Published in final edited form as:

J Speech Lang Hear Res. 2014 April 1; 57(2): 439–454. doi:10.1044/1092-4388(2013/12-0224).

Delayed Stimulus-Specific Improvements in Discourse Following Anomia Treatment Using an Intentional Gesture

Lori J. P. Altmann¹, Audrey A. Hazamy¹, Pamela J. Carvajal¹, Michelle Benjamin^{1,2,3}, John C. Rosenbek¹, and Bruce Crosson^{1,2,4,5,6}

¹University of Florida, Gainesville, FL

²Malcom Randal VAMC, Brain Research and Rehabilitation Research Center, Gainesville, FL

³University of Alabama, Birmingham, Birmingham, AL

⁴RR&D Center of Excellence, Atlanta VAMC, Decatur, GA

⁵Emory University, Atlanta, GA

⁶Georgia State University, Atlanta, GA

Abstract

Purpose—This study assesses how the addition of intentional, left hand gestures to an intensive treatment for anomia affects two types of discourse, picture description and responses to open-ended questions.

Methods—Fourteen people with aphasia completed treatment for anomia comprising 30 treatment sessions over three weeks. Seven participants also incorporated intentional left hand gestures into each treatment trial.

Results—Both groups demonstrated significant changes in trained items and improved naming of untrained items but no change in WAB-AQ scores. Changes in discourse were limited to the 3-month follow-up assessment. Several discourse measures showed significant improvements in the picture description task and declines during question responses. Additionally, the Gesture group produced more words at each assessment, while the No-Gesture group produced fewer words at each assessment. These patterns led to large improvements in picture descriptions and minimal declines in question responses in the Gesture group. In contrast, the No-Gesture group showed minimal improvements in picture descriptions and production declines in question responses relative to pre-treatment levels.

Conclusions—The intensive treatment protocol is a successful method for improving picture naming even of untrained items. We hypothesize that the intentional left hand gesture played an essential role in enabling generalization of treatment to discourse.

Keywords

Aphasia treatment; intention; gesture; discourse; picture description; generalization

Anomia is the most common affliction in persons with aphasia due to stroke or other brain damage (Nickels & Berndt, 2001; Raymer & Rothi, 2001; Wilshire & Coslett, 2000). Therefore, it is no surprise that a wide variety of treatment protocols have been developed to treat word retrieval difficulties (for an extensive list of treatment studies see <http://aphasiatx.arizona.edu/>). While all treatment studies assess treatment specific outcomes, such as improved naming of trained pictures, increasingly studies also assess generalization to untrained items and tasks, such as accessing untrained words and producing discourse (Raymer et al., 2008; Thompson, 1989). In the current study, we report the effects of two variations of an intensive anomia treatment, one incorporating an intentional left hand movement (the Gesture treatment¹) and the other identical but without the hand movement (the No Gesture treatment) on naming untrained pictures, aphasia severity and changes in two types of discourse, picture descriptions and responses to open-ended questions.

Restitutive versus Substitutive Treatments

While some treatments focus on reestablishing or strengthening the mechanisms associated with poor word retrieval to support the impaired word retrieval system (i.e., restitutive or rehabilitative treatments), others center on recruiting alternative systems left undamaged by the stroke (i.e., substitutive approaches) (Maher & Raymer, 2004; Nickels, 2002). Restitutive treatments are designed to reactivate and strengthen damaged language mechanisms or to entirely reteach language functions, often employing lexically-based activities directed toward the phonological (sound) and/or semantic (meaning) representations of words (Maher & Raymer, 2004). Phonological restitutive treatments emphasize rebuilding the phonological system using tasks focusing on number of syllables and/or phonemes, phoneme segmentation, initial phoneme cueing, rhyming judgments, or repetition. For instance, Kendall and colleagues (2008) reported a treatment based on the Lindamood Phoneme Sequencing Program (LiPS) (Lindamood, Bell, & Lindamood, 1992), which focused on explicitly connecting the sounds used in language to how they were articulated and written. Thus, by focusing on rebuilding lost connections within the phonological system, this treatment led to improvements in confrontation naming, nonword repetition and phonological production of the trained phonemes, as well as discourse changes.

Rather than centering therapy on phonology, semantic restitutive treatments emphasize strengthening connections within the semantic network and the semantic representations themselves. One commonly used aphasia treatment, Semantic Feature Analysis (Boyle, 2004; Coelho, McHugh, & Boyle, 2000), entails that patients be presented a picture which they attempt to name. Subsequently, they are guided through a chart presenting questions regarding the target concept's group membership, function, action, properties, location and associations to other concepts. Boyle and Coelho (1995) reported that Semantic Feature Analysis training in a participant with Broca's aphasia improved confrontation naming of both trained and untrained items and communicative effectiveness. Coelho et al. replicated these findings in another individual with a moderate fluent aphasia. In another study using

¹In previous publications (Crosson et al., 2005, 2007, 2009; Richards 2002), the Gesture treatment has been referred to as the "Intention treatment." We have chosen the former in this publication, because it is more descriptive of the added component.

four patients with aphasia, Kiran and Thompson (2003) employed a semantic feature training paradigm to train items that were either typical or atypical exemplars within a category. They reported that patients trained on typical items showed no generalization to untrained items, but patients trained using atypical items demonstrate improved naming of untrained items that are typical. Given that there have also been some reports of generalization to other tasks, discussed below, studies such as these demonstrate the usefulness of training semantic features during therapy. Other researchers have advocated combined semantic and phonologic treatments during therapy as the most effective treatment approach for many patients (e.g., Drew & Thompson, 1999).

While the above treatments tend to focus on remediation or restitution of phonological and/or semantic representations, others encourage the recruitment of alternative, intact systems which, in turn, support the impaired word retrieval system (Maher & Raymer, 2004; Nickels, 2002). In particular, substitutive treatments that encourage the use of gesture strategies during word retrieval have been examined by several researchers (e.g., Pashek, 1997; Raymer et al., 2006; Rose & Douglas, 2008). One of the first to use this approach, Pashek contrasted combining semantic treatments with either left (non-dominant) or right (dominant) hand pantomime gestures and found equal benefits for lexical access regardless of which hand was used. Rose and Douglas (2001) further noted that the use of iconic pantomime gestures provided greater benefits to an individual with phonological deficits than one with semantic impairments. Raymer et al. also employed a treatment using pantomime gestures paired with verbal responses that improved word retrieval for both nouns and verbs in persons with aphasia and increased use of gestures. Other studies demonstrated the efficacy of implementing combined gesture and semantic treatments for verb retrieval deficits in persons with Broca's aphasia (Rose & Sussmilch, 2008; Boo & Rose, 2011). All of these studies employed "iconic" gestures that pantomimed the action associated with the words being trained. The current study adopted a different approach to remediation by specifically attempting to relocate lexical access processes to the right hemisphere using unrelated intentional gestures made with the non-dominant left hand.

Generalization Following Treatment

While many anomia treatments result in long-term improvements in naming trained items, documented reports of generalized improvement extending beyond trained material or contexts have been variable within the literature (Nadeau, Rothi, & Rosenbek, 2008). A number of studies have reported generalization to naming of untrained items following anomia treatment (e.g., Crosson et al., 2007; Kendall et al., 2008; Kiran & Thompson, 2003; Nadeau et al., 2008). Although generalization to untrained items is an important metric of treatment success, the ultimate goal for anomia treatment is not improved picture naming, per se, but rather improved overall communication (Wright & Newhoff, 2005). Therefore, a more stringent metric of treatment effectiveness assesses the extent to which treatment generalizes to tasks beyond those targeted in therapy, such as discourse production, an analog of real-life communication. Consequently, several researchers have included assessments of discourse production in treatment studies as an important aspect of assessing generalization of treatment effects (Ballard & Thompson, 1999; Boles, 1998; Boyle, 2004; Conroy, Sage, & Ralph, 2009; del Toro et al., 2008; Edmonds, Nadeau, & Kiran, 2009;

Goral & Kempler, 2009; Hickin, Best, Herbert, Howard, & Osborne, 2001; Hong & Law, 2009; Kendall et al., 2008; Raymer et al., 2007; Wambaugh & Ferguson, 2007). Some researchers have lamented that few anomia treatments lead to overall generalized improvement in communication skills (e.g., Boyle, 2004; Conroy et al., 2009; Nickels, 2002; Raymer, et al., 2007). However, there are many exceptions to this statement (e.g., Edmonds et al., 2009; Kendall et al., 2008; Kiran & Thompson, 2003; Nadeau et al., 2008; Raymer & Ellsworth, 2002, Conroy et al., 2009). Gesture treatments, such as those used by Rose and Sussmilch (2008) and Boo and Rose (2011), discussed above, reported generalized improvement in the use of trained items during picture description and conversation tasks following a treatment for verb retrieval deficits. Similarly, a treatment focusing on verb semantics and the activation of various noun arguments associated with particular verbs elicited improvements in the number of utterances produced and the proportion of utterances with appropriate actions, agents and themes, directly reflecting the treatment goals (V-Nest, Edmonds et al). Generalized improvements in discourse, such as increases in lexical retrieval (e.g., Antonucci, 2009) and informativeness within discourse (e.g., Peach & Reuter, 2009; Wambaugh & Ferguson, 2007), were also frequently reported following Semantic Feature Analysis training (e.g., Antonucci, 2009; Boyle, 2004; Peach & Reuter, 2009; Coelho, McHugh, & Boyle, 2000; Wambaugh & Ferguson, 2007). One of the primary questions of the current research was whether an intentional left-hand gesture would lead to significant group differences in generalization to discourse production.

