFG, S2's activation pattern is the same as the two participants who demonstrated increased correct responses in the scanner with an increase in activation corresponding with each improvement. Like M1, S2 maintained her Post Treatment Period II (Scan Three) naming performance at Follow-up (Scan Four) but at this time, activation was decreased back to Post-Treatment Period I (Scan Two) levels. The greatest increase in activation followed Treatment Period I with percent signal increase of 0.20, followed by an additional increase of 0.10 after Treatment Period II and of -0.07 at Follow-up Scan Four (see Figure 48).

The left IFG had a generally similar pattern but with a more attenuated response following Treatment Period II. The Untrained Easy condition was the only condition for which activation did not increase after Treatment Period I. It did increase in line with all other conditions following Treatment Period II. As with S1, the Trained Easy condition may be the one for which maximum effort was allocated as activation levels increase following Treatment Period I (Scan Two) and again following Treatment Period II (Scan Three) in both hemispheres.

Again, like S1, the Hard conditions (Trained and Untrained) were the least activated relative to the other conditions and therefore it may have been the case that S2 "passed" on verbal attempts knowing that production was unlikely within the 4.5 second time limit.

If the right IFG is the best differentiator for conditions as it has been for the other participants, particularly following Treatment Period II (Scan Three), it would appear that the effort level for S2 was as follows, listed from most to least effortful: Trained Easy, Control, Untrained Hard, Untrained Easy, Trained Hard. In this case, the "least effortful" likely corresponds to material that was too difficult and therefore dismissed.

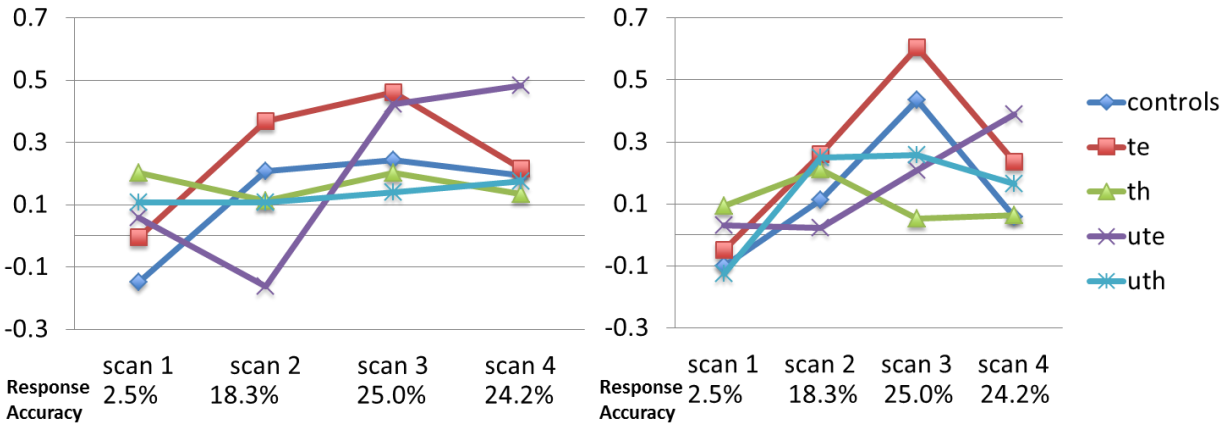


Figure 48. S2- Percent Signal Change in the Left IFG (on left) and right IFG (on right Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Right and left STG also showed similar activation patterns but patterns were much stronger on the right (see Figure 49). Bilateral increase activation occurred at Scan Two for the Trained Easy and Control conditions and decrease for all other conditions, following Treatment Period I. There was an increase in the Untrained Condition (both Hard and Easy) following Treatment Period II at Scan Three in the RIFG only and continued decrease in activation for Trained Hard at this same time point. The Trained Easy condition increased by 0.20 percent

activation following Treatment Period I and then decreased first by 0.38 after Treatment Period II and then by another 0.32 at Follow-up Scan Four. In the STG the mean percent activation change for all conditions from Pre-Treatment (Scan One) to Post-Treatment Period I (Scan Two) was -0.06 on the right and -0.01 on the left. From Post-Treatment Period I (Scan Two) to Post-Treatment Period II (Scan Three), the mean percent change was -0.19 on the right and there was no change on the left. At follow-up the changes were -0.08 on the right and -0.005 on the left. As with all four participants, S2's left STG was a main site of lesion. 20% of her lesion overlapped with this region and may account for the decreased activation on left. Unlike S1, it is unclear whether it is responding appropriately to stimuli since patterns of activation are difficult to interpret given the response from the other three participants.

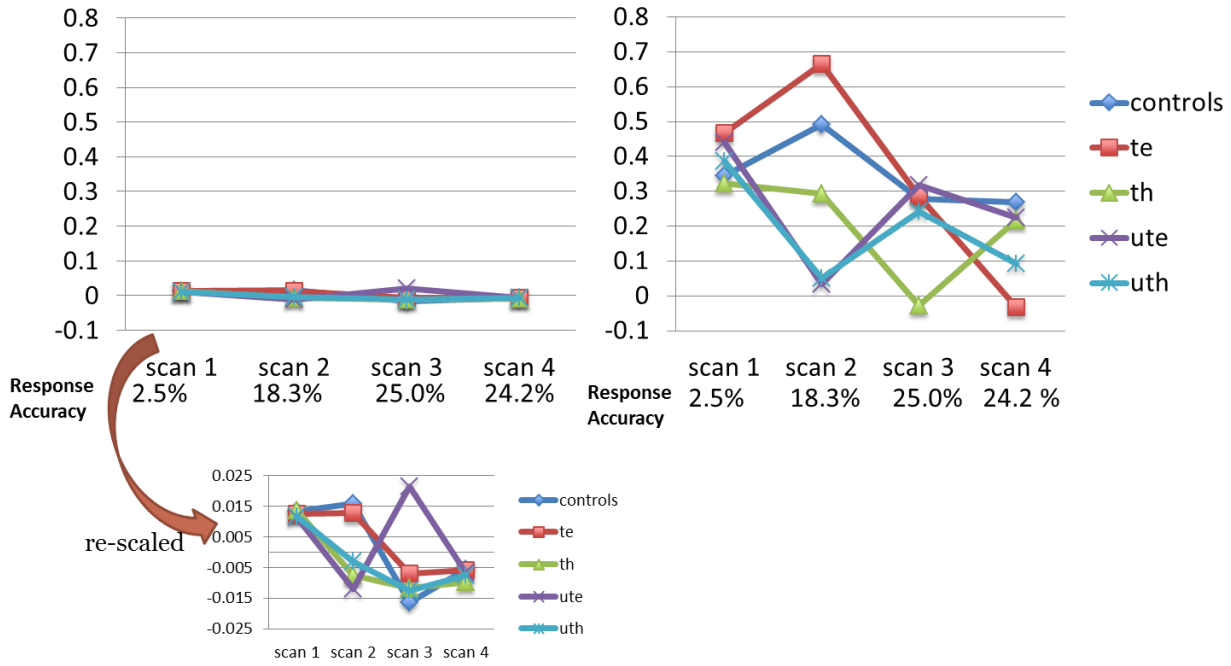


Figure 49. S2- Percent Signal Change in right (on right) and left (on left) STG with re-scaled left below in order to visualize changes. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Right and left MTG showed similar patterns for S2 as did the other ROIs, and like the other ROIs, greater changes were observed on the right (see Figure 50). The most pronounced shift was again seen in the Trained Easy condition following Treatment Period II with a 1.63 percent decrease in activation in RMTG. Mean change in activation across all conditions was at its maximum at Scan Three, following Treatment Period II with a .71 percent decrease in the RMTG and 0.28 percent decrease in the LMTG. Activation patterns for S2 in the MTG are difficult to interpret as they were for the STG. Nine percent of S2's total lesion was located in this area. Her IFG was the most consistent with the other participants' activation patterns. This may be due to the fact that this area was a spared area.

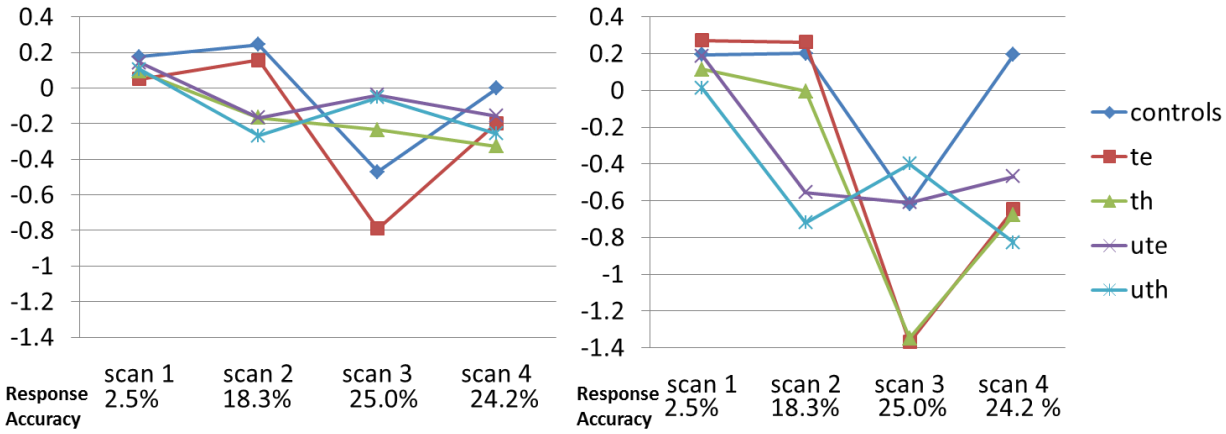


Figure 50. S2- Percent Signal Change in Right (on right) and Left (on left) Middle Temporal Gyrus. Stimuli types: TE-trained easy; UTE-untrained easy; TH-trained hard; UTH-untrained hard. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Laterality.

The laterality index was used to plot changes in the relative strength of activation in the entire right vs. the left hemisphere. Laterality in the three ROIs were also calculated and are plotted against the whole brain as reference (see Figure 51). Again, contributions of the three ROIs to the whole brain laterality are not obvious. The whole brain activation does not change

dramatically over time. There is a slight RH bias at Scan One and a slight LH bias by Scan Four. The STG, the most impacted area, is strongly right lateralized at Pre-treatment (Scan One) and this is maintained over time. The MTG, also implicated in the lesion, but to a lesser extent shifted strongly leftward following Treatment Period I (Scan Two) and remained relatively stable in subsequent scans. The IFG, the one spared area, showed strong leftward activation which shifted strongly rightward following Treatment Period I. After Treatment Period II, there was a slight shift back to left which was maintained at Follow-up (Scan Four).

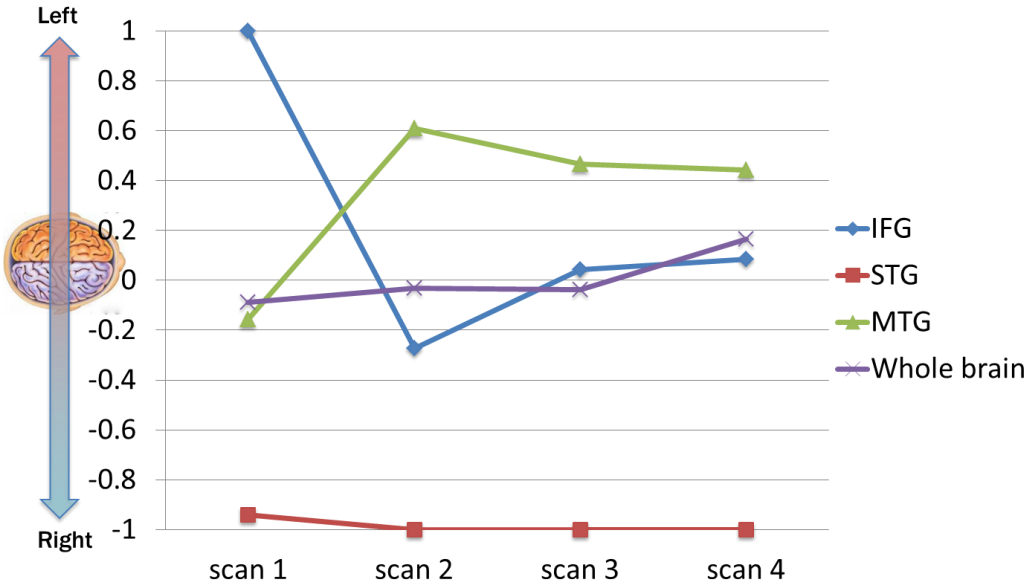


Figure 51. S2 laterality changes in the IFG, STG, MTG and the whole brain. Mean percent signal change across all conditions was used to calculate laterality. Mean percent signal change (MPSC) of the left hemisphere (LH)- MPSC of the right hemisphere (RH)/ (|MPSC LH|+|MPSC RH). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Summary of Behavioral Results

All individuals are considered separately, as his or her own control, however there is an overall picture of increased oral verbal expression that spans both treatment periods and of either increased efficiency or productivity of narrative discourse for which the greatest changes tend to be observed following Treatment Period II.

Standardized test measures.

Participants scores on the WAB-AQ increased over a range of 2.6-14 points after Treatment Period I and a range of 4.6-14.4 points after both treatment periods as shown in Table 24. Gains tended to be generalized across subtests with the most consistent trends observed on confrontational naming measures and fluency subtests. Gains were not attributable to comprehension subtests as has been reported following some previous studies (Johnson et al., 2013; Mozeiko et al., 2011; Szaflarski et al., 2008) as was reflected in the C-RTT results. None of the participants demonstrated appreciable change on this test. Naming on the BNT increased a range of 1-200% following Treatment Period I and 2.2%- 550% following both treatment periods as shown in Table 25.

Table 24

WAB-AQ summary scores and percent change

	ID	Pre-tx	Post-TPI	Pre-post-TPI change	% change	Post-TPII	Pre-post-TPII change	% change	F/U	Pre-F/U change	% change
Mild	M1	95	98	3	3%	100	5	5%	98	3	3%
	M2	88	94	6	7%	96	8	9%	94	6	7%
Severe	S1	39	53	14	36%	53	14	37%	52	14	36%
	S2	52	64	12	24%	63	11	21%	64	13	25%

Note. Pre-post TPI- refers to change from pre-treatment to post-Treatment Period I. Pre-post TPII refers to change from pre-treatment to post-Treatment Period II. Pre-F/U change refers to change from pre-treatment to eight weeks post treatment. AQ scored out of 100 total points.

Table 25

BNT summary scores and percent change

ID	Pre-tx	Post-TPI	Pre-post- TPI change	% change	Post-TPII	Pre-post- TPII change	% change	F/U	Pre-F/U change	% change
M1	92	93	1	1%	94	2	2%	97	5	5%
M2	77	85	8	10%	87	10	13 %	90	13	17%
S1	2	6	4	200%	13	11	550%	11	9	450 %
S2	3	9	6	200%	12	9	300 %	12	9	300%

Note. Pre-post TPI- refers to change from pre-treatment to post-Treatment Period I. Pre-post TPII refers to change from pre-treatment to post-Treatment Period II. Pre-F/U change refers to change from pre-treatment to eight weeks post treatment. BNT is scored out of 60 total points.

Probes of trained and untrained material.

The two participants with mild aphasia reached criterion at treatment levels four and five by the end of the first two weeks of treatment and of levels six and seven after the second treatment period. One participant also reached criterion for Level Eight. The two severe participants approached criterion for Level One and one did reach 90% accuracy in follow-up testing but never attained it during treatment and therefore Level Two was initiated and trained simultaneously starting at week two.

Probes of untrained equivalent material tended to result in longer response times but equivalent or near equivalent accuracy for M1 and M2. S1 and S2 demonstrated lower accuracy on untrained compared to trained items but demonstrated continued improvement over the course of treatment, in parallel with improvement on trained items. These two participants also learned to benefit from a minimal phonemic cue by the end of Treatment Period II.

Generalization probes for narrative discourse.

Those with severe aphasia demonstrated a large percent change in productivity during performance on picture description probes. Those with mild aphasia showed consistent,

unchanged productivity, as anticipated, but improved in discourse efficiency, particularly after the second treatment period. For these individuals, CIUs/minute were highly variable but with a visible and consistent slope increase. The second measure of efficiency, proportion of CIUs in total word count was slightly less variable and also showed consistent slope increase.

Effect sizes based on benchmarks provided by Beeson & Robey (2006) were calculated for treatment levels that comprised the majority of treatment time. A summary of effect sizes is shown in Table 26. For the participants with severe aphasia, Levels One and Two were the targets for both treatment periods. For the participants with mild aphasia, Level Five comprised the majority of Treatment Period I and Level Eight was the focus of Treatment Period II. All effect sizes are based on a minimum of two follow-up data points, and up to four, when available.

In general, effect sizes tended to be larger after both treatment periods than for the first treatment period alone. Also, in general, effect sizes tended to be larger for trained than for untrained materials. It is important to interpret effect sizes cautiously and in conjunction with the visual inspection of individual figures as baseline variability, or lack thereof, greatly influences the quotient. For example, visual inspection shows a modest rise in slope for discourse efficiency for both M1 and M2 but due to greater variability in M1's baseline and the very consistent baseline data for M2, effect sizes were 3.2 and 31.4, respectively. The more stable the baseline, the more likely any increase can be attributed to treatment. However, in the case of M1, visual inspection showed that variability persisted for this participant throughout treatment but that the range of variability was visibly higher throughout treatment and during follow-up testing.

