



Published in final edited form as:

J Speech Lang Hear Res. 2011 August ; 54(4): 1101–1117. doi:10.1044/1092-4388(2010/10-0117).

Treatment of category generation and retrieval in aphasia: Effect of typicality of category items.

Swathi Kiran and Chaleece Sandberg

Department of Speech Language and Hearing Sciences, Boston University

Rajani Sebastian

Department of Communication Sciences & Disorders, University of Texas at Austin

Abstract

Purpose: Kiran and colleagues (Kiran, 2007, 2008; Kiran & Johnson, 2008; Kiran & Thompson, 2003) have previously suggested that training atypical examples within a semantic category is a more efficient treatment approach to facilitating generalization within the category than training typical examples. The present study extended our previous work examining the notion of semantic complexity within goal-derived (ad-hoc) categories in individuals with aphasia.

Methods: Six individuals with fluent aphasia (range = 39-84 years) and varying degrees of naming deficits and semantic impairments were involved. Thirty typical and atypical items each from two categories were selected after an extensive stimulus norming task. Generative naming for the two categories was tested during baseline and treatment.

Results: As predicted, training atypical examples in the category resulted in generalization to untrained typical examples in five out of the five patient-treatment conditions. In contrast, training typical examples (which was examined in three conditions) produced mixed results. One patient showed generalization to untrained atypical examples, whereas two patients did not show generalization to untrained atypical examples.

Conclusions: Results of the present study supplement our existing data on the effect of a semantically based treatment for lexical retrieval by manipulating the typicality of category exemplars.

Introduction

A central theme in categorization research has been the examination of graded structure in categories. Graded structure refers to the continuum of category representativeness, beginning with the most typical members of a category and continuing through its atypical members to those nonmembers least similar to category members. This idea of “graded membership” within categories has been supported by several studies that have shown differences in lexical processing between typical and atypical exemplars of a category. Typical examples generally receive preferential processing compared to atypical examples, and this phenomenon has been labeled the typicality effect (Hampton, 1979; Posner & Keele, 1968; Rosch, 1973; 1975). It turns out that common taxonomic categories such as *fruits* and *birds* all have graded structures. In fact, Vigliocco and colleagues have argued that semantic distance (determined by the amount of feature overlap) is a stronger predictor of category organization and lexical access than category boundary definitions (Vigliocco, Vinson, Damian, & Levelt, 2002). Like common categories, goal-derived ad-hoc categories

also possess graded structures in which typicality can be determined for members of a particular category (Barsalou, 1983; 1985). These categories are instrumental to achieving goals, particularly goals of daily living, such as *'things to sell at a garage sale'*. However, ad-hoc categories are more graded than common categories because they do not have rigidly defined features that constitute category membership. Instead, category members follow a loosely combined thread of common features. In addition, these categories are not as established in memory as common categories because people have had less experience with them as categorical concepts. That is, one does not specifically think about *things to sell at a garage sale* as a discrete group of instances or a defined entity very often. Further, typical examples in goal-derived categories are those that are most suited to achieving the specific goal and not necessarily the ones that share the most common properties. Despite this, typicality effects have been observed in ad-hoc categories in normal individuals as well as in patients with aphasia. In a series of studies, Hough (Hough, 1993; 1989) investigated patients' awareness and knowledge of goal-derived and common category structure. Hough found that fluent and non-fluent patients with aphasia exhibited similar typicality patterns as controls on a category generation task for goal-derived categories; however, these patients were more anchored to the central portion of a category's referential field for common categories. On a category verification task, Hough found that fluent and non-fluent patients required significantly more time to identify category exemplars than the non-brain-damaged individuals; however, no differences between these two groups were observed in their overall typicality pattern. Therefore, these results suggest that patients with aphasia exhibit relatively unimpaired representations of goal-driven categories and are sensitive to the graded effects within such categories. Additionally, in previous work, we have also found that patients with aphasia respond faster to typical examples than to atypical examples in ad-hoc categories. Normal young, normal elderly and patients with aphasia participated in an online category verification task in which primes were ad-hoc category labels while targets were typical members, atypical members, nonmembers, or nonwords. All three groups were significantly faster and more accurate on typical examples than on atypical examples, however, patients with semantic impairment differed from their non-semantically impaired counterparts and normal controls by showing abnormal typicality effects for ad-hoc categories (Sandberg, Sebastian, & Kiran, submitted). These studies, therefore, seem to be inconclusive about the representation of typicality in ad-hoc categories across different types of patients with aphasia. Further, the extent to which the graded nature of category representation can be exploited in treatment for lexical retrieval in patients with aphasia has not been examined. This is an important empirical question as ad-hoc categories have a distinct advantage in that the number of items that can be retrieved for a category label is potentially endless, and therefore may be more conducive to examining lexical retrieval than a traditional picture naming task.