Right Hemisphere and Recovery of Language

In most people, damage to the left hemisphere results in the disruption of language functions known as aphasia. Research suggests that the right hemisphere contributes to language recovery in persons with aphasia. For example, researchers found that persons with aphasia engaged right hemisphere homologues of left hemisphere regions more often during speech production than healthy adults (Blank, Bird, Turkheimer, & Wise, 2003). Furthermore, compared to people with no brain damage, persons with aphasia who experienced substantial recovery of language abilities showed increased activation in the right hemisphere during silent picture naming and verb generation tasks without showing a significant decrease in left hemisphere activation (Cao, Vikingstad, George, Johnson, & Welch, 1999). Crosson and colleagues (2007) suggested that, in cases involving left hemisphere damage and persistent aphasia, the left hemisphere alone might not be sufficient to support language recovery. They proposed that, in these cases, optimal therapy outcomes might require the recruitment of language production mechanisms in the right hemisphere.

Based on the literature suggesting the advantages of recruitment of right hemisphere mechanisms during language recovery in aphasia, Crosson and colleagues (Crosson et al., 2007; Crosson et al., 2005; Crosson et al., 2009) devised an intention-based treatment technique that engaged the right hemisphere by shifting intention and language production mechanisms to homologous right-hemisphere regions. Crosson et al. (2007) defined intention as the ability to select and initiate an action from many possible competing actions. Because the intentional circuits for volitional hand movement overlap those for word generation in the pre-supplementary motor area (pre-SMA), Crosson et al. (2007) reasoned that preceding a naming attempt with a volitional hand movement could facilitate picture

naming. That is to say, the use of a complex left hand movement might serve to support word generation via priming activation of the right hemisphere pre-SMA. This in turn would potentially engage those frontal regions of the right hemisphere that support language functioning. In a phase 1 study, Crosson et al., (2007) demonstrated the advantage of using an intention treatment (i.e., the left hand movement) compared to an attention treatment which involved the presentation of pictures to patients in left hemi-space in an attempt to recruit right hemisphere attention mechanisms for word retrieval. Results suggested that, while there were no differential effects between treatments for persons with profound anomia, persons with moderate to severe anomia demonstrated greater cumulative improvements from the intention treatment. In an fMRI study of five patients from the 2007 article, Crosson and colleagues (2009) compared pre- and post-intention treatment cortical activity during word generation to that of controls. They reported that the four participants who improved following the intention treatment demonstrated a significant shift in lateral frontal activity to the right hemisphere. Lateralization of frontal cortical activity in these four participants did not differ from normal controls before the intention treatment but, after treatment, was significantly more right-lateralized than the cortical activity of controls. This research suggested that treatments incorporating an intentional gesture encouraged recruitment of right hemisphere mechanisms for language production. However, neither study (Crosson et al., 2007, 2009) isolated the effect associated with the left hand gesture, nor did they include outcome measures beyond naming and category generation. The current study addressed these concerns by contrasting the effects of two treatments for aphasia that differed only in the inclusion of the intentional gesture and also included discourse measures as outcome variables.

Outcome Measures

In the current study, a number of outcome measures were used to assess the generalization of treatment effects, including naming of pictures from the Boston Naming Test (BNT; Kaplan, Goodglass & Weintraub, 2001), the Aphasia Quotient from the Western Aphasia Battery (WAB-AQ; Kertesz et al. 1982), as well as discourse production. We assessed discourse with a number of different measures. Given that nouns were the focus of treatment, the number of nouns produced was included in the analyses as a quantitative measure of change. Considering that improved noun access may, in turn, lead to changes in verb activation as a result of priming (e.g., Edmonds et al., 2009; McRae, Hare, Elman, & Ferretti, 2005), verbs were also included as an outcome measure, as were total number of words and utterances since improving lexical access could also increase the quantity of output. Two methods for assessing information content within the aphasia literature were included as measures of qualitative improvements in discourse: Correct Information Units (CIUs; Nicholas & Brookshire, 1993; see also Doyle, 1995), focusing on word level information, and Utterances with New Information (UNIs; Del Toro et al., 2008), focusing on utterance level information. Finally, propositional analysis is the standard method of assessing information content in healthy adult narratives; thus, we included this analysis measure as an experimental qualitative measure of information content to determine if it was feasible to use in aphasic discourse. Grammaticality was included as the final qualitative measure, because grammaticality is an important component of normal language production

that relies on access of both lexical and grammatical information and can be very demanding of cognitive resources (e.g., Troche & Altmann, 2011).

The current study

In contrast to previous studies from the Crosson group (2007, 2009), the current study employed two treatments that were identical except for the addition of an intentional left hand gesture in the Gesture group. The treatment trained both picture naming and category generation. Generalization measures included scores on the BNT and the WAB-AQ assessed at three time points for evidence of generalization. Additionally, the current study contrasted the effects of the Gesture and the No Gesture treatments on discourse production during two tasks, describing Norman Rockwell pictures and answering open-ended questions. These tasks were chosen as discourse analogs of the primary treatment targets, picture naming and category generation, assessing language production from visual stimuli as well as endogenously generated language, respectively. The current study addressed the following aims:

1. Determine whether the treatment protocol led to acquisition of the target items and whether the addition of the left-hand gesture improved acquisition of target items. We predicted that participants would be able to name a greater number of the items in the treatment set post-intervention, and that changes would be maintained 3 months later. Additionally, we predicted that the magnitude of change would be greater for those using a left hand gesture as part of their training.
2. Determine whether the treatment protocol led to generalized improvement in the naming of untrained items and in overall aphasia severity, as well as whether the addition of the left-hand gesture increased the magnitude of change in these measures. We predicted that the treatment would generalize to untrained pictures and would lead to improved performance on a test of general aphasia severity post-intervention and that these changes would be maintained 3 months later. Additionally, we predicted that magnitude of change would be greater for those using a left hand gesture as part of their training.
3. Determine whether the treatment protocol affected discourse production elicited in different ways, as well as whether the addition of the left-hand gesture increased the magnitude of change in discourse measures. We predicted that the treatment would improve aspects of discourse specifically related to lexical access, including number of words produced, numbers of nouns and verbs, and informativeness of the discourse, and that these effects would be maintained 3 months later. Furthermore, we predicted that the Gesture group would achieve greater gains in performance, on measures of lexical access and informativeness.

Method

Participants

Nineteen participants were recruited for the study; however, five did not pass all pre-treatment screening procedures, leaving 14 participants for the current analysis. Participant

demographics are shown in Table 1. Participant S19 was African-American, and the other participants were Caucasian. Eleven participants in the current study had had ischemic strokes, and 3 participants (21%) had had hemorrhagic strokes. All participants were at least 6 months post stroke and had single or multiple lesions limited to the left hemisphere that included the precentral gyrus or underlying white matter which was confirmed by medical records and MRI. All participants signed consent forms approved by the University of Florida Health Science Center Institutional Review Board.

Inclusion criteria for study participants were as follows: 1) a standard score greater than 69.92 on the Peabody Picture Vocabulary Test-IV (PPVT-IV, Dunn & Dunn, 2007) ; 2) an ability to consistently follow one step commands; 3) scores on the WAB-AQ (Kertesz, 1982) below the aphasia cutoff of 93.8; 4) right-handed prior to the stroke, as determined by the Edinburgh Handedness Inventory (Oldfield, 1971); and 5) English as their first language. Potential participants were excluded from the study if they were not eligible for participating in MRI or if they had a history of head trauma, neurological disorder other than stroke (e.g., Alzheimer's disease), learning disability (e.g., dyslexia), psychiatric disorder (e.g., schizophrenia), drug or alcohol abuse, or chronic medical conditions likely to impair cognition (e.g., renal or hepatic failure).

Pre-treatment speech samples using the Cookie Theft Picture from the Boston Diagnostic Aphasia Examination—Third Edition (Goodglass, Kaplan, & Barresi, 2000) and sentence repetitions (Duffy, 1995; Wertz, LaPointe, & Rosenbek, 1984) were assessed by the expert speech/language pathologist serving as co-investigator on the treatment study (the fifth author) using perceptual criteria for apraxia of speech developed by McNeil, Robin and Schmidt (2009) and for aphasia by Goodglass (1981). Based on WAB criteria, three participants had Broca's aphasia, one had transcortical motor aphasia, five had conduction aphasia and five had anomic aphasia (see Table 1).