Table 26

Summary of Effect Sizes

Outcome Measure	Analysis	Participant ID S1 Post Treatment Period I	S1 Post Treatment Period II	S2 Post Treatment Period I	S2 Post Treatment Period II
Trained	Level One	2.6	**7.8**[b]	**6.4**[b]	**8.2**[b]
	Level Two	2.4	1.8	1.9	**3.8**[a]
Untrained	Level One	2.6	**6.0**[b]	2.7	**3.7**[a]
	Level Two	2.4	0.8	1.0	**3.8**[a]
Discourse	CIU:WC	0.9	0.9	0.9	0.9
	CIUs/min	1.8	**3.1**[a]	**3.6**[a]	**3.5**[a]
	TOTAL CIUs	**6.6**[b]	**13.7**[c]	**10.3**[c]	**15.9**[b]

Outcome Measure	Analysis	M1 Post Treatment Period I	M1 Post Treatment Period II	M2 Post Treatment Period I	M2 Post Treatment Period II
Trained	Level Five	**24.3**[c]		**13.9**[c]	
	Level Eight		**14**[c]		**7.4**[b]
Untrained	Level Five	**25.6**[c]		**9.8**[c]	
	Level Eight		**12.3**[c]		**6.5**[b]
Discourse	CIU:WC	1.1	**3.2**[a]	-0.88	**5.76**[a-b]
	CIUS/min	0.13	2.74	-4.3 [a]	**31.44**[c]
	TOTAL CIUs	-0.8	-0.07	-0.08	-0.54

Note. Positive effect sizes denoted in bold print: a-small (4); b-medium (7); c-large (10.1) Benchmarks are according to Beeson & Robey, 2006. Tx I-effect sizes of Treatment Period I. Tx II- effect sizes following Treatment Period I and Treatment period II.

Summary of Neuroimaging Results

All four participants participated in each of the four scheduled scans with successful completion of all six runs for each scanning session yielding a total of 24 runs per participant. Some TRs were not usable due to excessive head movement by one participant, S2. Head motion can lead to statistical artifacts in the dataset which can appear as a false positive activation. Like the behavioral results, neuroimaging results are also analyzed individually per participant, each serving as his or her own control however when patterns were observed in the data, repeated measures analysis of variance (ANOVAs) were used as another measure of analysis.

Overt naming in the scanner.

A summary of the behavioral results during scanning is below in Figure 52 but group averaging of fMRI data from participants was not performed given the known individual differences in lesion size and location, and likely individual cognitive and neural strategies for recovery. All four participants made gains in naming accuracy. Those with mild aphasia tended to improve most on Trained and Untrained Hard words. Both were near ceiling on Trained and Untrained Easy words. Participants with more severe aphasia made greater gains on both Trained and Untrained Easy words. Gains in the scanner tended to be less than those observed in behavioral naming but tended to follow the same pattern with the greatest gains observed for Trained materials but also for Untrained materials

A one-way within subjects ANOVA was conducted to compare the effect of treatment accuracy of scanner naming at Scan One, Scan Two, Scan Three and Scan Four. There was a significant main effect of time at the $p < 0.05$ level for the four scans [$F(1.72, 5.17) = 5.80$, $p =$

.017]. Post hoc tests using a Bonferroni correction revealed no significant differences between successive timepoints, which is perhaps unsurprising given the small sample size.

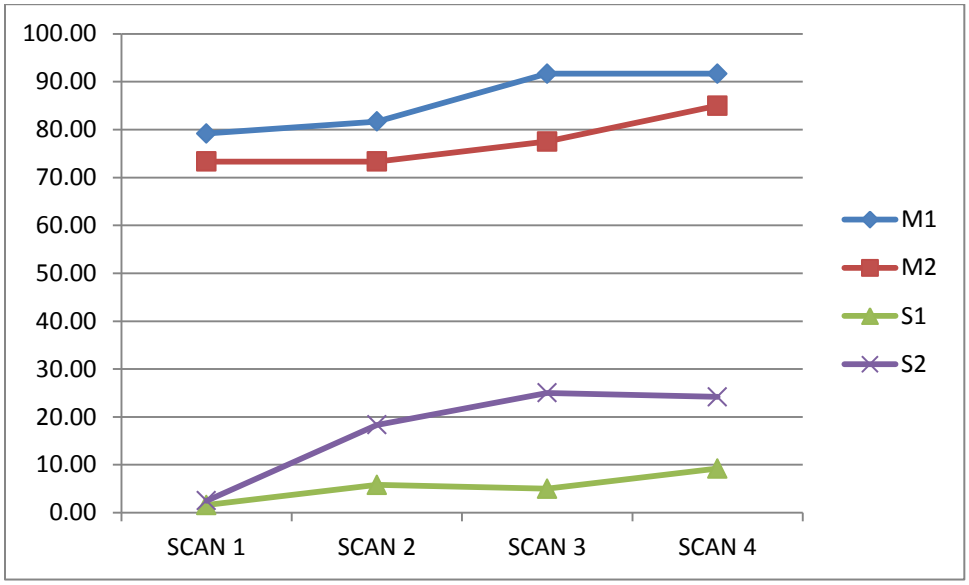


Figure 52. Mean percent accuracy of naming in the scanner across all stimulus types (Trained Easy, Trained Hard, Untrained Easy and Untrained Hard). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four

Functional neuroimaging data.

All participants showed significant activation for all contrasts after multiple corrections (refer to Appendix D). Greatest activation for all was seen in the Pre-Treatment- Post-Treatment Period II (Scan One-Scan Three) contrast and second for the Hard-Easy contrast. The least activation was seen for the Untrained- Trained contrast. Group statistics could not be performed on these measures due to the small number of participants but visual inspection of the contrasts for each participant appear to point a stronger effect of Time than of Condition. Of the two conditions, a stronger effect was observed for Difficulty than for Training. These trends were consistent for all four participants.

Regions of interest

Although this is a qualitative review of four different participants with highly variable behavioral and lesion patterns, there were some patterns in the ROIs in regard to response to treatment as shown in Figure 53-55. This pattern was best observed in both the right and left IFG where activation tended to increase in a pattern that followed increases in naming accuracy in the scanner. It was only in this ROI that activation tended to be greater in the left hemisphere than in the right despite the fact that the RIFG was implicated in three of the four participants' lesions (refer to Table 27 for main area of lesion for each participant including all ROIs).

Table 27

Percent of each individual lesion within a brain area

	M1	M2	S1	S2
Left Superior Temporal Gyrus	13.30%	21.00%	21.10%	20.10%
Left Middle Temporal Gyrus	5%	23.20%	14%	9%
Left Inferior Frontal Gyrus	11.10%	0%	6%	0%
Left Inferior Parietal Lobe	0%	9.40%	0%	14.60%
Left Insula	10.90%	0%	11.90%	12.60%
Left Middle Occipital Gyrus	0%	13.20%	0%	0%

Note. Any area that constituted at least 10% of a participant's lesion for a participant is included here.

A one-way within subjects ANOVA was conducted to compare the effect of treatment on neural activation in each ROI at Scan One, Scan Two, Scan Three and Scan Four. Mean activation of all conditions (TE, TH, UTE, UTH) was entered for each ANOVA. There was a

statistically significant main effect of time in both the left and right IFG but not for the other ROIs. The RIFG and LIFG each showed a significant effect of treatment over time ($F(3,9) = 3.875$, $p = .05$) and ($F(3,9) = 8.451$, $p = .006$), respectively. Post hoc tests again revealed no significant differences between successive time points, perhaps due to the small sample size.

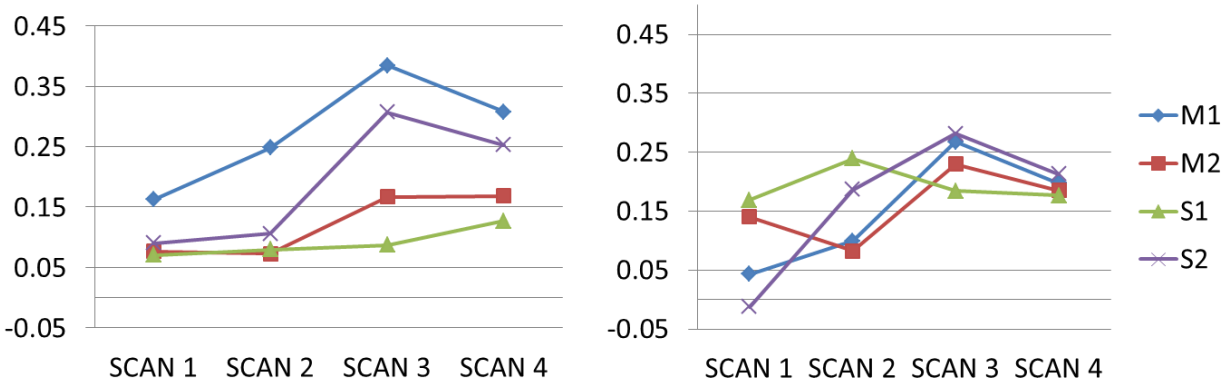


Figure 53. Summary Mean Percent Activation Change in LIFG (on left) and RIFG (on right) across all stimulus types (TE, TH, UTE, UTH) for each participant M1, M2, S1 and S2. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II-Scan Three; Follow-up eight weeks post-Scan Four

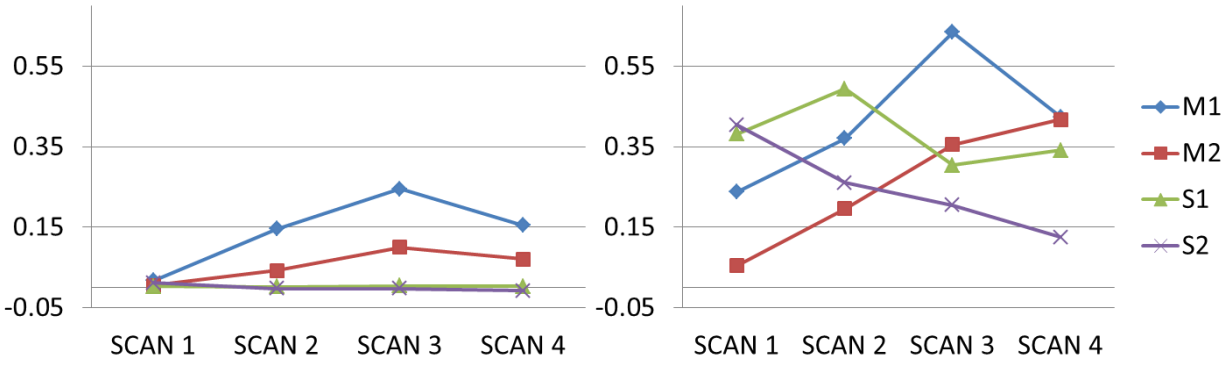


Figure 54. Summary Mean Percent Activation Change in LSTG (on left) and RSTG (on right) across all stimulus types (TE, TH, UTE, UTH) for each participant M1, M2, S1 and S2. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II-Scan Three; Follow-up eight weeks post-Scan Four.

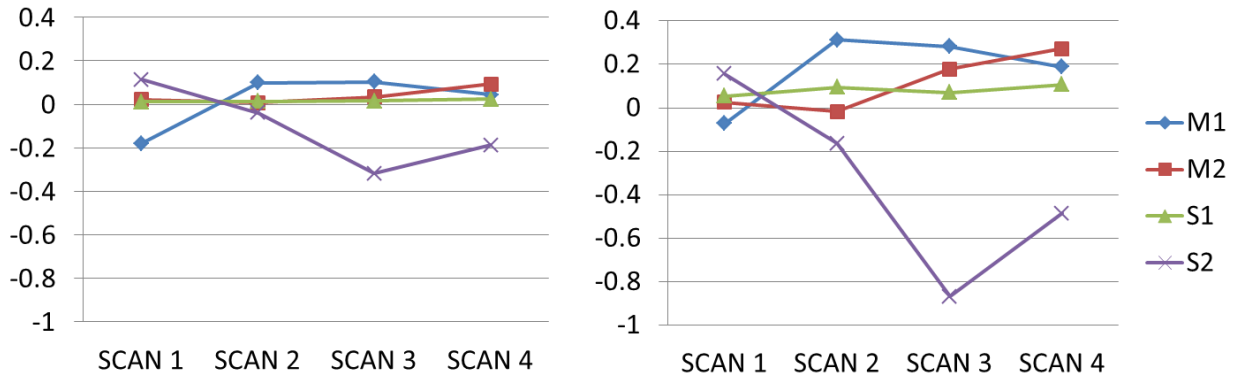


Figure 55. Summary Mean Percent Activation Change in LMTG and RMTG across all stimulus types (TE, TH, UTE, UTH) for each participant M1, M2, S1 and S2. Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Laterality

All four participants had sizeable, left hemisphere lesions (M1-75, 475 mm^3, M2- 64,869 mm^3, S1-99,671 mm^3, S2-67,250 mm^3). Three of the four participants did not demonstrate strongly right or left hemisphere dominance for the naming task prior to treatment as depicted in Figure 56. M1, however, was strongly left hemisphere dominant prior to treatment. Response to treatment varied and for S2 there was an activation shift leftward and for the other three it was rightward. Following Treatment Period I (Scan Two), the shift continued rightward for M1 but turned leftward for the other three participants to the point where no participant was strongly lateralized to either hemisphere and this hemispheric balance was maintained at Follow-up (Scan Four). General, whole-brain laterality did not appear to have any bearing on success with overt naming in the scanner.

A one-way within subject ANOVA was conducted to compare the effect of treatment on laterality at each time point (Scans 1-4). There was no significant main effect, Wilks' Lambda=.501, F(3,1)=.332, p = .819. These results are another indication that treatment with CILT in not responsible for a whole-brain laterality shift to a specific hemisphere.

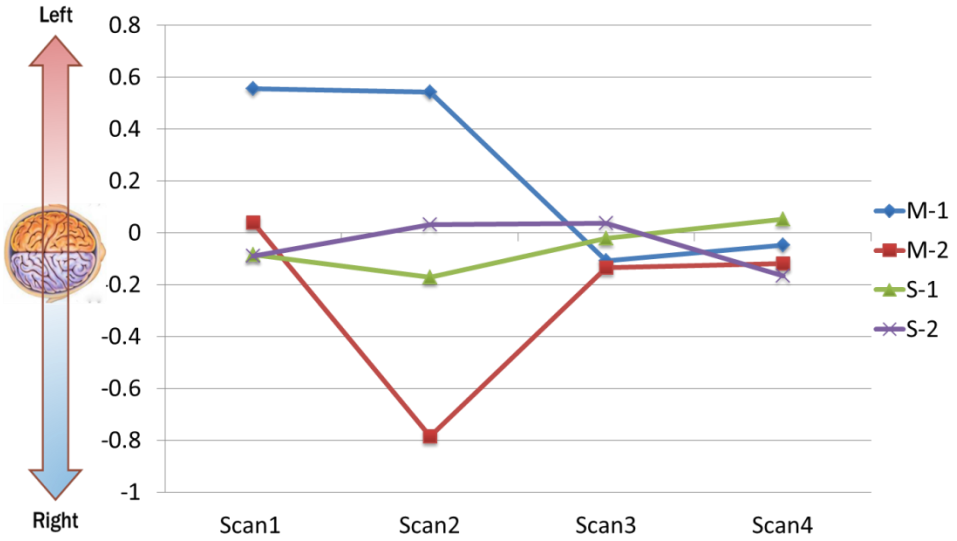


Figure 56. Whole-brain laterality for each participant. Mean percent signal change (MPSC) across all conditions was used to calculate laterality. MPSC of the left hemisphere (LH)- MPSC of the right hemisphere (RH)/ (|MPSC LH|+|MPSC RH). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Somewhat more consistent lateralization patterns were seen within the IFG, STG and MTG ROIs as shown in Figures 57-59. IFG activation tended to shift rightward throughout treatment resulting in equivalent left and right IFG contributions. This pattern was slightly different for M2 for whom a leftward shift was observed following Treatment Period I and then back to non-dominance. The STG tended to shift subtly leftward though still remained solidly right lateralized for all participants. MTG right dominance was stable for the two participants with more anterior lesions (M1 and S1) but shifted left for the two with more posterior lesions (M2 and S2) following Treatment Period I. Leftward shift was observed for M2 only following Treatment Period II.