The present study was therefore aimed at manipulating typicality (or gradedness) as a treatment variable to facilitate lexical retrieval in individuals with aphasia. In our previous work, we have suggested that training the more complex atypical examples in semantic categories results in generalization to the less complex typical examples because although atypical items are less representative of their category than typical items, features common to both typical and atypical items are trained. These findings have borne out in three studies examining generalization from atypical to typical examples and vice versa using animate categories (*birds, vegetables*) (Kiran & Thompson, 2003), inanimate categories (*clothing, furniture*) (Kiran, 2008) and in well defined categories (*shapes*) (Kiran & Johnson, 2008) and have formed the basis for the complexity hypothesis (Thompson, 2007) within the semantic domain (Kiran, 2007). This idea of complexity extends into domains other than categories of concrete objects. Similar findings were observed during treatment of category generation of abstract or concrete words by utilizing abstractness as a marker of complexity. Three of four patients with aphasia showed improvement on trained abstract words (e.g.,

prayer, within the context of *church*) and generalization to untrained concrete words (e.g., *candle* within the context of *church*). Two of the four patients showed improvement on trained concrete words, but no generalization was observed to untrained abstract words (Kiran, Sandberg, & Abbott, 2009).

In the present study, it was hypothesized that, similar to semantic categories such as *birds* and *furniture*, ad-hoc/goal-derived categories such as *things to sell at a garage sale* are represented in terms of typicality, with typical items (e.g., *cups, clocks*) in the center of the semantic space and atypical items (e.g., *candles, pens*) at the periphery of this space. Therefore, training atypical items in a category was hypothesized to result in generalization to untrained target typical items in the same category. Training typical items in a category, however, was hypothesized to result in the retrieval of trained typical items, but not generalization to target atypical items. An important assumption here is that words in ad-hoc categories are represented as semantic features (Mc Rae, de Sa, & Seidenberg, 1997; Pexman, Holyk, & Monfils, 2003) and that training semantic features for a specific word should improve access to the phonological form of the word and its semantically related neighbors (Kiran & Bassetto, 2008). With that in mind, the theoretical premise proposed here is that training items at the periphery would strengthen a more distributed set of featural representations of items that help fulfill the goal of the category, whereas training featural representation of items at the center of the category would only reinforce the core features that fulfill the goal but not the featural variations. For instance, semantic features relevant to atypical examples such as *used as protection from weather, multipurpose, and used for personal hygiene* fulfill the goals of a goal-derived category like *things to take camping* and also represent the featural variation of the category. Training such semantic features for relevant atypical examples is expected to impact a broader range of examples within the category than training features such as *used for setting up camp* which only fulfill the core goal of *things to take camping*.

As another goal of the study, we examined if training patients to generate typical or atypical examples for a category would likely influence their ability to generate other semantically related examples within the category. Therefore, two questions were posed. First, strengthening semantic features relevant to atypical examples results in strengthening access to words that are represented by a greater variety of features, it could be speculated that training access to atypical examples within ad-hoc categories result in improved access to a larger number of exemplars within the category than training typical examples? To address this question, we evaluated the total number of responses produced as a function of treatment (typical versus atypical treatment). Second, will patients develop and implement different strategies during improved category generation as a function of treatment? To address this question, we categorized responses that were generated into subgroups (clusters) and examined whether or not patients showed trends in generating similar semantic clusters.