Procedure

Materials—Prior to treatment, all participants named a set of over 400 black and white line drawings of objects and generated members of 120 categories twice. Items missed consistently were selected for treatment, beginning with the highest frequency items and progressing to lower frequency items until enough items were identified to construct the counterbalanced treatment lists and probe stimuli. Specifically, from the set of 400 pictures and 120 categories, 120 pictures and 60 categories were individually selected for each participant, with the selected picture and category sets each containing 25% consistently correct and 75% consistently incorrect at pre-treatment testing. Of these, 20 pictures and 20 categories were never trained and incorporated into sets of probe items (see below). 50 pictures were trained in phase 1, 50 different pictures were trained in phase 2, and 40 categories were trained in phase 3. Picture probes consisted of 20 items from phase 1, 20 items from phase 2, and 29 never-trained items. Category probes consisted of 20 items from phase 3 and 20 never-trained items. Participants underwent baseline testing using those probe items for a minimum of 8 sessions or until baseline treatment stability was obtained using Tryon's C-statistic (Tryon, 1982).

Participants were assigned to either the No Gesture or Gesture treatment group after baseline testing using stratified random sampling to equalize groups on picture naming ability using BNT scores. Groups were also matched on the number of subjects whose lesions extended anteriorly beyond the precentral sulcus. As shown in Table 1, there were no significant differences between groups for age, education, WAB-AQ or BNT scores. However, the education level showed marginal differences between groups, $t(12) = 1.917, p = .08$; the Gesture group had two years more education than the No Gesture group. Additionally, there were no significant between-group differences in the components of the WAB: WAB speech production $t(12) = 0$; WAB comprehension $t(12) = .680, p > .51$; WAB naming $t(12) = .898, p = .39$; however, WAB repetition showed marginal differences between the groups, $t(12) = 2.039, p = .064$, in favor of the No Gesture group.

Treatment—Treatment was delivered in 3 phases (10 sessions per phase), with two 1-hour treatment sessions per day, 5 days a week for a total of 30 treatment sessions. The two sessions each day were at least a half hour apart. Phase 1 consisted of treatment sessions 1–10 and focused on naming 50 pictures. Phase 2 consisted of treatment sessions 11–20 and trained naming of 50 different pictures. Within-subject, the items for the two naming phases were equivalent in word frequency. Phase 3, treatment sessions 21–30, required the generation of an exemplar of each of 40 different categories.

Naming trials in Phases 1 and 2 consisted of the presentation of a picture on a computer monitor for naming. In Phase 3, trials consisted of auditory and orthographic presentations of a category name for which the participant generated one category member. For all trials in all phases, a therapist verified response accuracy. If treatment trials were completed correctly (i.e., a picture was named correctly or a correct category exemplar was generated), participants began the next trial. If an item was not named correctly, the therapist would provide the correct name, and participants would then practice saying the correct response. Similarly, if a participant was unable to generate a member of a category, the therapist would provide an example, and the participant would practice saying this correct response. This correction procedure was repeated up to three times maximum or until the item was named correctly by the participant. The number of times a participant repeated the correct response was not regulated.

There were only two differences in the treatment protocol for the Gesture and No Gesture groups. First, the Gesture group initiated each treatment trial with their left hand by opening and reaching into a box and pushing a red button. Second, during each correction procedure the Gesture group also made a non-meaningful circular gesture with their left hand. In contrast, for the No Gesture group, a therapist pushed a button to initiate each treatment trial and no hand movement was required during the correction procedure. The therapist ensured that these protocols were followed. The same therapists administered both the Gesture and the No Gesture treatments. To ensure protocol compliance and treatment equivalency across sites, phases and subjects, a research assistant trained in both treatments who was not administering treatment at any of the sites evaluated one session per treatment phase (i.e., once a week) per participant for correct delivery of the assigned treatment and subsequent correction procedures.

From the end of treatment until the 3-month follow-up assessment, participants did not engage in any additional speech therapy, experimental treatments, or additional conversational groups. However, no attempt was made to control or record in what other activities they may have participated.

Discourse measures

Before intervention (pre-test), immediately after the intervention (post-test) and 3 months following the intervention (3-month follow-up), discourse samples were elicited in response to two pictures and two open-ended questions to sample language that was both exogenously-generated (in response to external stimuli, the pictures) and endogenously-generated (in response to one's own ideas cued by the questions). The pictures were reprints of paintings by Norman Rockwell that have been frequently used in the literature (e.g., *Easter Morning* and *Traffic Problems*) or that showed individuals exhibiting strong emotions (e.g., *In the Principal's Office*, *The Dugout*) (Olness, Ulatowska, Wertz, Thompson, & Auther, 2002; Olness, 2006; Coelho, Youse, Le & Feinn, 2003; Coelho, Grela, Corso, Gamble, & Feinn, 2005). Picture set A included *Easter Sunday* and *The Dugout*; picture set B included *Traffic Problems* and *In the Principal's Office*. Participants were asked, "Look at this picture carefully. Make up a story with a beginning, middle, and an end that tells me what happened here (Olness, 2006)." When responses faltered, participants were prompted with, "What do you think will happen next?" or "Why do you think s/he has that look on her/his face?" Open-ended questions were taken from published studies of discourse in older adults (Kemper, Herman, & Lian, 2003; Plummer-D'Amato et al., 2008), and included two personal questions ("Please tell me about a person who had a positive influence on your life" and "Please tell me about an event that had a positive influence on your life") and two general questions, ("Please tell me what you think was the most important invention of the last 100 years" and "Please tell me what you think was the most important event of the last 100 years"). Questions were paired so there was one personal and one general question asked at each session. Question set A included ("Please tell me about a person who had a positive influence on your life" and "Please tell me what you think was the most important event of the last 100 years." Question set B included the questions, "Please tell me about an event that had a positive influence on your life" and "Please tell me what you think was the most important invention of the last 100 years." If prompts were needed, the examiner asked "How did this affect you?" "How did this affect the world/the way we do things?" or "Can you tell me more about this/him/her?" Each discourse task ended with the question, "Do you have anything more you want to say about this/him/her?"

Unfortunately, test-retest reliability data on the measures we planned to use have not been published for discourse elicited by these stimuli. Our lab does have unpublished data on the particular questions used. Twenty-two healthy older adults produced answers to these questions four times, the first two within two weeks of each other and then twice more four months later. There were no significant differences in the number of utterances, words, verbs, grammatical sentences or utterances with new information produced between questions or over time, and measures also correlated significantly across repeated administrations, even in dual task situations (Altmann, Roberts, Jessup, Thomas &

Marsiske, 2008; Plummer-D'Amato, Altmann, & Reilly, 2011). Nonetheless, there is no data available on descriptions of the pictures used here or on other discourse variables, such as propositions and CIUs. To control for this potential confound and minimize differences between individual questions or pictures, eight lists of stimuli were prepared. In these eight lists, the two sets of pictures and questions described above were counterbalanced for order between pre- and post-treatment assessments and for the order of administration of pictures and questions (i.e., pictures first or questions first). Thus, all participants responded to all eight stimuli by the end of the post-intervention assessment, but whether a stimulus set was presented pre- or post-intervention varied across participants. No list was assigned to more than two participants. Finally, to control for the recency of presentation, the pictures and questions presented at follow-up were the same for all participants and consisted of one picture and one question from each of the previous assessments. Specifically, at the follow-up assessment, all participants described the *Easter Morning* and *Traffic Problems* pictures and answered the questions, "What was the most important invention of the last 100 years" and "What was the most important event of the last 100 years." By ensuring that no individual question or picture stimulus dominated pre- or post-testing, and standardizing stimuli at the follow-up assessment, this approach minimized potential, systematic confounding effects that might compromise the findings of the study.

All discourse responses were digitally recorded and transcribed by trained research assistants and then checked for accuracy by a second trained research assistant. Because discourse is not a unitary construct and has many different dimensions, several measures of discourse quantity and quality were scored, as described below.

Measures of Discourse Quantity—Responses were divided into utterances based on syntactic boundaries and prosodic contours (Doyle, Goda, & Spencer, 1995). Word count was obtained by following the procedures detailed in Nicholas and Brookshire (1993). Briefly, statements made before (e.g., Do you want me to start?) and after the actual description of the picture or response to the open-ended question (e.g., That's all) were removed from the transcript. Non-word fillers (e.g., um, ugh) as well as neologisms, unintelligible words and stutters were also removed. Extraneous words such as interjections and commentary made during the task (e.g., I don't know, this is hard) and lexical fillers (e.g., oh boy, you know) occurring in the middle of the discourse were included in the total word count (For more detail, please see the Appendix to Nicholas and Brookshire, 1993). Word and utterance counts were then obtained using the Systematic Analysis of Language Transcripts (SALT; Chapman & Miller, 1984).

Before scoring nouns and verbs, all comments on the task (e.g., this is hard), filler statements (e.g., you know), jargon, false starts, automatic speech and repeated words were excluded or "mazed" from the transcript using SALT. The third author then hand-coded all nouns and main verbs in the remaining discourse, even if it was difficult to discern how a word fit the context. The first author recoded nouns and verbs in thirty percent of these transcripts to assess inter-rater reliability, which was good, (both Cronbach's $\alpha = .90$, $p < .0001$).

Measures of Discourse Quality—Four measures were used to assess quality of discourse: the number of Correct Information Units (CIUs; Nicholas & Brookshire, 1993), Utterances with New Information (UNIs; del Toro, et al., 2008), propositions, and grammatical utterances. To identify CIUs, we closely followed the instructions provided in Nicholas and Brookshire (1993). Nicholas and Brookshire defined CIUs as all content and function words that were “accurate, informative, and relevant to the eliciting stimulus.” Thus, irrelevant words, false starts, restarts and words that provided incorrect information were excluded from the CIU count, but the rest of the sentence was included. CIUs were originally identified by the first author, and the second author rescored thirty percent of the transcripts. Reliability on CIU identification was good, (Cronbach’s $\alpha = .91$, $p < .0001$).