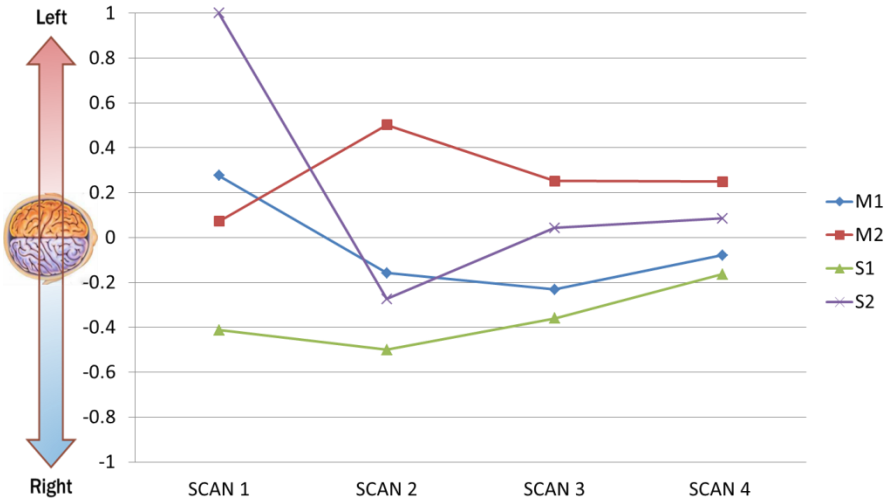


Figure 57. IFG laterality for each participant. Mean percent signal change (MPSC) across all conditions was used to calculate laterality. MPSC of the LIFG- MPSC of the RIFG/ (|MPSC LIFG|+|MPSC RIFG|). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

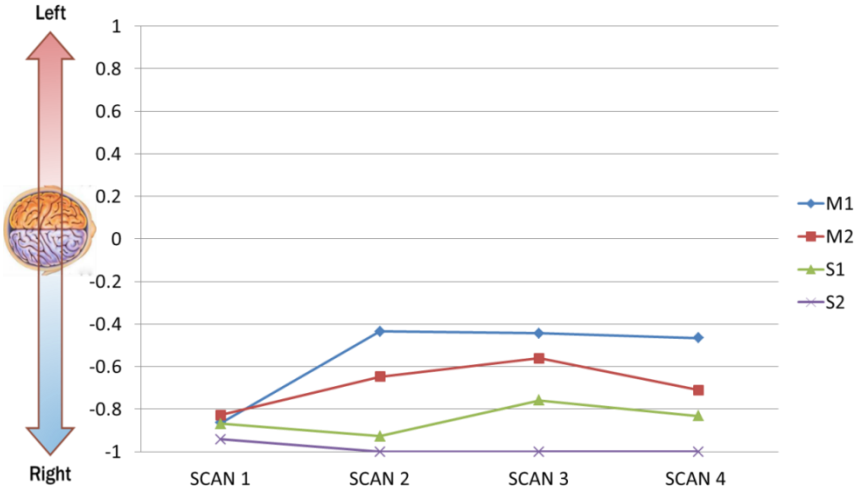


Figure 58. STG laterality for each participant. Mean percent signal change (MPSC) across all conditions was used to calculate laterality. MPSC of the LSTG- MPSC of the RSTG/ (|MPSC LSTG|+|MPSC RSTG|). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

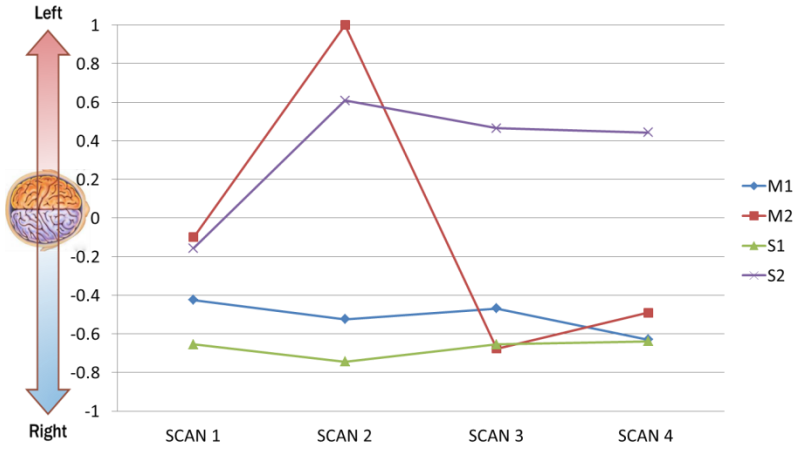


Figure 59. MTG laterality for each participant. Mean percent signal change (MPSC) across all conditions was used to calculate laterality. MPSC of the LMTG- MPSC of the RMTG/ (|MPSC LMTG|+|MPSC RMTG|). Scan periods: Pre-Treatment- Scan One; Post Treatment Period I-Scan Two; Post Treatment Period II- Scan Three; Follow-up eight weeks post-Scan Four.

Chapter VI

Discussion

This study investigated four PWA's treatment response given a double dose of CILT administered for three hours per day over four weeks. Each of these participants was more than 2.5 years post onset at the time of treatment initiation, well past the point of spontaneous recovery. Of primary interest was whether doubling an intensive "dose" (30 hours over two weeks) would result in behavioral gains above and beyond those seen in other treatment studies that have used CILT. CILT is based on the neuroplastic principles underpinning the successful motor treatment CIMT. Therefore it was expected that the neural change anticipated would be demonstrated in the successful generalization to untrained words and to connected speech, also never trained. Language gains were predicted to be maintained as has been reported for some participants in some studies using CILT (Barthel et al., 2008; Maher et al., 2006; Rose et al., 2013). Though often overlooked in treatment studies, durability of treatment gains is a critical component in assessing a program for clinical practice.

Functional magnetic resonance imaging (fMRI) was used to assess neural change in three main language regions, the IFG, STG and MTG and to investigate changes in laterality over the course of aphasia treatment. Four stimuli conditions were used to help characterize potential neuroplastic change following treatment. fMRI has emerged as a useful tool in conjunction with aphasia treatment studies but, at this time, it is not possible to make generalizations to the wider aphasia population based on results. The number of participants in such studies tends to be small and exclusion criteria, treatments and scanner tasks vary among studies. To date there have been

five fMRI studies that have used CILT as a treatment with a total of 31 participants. Twenty-seven of these participants were from two studies (Meinzer et al., 2008; Richter et al., 2008) and these two studies used different scanner tasks. Meinzer and colleagues (2008) used a confrontation naming task and Richter's group (2008) used sentence completion and word-reading tasks. Different tasks are likely to result in different activation responses even if the treatment the received was identical. Moreover, both of these studies took place in Germany and neither study specified whether monolingual participants were excluded, as tends to be the case in fMRI studies, particularly in investigations of language (Willems, Van der Haegen, Fisher, & Francks, 2014) since language processing is known to be a LH dominant function for most non brain injured people. Bilingual speakers are thought to process language more efficiently and perhaps have a different recovery process than those who are monolingual. Putting the aforementioned studies in the context of other aphasia treatment studies using fMRI, it appears that for production, the best recoveries tend to be associated with increases in left perilesional activation.

In order to conduct aphasia treatment studies using fMRI that are easier to compare and to interpret, more guidelines are necessary to add experimental control. Important considerations when planning aphasia treatment studies were discussed at a consensus conference at Northwestern University, Chicago. Conference topics included choice of experimental behavioral designs (Kiran et al., 2013), experimental tasks when investigating neural change (Rapp, Caplan, Edwards, Visch-Brink, & Thompson, 2013) and also fMRI data analysis (Meinzer et al., 2013). In addition, studies need to begin using standard terminology in regard to dosage parameters such as those proposed by Warren (2007) and modified by Cherney

(2012). The current study implemented many of the recommended best practices and used fMRI to investigate neuroplastic change resulting from a double dose of CILT for chronic aphasia.

Research Questions

Results of this study provided data that contribute to answers for the four research questions posed. Hypotheses were centered on questions of 1) aphasia treatment dosage 2) generalization of effect 3) durability of effect and 4) anticipated changes in brain activation.

Dosage

What is the effect on response accuracy and response time of CILT for trained material after one and two treatment doses (one dose=30 hours over 2 weeks)?

Treatment Period I. Confirmation of efficacy of 30 hour dose of CILT.

Increases in response accuracy were predicted for all participants and decreases in response times were predicted on trained levels in which criterion was reached (80% accuracy). Neural activation was predicted to increase in at least one of the ROIs, corresponding with behavioral change observed in the scanner. Medium to large effect sizes were observed for three of the four participants (S2, M1, M2) in response accuracy. Decreased response times were observed by visual inspection for M1 and M2. Though response accuracy is more important, response time is a useful metric for the participants with mild aphasia as a way to monitor progress once accuracy is consistently high and maintained.

Neural activation increases were observed for all participants between Scan One and Scan Two (pre-treatment vs. post-Treatment Period I) in the Trained Hard condition, as predicted, but also in the other non-control conditions. These results support the hypothesis that a single dose of CILT was effective for three of the four participants (M1, M2, S2). They only partially support its efficacy for one of the participants (S1). Activation that corresponds with

positive treatment responsiveness tends to be attributed to the remodeling or regrowth of a particular brain area or network (Saur & Hartwigsen 2012). Other factors, beyond task accuracy and latency also contribute to what can be determined a positive treatment response. Outcomes on measures of generalization and maintenance of gains are discussed below as they are relevant to the question of dosage but are also discussed unto themselves later in this chapter.

As predicted, a single dose of CILT resulted in language gains seen in previous studies that utilized CILT for approximately 30 hours over two weeks. As in previous studies, participants tended to have AQ scores that were at least five points above those seen in pre-treatment testing. Also like previous studies, greater gains on these tests were observed for those with more severe aphasia. Only M1, who at pre-treatment scored at ceiling on the WAB AQ (> 93) did not demonstrate a gain of five points at this time. After treatment Period I, those with severe aphasia also made greater gains in measures of connected speech than those with mild aphasia. S1 and S2 were both more productive and more efficient with oral verbal output as measured by gains in total CIUs. Gains in narrative efficiency measures (CIUs per minute and CIUs as a proportion of total words) were also observed for those with mild aphasia but effect sizes tended to be smaller due to great variability in performance. Gains in trained materials were much more evident for the two participants with mild aphasia. M1 was quicker than M2 in reaching criteria for a level. As a result he was often less challenged than M2 throughout most of Treatment Period I. Both participants met criterion for Levels Four and Five by the end of Treatment Period I. In contrast, neither S1 nor S2 ever met the criterion for Level One. Although confrontational naming gains were attenuated and performance remained inconsistent for these two participants, other changes were evident that were not captured in testing. For S1, this included a dramatic decrease in the use of his stereotyppie, “zertey bezert” though this was

never addressed in treatment. Where it was used previously as filler, S1 appeared to become more comfortable with silently working toward finding a word, though, usually, still unsuccessful in this regard. Both participants derived greater benefit from cueing than they did pre-treatment. Family members of S1 and S2 remarked on increased spontaneous word use at home but did not report much increase in other communication activities of daily living as reported on the CETI. Positive change on all primary and secondary measures would indicate that Treatment Period I was efficacious for S1, S2 and M2 and for M1 but, perhaps, to a lesser degree. This is further confirmation that language treatment provided at 30 hours per week over 10 consecutive working days can be beneficial to individuals with highly variable aphasia symptoms.

Changes in neural activation following Treatment Period I compared to baseline were robust for all participants with statistically significant activation in several brain areas. Though activation patterns varied from participant to participant, the strength of change appeared to relate to increases in behavioral response in the scanner. M1 and S2 made larger gains in mean accuracy across conditions than the other two participants at this time point and for both there were increases in activation during this period that exceeded that seen from the other two participants, in bilateral IFG, more notably on the right.

Treatment Period II. Will a double dose result in additional gains?

Additional increases in response accuracy were predicted as were decreases in response times on those levels in which criteria (80% accuracy) was reached. Additional increases in neural activation at least one of the ROIs were also anticipated for those who continued to make gains in accuracy. Effect sizes of treatment accuracy were medium to large for all four participants following 60 hours of treatment. This is particularly notable for S1 and S2 who

received an identical treatment for all of those 60 hours and for S1 who had only a minimal effect size following Treatment Period I. For him, the additional 30 hours was important in demonstrating significant effect of treatment. There was a significant main effect of time for naming in the scanner, as well, lending further support for a second treatment dose. Neural activation continued to increase at Scan Three (Post Treatment II) for all participants, as predicted, except for S1 who did not demonstrate additional behavioral gains within the scanner at this time point.

As stated previously, the dosage hypotheses were specific to accuracy and corresponding neural activation however, a discussion of dosage is more relevant within the context of all outcome measures and extant neuroplasticity literature. Increased gains were predicted to result from a second dose for most PWAs. In the motor literature, the concept of increased repetition is straightforward-- more repetition produces more neuroplastic change. This principle is more complex when applied to language. It is most likely not enough to repeat a stimulus item over and over. In fact, there is evidence that more stimuli are better than fewer in re-learning word lists (Snell et al., 2010). "Repetition" in language production may mean repeatedly engaging all brain areas contributing to oral verbal production language process, and it follows that an additional 30 hours of this would be more effective. It is also possible that in the chronic phase, intensive treatment provides a boost to the damaged system but that there are inherent limitations to that system such that a second dose would not be beneficial.

According to the results of the WAB AQ alone, one might conclude that there was almost no benefit to providing another two weeks of treatment for any participant. If this was the only measure, we might determine the second dose ineffective, when in fact, results for naming and discourse paint a slightly different picture, highlighting the need for multiple outcome measures

in aphasia treatment studies. For S2 there were continued gains on the WAB AQ in object naming but the decrease in repetition rendered the average AQ unchanged. On the BNT, S1 and S2 demonstrated strong, continued gains but M1 and M2 demonstrated virtual plateaus on this measure as well.

M1 and M2 continued to progress through all trained materials to the point where generalization was occurring on some subsequent levels prior to training. S1 and S2 however still did not achieve criteria for Level One although, again, changes within the session were evident for both participants. By the end of Treatment II, both of these participants could name 100% of the words given a phonemic cue.

Positive changes in discourse production were also evident for all four participants following Treatment Period II, though the effect sizes are misleading for some. Since the main outcome measure was performance on trained materials, treatment initiation was based on achievement of a stable naming accuracy baseline and not discourse performance. Discourse performance data were collected but stability was not required prior to treatment initiation. Therefore the baseline period shows great variability yielding effect sizes that do not appear to accurately reflect the upward trend that is apparent upon visual inspection.

When the outcome data is viewed all together for each participant, it is apparent that the second treatment was of value, though quantification of that value is difficult. It is not simply double the value attained from the single treatment session. A recent study by Rose and colleagues (2013) compared CILT to an equally intensive aphasia treatment where verbal production was the target. To do so, they used a cross-over design such that each of the 11 participants received a session of CILT and a session of the other treatment (multi-modality aphasia therapy; (Rose & Attard, 2011). In order to control for order effect, half of the

participants received CILT first and the other half second. Seven of the 11 participants achieved greater WAB AQ scores following the first treatment than the second demonstrating the importance of that first intensive treatment.

In addition, some participants in the aforementioned study continued to show increases a month or more after treatment was completed. This has also been observed in other studies using a single dose of CILT suggesting that the time between doses may play an important role in optimizing gain. In the current study, the second dose was administered five weeks following the first treatment session based on observations in previous studies including data from Study One (Mozeiko et al., 2011). For S1 and S2, there were generally positive results following both treatment periods but perhaps an even longer time between sessions would have allowed for stabilization of the initial boost making the language system more receptive to a second dose. It is also possible that having a shorter break between treatment periods or eliminating the break altogether would have been beneficial for S1 and S2, though it seems unlikely given reports of participants' fatigue at the end of Treatment Period I and given that naming performance was well-maintained just prior to Treatment Period II. For M1 and M2 the primary outcome measure of performance on trained materials continued to increase dramatically over Treatment Period II. If a plateau on trained materials warrants a longer break, increases might point to inclusion of continued treatment. For these two participants a third treatment period may have been warranted. Timing of the administration of consecutive intensive blocks is an area warranting future research. Is the break important for new neural processes to become fully instantiated or might these changes decay without the intensive practice?

Neuroimaging results would appear to support the efficacy of the second dose for M1 and M2. Following Treatment Period II, in which both participants were maximally challenged,

naming accuracy increased in the scanner, most notably on Untrained Hard words. In addition, mean percent activation changes were generally greater at this point in all regions of interest thought to be important to language production. This was particularly evident in the RH, though activation increases also took place in the LH. If the task in the scanner had changed, results might be attributed to increased effort but since the task was identical, and corresponded with increases in naming performance for both participants, it is reasonable to attribute changes in activation as something brought about by Treatment Period II. More evidence that treatment changes are responsible for increased activation levels in M1 and M2 are supported by the fact S1 made almost no change in scanner naming performance and activation levels were relatively stable over time. S2 made small, incremental positive changes in scanner naming performance but activation levels were less predictable and increased in some brain locations while steadily decreasing in others. Suppression of some brain areas may have been the result of activation of other areas for this participant.

It was predicted that activation changes would be best observed during the production of the more challenging stimuli and that if attenuation of activation was seen, it would likely be observed during the production of trained and challenging stimuli once mastery had been achieved. Attenuation was not expected for the production of untrained and challenging stimuli as it was expected that, without training, these words would be equally effortful compared to pre-treatment and would elicit the same BOLD response. Stability or a decline in activation was anticipated on trained and on over-trained stimuli. This did turn out to be the case, generally speaking, but "challenging" stimuli was not always the Hard stimuli customized for each participant and fMRI condition-contrasts reflected this. For M1 and M2 the Hard stimuli were actually the most challenging, and these conditions showed greater changes in activation than did

Easy stimuli. For S1 and S2, the Easy stimuli were those that posed the challenge. The Hard words, particularly the Untrained Hard words, were too difficult and were dismissed with an immediate "pass" response as was requested of the participant if he or she did not believe it possible to produce the word within 4.5 seconds. The Hard words for M1 and M2 activated many of the same brain regions as did the Easy words for S1 and S2. This should be taken into consideration for future studies such that "Hard" is achievable for all participants. Differences in task difficulty have not been compared in the previous studies, however challenging tasks have been shown to elicit a response in various LH language areas (e.g., Fridriksson et al., 2011; Kurland et al., 2012). In almost all comparisons of pre- and post-treatment measures in which participants have made behavioral gains, activation is greater post-treatment and tends to be maintained at follow-up (e.g., Breier et al., 2006). This was the case in this study for three of the four participants. M1, however, maintained gains and yet demonstrated reduced activation at Follow-up (Scan Four), perhaps demonstrating that the same task now requires less effort and perhaps more automaticity. If so, this would suggest that the ideal pattern of neural recovery is an increase in activation followed by a reduction in activation in the same brain region.