From a clinical standpoint, examining ad-hoc (or goal-derived) categories also permitted us to extend the typicality treatment protocol beyond confrontation picture naming to a category generation task. We have successfully implemented the category generation task as a dependent variable in a previously described study examining complexity (Kiran, et al., 2009). Because ad-hoc categories differ from common language categories in that they lack distinct category boundaries, a potentially vast list of examples can be generated for each category examined. Consequently, a category generation task is more conducive to examining lexical retrieval within such categories as compared to picture naming. Interestingly, Hough (2007) found that both middle-aged and elderly normal individuals generated similar numbers of responses for goal-derived categories (but not common language categories) as compared to young participants, reflecting the fact that the ability to

generate associative connections for goal-derived categories does not decline with increasing age. Therefore, employing a category generation task for ad-hoc (goal-derived) categories provides a practical and naturalistic opportunity to assess lexical retrieval in patients with aphasia. To summarize, the present study used a category generation task to facilitate lexical retrieval in patients with aphasia and examined whether or not manipulating typicality of items within specific goal-derived categories resulted in selective acquisition and generalization patterns for trained and untrained items.

Methods

Participants

Six monolingual, English-speaking individuals with aphasia were recruited from local hospitals within the Austin area to participate in the study. Several initial selection criteria were met, including (a) a single left hemisphere stroke in the distribution of the middle cerebral artery confirmed by a CT/MRI scan, (b) onset of stroke at least six months prior to participation in the study, (c) premorbid right-handedness as determined by a self-rating questionnaire, and (d) at least a high school diploma (see Table 1). All participants also passed an audiometric hearing screening at 40 db HL bilaterally at 500, 1000 and 2000 Hz and showed normal or corrected-to-normal vision as measured by the Snellen chart. All participants had received varying amounts of traditional language treatment during the initial months following their stroke, but were not involved in any concurrent therapy during the study. All participants provided written consent approved by the University of Texas Institutional Review Board.

The diagnosis of aphasia was determined by administration of the *Western Aphasia Battery* (WAB) (Kertesz, 1982). All participants were fluent, anomie, or conduction aphasic (See Table 1 for details). Several other inclusionary criteria were employed for participation in the study. First, performance on the *Boston Naming Test* (BNT) (Goodglass, Kaplan, & Weintraub, 1983) was required to be below 65% (40/60) accuracy to ensure that participants showed lexical retrieval impairments. However, all participants could name at least some of the pictures on the BNT, indicating that they did not demonstrate a severe naming impairment. Next, all participants demonstrated impaired category generation on the WAB category fluency task which was deemed to be similar to the treatment dependent variable (See Table 2). Except for P4, all participants demonstrated mild semantic impairments on four semantic processing subtests that were administered from the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) (Kay, Lesser, & Coltheart, 1992) and the *Pyramids and Palm Trees* test (PAPT) (Howard & Patterson, 1992). With the exception of P4, all patients showed mild impairments in written naming but fairly accurate reading and repetition skills, validating the hypothesis that patients demonstrated lexical retrieval impairments and not impairments in the phonological output lexicon. P4 presented with a combination of semantic and phonological impairments.

With the exception of P1, all participants were also administered the *Cognitive Linguistic Quick Test* (CLQT) (Helm-Estabrooks, 2001). All participants tested performed either within normal limits or with mild impairments on all components of the test. Exceptions to this include moderate impairment on the language component for P5 and P4 had significant difficulty with the memory, language, and attention components (see Table 3).

To assist in the development of norms for the stimuli employed in the study, 20 young (age range = 21 - 40 years) and 20 older individuals (age range = 41 - 75 years) were recruited from the University of Texas at Austin. All participants had normal or corrected-to-normal vision, normal hearing, and had at least a high school diploma. Exclusionary criteria

included history of neurological disorders, psychological illnesses, alcoholism, learning disability, seizures and attention deficit disorders.