Utterances with New Information (UNIs) were coded and counted in SALT, using the criteria described in del Toro et al. (2008) and Doyle et al. (1995). To qualify as a UNI, an utterance had to be coherent and relevant, while adding novel information to the discourse and, crucially, the reader must have been able to identify what that new information was. Information did not need to be in the form of a grammatical sentence, because single word utterances can be UNIs, especially in conversational discourse, if they, for example, answer a question. Therefore, in response to the question, “What do you think was the most important invention of the last 100 years?” a response of, “The computer” would count as a UNI. The third author coded transcripts for UNIs and thirty percent were rescored by the first author for reliability, which was good (Cronbach’s $\alpha = .98$, $p < .0001$).

Proposition counts are standard measures in the adult literature and have been used in many studies employing open-ended questions such as those used here (e.g., Kemper et al., 2003). The number of propositions was obtained using CPIDR version 3.2 (available for free download at <http://www.ai.uga.edu/caspr/>; Covington, 2007). As with identifying CIUs, all utterances with interpretable meaning were isolated and extracted to a separate file called the Proposition Set. Within the Proposition Set, any transparent phonemic paraphasias and colloquialisms, such as “em” for them, were corrected so that they could be counted by the program. The Proposition Set was then entered into CPIDR using Speech Mode, which automatically filters word repetitions and rejects fillers such as, “you know,” to obtain a proposition count. Propositions were also hand-coded by the second author and a trained research assistant using the rules in Turner and Greene (1977). Reliability was excellent between hand-coded and CPIDR proposition counts at all three time points (for all, Cronbach’s $\alpha = .98$, $p < .0001$). This is consistent with the findings of Brown et al. (2008), who reported high reliability between experienced human propositional coding and CPIDR coding. Although Brown et al. also reported that CPIDR was consistently about 5% higher than trained, human coders, it was absolutely consistent in its coding, so that the same sentence is always scored the same way, thus eliminating a source of variability from analysis. The counts from CPIDR were employed in the analyses below.

To determine grammatical utterances, each utterance was scrutinized to determine whether, when mazed portions of the utterance were removed, the remaining words of the utterance constituted a grammatical sentence. Grammaticality was strictly defined, requiring subject-verb agreement and appropriate articles, for example. Reliability on grammatical sentence identification was good (Cronbach’s $\alpha = .97$, $p < .0001$).

Statistical analyses—Discourse samples elicited by pictures and open-ended questions were analyzed separately. Analyses of discourse variables were conducted using $2 \times 3 \times 2$ repeated measures ANOVAs. Task (i.e., picture narratives and questions) and Time (i.e., pre, post, 3-month) were within-subject variables, and Group (i.e., Gesture, No Gesture) was a between-subjects variable. In these analyses, the interactions of primary interest were interactions with time. A task by time interaction would indicate that the treatment protocol had differential effects on the two types of discourse over time, but were similar for both groups. A time by group interaction would indicate that the performance of the two treatment groups changed differently over time, while a three-way interaction would suggest that performance of the two groups differed over time across tasks. In the following, all analyses were evaluated for significance at an alpha level of .05 (two-tailed) due to the extremely small numbers of participants. All statistics were calculated using SPSS (version 18.0).

RESULTS

Treatment Effects

To assess treatment effects for picture naming, participants completed picture naming probes before the intervention, immediately post-intervention and at a 3-month follow-up test session. Group mean scores are presented in Table 2. A (3) Time (pre, post, 3 month) by (2) Group (Gesture, No Gesture) repeated measures ANOVA found no main effects or interactions associated with Group; however, they revealed a significant main effect of time, $F(2,24) = 45.786$, $p = .001$, $\eta^2 = .79$. Planned, paired comparisons revealed that participants performed worst at pre-test ($M = 41.08$, $SD = 12.46$), improved significantly at post-test ($M = 64.40$, $SD = 19.62$; $p = .001$), and dropped significantly at the 3-month follow-up ($M = 54.04$, $SD = 12.33$, $p = .02$). Even with the drop in performance at 3 month follow-up, however, scores remained significantly higher than at pre-test ($p = .001$). There were no other significant effects.

In the analyses of category member generation probes, again there were no significant effects of group. As illustrated in Table 2, only the main effect of time was significant, $F(2,24) = 14.044$, $p = .001$, $\eta^2 = .54$. Paired comparisons revealed that participants performed worst at pre-test ($M = 54.93$, $SD = 20.54$), improved significantly at post-test ($M = 71.96$, $SD = 23.48$; $p = .001$), and dropped at the 3-month follow-up ($M = 62.86$, $SD = 26.06$). Scores at follow-up testing decreased and did not differ significantly from pre- ($p = .106$) or post-intervention ($p = .095$) scores. These results are consistent with our prediction for Aim 1 that the treatment would have significant and lasting effects on naming pictures in the training set, but the effects were more limited for the category generation task. The predictions of greater gains in the Gesture group were not supported.

Generalization Effects

To assess response generalization (i.e., generalization to untrained items), participants completed the BNT at pre-test, post-test, and 3-month follow-up, as shown in Table 2. A (3) Time by (2) Group repeated measures ANOVA revealed a significant effect of time, $F(2,24) = 4.982$, $p = .015$, $\eta^2 = .29$. Performance improved from pre-test ($M = 27.79$, $SD = 10.56$) to

post-test ($M = 31.21$, $SD = 13.00$), and this level of performance was maintained 3 months later ($M = 30.57$, $SD = 13.35$). There were no significant effects of Group or interactions. This supports our prediction of generalized improvement in naming following training, but not the prediction of greater improvements in the Gesture group.

To assess stimulus generalization, or generalization to an untrained task, participants completed the WAB at each time point; group mean scores are shown in Table 2. A (3) Time by (2) Group repeated measures ANOVA with WAB-AQ scores as the dependent variable found no significant effects of either group or time. Thus, while the treatment affected naming and category generation, and generalized to naming untrained pictures, it did not have pervasive effects on other aspects of language assessed by the WAB, such as repetition and comprehension. Therefore, our prediction for Aim 2 was only partially supported, as the only generalization found was to picture naming, but not overall aphasia severity.

Discourse Quantity

Scores on discourse quantity measures, the number of utterances, words, verbs, and nouns produced, are presented in Table 3. Statistical values for significant interactions are presented in Table 4. The analysis of the number of utterances revealed no three-way interaction, but there was a significant interaction between task and time. This interaction, depicted in Figure 1a, was explored using paired sample t-tests comparing the two tests at each time point. There were no differences in the number of utterances produced with pictures versus questions at either pre- or post-testing (both $p > .4$). However, participants produced significantly more utterances when describing pictures than when answering questions at follow-up testing, $t(13) = 3.283$, $p = .006$.

The analysis of the total number of words produced yielded three two-way interactions, depicted in Figure 1b, but not a significant three-way interaction. First, the task by time interaction was significant. The number of words produced was relatively equivalent across tasks at pre-test and post-test (both $p > .7$). However, at follow-up, word production increased in the picture task and decreased in the question task in both groups leading to a significant difference between tasks at follow-up testing, $t(13) = 2.741$, $p = .017$. Secondly, group interacted with task. The No Gesture group produced similar numbers of words in the picture ($M = 179.52$, $SD = 133.00$) and the question task ($M = 186.24$, $SD = 124.32$), $t(6) = -.396$, $p > .7$. However, the Gesture group produced more words in the picture task ($M = 288.95$, $SD = 111.48$) than in the question task ($M = 213.10$, $SD = 89.92$), $t(6) = 2.189$, $p = .07$, although this only approached significance. Finally, the interaction between group and time was also significant. The two groups produced nearly identical numbers of words at pre-test and diverged over time, with the No Gesture group decreasing word output and the Gesture group gradually increasing output, leading to a significant difference between groups at 3-month follow-up, $t(12) = 1.83$, $p = .05$.

As shown in Figure 1c, verb production showed a similar pattern. While the three-way interaction was not significant, task interacted with time. The number of verbs produced in the two tasks did not differ at pre- or post-test (both $p > .4$); however, overall participants produced more verbs at follow-up in the picture description task than in the question task,

$t(13) = 3.126, p = .008$. Additionally, group also interacted with task. The No-Gesture group produced similar numbers of verbs in both tasks ($p > .6$), but the Gesture group produced more verbs in the picture description task than the question task, $t(6) = 3.876, p = .008$.

The patterns of noun production were similar to those for verbs. Similar numbers of nouns were produced in the two tasks at pre- and post-intervention, with a divergence at the 3-month follow-up. Additionally, the Gesture group tended to produce more nouns during picture description than in questions or than the No Gesture group produced in either task. However, the interaction of task and time was not significant ($p = .20$), and the group by task interaction only approached significance ($p = .064$).