Generalization of effect

What is the effect on response accuracy and response time of CILT for untrained material and for functional communication outside of the clinic after one and two treatment doses?

It was predicted that there would be an increase in response accuracy and a decrease in response times for untrained materials; an increase on standardized test scores; an increase in functional communication and an increase in productivity and/ or efficiency after each Treatment Period. In addition, it was hypothesized that there would be an increase in activation after each Treatment Period (Scan One vs. Scan Two and Scan Two vs. Scan Three) for the production of

Hard, Untrained words, corresponding with improvements in behavioral naming of all Trained and Untrained words. No difference in activation was predicted for those who did not improve on Trained words.

Generalization performance on untrained exemplars and to connected speech is arguably more important than any other outcome measure. The goal of any treatment is the ultimate use of trained skills in everyday life, extending beyond the materials from the clinic, yet measures of generalizability are rarely included in treatment studies.

In the current study, generalizability to untrained items occurred beyond what was predicted for the mild participants who demonstrated nearly equivalent, large effect sizes on stimuli to which they had never been exposed. This occurred either at the same time as criteria was reached for trained materials or else closely following that time period. It is important to recall that M1 and M2 were exposed to hundreds of trained items for each level of treatment. This approach was based on the rationale that the training goal was to stimulate language processes rather than to memorize lists of words. M1 and M2 also improved in measures of discourse efficiency. This generalization to an untrained and more complex task may have been the result of "exercised" language circuits in the brain but it is also possible that training targeted these responses more directly than originally intended. Treatment practice never extended beyond the sentence level; however, the expectation during treatment was for highly structured, grammatically accurate sentences. The complexity of the required requests and responses increased over time, particularly during Treatment Period II which is when the slope for discourse efficiency was greatest for both M1 and M2. Increased efficiency, then, could also be a result of newly learned "formulas" for structuring a complex sentence. In other words, it is not clear whether efficiency increases in narrative discourse were actually a result of generalization

brought on by neural change or if discourse was actually indirectly trained. Although discourse efficiency was highly variable, lowest performances at two months post- treatment still exceeded high performances observed pre-treatment.

Less generalization to untrained words was observed for S1 and S2 than for M1 and M2 with small to medium effect sizes demonstrated only after Treatment Period II. This suggests that either M1 or M2 were more responsive to this particular treatment or that S1 and S2 needed more treatment to achieve the same level of gain since pre-treatment oral-verbal deficits were so much greater. For all participants, follow-up probes of untrained materials exceeded baseline. It's also of note that for S2, performance on Level Two increased prior to treatment initiation at this level demonstrating an effect of generalization from Level One training. Generalization effect was also seen at the discourse level for S1 and S2 with increases most pronounced on the productivity measure of narrative discourse. As with M1 and M2, performance varied from day to day for S1 and S2 however, follow-up productivity was consistently higher compared to baseline.

Of interest in this discussion of generalization is the increase in performance on the writing subtest of the Western Aphasia Battery (WAB) for all participants and on the Raven's Coloured Progressive Matrices (RCPM) for three of the four participants. In the case of the M1 and M2, maximum changes were observed at the follow-up testing, though all four participants showed changed following Treatment Period II. Writing subtests all improved by a minimum of ten percentage points and by 18 points for both M1 and S1. Also of note was that RCPM scores increased by nine points for M1 and M2 and by 12 points for S2. These were unanticipated changes, not previously reported in studies using CILT. Previous studies have found changes in auditory comprehension, however, with fair consistency. This was not the case for any of these

four participants. CILT requires appropriate auditory comprehension so changes in receptive language would not be surprising. It is not clear why there were improvements in reading and writing, two modalities that were never trained.

Generalizability is considered evidence that sufficient neural stimulation has occurred such that the re-learning process can occur more organically and as a result of a series of successes rather than a trained moment in the clinic. Measures of generalization are important to include in all treatment studies not only to test for potential neural change but also as another outcome measure if direct measures of treatment are insufficient as was the case for standardized batteries following Treatment Period II. Generalization to untrained materials may also be a predictor of outcome and is worth more focused attention in the aphasia treatment literature.

It was predicted that activation for Untrained materials would correspond with behavioral improvements but not necessarily in the same regions activated for Trained materials. In two other studies that included an investigation of untrained words, one found no behavioral improvements and no difference in activation (Kurland, 2012) and one found behavioral improvements but they were not correlated with the increased perilesional activity that correlated with trained words (Meinzer, 2008). Results from the latter study suggest there may be a different mechanism for generalization. Contrasts in the current study showed very little difference between the Trained and Untrained condition for any of the participants. There were slight increases in occipital and fusiform areas for M1, S1 and S2 for Untrained words compared to Trained words. Trained words showed small increases activation over Untrained words in a variety of locations for all four participants including bilateral IFG and precuneus, right superior, medial and middle frontal gryi among others. Examination of the three main ROIs tend to show parallel activation for Trained and Untrained versions of the same condition. Differences from

Meinzer's (2008) findings may lie in the fact that his Trained condition consisted of 40 words and the Trained condition in the current study consisted of 120 for S1 and S2 and for several hundred for M1 and M2. If, in fact, there are different processes for generalization and trained materials, it may not be revealed without a greater number of repetitions of identical stimuli.

Functional Communication

The ultimate goal of this and any treatment is generalization to functional gain in communication skills. To measure changes in this study, discourse measures were used and are discussed in relation to dosage above. In addition, family members were asked to provide input using the Communicative Effectiveness Index (CETI; Lomas et al., 1989). Increases were noted on several items from pre to post Treatment Period I. Fewer increases were noted following Treatment Period II. It is not clear whether this measure is effective since participants, family and the clinician may be biased in that they all *want* to see change, particularly following those treatments requiring a large time investment. In addition, participants and family members may feel some obligation to report positive change to please the clinician who administered treatment. Finally, day to day variability in function may bear on the results depending on the function on day of filling out the CETI. Measuring functional improvement is critical as outcomes here are the reason for performing treatment but questionnaires leave too much room for interpretation and are subject to loss of experimental control. Unsolicited reports of changes in life activities may be more informative. In addition, individuals uninvolved with the family or the treatment may make better reporters.

In the current study, all participants reported increases on several questions from the CETI. Additional evidence of functional life changes included the following. These are not

necessarily due to treatment just as change on the CETI may not be due to treatment, but they are noteworthy in that they reflect a change since the completion of the study.

M1. Following treatment, M1 declined participation in an aphasia group he had attended for more than four years, stating that he was ready to focus on his goal of re-employment.

S1. S1 reported being able to listen to and enjoy audio books for the first time since his stroke. His wife confirmed that he has attempted this multiple times but that he was unable to follow even simple audio books previously. S1 is also currently producing spontaneous and accurate words during participation in aphasia group. If he uses his old stereotypic utterance, "zertey bezert," it is no more than once per hour. His wife indicated an overall better quality of life following Treatment Period II for both of them.

S2. S2 reported that she was "getting worse" and family members of S2 reported that her frustration levels have increased since the completion of treatment with often multiple attempts at sentence production. Previously, S2 produced unintelligible utterances with little awareness of errors. Since productivity and efficiency of discourse improved, it is possible that her perceptions may be based on increased error awareness.

M2. M2 is very independent and self-sufficient and was never accompanied by family members who might comment on changes. His performance in the aphasia group appears to have changed but this has not been confirmed by someone uninvolved in the treatment process. M2 has taken more of a leadership role in group, often suggesting to students the types of activities he would like to practice and drawing in new members to the conversation. He participates in debates and is able to argue his points in more depth than observed previously, even addressing other participants' counterpoints.

Durability of effect

Will treatment effects be maintained at follow-up assessment eight weeks after treatment completion?

It was predicted that gains in response accuracy for trained material and untrained materials, decreases in response times for trained and untrained materials, gains on standardized test scores and in discourse productivity and efficiency would all be maintained eight weeks post Treatment Period II. In addition, no change or a decrease in neural activation in ROIs was predicted for all conditions, two months post-Treatment II (Scan Three vs. Scan Four) corresponding with maintained gains in naming accuracy in the scanner.

Durability of effect, also known as maintenance of gains, is an encouraging outcome observed following intensive treatment protocols such as this. When Maher (2006) contrasted two equally intensive treatments, she noted that those who participated in CILT tended to maintain language gains better than those who participated in a multimodality treatment. Animal studies have shown that changes to the uninjured motor cortex requires hundreds of repetitions (Nudo & Milliken, 1996) and musician studies show that practice is necessary to maintain neural change (Pascual-Leone, 2005). Once again, translation to language re-learning in the injured brain is not straightforward; however, it follows that increased accurate, oral verbal productions will result in increased neural change. If it also results in functional change in the form of increased verbal output, "practice" can continue naturally and will be maintained.

In the current study, maintenance was observed on nearly all measures including primary measures of performance on trained exemplars, secondary measures of standardized tests and measures of generalization. Some subtests for some participants were maximized following Treatment Period I, some following Treatment Period II. On subtests where a decrease was seen

between these two time periods, it tended to be recouped at the follow-up assessment such that nearly all gains made at either treatment period were maintained post-treatment. Performance on the repetition subtest was one exception for both S1 and S2. S1 made incremental progress on this subtest over each treatment period but follow-up scores dipped down close to baseline. S2 decreased in repetition proficiency during each treatment period and appeared to regain some at follow-up though not back to baseline levels. It is not clear on how treatment could contribute to negative change in this one area but it is possible that the injured system is competing for limited resources and impacting areas that have not been the subject of focus. (For a discussion of resource allocation in aphasia and inefficient allocation of attention, refer to McNeil, Odell, & Tseng, 1991).

Maintenance of gains is not always the case following CILT. In response to a participant's drop in language gain seven months post treatment Kurland (2012, p. S82) postulated that perhaps an intensive short-term treatment "provides a spark, but not continuous fuel, for ongoing recovery in some individuals." It is also possible that the initial spark needs to be stronger for some individuals. Although the current treatment results cannot be generalized to others with aphasia, these data are promising and it would be useful to test the double dose on a larger sample to test whether sixty hours of treatment might be the stronger spark needed for insured maintenance of gains.

Results in the scanner mirrored behavioral results for durability as they did for generalizability. Gains in naming in the scanner were maintained for all four participants eight weeks following treatment completion. It was predicted that activation levels would be maintained or decrease for PWAs who achieved mastery on trained, challenging materials. With mastery of each word, use of that word in the real word is possible. With use, activation levels

would be expected to remain the same or else decline as proficiency increases. It was predicted that activation levels would increase for PWAs who did not achieve mastery. Without continued use, the effort required for production might increase and result in increased activation levels at Follow-up Scan Four.

M2 was the only participant to achieve "mastery" or 100% on the Easy materials but there was no difference in post-treatment response to these words compared to the Hard. M2 was also the only participant to continue to increase accuracy on the scanner naming task at this time. There was no one pattern to activation levels however for this participant. M1 did not achieve 100% accuracy for any condition but had the highest rate of correct responses (>90%) by the end of treatment. Activation in all ROIs decreased for all conditions at the final scan even though accuracy was maintained. S1 and S2 both with the least success in naming in the scanner did not consistently increase or decrease in activation in the various ROIs in the various conditions.

Neural activity

Will neural activity changes impact the language lateralization in the IFG, STG, MTG and whole brain laterality over the course of treatment?

It was predicted that activation prior to treatment (Scan One) would be observed in the RH for those with the most extensive lesions and in the LH for those with the most spared LH tissue. Over time, LH activation was expected to increase for all participants who experienced positive language gains. LH lateralization was predicted to be the result of a shift from RH to LH for some or just increased activation in LH regions for others and no change in the RH. This prediction was not borne out.

Ways in which neural activation corresponded with treatment was a primary question of this study and has been discussed above in terms of its relationship with dosage, generalizability and durability of treatment. A second important question relates to the brain areas or networks responsible for change. Understanding patterns of recovery and how the brain responds to various treatment types following injury will allow for eventual better implementation of current language programs and adjunctive treatments such as rTMS. Recovery patterns also have implications for prognosis and potentially for treatment selection. Thus far, research has tended to focus on regions of the brain in which functional compensation takes place by looking at two broad areas, a) the right hemisphere and b) the viable tissue in the left hemisphere.

The intact RH has been the focus of several studies. Temporary disruption of Broca's area using TMS has been associated with increased activity in the right hemisphere homologue (Naeser et al., 2005) but it has not been clear whether that activation is maladaptive and reflects a release of transcallosal inhibition or whether the RH is actively supporting the disrupted processing in the LH.

Hartwigsen et al. (2013) recently tested this in healthy participants by applying continuous theta burst stimulation (cTBS) to the LIFG and then using fMRI to investigate changes in connectivity between the left and right hemispheres during word repetition. These investigators found that the right hemisphere homologues actively contributed to language function lending support to the relatively few studies that report that the RH does contribute to aphasia recovery after LH damage (e.g., Raboyeau et al., 2008; Saur et al., 2006; Winhuisen et al., 2005). Hartwigsen and colleauges (2013) also demonstrated that increased activity in the contralateral homologue was associated with a stronger facilitatory drive from the RIFG to the LIFG as responses became faster with increased influence of the RIFG on the LIFG. This is an

important study because it demonstrates the way the brain shifts to accommodate change, something more difficult to observe in regard to the heterogeneous lesions investigated in aphasia treatment studies.

There are other studies that suggest that the right hemisphere plays a crucial role in aphasia recovery, however the vast majority of studies provide evidence that the ability to make use of residual left hemisphere tissue are those who make the best recoveries (Fridriksson et al., 2010b, 2011; Heiss, Kessler, Thiel, Ghaemi, & Karbe, 1999; Meinzer et al., 2008).

These too are still subject to some interpretation. Fridrikson's (2010) initial study's findings were the result of one time-point in the scanner looking at gross naming ability compared with activation. His follow-up (Fridriksson et al., 2011) included a treatment component and looked at one time point immediately following treatment as did Meinzer and colleagues (2008) who found similar results of a strong relationship of left perilesional areas activated post treatment. Although there were clear changes post treatment, it is difficult to determine with certainty that changes were due to treatment or to draw conclusions based on one post-treatment scan. Intensive treatments, most often used in studies using fMRI, are often reported to result in behavioral changes more than a month following the completion of treatment(Maher et al., 2006; Mozeiko et al., 2011; Rose et al., 2013; Szaflarski et al., 2008) suggesting that treatment induced neural change is still in progress. Thus, immediate post-treatment scans may only reveal a partial story.

Few studies report changes in lateralization, the relative differences in activation between the RH and LH, and instead focus on each hemisphere as separate units. Richter et. al (2008) proposed that the decreased activity observed in the RH corresponding with increased language behavior was the result of increased efficiency of the RH. Another possibility, however, is that

the decreased RH activity indicates a leftward shift in lateralization. This may mean that the LH required less input from the LH homologue following treatment, supporting studies that have indicate the recovering brain's increased reliance on LH perilesional regions. Kurland and colleagues (2007) reported continued increase in activation (in the LH) even after *overlearning* of words in which reading accuracy was 95% prior to a final training session. It is in overlearning that decreased activation would seem most appropriate and yet it was not the case.

The current study included analysis of neural activation in three regions of interest within each hemisphere and then showed changes in laterality in these three areas as well as in the whole brain. In addition, four scans per participant allowed for close examination of change in response to treatment following a first dose, a second dose and following a period of no treatment. The results of this study indicate that activation patterns immediately following treatment may vary significantly from those seen at later time points. M2 is a good example in that he showed what looks like mild deactivation in the RIFG for all conditions following treatment. His LIFG, on the other hand, showed activation for most conditions. IFG laterality then looks to shift strongly left and, if these were the only two data points, M2 would look to have shifted leftward as a result of treatment. This trend does not continue after the second treatment, at which time activation in the RH exceeded that in the LH, thus laterality shifted strongly rightward. This same pattern was even stronger in the MTG. What looked initially like a strong left shift, changed to a right shift following Treatment Period II and this was maintained at the eight-week follow-up. M2's STG however, was more predictable in its response to treatment with increasing RH activation over both treatment periods but there was a shift back to initial RH levels at follow-up.

This study does not add to data favoring either right or left hemisphere activation following treatment. In the IFG, perhaps the most well studied area since it often is implicated in lesions following stroke, we saw four individuals all shift to what appeared a more balanced use of both the right and left regions. Three participants showed greater left than right activation in this area pre-treatment, one of them very strongly LH dominant. A fourth participant showed moderate RH dominance. All four shifted such that at Follow-up Scan Four there was a greater balance between hemispheres than pre-treatment, perhaps emphasizing the interdependency of the LH spared tissue and the supporting RH homologue during the treatment process.

Individual differences were anticipated due to the range of lesion severity. Those with less extensive damage were predicted to have the greater leftward shift based on the supposition that less damage means there is more healthy tissue available to recruit in the LH. All four individuals had extensive damage despite the wide variability in behavioral performance. The person with the least damage was not the one to shift most leftward, nor was the person with the mildest aphasia symptoms. Thus, the prediction was not borne out.