Stimuli

Two ad-hoc categories and their exemplars were developed for utilization in the experiment. In the stimulus development and norming phase, 20 normal young and elderly individuals generated as many items as possible for five ad-hoc categories (things to take camping, things at a grocery store, things at a garage sale, things that fly, things that smell). The items generated were entered into a word database using the software Linguistic Inquiry and Word Count (LIWC) (<http://www.liwc.net/index.php>) that is described in greater detail below. These categories have been used in previous ad-hoc category studies (Barsalou, 1983, 1985; Hough, 1989, 1993). A separate group of 20 normal young and elderly individuals rated the previously-generated items for each of the five ad-hoc categories on a seven point scale. A rating of one corresponded to the item being a very good example or fit of the category; a rating of seven indicated that item was considered a very poor example; a rating of four indicated a moderate fit (Rosch, 1975). Once the participants completed the task, the average rating score, standard deviations, and z-scores for each item were calculated across the 20 participants. Items were eliminated (a) whose average typicality rating occurred with a standard deviation of two or more, (b) that consisted of two or more synonyms in the list of which one was a superordinate label, (c) that were both atypical and unfamiliar, and/or (d) that were verbs. After deleting specific items in each category, some categories were left with too few items to separate into typical and atypical groups. As a result, the following categories were eliminated: *things that fly*, *things at a grocery store* and *things that smell*. The two remaining categories (*things at a garage sale* and *things to take camping*) were selected for treatment.

Stimuli for each experimental category were selected based on the z-scores of the average typicality ratings for each rated item. The 15 items with the lowest z-scores were considered typical examples (*Garage sale* = -0.50, *Camping* = -0.55) and the 15 items with highest z-scores were considered atypical examples (*Garage sale* = .62; *Camping* = .68). We made every effort to ensure that there were no differences in the written word frequency between the typical and atypical examples (*Garage sale* typical $M = 21.84$; *Garage sale* atypical $M = 19.28$, $t = .37$, $p = .71$; *Camping* typical $M = 17.36$; *Camping* atypical $M = 12.11$, $t = 1.06$, $p = .30$) and familiarity (*Garage sale* typical $M = 527$, *Garage Sale* atypical $M = 554$, $t = -.57$, $p = .57$; *Camping* typical $M = 540$; *Camping* atypical $M = 536$, $t = .14$, $p = .88$) (MRC Psycholinguistic Database, Coltheart, 1981). In some cases, we kept two-word phrases because they were an integral part of the representation of the category (e.g., *sleeping bag* for *camping*). Individual typed cards were printed for all words.

Development of semantic features for treatment.—For each of the two categories, ten normal participants listed as many semantic features as possible. Instructions to participants were “In this experiment, I will give you a written list of items and ask you to write down as many attributes as you can that you think are applicable for the given examples. Please keep in mind that there is no right or wrong response. Please provide at least 15-20 attributes that are relevant for all or some of the examples provided.” As expected, the category *garage sale* elicited many more features ($N = 46$) than *camping* ($N = 22$). Then, the number of items that each feature applied to was tabulated. Certain features were applicable to all examples and fulfilled the goal of the category (e.g., *unwanted/ unneeded* for *garage sale*) and were hence labeled **core** features. Others were applicable to typical examples in the category but were not considered integral to fulfilling the goal (e.g., *buy at a sporting goods store* for *camping*). Still others were applicable only for atypical examples in the category (e.g., *things needed for cooking* for *camping*). Similarly, the

number of features that each example possessed or did not possess was also tabulated ensuring that there was a relatively even distribution of features across the examples generated for the category. See appendix A for a representative list of semantic attributes used for each of the two treatment categories. Fifteen distracter features for each category were developed that did not apply to any of the examples in the category. Consequently, there were a total of 58 features for the category *garage sale* and a total 31 features for the category *camping*. It was reasoned that the differential number of features selected for each category would not affect treatment outcomes; however, administration of a multiple baseline across categories treatment design allowed for the systematic examination of this issue.

Design

This study used a single-subject multiple baseline across categories experimental treatment design with the order of category and typicality counterbalanced across the six participants (see Table 4). Criterion for switching treatment from one category to the next was set at either 80% accuracy generating target items in the trained category on two consecutive treatment probes or the completion of 20 training sessions.