Discourse Quality

The measures of discourse quality assessed included grammatical sentences, UNIs, CIUs and propositions; group means for these are provided in Table 5. Statistical information for significant interactions is presented in Table 4. Regarding grammatical sentences, the three-way interaction was not significant, although the task by time interaction was, as shown in Figure 1d. As in previous analyses, the number of grammatical sentences produced did not differ at pre-and post-test (both $p > .3$). In contrast, at follow-up participants produced more grammatical sentences during picture descriptions than during question responses, $t(13) = 2.328, p = .037$. There were no other significant effects in the analysis of the number of grammatical sentences.

The ANOVA investigating CIU production revealed a significant task by group interaction, shown in Figure 1e. The Gesture group produced more CIUs in picture description ($M = 81.00, SD = 59.97$) than during question responses, ($M = 47.05, SD = 30.45$), $t(6) = 2.330, p = .059$, although the difference only approached significance. In contrast, the No-Gesture group produced similar numbers of CIUs in picture descriptions ($M = 48.00, SD = 54.27$), $p > .30$, and in questions ($M = 54.67, SD = 69.45$), $t(6) = -1.081, p > .32$. Additionally, the time by group interaction approached significance, $p = .06$, echoing the same effect in the analysis of the number of words. As shown in Table 5, the Gesture group increased mean production of CIUs across tasks at each successive assessment, while the No Gesture group produced, on average, fewer CIUs at each assessment. Examination of the means suggested that the Gesture group's increased CIU production was limited to the picture task; however, the three-way interaction was not significant. There were no other significant effects in this analysis.

The ANOVA investigating the number of propositions produced revealed only a significant task by group interaction. The No Gesture group, in general, produced more propositions in the question task ($M = 37.52, SD = 37.56$) than when describing pictures ($M = 30.81, SD = 31.45$), $t(6) = 2.514, p = .046$. In contrast, the number of propositions produced by Gesture group when describing pictures ($M = 49.43, SD = 40.39$) did not differ significantly from the number produced answering questions ($M = 36.19, SD = 26.78$), $p = .13$. There were no significant effects in the analysis of UNIs.

Summary of Discourse Findings

Taken individually, these results seem very distinct and disconnected. As shown in Figures 1, however, the changes in the various measures showed similar patterns. Although none of the expected three-way interactions were significant, there were many two-way interactions affecting performance. First, the Task by Time interactions reflected the fact that, in general, performance in the picture description task tended to improve in both groups at follow-up, while performance in the question task tended to decline. However, these effects were modulated by two group interactions. The Group by Task interactions arose because participants in the Gesture group performed better at all time points on the picture description task than the question task, while the No Gesture group showed few differences in performance between tasks averaged over assessments, except in propositions. Additionally, the Group by Time interactions in words and CIUs captured the fact that changes in performance of the two groups were limited to the 3-month follow-up assessment. From post-test to follow-up, the word production of the Gesture group improved and the No Gesture group declined leading to group differences at follow-up that were significant for words and approaching significance for CIUs. Consequently, many of the reported interactions were driven primarily by improvements in the performance of the Gesture group in the picture description task at the 3-month follow-up, coupled with the poorer performance of both groups in the question task at 3-month follow-up. Despite this pattern, it is important to reiterate that none of the three-way interactions were significant.

Discussion

This study documents changes in picture naming, category generation and discourse following an intensive three week protocol (two 1-hour sessions per day, 5 days a week, for 3 weeks) which trained naming of 100 pictures and word generation in response to 40 category prompts. There were significant improvements in naming trained pictures as well as generalization to untrained pictures on the BNT, and these treatment gains persisted three months following the end of treatment. Category generation also improved significantly following treatment, although at the three month follow-up, improvements were no longer significant. Overall aphasia severity as measured by the WAB-AQ was unaffected by the treatments. Importantly, there were no differences in performance on any of these measures that could be attributed to the addition of the intentional gesture to the treatment. Thus, all of these effects appear to be due to the treatment protocol itself and the intensive training of noun production via picture naming and category generation, rather than to the intentional gesture.

One unexpected finding, considering how well the groups were matched, was that the Gesture group produced more words, CIUs, verbs, and nouns during picture descriptions than during questions, even at the pre-intervention assessment. The No Gesture group produced similar numbers of these measures across tasks, resulting in task by group interactions. A potentially contrasting pattern was shown by the group by task interaction in propositions, where the significant difference stemmed from the No Gesture group performance across tasks. The No Gesture group produced significantly more propositions in questions than picture descriptions; whereas the Gesture group showed a trend to produce

more propositions during picture descriptions. The etiology of these differences is unclear, since the groups were well-matched on naming and WAB scores, age, and education. It is possible that aphasia type played a role. The Gesture group contained a majority of participants with conduction aphasia, while the No Gesture group contained a majority of participants with anomic aphasia. This is a question for future research.

Elicitation-specific improvements

The common treatment shared by both groups, as reflected in task by time interactions, led to improvements in lexical access that were strongly influenced by the manner of language elicitation. The quantity and quality of the responses to pictures versus open-ended questions diverged over time, affecting four of eight measures of discourse quantity and quality. Specifically, mean quantities of utterances, words, verbs, and grammatical sentences increased from post-test to 3-month follow-up when describing pictures and simultaneously declined in the open-ended question task for both groups, leading to a significant difference in the number of each produced at the 3-month follow-up. We propose that these improvements in performance in the picture task may be attributable directly to the nature of the training regimen that strongly focused on lexical access from visual stimuli. Participants engaged in 20 picture-naming training sessions within a two week period, practicing naming 100 pictures. Indeed, even the subsequent category generation task used visual, orthographic stimuli in addition to auditory stimuli. We speculate that this intense emphasis on lexical access from visual input led to the strengthening of connections between visual representations and phonological word forms. These strengthened connections would not be word-specific but, instead, likely participate in the representations of thousands of words. Consequently, access to many untrained words might improve. Supporting evidence for this line of reasoning comes from the significant improvement in BNT scores post intervention that were maintained nearly unchanged 3 months later.

In contrast to the widespread improvement found in discourse produced in response to pictures, performance in open-ended question responses became noticeably sparser at follow-up testing. The etiology of the decline in production on open-ended questions is difficult to determine. The purpose of the category generation task in the treatment protocol was specifically to bolster lexical access from an idea or concept. However, stimuli in the category generation task were also visual (i.e., the orthographic forms of the words), which may have thwarted this goal somewhat. It is also possible that the abbreviated length of the category generation training, 10 training sessions versus 20 for picture naming, may have limited its effects. Alternatively, the category generation task may have simply been less effective as a treatment method for rebuilding connections between conceptual areas and phonological representations. Another possible explanation is that the word association-based strategy for successful completion of a category generation task may not be optimal for producing conceptually-driven, grammatical speech of the type required by the open-ended questions used here. The nonsignificant declines in production of verbs, CIUs and grammatical sentences in question responses and marginal changes in noun production support this reasoning. Nonetheless, it is reassuring that there were no significant WAB-AQ changes or changes in the amount of propositional speech in question responses, despite declines in other measures.

Paradoxically, the number of nouns produced in discourse did not change significantly despite the intense treatment they received, but verb production did show improvement. This is likely due to low power resulting from the small size of the study, since the significance level of the time by group interaction in noun production approached significance ($p < .07$). This hypothesis is supported by the fact that the observed power for the task by time interaction for nouns was .43, whereas it was greater than .6 for all significant interactions. One factor potentially contributing to the lack of significance in this interaction might have been the intensity of the training protocol, which could result in increased activation throughout the noun semantic system. This wide-spread activation could lead to somewhat diminished distinctiveness of individual lexical representations of nouns due to activation of overlapping semantic representations and spread of activation, making access of nouns more difficult.

In contrast, although verbs may have partially overlapping semantic representations with their typical noun arguments (MacRae et al., 2005), verb representations do not encompass the core semantic features of nominal categories that would link related nouns. For example, the semantic representations of all birds would include features related to feathers, beaks, nests and eggs, but common verbs associated with birds like flying and singing would not share any of these features. Supporting evidence for this comes from the finding that activation of a noun argument can lead to improved access to an associated verb, and vice versa (Bicknell, Elman, Hare, McRae, & Kutas, 2010; Edmonds, et al., 2009; Edmonds & Babb, 2011; Matsuki et al., 2011; Metusalem et al., 2012). Thus, the intensive training that could have led to less distinctiveness in noun representations could still improve verb access via the overlapping noun to verb representations. However, as time passed after the intervention and activation within the language system settled as people continue to converse at home, the improved connectivity underlying improved lexical access could manifest, leading to the changes in verbs and CIUs at the 3-month follow-up, while access to some nouns would continue to be difficult due to loss of semantic distinctiveness.

Effects of the Gesture treatment

The Gesture and No Gesture treatments were associated with different patterns in word and CIU production over time. At pre-intervention, the two groups produced nearly identical amounts of words and CIUs. However, the Gesture group increased production of words and CIUs over time, while the No Gesture group decreased production of words and CIUs, leading to significant or nearly significant differences, respectively, at the 3-month follow-up. Thus, the addition of the intentional gesture affected the volume of discourse production, although not significantly until the 3 month follow-up assessment. Nevertheless, it is important to remember that the changes in words and CIUs from pre-intervention to 3-month follow-up were not significant in either group.