Though patterns of activation and lateralization reflect inter-participant variability, the RH, in all cases, appears to respond to treatment sensibility, appropriately reflecting the level of difficulty of stimuli for each participant. Considering that improvements in naming correspond with this increased activation, it can be inferred that the RH is then responding in a functional way, supporting the results seen following cTBS in the recent study by Hartwigsen and colleagues (2013).

Most importantly, this study highlights that a) neuronal change continues to be possible in people with chronic aphasia after a second treatment of CILT b) a pre- and post- scan alone may be misleading, especially following an intensive treatment where positive behavioral

changes continue once treatment has ended and c) single subject design allows for the analysis of patterns of four individuals revealing information that is often lost when analyzing at group data alone. It is likely that recovery patterns vary widely and may depend more on the individual, lesion and premorbid language organization (cf. Saur et al., 2010), than on treatment.

The Role of Constraint Induced Language Therapy

CILT (also referred to as CIAT or ILAT) has been the subject of some controversy. Opponents have taken issue with "constraint" to the verbal modality which may have prompted the name change from "Constraint Induced Language Therapy" to "Intensive Language-Action Therapy." Disallowing any form of expressive communication in an individual with aphasia runs contrary to the clinical mindset. In reality, CILT's emphasis on the verbal modality is no different from any other treatment of oral verbal expression. Response elaboration therapy (RET; Kearns, 1985) and Semantic Features Analysis (SFA; Boyle & Coelho, 1995) are two examples of treatments in which the oral verbal language is produced repeatedly in order to improve in this specific modality. Both of these treatments have been demonstrated to be effective and may be considered superior to CILT as they are better tailored to the individual's needs and since treatment takes place individually potentially allowing more opportunity for repetition and thus, neural change. CILT's group design is more focused on interactive productivity and is not customized to an individual. Despite this, no other treatment reports the consistent positive changes in pre-post standardized test scores, overall generalizability, and maintenance of gains seen following CILT. This is likely due to the fact that they are rarely administered at the same consistently high dose.

Another important aspect of CILT is the group effect. The potential drawback of shared time for verbal productions appears to be outweighed by the positive effects of peers working

together. The "Go-Fish" game aspect of CILT is repetitive and maintaining focus for three hours would be more difficult if not for the fact that there is an aspect of competition to the game keeping participants motivated, engaged and putting forth maximal effort. Motivation and salience are two other important neuroplastic factors that contribute to treatment gains. The support and encouragement from peers is equally important and perhaps more effective than when coming from the SLP. Group work with SFA, which tends to be administered in individual treatment, has shown promising results (Antonucci, 2009) in support of the idea that it is another factor that may contribute to outcomes seen following CILT. Group work and intensity appear to play significant roles, however in the few studies that have administered another treatment and CILT at equal intensities and in group settings, some still tend to find at least slight advantage with CILT. Maher (2006) reported better maintenance; Kurland (2012) reported better naming and only Rose (2013) reported no difference at all.

The role of intensity and group effect of various treatments should continue to be empirically tested but in the meantime CILT remains an effective tool for both treatment and research purposes, offering consistently positive outcomes for participants. It can also be a time-effective solution for clinical researchers compared to more drawn out, individual protocols in which several months may be necessary for data collection.

Limitations of the Study

Progress in naming was not adequately captured for S1 and S2. The participants with more severe aphasia did not meet criterion on the starting levels but it became apparent that progress was occurring as participants became more responsive to cues. Tracking success at a lower starting level such as accuracy given a phonemic or semantic cue, for example, would have provided more information for these participants.

Generalization to subsequent treatment levels weakens design. Ideally, in a multiple probe design, the various levels of treatment should not be influenced by the training of earlier levels (Thompson, 2006). There were regular increases in subsequent levels showing generalization of treatment for M1 and M2. This kind of generalization is positive in terms of outcome for the participants but calls into question the experimental control of the design.

More homogeneity between participants would increase interpretability of the data. The heterogeneity of individuals with aphasia is well-observed in the treatment literature. Four participants of similar severity would have been preferred but it became clear that this was not a possibility after protracted recruitment.

Inclusion of both correct and incorrect responses in neuroimaging data. This may be considered a limitation by some, since studies have reported differences in neural activation for correct and incorrect responses (e.g., Fridriksson et al., 2009; Kurland et al., 2012; Meinzer et al., 2006). Analyzing both may lead to questionable findings because error responses likely reflect increased processing demands. However, others take the position that participants use whatever processing resources are available to them when performing a given linguistic tasks and changes in language ability will be reflected by brain activation changes from pre-treatment (for more discussion, cf. Meinzer et al., 2013)

Conclusions and Future Directions

Evidence suggests that a double dose of CILT confers advantages over the single dose but optimal schedule of dosing is still unclear. Thirty hours, provided daily (Session Frequency), in three hour increments (Session Duration) and over two weeks (Total Intervention Duration) appears to be an effective, if not optimal, dose as positive results are the consistent result of administration of CILT and other intensive therapies. Increasing the Total Intervention Duration

while keeping other dosage parameters constant has shown some benefit when a new treatment type was provided immediately after the first as in Kurland (2012) and Rose (2013). In this dissertation study, when a second dose of the same treatment (CILT) was provided five weeks following completion of the first, there was benefit observed in both primary outcome measures and at least one secondary outcome measure. Gains were observed for all four participants despite the wide range of severity level. Neuroimaging data supported the utility of the second dose as greatest increases in neural activation in the IFG occurred following Treatment Period II for the three of four participants who were continued to demonstrate gains in naming accuracy. Neuroimaging data from this study also highlighted the importance of single subject design, the utility of scans from multiple time points and in analyzing RH activation as it relates to the recovery process rather than as a maladaptive process to be overcome.

Future research should address how to time dosages such that gains may be optimized. Study should first target periods of consolidation following intensive treatment investigating whether changes in neural activation strength or activation location correspond with continued behavioral changes. The next question should then address whether it is better to initiate a second block of treatment while this activation is still peaking or after it has plateaued.

Individuals with mild aphasia tend not to be the focus of research studies and may be discharged from services prematurely due to high achievement. This is unfortunate since this may be the group with the greatest likelihood of regaining employment. Results of correlation of initial severity of aphasia with improvements on test scores indicated that this population is less likely to benefit from CILT (Meinzer, Elbert, Djundja, Taub, & Rockstroh, 2007). Higher scores on standardized tests however, may leave less room for improvement on these measures.

Finally, more research is still needed comparing different treatment types using the same intensity level to determine whether intensity will actually strengthen the effect of any treatment type. In addition, the effect of using small groups to deliver other treatment types would be worthy of research to determine whether benefits are similar to those seen during CILT.

The current state of aphasia treatment is one of compensation and one that prepares patients and families for the limitations they will face. Emerging adjuvant therapies in conjunction with optimized treatment protocols may mean it is time for a paradigm shift in how we view aphasia rehabilitation. Optimization of various treatments require careful study of each of the parameters contributing to its success in bringing about neuroplastic change and a goal of full remediation should be the ultimate goal of aphasia researchers.

References

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

Aten, J. L., Caligiuri, M. P., & Holland, A. L. (1982). The efficacy of functional communication therapy for chronic aphasic patients. *Journal of Speech and Hearing Disorders*, *47*(1), 93–96.

Bakheit, A. M. O., Shaw, S., Barrett, L., Wood, J., Carrington, S., Griffiths, S., … Koutsi, F. (2007). A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. *Clinical Rehabilitation*, *21*(10), 885–894.

Bakheit, A. M. O., Shaw, S., Carrington, S., & Griffiths, S. (2007). The rate and extent of improvement with therapy from the different types of aphasia in the first year after stroke. *Clinical Rehabilitation*, *21*(10), 941–949.

Barthel, G., Meinzer, M., Djundja, D., & Rockstroh, B. (2008). Intensive language therapy in chronic aphasia: Which aspects contribute most? *Aphasiology*, *22*(4), 408–421.

Barthel, G., Meinzer, M., Djundja, D., & Rockstroh, B. (2008). Intensive language therapy in chronic aphasia: Which aspects contribute most? *Aphasiology*, *22*(4), 408–421.

Basso, A. C. (2001). Aphasia therapy or the importance of being earnest. *Aphasiology*, *15*(4), 307–332.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, *16*(4), 161–169.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, *16*(4), 161–169.

Berthier, M. L., Green, C., Lara, J. P., Higueras, C., Barbancho, M. A., Dávila, G., & Pulvermüller, F. (2009). Memantine and constraint-induced aphasia therapy in chronic poststroke aphasia. *Annals of Neurology, 65*(5), 577–585.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*, 987–992.

Bonakdarpour, B., Parrish, T. B., & Thompson, C. K. (2007). Hemodynamic response function in patients with stroke-induced aphasia: Implications for fMRI data analysis. *Neuroimage*, *36*(2), 322–331.

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.

Breier, J. I., Juranek, J., Maher, L. M., Schmadeke, S., Men, D., & Papanicolaou, A. C. (2009). Behavioral and Neurophysiologic Response to Therapy for Chronic Aphasia. *Archives of Physical Medicine and Rehabilitation*, *90*(12), 2026–2033.

Breier, J. I., Juranek, J., & Papanicolaou, A. C. (2011). Changes in maps of language function and the integrity of the arcuate fasciculus after therapy for chronic aphasia. *Neurocase, 17*(6), 506–517.

Breier, J. I., Maher, L. M., Novak, B., & Papanicolaou, A. C. (2006). Functional imaging before and after constraint-induced language therapy for aphasia using magnetoencephalography. *Neurocase*, *12*(6), 322–331.

Breier, J. I., Maher, L. M., Schmadeke, S., Hasan, K. M., & Papanicolaou, A. C. (2007). Changes in language-specific brain activation after therapy for aphasia using magnetoencephalography: A case study. *Neurocase*, *13*(3), 169–177.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research.

Carlomagno, S., Pandolfi, M., Marini, A., Di Iasi, G., & Cristilli, C. (2005). Coverbal gestures in Alzheimer's type dementia. *Cortex*, *41*(4), 535–546.

Cherney, L., Patterson, J., Raymer, A., Frymark, T., & Schooling, T. (2010). Updated evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia, *17*, 2011.

Cherney, L. R. (2012). Aphasia treatment: Intensity, dose parameters, and script training. *International Journal of Speech-Language Pathology*, *14*(5), 424–431.

Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1282.

Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of Aphasia Therapy: Evidence and Efficacy. *Current Neurology and Neuroscience Reports*, 1–10.

Code, C., Torney, A., Gildea-Howardine, E., & Willmes, K. (2010). Outcome of a one month therapy intensive for chronic aphasia: Variable individual responses. *Seminars in Speech and Language*, (31), 21–33.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

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

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173.

Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, *42*(6), 1014–1018.

Crosson, B., Moore, A. B., McGregor, K. M., Chang, Y. L., Benjamin, M., Gopinath, K., … others. (2009). Regional changes in word-production laterality after a naming treatment designed to produce a rightward shift in frontal activity. *Brain and Language*, *111*(2), 73–85.

Davis, G. A. (2003). PACE revisited. *Aphasiology*, *19*(1), 21–38.

Difrancesco, S., Pulvermüller, F., & Mohr, B. (2012). Intensive language-action therapy (ILAT): The methods. *Aphasiology*, *26*(11), 1317–1351.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*, 795–805.

Duffy, J. R. (2005). *Motor speech disorders* (2nd ed.). St. Louis, MI: Elsevier Mosby.

Duffy, J. R. (2012). *Motor speech disorders: Substrates, differential diagnosis, and management*.

Faroqi-Shah, Y., & Virion, C. R. (2009). Constraint-induced language therapy for agrammatism: Role of grammaticality constraints. *Aphasiology*, *23*(7-8), 977–988.

Flanagan, J. L., & Jackson, S. T. (1997). Test-retest reliability of three aphasia tests: Performance of non-brain-damaged older adults. *Journal of Communication Disorders*, *30*(1), 33–43.

Francis, W., & Kucera, H. (1982). Frequency analysis of English usage.

Frank, A. F., Tanenhaus, M. K., Aslin, R. N., & Salverda, A. P. (n.d.). Frequency, Neighborhood Density, and Phonological Similarity Effects in Picture Naming: An Artificial Lexicon Study.

Fridriksson, J., Baker, J. M., Whiteside, J., Eoute Jr, D., Moser, D., Vesselinov, R., & Rorden, C. (2009). Treating visual speech perception to improve speech production in nonfluent aphasia. *Stroke*, *40*(3), 853–858.

Fridriksson, J., Bonilha, L., Baker, J. M., Moser, D., & Rorden, C. (2010a). Activity in preserved left hemisphere regions predicts anomia severity in aphasia. *Cerebral Cortex*, *20*(5), 1013–1019.

Fridriksson, J., Bonilha, L., Baker, J. M., Moser, D., & Rorden, C. (2010b). Activity in preserved left hemisphere regions predicts anomia severity in aphasia. *Cerebral Cortex*, *20*(5), 1013–1019.

Fridriksson, J., Moser, D., Bonilha, L., Morrow-Odom, K. L., Shaw, H., Fridriksson, A., … Rorden, C. (2007). Neural correlates of phonological and semantic-based anomia treatment in aphasia. *Neuropsychologia*, *45*(8), 1812–1822.

Fridriksson, J., Richardson, J. D., Fillmore, P., & Cai, B. (2011). Left hemisphere plasticity and aphasia recovery. *NeuroImage*.

Goral, M., & Kempler, D. (2009). Training verb production in communicative context: Evidence from a person with chronic non-fluent aphasia. *Aphasiology*, *23*(12), 1383–1397.

Hadar, U., Burstein, A., Krauss, R., & Soroker, N. (1998). Ideational gestures and speech in brain-damaged subjects. *Language and Cognitive Processes*, *13*(1), 59–76.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., … Bowtell, R. W. (1999). Sparse temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223.

Hartwigsen, G., Saur, D., Price, C. J., Ulmer, S., Baumgaertner, A., & Siebner, H. R. (2013). Perturbation of the left inferior frontal gyrus triggers adaptive plasticity in the right

homologous area during speech production. *Proceedings of the National Academy of Sciences*, *110*(41), 16402–16407.

Heiss, W. D., Kessler, J., Thiel, A., Ghaemi, M., & Karbe, H. (1999). Differential capacity of left and right hemispheric areas for compensation of poststroke aphasia. *Annals of Neurology*, *45*(4), 430–438.

Holland, A. L., Frattali, C., & Fromm, D. (1999). *Communication Activities of Daily Living: CADL-2*.

Hula, W., Donovan, N. J., Kendall, D. L., & Gonzalez-Rothi, L. J. (2010). Item response theory analysis of the Western Aphasia Battery. *Aphasiology*, *24*(11), 1326–1341.

Jenkins, T., Coelho, C., & Coppola, M. (2014). *Effects of gesture restriction on discourse quality*. Presented at the Cognitive Neuroscience Society Annual Meeting, Boston, MA.

Johnson, M. L., Taub, E., Harper, L. H., Wade, J. T., Bowman, M. H., Bishop-McKay, S., … Uswatte, G. (2013). An Enhanced Protocol for Constraint-Induced Aphasia Therapy II: A Case Series. *American Journal of Speech-Language Pathology*, 1–13.

Kaplan, E., Goodglass, H., Weintraub, S., & others. (1983). Boston naming test.

Kearns, K. P. (1985). Response elaboration training for patient initiated utterances.

Kelly, H., Brady, M. C., & Enderby, P. (2010). Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev*, (5).

Kempler, D., & Goral, M. (2011). A comparison of drill- and communication-based treatment for aphasia. *Aphasiology*, *25*(11), 1327–1346.

Kempler, D., Goral, M., & Tison, K. (2006). Structured language treatment improves social communication in chronic aphasia. *Brain and Language*, *99*(1), 49–50.

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

Kiran, S., Ansaldo, A., Bastiaanse, R., Cherney, L. R., Howard, D., Faroqi-Shah, Y., … Thompson, C. K. (2012). Neuroimaging in aphasia treatment research: standards for establishing the effects of treatment. *Neuroimage*.

Kiran, S., Ansaldo, A., Bastiaanse, R., Cherney, L. R., Howard, D., Faroqi-Shah, Y., … Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: standards for establishing the effects of treatment. *Neuroimage, 76*, 428–435.

Kiran, S., Sandberg, C., & Sebastian, R. (2011). Treatment of category generation and retrieval in aphasia: Effect of typicality of category items. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1101–1117.

Kirmess, M., & Maher, L. M. (2010). Constraint induced language therapy in early aphasia rehabilitation. *Aphasiology*, *24*(6-8), 725–736.

Kleim, J. A., Barbay, S., Cooper, N. R., Hogg, T. M., Reidel, C. N., Remple, M. S., & Nudo, R. J. (2002). Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. *Neurobiology of Learning and Memory*, *77*(1), 63–77.

Kleim, J. A., & H. (2004). Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning. *J Neurosci*, *24*, 628–633.

Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research*, *51*(1), S225–S239.

Kunkel, A., & K. (1999). Constraint-induced movement therapy for motor recovery in chronic stroke patients. *Archives of Physical Medicine and Rehabilitation*, *80*, 624–628.