Baseline naming procedures.—Generative naming for the two categories was tested during baseline. The number of baselines was varied across participants to evaluate the stability of the dependent measure. Participants were instructed to name as many words associated with each category/location as they could without a time limit. The number of (a) target typical words (e.g. *tent, flashlight* for *camping*) and (b) target atypical words (e.g. *pen, pillow* for *garage sale*) were tabulated. Target typical and atypical words were category exemplars that were normed based on the procedures described above. They were marked as generated (1) or not generated (0) and were considered correct if they were clear and intelligible productions of the target words, semantically similar variations of the target word, or a very close synonym of the target word. We also kept track of (1) untrained typical words (e.g., *bread maker, skillet* for *garage sale*) and (2) untrained atypical items (*necklace* for *garage sale*) which were category exemplars that we had typicality norms for and that were spontaneously generated by each participant. These responses were only considered correct untrained words if they were intelligible productions of words that were appropriate for the category, and that were identical or semantic variations of items that we had typicality z ratings from our normed data set (e.g., *DVDs* counted as a correct other response for *movies*). We should note that we could not classify all responses produced by each participant as untrained typical/atypical examples because our norms on typicality for items in the category are for approximately 50 items whereas the participants generated many more items during the sessions. Additionally, participants produced several items that were similar to the norms collected but did not carry the same meaning in the context of the category (e.g., *shoes/hiking boots*). Consequently, the data reported as untrained typical or atypical is somewhat subjective and should be interpreted in the context of this limitation. To circumvent this issue, all spontaneous generations that did not belong to the target typical/atypical set but were considered acceptable members of a category were also tabulated.

Qualitative analysis of responses.—We also conducted a qualitative analysis of the responses generated. First, all responses generated by patients and normal controls during the norming tasks were entered into a database using the software Linguistic Inquiry and Word Count (LIWC) (<http://www.liwc.net/index.php>). LIWC was used to count the total number of items produced by all normal controls and all patients for each category and assembled into a ‘category dictionary’. In this way, two dictionaries were created, (a) *garage sale* with 596 unique words and (b) *camping* with 469 unique words. For each

category, dictionary responses were categorized into subcategories (camping: supplies/tools, clothing/personal care, food/cooking, games/entertainment, transportation, wildlife/animals and *other*; *garage sale*: clothing, kitchen, electronics, fruits/vegetables, furniture, entertainment, prepared food/drinks, home/garden, and *miscellaneous*). Then, LIWC was used to count the number of times a word in a particular subcategory was produced. Responses that did not fall under these categories were classified as production errors.

Treatment.—Each treatment session was carried out in four steps: (1) category generation, (2) category sorting, (3) feature generation/selection, and (4) yes/no feature questions (see Appendix B). Patients were seen two times per week for two hours each session.

Treatment probes.—Throughout treatment, the same generative naming probes used in the baseline condition were presented every second treatment session to assess retrieval of the trained and untrained items. Generalized retrieval of untrained items was considered to have occurred when levels of performance changed by at least forty percentage points over baseline levels.

Reliability.—All baseline sessions and treatment sessions were recorded on videotape. Reliability on the dependent variable for participants was calculated for 75% of the probe sessions, resulting in 100% agreement. Reliability on the independent variable (i.e., presentation of the treatment protocol) was calculated for 50% of treatment sessions, resulting in 100% agreement.

Data analysis.—To calculate effect sizes (ES), the average baseline probe scores were subtracted from the average post-treatment scores and the result was divided by the standard deviation of the baseline scores (Beeson & Robey, 2006). In cases where treatment was provided for a second category, all pre-treatment sessions were entered into baseline calculation (e.g., for P4 *garage sale* treatment, 15 data points were entered into baseline calculation). For P1 and P3, post-treatment probe scores could not be obtained due to patient evaluation scheduling issues. Consequently for these two patients, the average of the final two treatment probe scores was used. Beeson and Robey (2006) recently updated the benchmarks for direct treatment of naming deficits and generalization of treatment (acquisition of trained items, 6.5 = small ES, 8.0 = medium ES, 9.5 = large ES and for generalization of treatment, 2.0 = small ES, 5.0 = medium ES, 8.0 = large ES). In order to consider the treatment effective, we set the benchmark of an effect size of 6.5 for the trained items and an improvement to 80% accuracy for two consecutive sessions. Likewise, for generalization to be considered positive, we set a benchmark of an effect size of 2.0 and an improvement of 40% accuracy over baseline levels for the untrained items (see Table 4).

Results

Category generation treatment results

Participant 1.—P1 received treatment for atypical examples of *garage sale* for seven weeks. These items improved to criterion (see Figure 1, Table 4) and generalization occurred for untrained typical words. Treatment was not provided for the second category due to personal health issues.