While within-group changes were not significant for either group, the different patterns of change shown by the groups in Figure 1 are striking. Accounting for these different patterns of change is not easy. One possibility is that the addition of the intentional left hand gesture was successful in redirecting lexical access procedures to the right hemisphere (e.g., Crosson et al., 2009). Subsequently, the effects of this relocation of function became evident after the

postulated hyperactivation of the semantic system had settled after the intervention ended. However, other explanations are also possible. There was only one discourse sample obtained before the intervention, so some differences in discourse may have existed pre-intervention and been missed due to the absence of multiple assessments. However, there is no evidence for this in the changes in the training probes, BNT or WAB, as noted in Table 2. Another possibility is that the use of repeated stimuli from previous assessments affected responses at the 3-month follow-up, but since all participants responded to the same follow-up stimuli, this would not explain why the repetition affected the two groups differently. For whatever reason, for those participants who included an intentional left-hand gesture with the intensive treatment, word production improved at follow-up testing, particularly in discourse about pictures, relative to that of participants who received the same treatment without the gesture whose word production remained constant in picture description and declined nonsignificantly in questions.

There are several limitations to this study. First, the groups were very small, and variability in severity and diagnosis within groups was high as a result of requiring participants to be eligible and willing to undergo an MRI scan. Future replication studies without MRI components should focus recruitment on more limited ranges of severity and diagnosis. One result of the extremely small group size is that the study was underpowered for identifying three-way interactions. Considering the pattern of results found, this is a serious limitation and should be of high importance in future studies. Additionally, repeated discourse samples at baseline are recommended for future studies. The lack of test-retest reliability data for the particular discourse measures used with these stimuli presents another potential limitation. However, poor test-retest reliability of measures would likely have added to the variability in outcomes, making the identification of significant results more difficult, and would have affected all participants not just one group. Further explorations of the test-retest reliability of these discourse measures with additional discourse tasks are needed to expand the tools available to clinicians and researchers. Another limitation, which has plagued many studies examining discourse, was the large number of independent variables and analyses, which increase the probability of finding spurious results. On the other hand, the results of the study were remarkably consistent across independent variables, a finding that raises our confidence in these results.

In conclusion, the findings presented here support the efficacy of this intensive lexical access training protocol for achieving long term improvements in picture naming of trained items and category member generation, as well as improving access to untrained words. Furthermore, the training protocol was also associated with improvements at follow-up testing in the number of utterances, words, verbs and grammatical sentences produced that were limited to the picture description task. The addition of the intentional gesture to the treatment led to increases in word and CIU production, particularly affecting picture descriptions, while the No Gesture treatment group showed minimal increases word and CIU production in picture descriptions and nonsignificant declines when responding to questions, resulting in significant group differences at follow-up. Importantly, all changes were found 3-months after the end of treatment, thus, emphasizing the importance of follow-up testing. We conclude that the addition of the intentional gesture to this treatment contributed to generalization of improved lexical access to discourse contexts when the

discourse involved concrete, imageable visual stimuli. Improving lexical access in people with aphasia should have positive functional consequences, such as facilitating everyday conversation around the home about topics of immediate relevance. Therefore, the inclusion of intentional left hand gestures as an adjunct to this and other anomia treatments shows promise as a potential means of achieving generalized improvement in language production and, consequently, needs more research.

Acknowledgments

We thank JoEllen Gilbert, Ceil Brooks, Flo Singletary, Zvinka Zlatař and Stacy Harnish for treatment delivery and assessment data collection. We also thank Amanda Garcia, Jillian Green, and James Chastain for transcription assistance, and Ashley Atkinson and Analise Wise for assistance coding manuscripts. We would also like to thank Heather Harris Wright, for insightful comments on previous versions of this manuscript. This study was supported by grants # R01DC007387 from the National Institute of Deafness and Other Communication Disorders and #B6364L from the Department of Veterans Affairs Rehabilitation Research and Development Service to Bruce Crosson, and by grant # R21AG033284 from the National Institute on Aging to Lori Altmann.

References

- Altmann, LJP.; Roberts, B.; Jessup, JV.; Thomas, LA.; Marsiske, M. Effects of Tai Chi training on speech and cognition in older adults; Presented at the 2008 Cognitive Aging Conference; Atlanta, GA. April 2008; 2008.
- Antonucci SM. Use of semantic feature analysis in group aphasia treatment. *Aphasiology*. 2009; 23(7–8):854–866.10.1080/02687030802634405
- Ballard KJ, Thompson CK. Treatment and generalization of complex sentence production in agrammatism. *Journal of Speech, Language, and Hearing Research*. 1999; 42(3):690–707.
- Bicknell K, Elman JL, Hare M, McRae K, Kutas M. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*. 2010; 63(4):489–505. doi: <http://dx.doi.org/10.1016/j.jml.2010.08.004>. [PubMed: 21076629]
- Blank SC, Bird H, Turkheimer F, Wise RJS. Speech production after stroke: The role of the right pars opercularis. *Annals of Neurology*. 2003; 54(3):310–320.10.1002/ana.10656 [PubMed: 12953263]
- Boles L. Conversational discourse analysis as a method for evaluating progress in aphasia: A case report. *Journal of Communication Disorders*. 1998; 31:261–274.10.1016/S0021-9924(98)00005-7 [PubMed: 9621907]
- Boo M, Rose ML. The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology*. 2011; 25(2):154–175.10.1080/02687031003743789
- Boyle M. Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*. 2004; 13:236–249.10.1044/1058-0360(2004/025) [PubMed: 15339233]
- Boyle M, Coelho CA. Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*. 1995; 4(4):94–98.
- Brown C, Snodgrass T, Kemper S, Herman R, Covington M. Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*. 2008; 40(2):540–545.10.3758/brm.40.2.540 [PubMed: 18522065]
- Cao Y, Vikingstad EM, George KP, Johnson AF, Welch KM. Cortical language activation in stroke patients recovering from aphasia with functional MRI. *Stroke*. 1999; 30(11):2331–2340.10.1161/01.STR.30.11.2331 [PubMed: 10548667]
- Chapman, R.; Miller, J. *SALT: Systematic analysis of language transcripts*. Madison, WI: University of Wisconsin; 1984.
- Coelho C, Grela B, Corso M, Gamble A, Feinn R. Microlinguistic deficits in the narrative discourse of adults with traumatic brain injury. *Brain Injury*. 2005; 19(13):1139–1145.10.1080/02699050500110678 [PubMed: 16286327]
- Coelho CA, McHugh RE, Boyle M. Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology*. 2000; 14(2):133–142.10.1080/026870300401513

- Coelho C, Youse K, Le K, Feinn R. Narrative and conversational discourse of adults with closed head injuries and non-brain-injured adults: A discriminant analysis. *Aphasiology*. 2003; 17(5):499–510.10.1080/02687030344000111
- Conroy P, Sage K, Ralph ML. Improved vocabulary production after naming therapy in aphasia: Can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders*. 2009; 44(6):1036–1062.10.1080/13682820802585975 [PubMed: 19294554]
- Covington, MA. CPIDR 3 User Manual. A. I. Center. , translator. Vol. 2007-3. 2007. CASPR Research Report
- Crosson B, Fabrizio KS, Singletary F, Cato MA, Wierenga CE, Parkinson RB, Rothi LJG. Treatment of naming in nonfluent aphasia through manipulation of intention and attention: A phase 1 comparison of two novel treatments. *Journal of the International Neuropsychological Society*. 2007; 13(4):582–594.10.1017/S1355617707070737 [PubMed: 17521480]
- Crosson B, Moore AB, Gopinath K, White KD, Wierenga CE, Gaiefsky ME, Rothi LJG. Role of the Right and Left Hemispheres in Recovery of Function during Treatment of Intention in Aphasia. *Journal of Cognitive Neuroscience*. 2005; 17(3):392–406.10.1162/0898929053279487 [PubMed: 15814000]
- Crosson B, Moore AB, McGregor KM, Chang YL, Benjamin M, Gopinath K, White KD. Regional changes in word-production laterality after a naming treatment designed to produce a rightward shift in frontal activity. *Brain and Language*. 2009; 111(2):73–85.10.1016/j.bandl.2009.08.001 [PubMed: 19811814]
- del Toro CM, Altmann LJP, Raymer AM, Leon S, Blonder LX, Rothi LG. Changes in aphasic discourse after contrasting treatments for anomia. *Aphasiology*. 2008; 22(7):881–892.10.1080/026870300401513
- Doyle PJ, Goda AJ, Spencer KA. The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology*. 1995; 4:130–135.
- Drew RL, Thompson CK. Model-based semantic treatment for naming deficits in aphasia. *Journal of Speech, Language, and Hearing Research*. 1999; 42(4):972–989.
- Duffy, JR. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St. Louis, MO: Mosby; 1995.
- Edmonds LA, Babb M. Effect of verb network strengthening treatment in moderate-to-severe aphasia. *American Journal of Speech-Language Pathology*. 2011; 20(2):131–145.10.1044/1058-0360(2011/10-0036) [PubMed: 21386047]
- Edmonds L, Nadeau S, Kiran S. Effect of verb networking strengthening treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*. 2009; 23(3):402–424.10.1080/02687030802291339 [PubMed: 19763227]
- Goodglass H. The syndromes of aphasia: Similarities and differences in neurolinguistic features. *Topics in Language Disorders*. 1981; 1(4):1–14.10.1097/00011363-198109000-00004
- Goodglass, H.; Kaplan, E. *Boston Naming Test*. New York: Lippincott, Williams & Wilkins; 2001.
- Goodglass, H.; Kaplan, E.; Barresi, B. *Boston Diagnostic Aphasia Examination*. 3. San Antonio, TX: Psychological Corporation; 2000.
- Goral M, Kempler D. Training verb production in communicative context: Evidence from a person with chronic non-fluent aphasia. *Aphasiology*. 2009; 23(12):1383–1397.10.1080/02687030802235203 [PubMed: 19911035]
- Hickin J, Best W, Herbert R, Howard D, Osborne F. Treatment of word retrieval in aphasia: Generalisation to conversational speech. *International Journal of Language & Communication Disorders*. 2001; 36(Suppl):13–18.10.3109/13682820109177851 [PubMed: 11340768]
- Hong APH, Law SP. A linguistic communication measure for monitoring changes in Chinese aphasic narrative production. *Clinical Linguistics and Phonetics*. 2009; 23(4):255–269.10.1080/02699200802673234 [PubMed: 19382013]
- Kemper S, Herman RE, Lian CHT. The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. *Psychology & Aging*. 2003; 18(2):181–192.10.1037/0882-7974.18.2.181 [PubMed: 12825768]