Kurland, J., Cortes, C., Sperling, A., Lott, S., Lacey, E., Orchinik, L., … Friedman, R. B. (2007). Functional reorganization supporting learning and maintenance in a case of phonologic alexia. *Brain and Language*, *103*(1), 99–100.

Kurland, J., Pulvermüller, F., Silva, N., Burke, K., & Andrianopoulos, M. (2012). Constrained vs. Unconstrained Intensive Language Therapy in Two Individuals with Chronic, Moderate-to-Severe Aphasia and Apraxia of Speech: Behavioral and fMRI Outcomes. *American Journal of Speech-Language Pathology*.

Kurland, J., Silva, N., Burke, K., & Iyer, P. (2011). Treatment-induced neuroplasticity following intensive constrained and unconstrained language therapy in a case of severe non-fluent aphasia. In *Procedia - Social and Behavioral Sciences* (Vol. 23, pp. 133–134).

Leon, S. A., Maher, L. M., & Gonzalez Rothi, L. J. (2011). Language therapy. In S. A. Raskin (Ed.), *Neuroplasticity and rehabilitation.* (pp. 209–232). New York, NY US: Guilford Press.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Developmental and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*(1), 113–124.

Mackenzie, C. (1991). An aphasia group intensive efficacy study. *International Journal of Language & Communication Disorders*, *26*(3), 275–291.

Maher, L. M., Kendall, D., Swearengin, J. A., Rodriguez, A., Leon, S. A., Pingel, K., … Rothi, L. J. G. (2006). A pilot study of use-dependent learning in the context of Constraint Induced Language Therapy. *Journal of the International Neuropsychological Society*, *12*(6), 843–852.

Maher, L., Singletary, F., Swearingen, M., Moore, A., Wierenga, C., Crosson, B., & others. (2002). An errorless learning approach to sentence generation in aphasia. *Proceedings. Rehabilitation Research for the 21st Century: The New Challenges*.

Mark, V. W., & Taub, E. (2004). Constraint-induced movement therapy for chronic stroke hemiparesis and other disabilities. *Restorative Neurology and Neuroscience*, *22*(3-5), 317–336.

McNeil, M., Pratt, S., Sung, J. E., Szuminsky, N., Ventura, M. B., Kim, A., … Musson, N. (2008). Test-retest reliability of the auditory Computerized Revised Token Test (CRTT) and three experimental reading CRTT-R versions in normal elderly individuals and persons with aphasia.

McNeil, M. R., Darley, F. L., Olsen, W. O., & Rose, D. E. (1983). Effects of Intensity Variations on Auditory Processing in Aphasia I. Equal Intensities at Each Ear. *International Journal of Audiology*, *22*(6), 560–581.

McNeil, M. R., Odell, K., & Tseng, C.-H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical Aphasiology*, *20*(21-39). Retrieved from http://aphasiology.pitt.edu/archive/00000128/01/20-03.pdf

McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. University Park Press Baltimore, MD.

Meinzer, M., Beeson, P. M., Cappa, S., Crinion, J., Kiran, S., Saur, D., … Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: consensus and practical guidelines for data analysis. *Neuroimage*, *73*, 215–224.

Meinzer, M., & Breitenstein, C. (2008). Functional imaging studies of treatment-induced recovery in chronic aphasia. *Aphasiology*, *22*(12), 1251–1268.

Meinzer, M., & D. (2005). Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke*, *36*(7), 1462–1466.

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.

Meinzer, M., Elbert, T., Djundja, D., Taub, E., & Rockstroh, B. (2007). Extending the Constraint-Induced Movement Therapy (CIMT) approach to cognitive functions: Constraint-Induced Aphasia Therapy (CIAT) of chronic aphasia. *NeuroRehabilitation*, *22*(4), 311–318.

Meinzer, M., Elbert, T., Djundja, D., Taub, E., & Rockstroh, B. (2007). Extending the Constraint-Induced Movement Therapy (CIMT) approach to cognitive functions: Constraint-Induced Aphasia Therapy (CIAT) of chronic aphasia. *NeuroRehabilitation*, *22*(4), 311–318.

Meinzer, M., Elbert, T., Wienbruch, C., Djundja, D., Barthel, G., & Rockstroh, B. (2004). Intensive language training enhances brain plasticity in chronic aphasia. *BMC Biology*, *2*(1), 20.

Meinzer, M., Flaisch, T., Breitenstein, C., Wienbruch, C., Elbert, T., & Rockstroh, B. (2008). Functional re-recruitment of dysfunctional brain areas predicts language recovery in chronic aphasia. *Neuroimage*, *39*(4), 2038–2046.

Meinzer, M., Flaisch, T., Obleser, J., Assadollahi, R., Djundja, D., Barthel, G., & Rockstroh, B. (2006). Brain regions essential for improved lexical access in an aged aphasic patient: a case report. *BMC Neurology*, *6*(1), 28.

Meltzer, J. A., Postman-Caucheteux, W. A., McArdle, J. J., & Braun, A. R. (2009). Strategies for longitudinal neuroimaging studies of overt language production. *Neuroimage*, *47*(2), 745–755.

Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgärtner, A., Schiffbauer, H., … Breitenstein, C. (2009). Imaging short- and long-term training success in chronic aphasia. *BMC Neuroscience*, *10*, 118.

Morris, D. C., & Bickel, C. S. (2011). Activity-Based Interventions for Neurorehabilitation. In S. A. Raskin (Ed.), *Neuroplasticity and rehabilitation.* (pp. 117–148). New York, NY US: Guilford Press.

Mozeiko, J., Myers, E., & Coelho, C. (2011). A comparison of intensity dosage for Constraint Induced Language Therapy in aphasia. Presented at the Clinical Aphasiology Conference, Ft. Lauderdale, FL.

Musso, M., Weiller, C., Kiebel, S., Muller, S. P., Bulau, P., & Rijnties, M. (1999). Training-induced brain plasticity in aphasia. *Brain; a Journal of Neurology*, *122*, 1781–1790.

Naeser, M. A., & M. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. *Brain & Language*, *93*(1), 95–105.

Nicholas, L. E., & Brookshire, R. H. (1993). A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults With Aphasia. *J Speech Hear Res*, *36*(2), 338–350.

Nudo, R. J., & Milliken, G. W. (1996). Reorganization of movement representations in primary motor cortex following focal ischemic infarcts in adult squirrel monkeys. *J.Neurophysiol*, *75*, 2144–2149.

Nudo, R. J., Plautz, E. J., & Frost, S. B. (2001). Role of adaptive plasticity in recovery of function after damage to motor cortex. *Muscle & Nerve*, *24*(8), 1000–1019.

Pascual-Leone, A. . A. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, *28*, 377–401.

Poeck, K., Huber, W., & Willmes, K. (1989). Outcome of intensive language treatment in aphasia. *Journal of Speech and Hearing Disorders*, *54*(3), 471.

Prabhakaran, R., Blumstein, S. E., Myers, E. B., Hutchison, E., & Britton, B. (2006). An event-related fMRI investigation of phonological-lexical competition. *Neuropsychologia*, *44*(12), 2209–2221.

Pulvermüller, F., Hauk, O., Zohsel, K., Neininger, B., & Mohr, B. (2005). Therapy-related reorganization of language in both hemispheres of patients with chronic aphasia. *Neuroimage*, *28*(2), 481–489.

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

Raboyeau, G., De Boissezon, X., Marie, N., Balduyck, S., Puel, M., Bézy, C., … Cardebat, D. (2008). Right hemisphere activation in recovery from aphasia: lesion effect or function recruitment? *Neurology*, *70*(4), 290–298.

Ramsberger, G., & Marie, B. (2007). Self-administered cued naming therapy: A single-participant investigation of a computer-based therapy program replicated in four cases. *American Journal of Speech-Language Pathology*, *16*(4), 343.

Rapp, B., Caplan, D., Edwards, S., Visch-Brink, E., & Thompson, C. K. (2013). Neuroimaging in aphasia treatment research: Issues of experimental design for relating cognitive to neural changes. *Neuroimage, 73*, 200–207.

Raven, J. C., Court, J. H., & Raven, J. (1988). Progressive Matrices Couleur/Colored Progressive Matrices. *Les Editions Du Centre de Psychologie Appliquée, Paris*.

Raymer, A. M., Beeson, P., Holland, A., Kendall, D., Maher, L. M., Martin, N., … others. (2008). Translational research in aphasia: from neuroscience to neurorehabilitation. *Journal of Speech, Language, and Hearing Research*, *51*(1), S259.

Richter, M., Miltner, W. H. R., & Straube, T. (2008). Association between therapy outcome and right-hemispheric activation in chronic aphasia. *Brain*, *131*(5), 1391–1401.

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

Robey, R. R., & Beeson, P. M. (2005). Aphasia treatment: Examining the evidence. In *Presentation at the American Speech-Language-Hearing Association Annual Convention. San Diego, CA.*

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

Rose, M., & Attard, M. (2011). *Multi-modality aphasia therapy (M-MAT): A procedural manual. Melbourne: La Trobe University.* Melbourne: La Trobe University.

Rose, M., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology*, *16*(10-11), 1001–1030.

Rose, M. L., Attard, M. C., Mok, Z., Lanyon, L. E., & Foster, A. M. (2013). Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study. *Aphasiology*, *27*(8), 938–971.

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

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

Saur, D., & Hartwigsen, G. (2012). Neurobiology of Language Recovery After Stroke: Lessons From Neuroimaging Studies. *Archives of Physical Medicine and Rehabilitation*, *93*(1, Supplement), S15–S25.

Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain*, *129*(6), 1371–1384.

Saur, D., Ronneberger, O., Kümmerer, D., Mader, I., Weiller, C., & Klöppel, S. (2010). Early functional magnetic resonance imaging activations predict language outcome after stroke. *Brain*, *133*(4), 1252–1264.

Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for Plasticity in White-Matter Tracts of Patients with Chronic Broca's Aphasia Undergoing Intense Intonation-based Speech Therapy. *Annals of the New York Academy of Sciences*, *1169*(1), 385–394.

Schneider, W., Eschman, A., Zuccolotto, A., & Guide, E.-P. U. (2002). Psychology Software Tools Inc. *Pittsburgh, USA.*

Schuhmann, T., Schiller, N. O., Goebel, R., & Sack, A. T. (2012). Speaking of which: dissecting the neurocognitive network of language production in picture naming. *Cerebral Cortex*, *22*(3), 701–709.

Sebastian, R., & Kiran, S. (2011). Task-modulated neural activation patterns in chronic stroke patients with aphasia. *Aphasiology*, *25*(8), 927–951.

Seghier, M. L. (2008). Laterality index in functional MRI: methodological issues. *Magnetic Resonance Imaging*, *26*(5), 594–601.

Shewan, C. M., & Donner, A. P. (1988). A comparison of three methods to evaluate change in the spontaneous language of aphasic individuals. *Journal of Communication Disorders*, *21*(2), 171–176.

Shewan, C. M., & Kertesz, A. (1980). Reliability and Validity Characteristics of the Western Aphasia Battery (WAB). *J Speech Hear Disord*, *45*(3), 308–324.

Sickert, A., Anders, L.-C., Münte, T. F., & Sailer, M. (2014). Constraint-induced aphasia therapy following sub-acute stroke: a single-blind, randomised clinical trial of a modified therapy schedule. *Journal of Neurology, Neurosurgery & Psychiatry*, *85*(1), 51–55.

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

Stark, R., Schienle, A., Walter, B., Kirsch, P., Sammer, G., Ott, U., … Vaitl, D. (2003). Hemodynamic responses to fear and disgust-inducing pictures: an fMRI study. *International Journal of Psychophysiology*, *50*(3), 225–234.

Szaflarski, J. P., Ball, A. L., Grether, S., Al-fwaress F., Griffith, N. M., Neils-Strunjas, J., … Reichhardt, R. (2008). Constraint-induced aphasia therapy stimulates language recovery in patients with chronic aphasia after ischemic stroke. *Medical Science Monitor*, *14*(5), CR243–CR250.

Tanner, D., & Culbertson, W. (1999). *The Quick Assessment of Apraxia of Speech*. Oceanside, CA: Academic Communication Associates, Inc.

Taub, E., Miller, N. E., & Novack, T. A. (1993). Technique to improve chronic motor deficit after stroke. *Arch Phys Med Rehabil., 74*, 347–354.

Teasell, R., Foley, N., Salter, K., Bhogal, S., Jutai, J., & Speechley, M. (2009). Evidence-based review of stroke rehabilitation: Executive summary, 12th edition. *Topics in Stroke Rehabilitation*, *16*(6), 463–488.

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders, 39*(4), 266–291.

Thompson, C. K., den Ouden, D. B., Bonakdarpour, B., Garibaldi, K., & Parrish, T. B. (2010). Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. *Neuropsychologia, 48*(11), 3211–3227.

Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, *16*(1), 30–42.

Wambaugh, J. L., Martinez, A. L., McNeil, M. R., & Rogers, M. A. (1999). Sound production treatment for apraxia of speech: Overgeneralization and maintenance effects. *Aphasiology*, *13*(9-11), 821–837.

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

Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience*.

Winhuisen, L., Thiel, A., Schumacher, B., Kessler, J., Rudolf, J., Haupt, W. F., & Heiss, W. D. (2005). Role of the Contralateral Inferior Frontal Gyrus in Recovery of Language

Function in Poststroke Aphasia A Combined Repetitive Transcranial Magnetic Stimulation and Positron Emission Tomography Study. *Stroke, 36*(8), 1759–1763.

Yang, G., Lai, C. S. W., Cichon, J., Ma, L., Li, W., & Gan, W.-B. (2014). Sleep promotes branch-specific formation of dendritic spines after learning. *Science, 344*(6188), 1173–1178.

Appendices

Appendix A: Communicative Effectiveness Index (CETI; Lomas et al., 19890

Communicative Effectiveness Index

At outset: mark an 'X' on each scale line to indicate how well the patient performs the task.
At end; repeat rating on a second, clean scoresheet after looking each time at the pre-therapy 'X' location.

Please rate _____'s ability at …

Not at All Able — *As Able As Before Stroke*

1. Getting somebody's attention.
2. Getting involved in group conversations that are about him/her.
3. Giving yes and no answers appropriately.
4. Communicating his/her emotions.
5. Indicating that he/she understands what is being said to him/her.
6. Having coffee-time visits and conversations with friends and neighbors (around the bedside or at home).
7. Having a one-to-one conversation with you.
8. Saying the name of someone whose face is in front of him/her.
9. Communicating physical problems such as aches and pains.
10. Having a spontaneous conversation (i.e., starting the conversation and/or changing the subject).
11. Responding to or communicating anything (including yes or no) without words.
12. Starting a conversation with people who are not close family.
13. Understanding writing.
14. Being part of a conversation when it is fast and there are a number of people involved.
15. Participating in a conversation with strangers.
16. Describing or discussing something in depth

Appendix B: Results of Monte Carlo Simulation for Each Participant

	Participant			
p value	M1	M2	S1	S2
.001	6; .033	6; .032	6; .034	6; .04
.005	11; .036	11; .035	11; .036	11; .041
.01	15; .042	15; .044	15; .041	15; .05
.025	28; .044	28; .044	28; .044	28; .049
.05	54; .048	54; .05	54; .05	55; .05

Note. Voxel cluster sizes and alpha value are provided for each participant for five p-values. Participant- M1 and M2- participants with mild aphasia; S1 and S2- participants with severe aphasia.