- Kendall D, Rosenbek JC, Heilman K, Conway T, Klenberg K, Rothi LJG, Nadeau S. Phoneme-based rehabilitation of anomia in aphasia. *Brain and Language*. 2008; 105:1–17. doi: <http://dx.doi.org/10.1016/j.bandl.2007.11.007>. [PubMed: 18237773]
- Kertesz, A. *Western Aphasia Battery*. San Antonio, TX: Psychological Corporation; 1982.
- Kiran S, Thompson CK. The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language and Hearing Research*. 2003; 46(4):773.10.1044/1092-4388(2003/061)
- Lindamood PC, Bell N, Lindamood P. Issues in phonological awareness assessment. *Annals of Dyslexia*. 1992; 42:242–259. [PubMed: 24233878]
- Maher LM, Raymer AM. Management of Anomia. *Topics in Stroke Rehabilitation*. 2004; 11(1):10–21.10.1310/318R-RMD5-055J-PQ40 [PubMed: 14872396]
- Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37(4):913–934.10.1037/a0022964
- McNeil, MR.; Robin, DA.; Schmidt, RA.; McNeil, M. Apraxia of speech: Definition, differentiation, and treatment. In: McNeil, M., editor. *Clinical Management of Sensorimotor Speech Disorders*. New York: Thieme Medical Publishers; 1997. p. 311-344.
- McRae K, Hare M, Elman JL, Ferretti T. A basis for generating expectancies for verbs from nouns. *Memory & Cognition*. 2005; 33(7):1174–1184.10.3758/BF03193221 [PubMed: 16532852]
- Metusalem R, Kutas M, Urbach TP, Hare M, McRae K, Elman JL. Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*. 2012.10.1016/j.jml.2012.01.001
- Nadeau, S.; Rothi, L.; Rosenbek, J. Language rehabilitation from a neural perspective. In: Chapey, R., editor. *Language Intervention Strategies in Aphasia And Related Neurogenic Communication Disorders*. Philadelphia: Lippincott, Williams & Wilkins; 2008. p. 689-734.
- Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech & Hearing Research*. 1993; 36(2):338–350. [PubMed: 8487525]
- Nickels LA. Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*. 2002; 16(10–11):935–979.10.1080/02687030244000563
- Nickels, LA.; Berndt, RS. Words fail me: Symptoms and causes of naming breakdown in aphasia. In: Berndt, RS., editor. *Handbook of Neuropsychology*, 2nd ed.: Vol 3 Language and Aphasia. Amsterdam Netherlands: Elsevier Science Publishers B.V; 2001. p. 115-135.
- Oldfield RC. The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*. 1971; 9:97–113. [PubMed: 5146491]
- Olness GS. Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*. 2006; 20(02–04):175–187.10.1080/02687030500472710
- Olness GS, Ulatowska HK, Wertz RT, Thompson JL, Auther LL. Discourse elicitation with pictorial stimuli in African Americans and Caucasians with and without aphasia. *Aphasiology*. 2002; 16(4–6):623–633.10.1080/02687030244000095
- Pashek GV. A case study of gesturally cued naming in aphasia: dominant versus nondominant hand training. *Journal of Communication Disorders*. 1997; 30(5):349–366. [http://dx.doi.org/10.1016/S0021-9924\(96\)00079-2](http://dx.doi.org/10.1016/S0021-9924(96)00079-2). [PubMed: 9309528]
- Peach RK, Reuter KA. A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. *Aphasiology*. 2009; 24(9):971–990.10.1080/02687030903058629
- Plummer-D'Amato P, Altmann LJP, Reilly K. Dual-task effects of spontaneous speech and executive function on gait in aging: Exaggerated effects in slow walkers. *Gait & Posture*. 2011; 33(2):233–237. doi: dx.doi.org/10.1016/j.gaitpost.2010.11.011. [PubMed: 21193313]
- Plummer-D'Amato P, Altmann LJP, Saracino D, Fox E, Behrman AL, Marsiske M. Interactions between cognitive tasks and gait after stroke: A dual task study. *Gait & Posture*. 2008; 27:683–688. doi: <http://dx.doi.org/10.1016/j.gaitpost.2007.09.001>. [PubMed: 17945497]
- Raymer AM, Beeson P, Holland A, Kendall D, Maher LM, Martin N, et al. Translational research in aphasia: From neuroscience to neurorehabilitation. *Journal of Speech, Language, and Hearing Research*. 2008; 51(1):S259–S275.10.1044/1092-4388(2008/020)

- Raymer AM, Ciampitti M, Holliway B, Singletary F, Blonder LX, Ketterson T, Rothi LJG. Semantic-phonologic treatment for noun and verb retrieval impairments in aphasia. *Neuropsychological rehabilitation*. 2007; 17(2):244–270.10.1080/09602010600814661 [PubMed: 17454696]
- Raymer AM, Ellsworth TA. Response to contrasting verb retrieval treatments: A case study. *Aphasiology*. 2002; 16(10–11):1031–1045.10.1080/026870401430000609
- Raymer, AM.; Rothi, LG. Cognitive approaches to impairments of word comprehension and production. In: Chapey, R., editor. *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders*. 4. Baltimore, MD: Lippincott, Williams & Wilkins; 2001. p. 524-548.
- Raymer AM, Singletary F, Rodriguez A, Ciampitti M, Heilman KM, Rothi LJG. Effects of gesture+verbal treatment for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society*. 2006; 12(6):867–882. doi: 10.1017/S1355617706061042. [PubMed: 17064449]
- Rose M, Douglas J. Treating a semantic word production deficit in aphasia with verbal and gesture methods. *Aphasiology*. 2008; 22(1):20–41.10.1080/02687030600742020
- Rose M, Douglas J. The differential facilitatory effects of gesture and visualisation processes on object naming in aphasia. *Aphasiology*. 2001; 15(10–11):977–990.10.1080/02687030143000825
- Rose M, Sussmilch G. The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology*. 2008; 22(7–8):691–706.10.1080/02687030701800800
- Thompson, CK. Generalization issues in the treatment of communication disorders. In: McReynolds, LV.; Spradlin, JE., editors. *Generalization in the Treatment of Aphasia*. Toronto: Decker; 1989. p. 82-115.
- Troche MS, Altmann LJP. Sentence production in Parkinson disease: Effects of conceptual and task complexity. *Applied Psycholinguistics*. 2011; 33(2):225. doi: <http://dx.doi.org/10.1017/S0142716411000336>.
- Tryon WW. A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis*. 1982; 15(3):423.10.1901/jaba.1982.15-423 [PubMed: 7142062]
- Turner, & Greene. The Construction and Use of a Propositional Text Base. 1977. from <<http://ics.colorado.edu/techpubs/pdf/87-02.pdf>>
- Wambaugh JL, Ferguson M. Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research and Development*. 2007; 77(3):381–394.10.1682/JRRD.2006.05.0038 [PubMed: 18247235]
- Wertz, RT.; LaPointe, LL.; Rosenbek, JC. *Apraxia of speech in adults: The disorder and its management*. Grune & Stratton; Orlando, FL: 1984.
- Wilshire, CE.; Coslett, HB. Disorders of word retrieval in aphasia: Theories and potential applications. In: Nadeau, S.; Rothi, LG.; Crosson, B., editors. *Aphasia and Language: Theory to Practice*. New York, NY: Guildford Press; 2000. p. 82-107.
- Wright, HH.; Newhoff, M. Pragmatics. In: LaPointe, LL., editor. *Aphasia and Related Neurogenic Language Disorders*. 3. New York: Thieme; 2005. p. 237-248.

Figure 1.

At the 3-month follow-up assessment, the number of utterances, words, verbs, grammatical sentences, and CIUs improved for descriptions of pictures in the Gesture group, while they declined nonsignificantly for responses to open-ended questions in the No Gesture group. Note that none of the three-way interactions were significant.