Appendix C: Voxel Cluster Tables (C1-4) Corresponding with Linear Contrasts and Lesion Tracings For Each Participant

Table C1

Significant Clusters For Three Contrasts for Participant M1, Thresholded at a Corrected Threshold of $p < .05$

		Maximum Intensity Coordinates (T-T)			
Anatomical localization (TT)	Number of activated voxels	x	y	z	Max-imum t Value
Main Effect of Time (Pre-Treatment- Post-Treatment Period I (Scan One- Scan Three)),Voxel-wise p < .001, Clusters of 20 Contiguous Voxels					
Right Middle Frontal Gyrus, Right Superior Temporal Gyrus, Right Superior Frontal Gyrus, Right Middle Temporal Gyrus	14676	-1.5	13.5	77.5	-10.1
Right Cingulate Gyrus, Left Cingulate Gyrus, Right Paracentral Lobule, Right Medial Frontal Gyrus	64	-1.5	-10.5	35.5	-6.5
Middle Frontal Gyrus, Left Superior Frontal Gyrus	46	40.5	-25.5	50.5	-5.4
Left Cerebellum	42	22.5	19.5	-42.5	-5.7
Left Superior Parietal Lobule, Left Precuneus, Left Inferior Parietal Lobule	40	40.5	61.5	56.5	-5.4
Left Culmen, Left Fusiform Gyrus	35	40.5	28.5	-27.5	6.7
Left Superior Temporal Gyrus, Left Transverse Temporal Gyrus	28	31.5	43.5	20.5	4.6

Left Medial Frontal Gyrus	27	4.5	1.5	-27.5	-6.2
Right Cingulate Gyrus, Right Posterior Cingulate, Left Cingulate Gyrus	27	-1.5	28.5	26.5	-5.1
Left Superior Parietal Lobule	24	28.5	61.5	59.5	-5.8
Right Superior Temporal Gyrus	22	-43.5	-25.5	-27.5	7.5
none	22	16.5	40.5	14.5	-7
Left Middle Frontal Gyrus, Left Inferior Frontal Gyrus	21	40.5	-37.5	14.5	-5.2
Left Middle Occipital Gyrus, Left Parahippocampal Gyrus	20	28.5	55.5	2.5	-4.3
Main Effect of Difficulty (All Hard Stimuli- All Easy Stimuli), Voxel-wise p < .001, Clusters of 6 Contiguous Voxels					
Right Inferior Frontal Gyrus, Right Thalamus, Right Lentiform Nucleus, Left Thalamus, Right Caudate, Right Insula, Left Parahippocampal Gyrus, Left Lentiform Nucleus	1171	-31.5	-10.5	12.5	6.13
Left Superior Frontal Gyrus, Right Medial Frontal Gyrus, Left Cingulate Gyrus, Left Medial Frontal Gyrus , Right Superior Frontal Gyrus, Right Cingulate Gyrus	512	7.5	-4.5	74.5	5.7
Left Cerebellar Tonsil, Right Cerebellar Tonsil, Right Culmen	339	-25.5	64.5	-12.5	6

Left Superior Frontal Gyrus, Left Middle Frontal Gyrus, Left Medial Frontal Gyrus, Right Medial Frontal Gyrus	279	16.5	-67.5	29.5	5.6
Right Precentral Gyrus, Right Middle Frontal Gyrus, Right Inferior Frontal Gyrus, Right Postcentral Gyrus	270	-55.5	1.5	50.5	5.2
Left Middle Frontal Gyrus, Left Precentral Gyrus, Left Superior Frontal Gyrus	244	31.5	-7.5	65.5	5.5
Right Cuneus, Right Precuneus, Left Precuneus, Left Cuneus	194	-19.5	88.5	41.5	5.5
Right Culmen	82	-16.5	25.5	-24.5	5
Right Superior Temporal Gyrus,Right Middle Temporal Gyrus	82	-55.5	67.5	14.5	5
Right Precuneus, Right Superior Parietal Lobule	78	-1.5	55.5	71.5	5.4
Right Parahippocampal Gyrus, Right Thalamus	62	-28.5	43.5	-0.5	4.1
Left Posterior Cingulate, Right Posterior Cingulate	59	1.5	55.5	8.5	4.1
none	50	-19.5	13.5	74.5	4.3
Left Inferior Frontal Gyrus, Left Middle Frontal Gyrus	43	46.5	-43.5	-0.5	4.4
Right Supramarginal Gyrus, Right Inferior Parietal Lobule	39	-64.5	40.5	26.5	4.3
Right Declive, Right Tuber, Right Uvula	33	-28.5	76.5	-18.5	-4.6

Left Superior Temporal Gyrus, Left Insula, Left Transverse Temporal Gyrus	31	37.5	22.5	14.5	4.8
Right Lingual Gyrus, Right Cuneus	30	-10.5	88.5	5.5	4.3
brain stem	28	1.5	25.5	-36.5	4.5
Left Precentral Gyrus, Left Postcentral Gyrus.	28	49.5	13.5	59.5	4.6
Left Superior Parietal Lobule, Left Inferior Parietal Lobule	28	34.5	58.5	53.5	4.4
Left Culmen, Left Fusiform Gyrus, Left Inferior Temporal Gyrus	27	40.5	43.5	-21.5	4.6
Right Superior Frontal Gyrus, Right Medial Frontal Gyrus, Right Middle Frontal Gyrus	26	-19.5	-61.5	-0.5	4
Left Superior Parietal Lobule, Left Postcentral Gyrus, Left Precuneus	26	13.5	61.5	68.5	4.8
Left Precentral Gyrus, Left Postcentral Gyrus.	23	55.5	10.5	29.5	4.3
Left Middle Frontal Gyrus, Left Superior Frontal Gyrus	23	34.5	-22.5	56.5	4.2
Right Inferior Occipital Gyrus, Right Middle Occipital Gyrus, Right Fusiform Gyrus	22	-34.5	85.5	-12.5	4.4
Right Precentral Gyrus, Right Postcentral Gyrus	22	-28.5	25.5	53.5	4.6
Left Culmen, Left Parahippocampal Gyrus	21	10.5	31.5	-18.5	4.4

Main Effect of Training (All Untrained Stimuli – All Trained Stimuli), Voxel-wise p < .025, Clusters of 28 Contiguous Voxels)					
Left Thalamus, Left Parahippocampal Gyrus, Right Parahippocampal Gyrus, Right Thalamus	99	1.9	31.6	0.9	4.1
Right Inferior Occipital Gyrus, Right Middle Occipital Gyrus, Right Lingual Gyrus, Right Fusiform Gyrus	84	-29.3	87.4	-5.1	3.7
Left Inferior Occipital Gyrus, Left Middle Occipital Gyrus, Left Fusiform Gyrus, Left Lingual Gyrus, Left Declive	81	34	81.1	-7.9	4
Right Lentiform Nucleus, Right Parahippocampal Gyrus	57	-19.1	0.8	-7.2	4.1
Left Declive, Left Uvula, Left Pyramis	44	23.5	66.7	-21.3	3.3
Left Lingual Gyrus, Left Inferior Occipital Gyrus	36	9.9	94.3	-9.8	3.2
Right Precentral Gyrus, Right Inferior Frontal Gyrus	36	-61.8	-2.4	36.4	3.4
Right Superior Frontal Gyrus, Right Middle Frontal Gyrus, Right Medial Frontal Gyrus	31	-18.1	-7.1	65.5	-2.9
Right Superior Temporal Gyrus, Right Inferior Frontal Gyrus	29	-46.1	-16.7	-19.4	3.1

Table C2

Significant Clusters For Three Contrasts for Participant M2, Thresholded at a Corrected Threshold of p < .05

Anatomical localization (TT)	Number of activated voxels	Maximum Intensity Coordinates (T-T) x	y	z	Maximum t Value
Main Effect of Time (Pre-Treatment- Post-Treatment Period I (Scan One- Scan Three)),Voxel-wise p < .001,Clusters of 20 Contiguous Voxels					
Right Cuneus, Right Precuneus,Left Cuneus, Left Lingual Gyrus, Right Lingual Gyrus, Right Middle Temporal Gyrus, Left Declive,Right Declive,Left Posterior Cingulate, Right Superior Parietal Lobule, Right Postcentral Gyrus, Right Middle Occipital Gyrus, Left Precuneus, Right Posterior Cingulate, Left Fusiform Gyrus	2413	-9	67.2	25	6.72
Right Precentral Gyrus, Right Superior Temporal Gyrus, Right Middle Temporal Gyrus, Right Postcentral Gyrus, Right Insula, Right Inferior Parietal Lobule	685	-52.2	15.1	13.1	-5.98
Left Inferior Frontal Gyrus,Left Middle Frontal Gyrus, Left Inferior Parietal Lobule, Left Insula, Left Superior Temporal Gyrus	417	58.1	-1.7	25.6	-4.67
Left Postcentral Gyrus,Left Paracentral Lobule,Left Precuneu, Left Medial Frontal Gyrus, Right Medial Frontal Gyrus, Left Precentral Gyrus,Right Paracentral Lobule, Left Superior Parietal Lobule	259	9.9	36.3	59.4	-3.43
Right Thalamus, Left Thalamus	151	-2.5	20.8	7.4	-4.96

Left Precuneus, Left Cuneus, Right Precuneus	136	14.9	70.6	51	4.03
Right Superior Frontal Gyrus, Right Middle Frontal Gyrus, Right Medial Frontal Gyrus	130	-19.4	-9.4	50.8	-4.85
Right Cingulate Gyrus, Left Cingulate Gyrus, Right Paracentral Lobule, Left Paracentral Lobule, Right Medial Frontal Gyrus, Left Medial Frontal Gyrus	105	-1.5	11.4	42.9	-5.05
Left Precuneus, Left Superior Parietal	100	22.8	66.4	47.3	-4.94
Right Inferior Parietal Lobule, Right Superior Temporal Gyrus, Right Supramarginal Gyrus,Right Middle Temporal Gyrus	96	-67.2	39	23	6.91
Left Medial Frontal Gyrus, Right Medial Frontal Gyrus, Left Cingulate Gyrus, Right Cingulate Gyrus, Left Middle Frontal Gyrus, Left Superior Frontal Gyrus, Left Anterior Cingulate, Right Anterior Cingulate	91	7.5	-35	27.3	-5.78
Right Middle Frontal Gyrus, Right Superior Frontal Gyrus	72	-38.3	-52.2	15.2	-3.8
Right Culmen,Right Fusiform Gyrus, Right Inferior Temporal Gyrus	60	-44.5	39.2	-22.2	-4.69
Right Middle Frontal Gyrus, Right Inferior Frontal Gyrus	58	-56.2	-36.4	14.8	-5.88

Left Middle Temporal Gyrus, Left Inferior Temporal Gyrus Left Fusiform Gyrus, Left Superior Temporal Gyrus	56	49.8	3.4	-24.4	-4.51
Left Pyramis,Left Uvula,Left Inferior Semi-Lunar	55	12.9	70.1	-28.8	-5.87
Right Middle Occipital Gyrus, Right Inferior Occipital Gyrus, Right Cuneus, Right Lingual Gyrus	49	-28.7	96.1	-0.7	5.57
Right Superior Frontal Gyrus	48	-13.6	-38.6	44.9	-4.49
Left Superior Frontal Gyrus, Left Middle Frontal Gyrus	44	24.2	-51.8	41.2	3.67
Left Superior Frontal Gyrus	44	-42.2	-0.3	39.5	-3.69
Left Cuneus	42	4.8	101.6	12.9	5.96
Right Caudate	42	-11.4	-11.1	13	-4.43
Left Transverse Temporal Gyrus	42	35.8	28.8	14.4	-5.11
Left Insula, Left Precentral Gyrus	41	38.7	-15.1	8.3	-5.63
Right Inferior Semi-Lunar Lobule	40	-17.1	80.5	-40.7	5.61
Right Subcallosal Gyrus	40	-13.9	-17.6	-5.9	3.8
Left Superior Frontal Gyrus	35	13.1	-43	42.2	-4.2
Right Middle Temporal Gyrus	34	-46.9	55.3	-5.1	-3.48
Right Precentral Gyrus	34	-20.4	27.6	72.4	5.35
Right Lingual Gyrus	33	-10.7	91.9	-5.9	-4.6
Right Precentral Gyrus	32	-54.6	-8.7	0.6	-3.38
Right Insula	32	-36.9	-13.6	11.4	-4.09
Right Precuneus, Right Inferior Parietal Lobule	32	-56.8	32.3	44.7	-5.12
Right Inferior Parietal Lobule	31	-46.3	73.4	40.8	3.57
Left Insula	29	37.4	9.9	5.4	-4.6
Left Superior Frontal Gyrus	28	4.5	-17.9	53.8	-3.84
Right Cerebellum	27	-31.6	64.4	-43.2	-4.38
Left Anterior Cingulate, Left Caudate	27	13.6	-20.2	-2.5	4.09

Left Caudate	26	11.9	-7.8	13.2	-4.51
Right Postcentral Gyrus, Right Inferior Parietal Lobule	26	-53.4	32.3	53	4.43
Left Middle Temporal Gyrus	24	60.9	64.3	3.1	5.59
Right Inferior Frontal Gyrus	24	-61.6	-9.1	13.9	-4.03
Left Superior Frontal Gyrus	24	7.9	-48.7	51.3	4.25
Right Inferior Frontal Gyrus, Right Precentral Gyrus	22	-54.9	-13.1	20.4	-3.48
Left Middle Frontal Gyrus, Left Superior Frontal Gyrus	21	24.9	-45.9	0	4.27
Right Anterior Cingulate, Left Anterior Cingulate	20	-4.1	-20.5	0.5	4.63
Left Postcentral Gyrus, Left Precuneus	20	8.4	55.3	69.2	4.39
Main Effect of Difficulty (All Hard Stimuli- All Easy Stimuli), Voxel-wise p < .001, Clusters of 6 Contiguous Voxels					
Left Anterior Cingulate, Right Anterior Cingulate, Left Medial Frontal Gyrus, Right Medial Frontal Gyrus	465	4.9	-35	3.5	-4.38
Left Lentiform Nucleus, Left Thalamus, Left Insula, Left Claustrum Left Parahippocampal Gyrus	317	19.8	3.4	7.5	3.314
Left Subcallosal Gyrus, Right Anterior Cingulate, Left Anterior Cingulate, Right Subcallosal Gyrus	302	0.8	0.4	-12.6	-5.42
Right Uncus, Right Inferior Temporal Gyrus, Right Fusiform Gyrus, Right Middle Temporal Gyrus, Right Parahippocampal Gyrus, Right Superior Temporal Gyrus	243	-38.5	4.5	-31.1	5.17
Right Inferior Semi-Lunar Lobule, Right Pyramis, Left Pyramis	229	-10.6	80.1	-39.5	5.35

Right Lentiform Nucleus, Right Caudate, Right Insul, Right Claustrum	188	-21.3	-7.4	5.6	5.525
Left Uncu, Left Inferior Temporal Gyrus, Left Fusiform Gyrus, Left Parahippocampal Gyrus, Left Middle Temporal Gyrus, Left Superior Temporal Gyrus	175	39.1	8.4	-29.4	4.15
Left Medial Frontal Gyrus, Left Superior Frontal Gyrus, Left Cingulate Gyrus	115	4.8	-4.2	53.7	3.68
Left Middle Frontal Gyrus, Left Superior Frontal Gyrus	81	36.3	-55.8	20.4	4.36
Right Middle Frontal Gyrus, Right Superior Frontal Gyrus	75	-34.3	-61.1	11	4.29
Right Precuneus, Left Precuneus, Right Cingulate, Left Cingulate	60	-3.3	52.3	31.6	-3.39
Left Middle Frontal Gyrus, Left Precentral Gyrus	55	51.3	-7.1	48.8	5.05
Left Precuneus,	51	18.5	53.4	45.1	-3.84
Right Inferior Frontal Gyrus	49	-41.9	-29.6	-20.1	-4.82
Left Superior Frontal Gyrus	44	17.5	-43.1	50.5	3.4
Right Lingual Gyrus, Right Cuneus, Right Inferior Occipital, Right Fusiform Gyrus	40	-17.2	98.1	-8.6	3.33
Right Thalamus, Left Thalamus	36	-0.8	22	-1.8	3.31
Left Superior Temporal Gyrus, Left Insula, Left Postcentral Gyrus, Left Transverse Temporal Gyrus	35	49	25	15.1	-4.03
Right Precentral Gyrus, Right Insula, Right Postcentral Gyrus	34	-51.7	8.8	13.1	-3.9
Right Cingulate Gyrus, Right Caudate	31	-23.5	36.6	21.2	-3.84
Right Inferior Semi- Lunar Lobule, Right Cerebellar Tonsil	23	-43.4	67.5	-35.6	4.51

Right Middle Temporal Gyrus, Right Superior Temporal Gyrus	23	-44.6	54.9	5.8	-5.49
Right Precuneus, Right Superior Parietal Lobule, Right Inferior Parietal	20	-21	51.2	42.7	-4.12
Right Middle Frontal Gyrus	20	-52	-14.4	43.6	4.47
Main Effect of Training (All Untrained Stimuli – All Trained Stimuli), Voxel-wise p < .025,Clusters of 28 Contiguous Voxels					
Right Superior Frontal Gyrus, Right Medial Frontal Gyrus,Right Middle Frontal Gyrus	119	-16.3	-45.5	35	-3.25
Right Precuneus, Right Paracentral Lobule, Right Superior Parietal Lobule	72	-20.3	45.7	49	-3.08
Left Inferior Frontal Gyrus, Left Precentral Gyrus	64	53.8	-6.8	24.3	-2.56
Left Cingulate, Left Posterior, Left Precuneus	54	7.8	47.9	25.7	-2.21
Left Postcentral Gyrus, Left Precentral Gyrus	51	52.6	18.4	40.6	-2.28
Right Inferior Semi-Lunar Lobule	42	-13.4	81.1	-41	2.78
Right Postcentral Gyrus, Right Precentral Gyrus, Right Inferior Parietal	35	-37.9	21.6	34	-3.13
Left Precentral Gyrus, Left Postcentral Gyrus	29	48.6	7	27.6	-2.96

Table C3

Significant Clusters For Three Contrasts for Participant S1, Thresholded at a Corrected Threshold of $p < .05$

Anatomical localization (TT)	Number of activated voxels	Maximum Intensity Coordinates (T-T) x	y	z	Maximum t Value
Main Effect of Time (Pre-Treatment- Post-Treatment Period I (Scan One- Scan Three)),Voxel-wise p < .001, Clusters of 6 Contiguous Voxels					
Right Precentral Gyrus	200	-24.6	3.5	68.1	-9.6
Right Post Central Gyrus	198	-10.7	43.2	68.7	-6.8
Left Cuneus, Left Posterior Cingulate, Right Posterior Cingulate	109	7.7	66.4	11.4	-5.3
Superior Frontal Gyrus	79	26.7	-32.9	12.6	-5.5
Right Declive, Left Declive	75	-1.1	67	-11.8	-5.3
Right Parahippocampal Gyrus	74	-11.8	33.6	5.8	-6.1
Right Thalamus	71	21.7	47.6	69.9	-10
Left Postcentral Gyrus, Left Superior Parietal Lobule	68	60.3	6.5	38.5	-7
Left Precentral Gyrus, Left Middle Frontal Gyrus	59	-28.3	63.1	56	-6.1
Right Superior Parietal Lobule, Right Precuneus	53	-13.5	19.2	79	-7
Right Cerebellar Tonsil, Right Culmen	51	-17.5	66.5	1.3	-4.6
Right Lingual Gyrus, Right Posterior Cingulate	50	66.5	37.7	23.1	8.4
Left Inferior Parietal Lobule, Left Superior Temporal Gyrus,	49	51.3	47.9	-27.4	-8.5
Left Tuber, Left Culmen	43	32.7	27.2	-35.2	-4.2
Left Precentral Gyrus, Left Postcentral Gyrus	42	48.6	10	55.2	-5.9
Right Middle Temporal Gyrus, Right Middle Occipital Gyrus	41	-35.1	68.2	19.3	-5
Right Cuneus, Right Lingual Gyrus	36	-7.6	81.6	10	-4.7