Descriptive information on participants in both groups before the intervention. Note that groups did not differ significantly on any variable. Group means and standard deviations are also shown

Table 1

Subject	Age ¹	Education ²	Gender	CVA Type	Aphasia Type	Treatment	WAB AQ Pre ³	BNT Pre ³
S01	79	12	Female	Hemorrhage	Broca's	Gesture	51.2	8
S03	92	12	Female	Ischemic	Conduction	Gesture	69.1	29
S05	73	14	Male	Ischemic	Anomic	Gesture	76.4	38
S06	62	18	Female	Hemorrhage	Conduction	Gesture	57.8	16
S11	68	14	Male	Ischemic	Conduction	Gesture	67.8	28
S12	63	18	Female	Ischemic	Conduction	Gesture	66.6	11
S15	68	16	Female	Ischemic	Broca's	Gesture	69.4	43
Mean	72.1 (10.5)	14.9 (2.5)	5/2	5/2			65.5 (8.3)	24.7 (13.4)
S04	53	12	Male	Ischemic	Conduction	No Gesture	65.8	27
S07	80	14	Female	Ischemic	Anomic	No Gesture	78.0	40
S08	55	14	Male	Ischemic	TC Motor	No Gesture	67.1	31
S10	69	12	Male	Ischemic	Broca's	No Gesture	52.2	21
S14	61	12	Male	Ischemic	Anomic	No Gesture	75.8	30
S16	59	14	Male	Hemorrhage	Anomic	No Gesture	74.5	30
S19	64	12	Male	Ischemic	Anomic	No Gesture	90.0	37
Mean	63.0 (9.2) ¹	12.9 (1.1) ²	(1/6)	6/1			71.9 (11.8) ³	30.9 (6.3) ³

¹ Groups do not differ, $p = .11$,
² Groups do not differ, $p = .08$,
³ Groups do not differ, $p > .25$

Table 2

Mean scores (SD) of the two groups for picture naming and category generation probe items (percent correct) and two generalization measures, the BNT (number) and WAB-AQ. Note there were no significant differences in performance between groups at any time point on any measure.

	Gesture Group			No Gesture Group		
	Pretest	Posttest	Follow-up	Pretest	Posttest	Follow-up
Picture Naming	38.31 (12.35)	61.91** (20.17)	49.04** (11.15)	43.86 (12.87)	66.89** (20.32)	59.04** (12.12)
Probes (%)						
Category Gen.	52.81 (18.51)	70.71** (26.01)	64.29 (28.46)	57.05 (23.68)	73.21** (22.67)	61.43 (23.36)
Probes (%)						
BNT (#)	24.71 (13.44)	28.57* (16.07)	27.29* (16.38)	30.86 (6.26)	33.86* (9.56)	33.86* (9.62)
WAB-AQ	65.47 (8.34)	67.09 (9.09)	69.79 (11.74)	71.91 (11.80)	72.89 (14.50)	73.33 (15.19)

* Performance is better than pretest at the $p < .05$ level.

** Performance is better than pretest at the $p < .01$ level.

Table 3

Means and (standard deviations) for Discourse Quantity measures of each treatment group over time.

Measure	Task	Gesture Group			No Gesture Group		
		Pre-test	Post-test	3-Month	Pre-test	Post-test	3-Month
Utterances							
	Pictures	30.9 (17.3)	37.0 (24.0)	43.9 (24.1)	25.4 (14.8)	22.3 (11.4)	24.4 (14.2)
	Questions	27.0 (17.9)	27.1 (16.0)	27.1 (9.0)	27.6 (11.6)	26.0 (7.8)	16.4 (5.7)
	Mean	28.9 (16.6)	32.1 (19.1)	35.5 (16.2)	26.5 (12.3)	24.1 (7.9)	20.4 (9.5)
Total Words							
	Pictures	213.7 (93.4)	240.6 (129.4)	334.6 (180.6)	188.1 (181.4)	134.4 (80.3)	144.9 (121.8)
	Questions	168.3 (99.6)	198.0 (96.0)	191.0 (95.0)	198.6 (169.1)	176.4 (131.3)	111.1 (69.6)
	Mean	191.0 (72.4)	219.3 (109.1)	262.8 (134.5)	193.4 (181.8))	155.4 (101.9)	128.0 (88.8)
Nouns							
	Pictures	22.0 (28.7)	21.0 (13.1)	34.7 (28.1)	11.3 (9.9)	13.3 (8.8)	11.4 (11.7)
	Questions	14.1 (7.9)	18.1 (10.3)	14.7 (12.0)	15.0 (17.5)	13.4 (11.6)	11.0 (8.0)
	Mean	18.1 (16.0)	19.6 (11.4)	24.7 (18.6)	13.4 (9.6)	13.4 (9.6)	11.2 (9.5)
Verbs							
	Pictures	26.0 (17.4)	27.3 (19.9)	37.0 (32.7)	18.4 (17.3)	16.4 (8.8)	17.6 (16.9)
	Questions	18.0 (14.6)	20.6 (15.2)	17.1 (17.6)	21.7 (23.5)	20.9 (14.6)	11.9 (10.3)
	Mean	22.0 (14.2)	23.9 (16.8)	23.9 (25.0)	20.1 (20.3)	18.6 (11.1)	14.7 (12.8)

Table 4

Summary of significant interactions on measures of discourse quantity and quality. All analyses comprised a 3-way repeated measures ANOVA, with Task and Time as within subjects variables and Group as a between subjects variable. Discussion of each effect is found in the text.

Measure	Group × Time	Group × Task	Task × Time
Discourse Quantity			
Utterances	<i>ns</i>	<i>ns</i>	$F(2, 24) = 4.525$, $p = .03$, $\eta^2 = .27$
Words	$F(2, 24) = 3.75$, $p = .05$, $\eta^2 = .22$	$F(1, 12) = 9.522$, $p = .01$, $\eta^2 = .44$	$F(2, 24) = 3.779$, $p = .04$, $\eta^2 = .24$
Verbs	<i>ns</i>	$F(1, 12) = 13.162$, $p = .004$, $\eta^2 = .52$	$F(2, 24) = 3.586$, $p = .05$, $\eta^2 = .23$
Nouns	<i>ns</i>	$F(2, 24) = 4.176$, $p = .06$, $\eta^2 = .23$	<i>ns</i>
Discourse Quality			
Grammatical	<i>ns</i>	<i>ns</i>	$F(2, 24) = 4.043$, $p = .03$, $\eta^2 = .25$
CIUs	$F(2, 24) = 3.109$, $p = .06$, $\eta^2 = .21$	$F(1, 12) = 6.590$, $p = .025$, $\eta^2 = .36$	<i>ns</i>
Propositions	<i>ns</i>	$F(1, 11) = 5.871$, $p = .03$, $\eta^2 = .35$	<i>ns</i>
UNIs	<i>ns</i>	<i>ns</i>	<i>ns</i>

Table 5

Means and (standard deviations) on Discourse Quality measures for each treatment group over time.

Measure	Task	Gesture Group			No Gesture Group		
		Pre-test	Post-test	3-Month	Pre-test	Post-test	3-Month
Grammatical Sentences							
CIUs	Pictures	6.1 (5.1)	7.0 (10.8)	9.3 (11.6)	4.7 (5.850)	3.4 (3.3)	3.4 (5.2)
	Questions	5.4 (7.4)	5.4 (7.5)	3.9 (5.9)	4.3 (5.56)	5.0 (4.1)	1.9 (2.3)
	Mean	5.8 (6.2)	6.2 (9.0)	6.6 (8.5)	4.5 (5.6)	4.2 (3.5)	2.6 (3.7)
	Pictures	71.4 (70.6)	67.0 (46.8)	104.6 (87.8)	52.9 (53.8)	43.6 (36.4)	47.6 (77.0)
	Questions	41.3 (32.1)	61.6 (48.8)	38.3 (20.8)	75.6 (111.8)	51.0 (66.0)	37.4 (38.1)
	Mean	56.4 (41.8)	64.3 (46.4)	71.4 (51.2)	64.2 (81.5)	47.3 (50.3)	42.5 (54.9)
	Propositions						
	Pictures	44.0 (39.2)	43.1 (36.3)	61.1 (48.7)	38.0 (29.7)	29.3 (20.5)	40.0 (48.7)
	Questions	31.9 (26.9)	42.3 (29.3)	34.4 (32.4)	58.3 (68.0)	39.3 (32.9)	33.1 (18.5)
Mean	37.9 (27.4)	42.7 (32.2)	47.8 (43.7)	41.4 (47.0)	29.6 (42.7)	31.4 (31.5)	
UNIs							
Pictures	10.4 (6.8)	11.3 (7.3)	16.4 (12.9)	8.7 (7.9)	5.6 (4.2)	7.1 (8.6)	
Questions	9.3 (9.3)	10.1 (6.1)	7.1 (4.4)	9.1 (11.5)	7.9 (6.7)	5.4 (3.8)	
Mean	9.9 (6.8)	10.7 (6.3)	11.8 (8.0)	8.9 (9.6)	6.7 (5.1)	6.3 (5.6)	