Left Superior Frontal Gyrus, Left Middle Frontal Gyrus	36	36.5	-20	54.3	-4.8
Left Caudate, Left Lentiform Nucleus	34	9.8	-11.3	-3.3	-4.7
Left Postcentral Gyrus, Left Superior Temporal Gyrus	33	66.4	15.8	17.3	8.8
Right Cuneus	33	-11.2	87.4	33.2	5.1
Right Cingulate Gyrus	33	-9.9	1.7	39.4	-5
Right Medial Frontal Gyrus, Right Superior Frontal Gyrus, Right Anterior Cingulate	28	-5.2	-53.9	-3.2	-4.6
Left Superior Parietal Lobule, Left Inferior Parietal Lobule	28	41.7	53.8	55.9	-4.8
Left Inferior Frontal Gyrus	27	27.1	-24.6	-15.4	-5.4
Right Postcentral Gyrus, Right Precentral Gyrus	27	-65.3	11.2	27.4	-6
Right Lentiform Nucleus	26	-17.9	0.4	-1.1	8.3
Left Middle Temporal Gyrus	26	68.7	41.5	0.1	-4.4
Left Caudate	26	14.1	-15.8	14.1	-6
Right Cuneus, Right Middle Occipital Gyrus	26	-5.4	93.6	15.6	-6
Left Postcentral Gyrus, Left Inferior Parietal Lobe	26	51.6	32.4	56.9	-5.8
Left Inferior Frontal Gyrus	25	46.9	-27.9	-2.6	-6.2
Left Lingual Gyrus, Left Cuneus	23	18.4	85.5	2.2	-7.2
Right Inferior Frontal Gyrus, Right Insula	23	-39.1	-25.5	7.5	-4.6
Left Superior Frontal Gyrus, Left Medial Frontal Gyrus	22	5.8	-59.9	31.4	-5
Left Middle Frontal Gyrus	22	32.1	0.8	67.2	-5.2
Right Superior Frontal Gyrus, Right Medial Frontal Gyrus	21	-23.3	-56.4	-3	-5.8
Main Effect of Difficulty (All Hard Stimuli- All Easy Stimuli), Voxel-wise p < .01, Clusters of 15 Contiguous Voxels					
Left Cuneus, Right Cuneus	115	1.7	80.1	19.7	-4.3
Left Lingual Gyrus, Left Cuneus	111	15.2	96.6	-3.1	4.4
Right Cuneus, Right Lingual Gyrus	36	-17.2	98.1	4.1	4.1
Inferior Frontal Gyrus, Left Middle Frontal Gyrus	30	52.7	-33.6	11.6	-3.7
Left Lentiform Nucleus	25	20.6	-10.9	-1.9	-4.3
Medial Frontal Gyrus	18	9.1	-50.7	10.8	-3.5

Right Middle Temporal Gyrus	16	-63.9	37.1	0	-3.3
Right Superior Temporal Gyrus	16	-58.6	-4.5	1.9	-3.8
Right Superior Temporal Gyrus	15	-49.5	16.4	-4.2	-4.2
Main Effect of Training (All Untrained Stimuli – All Trained Stimuli), Voxel-wise p < .025, Clusters of 28 Contiguous Voxels)					
Right Superior Temporal Gyrus, Right Parahippocampal Gyrus, Right Uncus, Right Middle Temporal Gyrus, Right Insula	526	-29.2	6	-20.2	-5.4
Left Parahippocampal Gyrus,Right Thalamus, Left Thalamus, Left Lentiform Nucleus	240	6.1	17.9	-4.7	-4.3
Left Caudate, Right Caudate, Left Anterior Cingulate	214	3.7	-21.2	5.3	4.2
Left Parahippocampal Gyrus, Left Uncus, Left Superior Temporal Gyrus, Left Lentiform Nucleus, Left Subcallosal Gyrus	127	26.9	-1.2	-17.3	-4.3
Right Lentiform Nucleus, Right Insula	107	-22.7	-0.8	10.2	-3.4
Right Culmen, Right Lingual Gyrus, Right Fusiform Gyrus, Right Declive	60	-13.3	64.9	-5.5	-3.9
Right Middle Frontal Gyrus, Right Superior Frontal Gyrus	55	-29.9	-57.1	8.4	-4.15
Right Cerebellar Tonsil	49	-18.8	17.7	-41.3	3.6
Left Medial Frontal Gyrus, Left Superior Frontal Gyrus, Left Anterior Cingulate	48	8.2	-52.1	-6.3	3.9
Left Cuneus, Left Lingual Gyrus	43	11.9	97.2	-0.3	-4.4
Right Precentral Gyrus, Right Postcentral Gyrus	42	-53.9	11.6	43.3	-3.9
Left Posterior Cingulate, Left Cingulate Gyrus, Left Precuneus	38	6.1	50.8	22.9	-3.6
Right Middle Frontal Gyrus, Right Inferior Frontal Gyrus	36	-53.6	-34.3	24.7	-3.6
Right Culmen, Right Cerebellar Lingual, Right Cerebellar Tonsil	35	-15.2	47.8	-21.7	-3.2
Right Parahippocampal Gyrus, Right Culmen, Right Fusiform Gyrus	34	-24	31.5	-12.1	-3.4
Left Culmen, Left Parahippocampal Gyrus	33	29.2	26.8	-23	-3.5

Table C4

Significant Clusters for three contrasts for participant S2

Anatomical localization (TT)	Number of activated voxels	Maximum Intensity Coordinates (T-T) x	y	z	Maximum t Value
Main Effect of Time (Pre-Treatment- Post-Treatment Period I (Scan One- Scan Three)),Voxel-wise p < .001, Clusters of 20 Contiguous Voxels					
Left Cuneus, Left Middle Occipital Gyrus, Left Middle Temporal Gyrus, Right Cuneus, Left Precuneus, Right Declive, Left Lingual Gyrus, Left Cerebellar Tonsil, Right Inferior Semi-Lunar Lobule	5530	10.6	71.4	-11.9	12.7
Right Middle Frontal Gyrus, Left Anterior Cingulate, Right Superior Frontal Gyrus, Right Anterior Cingulate, Left Medial Frontal Gyrus, Right Inferior Frontal Gyrus, Right Caudate, Right Medial Frontal Gyrus, Left Superior Frontal Gyrus	1086	-12.7	-38.5	-3.7	7.8
Right Middle Temporal Gyrus, Right Superior Temporal Gyrus, Right Inferior Temporal Gyrus, Right Parahippocampal Gyrus, Right Uncus, Right Fusiform Gyrus	897	-44.2	10	-23.1	6.6
Right Cingulate Gyrus, Right Precentral Gyrus, Right Medial Frontal Gyrus, Right Precuneus, Right Postcentral Gyrus, Right Insula	560	-22.5	22.5	40.5	-6.7

Left Middle Temporal Gyrus, Left Parahippocampal Gyrus, Left Uncus, Left Inferior Temporal Gyrus, Left Superior Temporal Gyrus, Left Fusiform Gyrus	436	43.2	1.6	-29.5	6.7
Left Middle Frontal Gyrus	292	39.6	-52	34.5	-6.2
Left Cingulate Gyrus	268	25	10.7	37.9	-6.8
Right Middle Frontal Gyrus	223	-48.2	-53.2	36	-5.5
Right Superior Frontal Gyrus	158	4.1	-72.4	35.4	-6.9
Left Cingulate Gyrus, Right Cingulate Gyrus, Left Posterior Cingulate, Left Precuneus, Right Posterior Cingulate	137	2.2	39.6	24.9	-5.7
Right Anterior Cingulate, Right Cingulate Gyrus, Left Cingulate Gyrus, Left Anterior Cingulate, Right Medial Frontal Gyrus	127	-9.9	-25.4	26.1	-4.9
Right Inferior Frontal Gyrus, Right Insula, Right Middle Frontal Gyrus, Right Precentral Gyrus	126	-41.8	-27.1	9.7	-7.6
Right Fusiform Gyrus, Right Declive, Right Culmen	109	-27.8	52.7	-10.9	-6.3
Left Medial Frontal Gyrus, Left Precentral Gyrus, Left Paracentral Lobule, Left Middle Frontal Gyrus, Left Superior Frontal Gyrus	75	13.4	18.1	57.9	-5.4
Right Culmen, Right Fusiform Gyrus, Right Parahippocampal Gyrus	74	-35.7	30.6	-28.8	-5.5

Left Cingulate Gyrus, Left Medial Frontal Gyrus	71	14.2	-21	33.9	-4.5
Right Lingual Gyrus, Right Cuneus	56	-15.4	73.1	1	-7.3
Left Superior Frontal Gyrus, Left Medial Frontal Gyrus	56	7.7	-54.4	37.2	6.1
Right Insula, Right Precentral Gyrus, Right Inferior Frontal Gyrus	52	-46.1	1.6	15	-5
0% accounted for	50	-16	-42.2	59.2	5.6
Right Culmen, Right Cerebellar Lingual, Left Culmen	48	0.2	37	-0.9	5.9
0% accounted for	48	-23.8	-27.1	64	5.1
Left Culmen, Left Culmen of Vermis, Right Culmen of Vermis,Right Culmen,Right Lingual Gyrus	43	0.7	63.4	-4.1	6.3
Left Parahippocampal Gyrus, Left Lingual Gyrus	43	17.6	37.7	0.4	5.1
Left Inferior Frontal Gyrus	42	34.8	-29.6	11.9	-5.8
Right Lentiform Nucleus	38	-25.2	-1.4	-1.4	-5
Right Culmen	36	-12.2	40.3	-17.7	4.7
Left Thalamus Left Lentiform Nucleus	36	10.7	3.6	8	7

Left Middle Occipital Gyrus, Left Middle Temporal Gyrus	35	38.3	65.3	2.6	-5.3
Left Superior Frontal Gyrus, Right Superior Frontal Gyrus	34	-0.9	-30.4	52.3	-5
Left Superior Frontal Gyrus, Left Medial Frontal Gyrus, Left Middle Frontal Gyrus, Left Anterior Cingulate	33	25.2	-41.4	8	-5.7
Right Superior Frontal Gyrus	33	-8.3	-1.4	73.3	4.6
Right Inferior Occipital Gyrus, Right Middle Occipital Gyrus	31	-39	89	-3.5	-7.3
Brain stem	30	-1.9	7.7	-32	-4.7
Left Insula	30	36.8	-12.3	15.1	-4.9
Right Precuneus, code 45 Right Posterior Cingulate, Right Cingulate Gyrus, Right Cuneus	29	-4.9	62.4	20.5	-5.8
Left Superior Frontal Gyrus, Left Middle Frontal Gyrus	29	11.1	-13.5	60.5	-5.2
Left Superior Frontal Gyrus, Left Middle Frontal Gyrus, Left Medial Frontal Gyrus	27	19.7	-51.3	25.5	5.2
Right Middle Temporal Gyrus, Right Inferior Temporal Gyrus	26	-52.9	54.4	-1.2	4.5
Left Fusiform Gyrus, Left Declive, Middle Occipital Gyrus	24	38	60.6	-13.8	-5.7
Right Postcentral Gyrus, Right Superior Parietal Gyrus	24	-23.9	42.2	61.4	-5.5

Right Middle Occipital Gyrus	23	-32	71.4	8	-6.1
Right Superior Frontal Gyrus	22	-20.4	-54.5	31.3	4.7
Right superior Frontal Gyrus	22	-16.2	-20.2	53.9	-5.3
Main Effect of Difficulty (All Hard Stimuli- All Easy Stimuli), Voxel-wise p < .025, Clusters of 28 Contiguous Voxels					
Left Superior Frontal Gyrus,Right Superior Frontal Gyrus, Right Medial Frontal Gyrus	224	8.6	-30.1	51.9	-4
Right Superior Frontal Gyrus, Right Middle Frontal Gyrus	196	-17.9	-30.8	51.2	-3.9
Left Fusiform Gyrus, Left Inferior Temporal Gyrus, Left Declive	170	54.9	40.6	-19.8	-3.9
Right Cingulate Gyrus, Left Medial Frontal Gyrus, Left Cingulate Gyrus, Right Medial Frontal Gyrus, Left Superior Frontal Gyrus	120	0.7	-12.4	42	-3.6
Right Cerebellum	119	-7.2	31.9	-55.3	3.6
Left Middle Temporal Gyrus, Left Superior Temporal Gyrus, Left Superior Occipital Gyrus	96	48.7	66	20.8	-3.8
Right Superior Frontal Gyrus, Right Medial Frontal Gyrus, Right Middle Frontal Gyrus, Right Anterior Cingulate	88	-21	-51.8	14.7	-3.2
Right Cuneus, Left Cuneus, Right Lingual Gyrus	84	-0.8	80.8	15.8	-4
Left Cerebellar Tonsil	82	11.8	48.3	-46.2	3.6

Left Caudate, Left Lentiform Nucleus, Left Claustrum	79	18.5	-12.6	10.1	-3.4
Left Middle Frontal Gyrus, Left Precentral Gyrus	72	42.2	-31.4	27.9	-4.2
Right Cerebellar Tonsil	64	-25.8	45.1	-47.9	3.6
Right Superior Frontal Gyrus, Right Middle Frontal Gyrus, Right Medial Frontal Gyrus, Right Precentral Gyrus	54	-14.2	10.6	65.7	-2.9
not accounted for	52	27.2	-70.2	33.8	3.2
Right Declive, Right Declive of Vermis, Left Declive of Vermis, Right Pyramis, Left Declive	50	-3.4	79.9	-19.4	-3.1
Left Middle Frontal Gyrus	48	43.6	-6.9	46.6	-3.5
Left Cingulate Gyrus, Left Anterior Cingulate, Left Medial Frontal Gyrus	46	9.3	-24.8	28.1	3.3
Right Middle Frontal Gyrus, Right Superior Frontal Gyrus	44	-41.6	-38.1	23.7	-3
Left Inferior Frontal Gyrus, Left Middle Frontal Gyrus	43	41.5	-35	-4.1	-3.8
Left Inferior Occipital Gyrus, Left Lingual Gyrus, Left Cuneus, Left Middle Occipital Gyrus	43	27.3	94.6	-3.7	-3.8
Left Parahippocampal Gyrus	41	34	33.6	-7	-3.4
Left Superior Frontal Gyrus, Left Medial Frontal Gyrus, Left Middle Frontal Gyrus	39	20.6	-51.5	11.6	-3.7

Right Anterior Cingulate, Right Caudate, Right Medial Frontal Gyrus	37	-8.9	-26.9	-1.8	-3.1
Left Lingual Gyrus, Left Uvula	34	7.3	93.2	-15.3	4.9
Right Uncus, Right Parahippocampal Gyrus	31	-18.9	10.6	-33.5	3
Left Superior Frontal Gyrus, Left Middle Frontal Gyrus	31	28.3	-45	28.3	-3.2
Right Declive, Right Fusiform Gyrus	30	-26.3	75.8	-15.8	-3.5
Right Medial Frontal Gyrus, Right Anterior Cingulate, Left Anterior Cingulate, Left Medial Frontal Gyrus	28	-2.2	-48.8	8.8	-3.1
Main Effect of Training (All Untrained Stimuli – All Trained Stimuli), Voxel-wise p < .025, Clusters of 28 Contiguous Voxels)					
Left Cingulate Gyrus, Left Medial Frontal Gyrus, Right Cingulate Gyrus, Right Medial Frontal Gyrus	98	3.6	-11.1	38.7	-3.7
Left Lingual Gyrus, Right Lingual Gyrus, Left Culmen, Left Declive, Right Declive, Left Fusiform Gyrus, Left Declive of Vermis, Left Parahippocampal Gyrus, Right Declive of Vermis, Left Culmen of Vermis	94	2.6	80.6	-8.3	4
Right Fusiform Gyrus, Right Middle Occipital Gyrus, Right Lingual Gyrus, Right Declive, Right Inferior Occipital Gyrus	49	-37.8	71.8	-10.4	3.7
Right Inferior Parietal Lobule, Right Superior Parietal Lobule	46	-43.7	52.2	49.2	-3.05
Right Precuneus, Left Precuneus, Right Cuneus, Right Cingulate Gyrus, Left Cingulate Gyrus, Left Cuneus	41	-1.7	56.5	32.6	-3.47

Right Paracentral Lobule, Right Precuneus, Left Paracentral Lobule, Left Precuneus	41	-7.9	36.7	48.8	-3.9
Right Lingual Gyrus, Right Fusiform Gyrus	40	-14.5	96.6	-16.9	-3.86
Right Anterior Cingulate	33	-11.4	-31.1	8.1	3.3