Participant 2.—P2 received six weeks of treatment for atypical examples of the category *camping*. Retrieval of the trained atypical items improved and generalization also occurred for the untrained typical items of the category (see Figure 2; Table 4). Treatment was then shifted to typical examples of *garage sale* which improved to criterion but did not meet our

a priori effect size criterion. Interestingly, some generalization was observed to untrained atypical examples.

Participant 3.—P3 received 15 weeks of treatment for atypical examples of the category *garage sale*. Treatment was extended from 10 to 15 weeks for this patient as he showed trends of improvement on the trained items but an overall variable performance. This patient did eventually improve to 87% accuracy on the trained atypical items (see Figure 3; Table 4). Generalization to the untrained typical items was modest in accuracy but yielded a large ES (see Table 4). This patient developed a health complication towards the end of treatment and chose not to continue in treatment.

Participant 4.—P4 was trained on typical examples of *camping* for 10 weeks. Retrieval of trained items improved but no generalization was observed for the untrained atypical examples. Treatment was then shifted to atypical examples of *garage sale*. Again, improvement was noted but the number of items retrieved did not reach criterion (see Figure 4, Table 4). Notably, generalization to the untrained typical examples was observed and importantly, P4 could retrieve more typical examples than the trained atypical examples.

Participant 5.—P5 was trained on typical examples of *camping* for 10 weeks. Retrieval of trained items improved but did not reach criterion. Little to no generalization was observed for the untrained atypical examples (see Figure 5, Table 4). P5 was not trained on the second category per the participant's wish to terminate treatment.

Participant 6.—P6 was trained on atypical examples of *garage sale* for 10 weeks. Retrieval of trained atypical items reached criterion. Generalization to untrained typical examples was also observed with retrieval of typical examples reaching consistently higher levels than trained atypical examples during the course of treatment (see Figure 6, Table 4). Participant 6 also expressed an interest in terminating treatment due to scheduling conflicts.

Quantitative and qualitative analysis of other responses generated for each category

The responses generated by each participant were further examined to see if therapy influenced the nature of retrieval. For this analysis, target typical and atypical words were eliminated from the data set as these responses were directly related to the therapy and are already illustrated in Figures 1-6. Next, production errors or phonological errors and/or neologisms were eliminated from the data set and not analyzed further. In order to understand whether or not the stimuli trained (either typical or atypical) had any effect on the number of responses produced as a consequence of treatment, an analysis of the overall number of responses in the category was conducted. Responses generated by each participant during the first two baseline sessions and the final two treatment/post – tx sessions were tabulated. For this analysis, we only included the total numbers of items in each category (and not the categorized numbers). A repeated measures ANOVA using the average number of responses produced during baseline and end of treatment probes as the dependent measure and participant and typicality of treatment stimuli as the independent variables showed significance [Wilks lambda = .02814, $F(14, 14)=4.9610$, $p=.00249$]. Figure 7 shows that, in general, participants who were trained on atypical examples in a category generated more items for that category at the end of treatment as compared to participants who were trained on typical examples of a category.

Next, we regrouped all responses (including target trained, target untrained, other responses and production errors) to see if there were changes as a function of treatment. These examples belonged to one of several subcategories for the category (e.g., for the category *camping*, other responses could be examples of supplies/tools, clothing/personal care, food/

cooking, games/entertainment, transportation, wildlife/animals and *miscellaneous*) with the exception of target typical and target atypical examples. For each patient, the proportion of responses within each subtype was computed for each session. Then, for each subtype of response, the average from the first two (baseline) sessions was subtracted from the average of the final two (treatment/post-tx) sessions in order to obtain a **difference score**. We decided to include data from two sessions in order to get a measure of consistency across error types. Negative values reflected a decrease in the proportion of response subtypes as a function of treatment whereas positive values reflect an increase in the response subtypes subsequent to treatment. We conducted a hierarchical cluster analysis to see if patients showed similar trends in their production of different response subtypes as a function of treatment. The difference score described above was entered into a hierarchical joining tree cluster analysis (Everitt, Landau, & Leese, 2001). For each category, patients were entered as variables and the response subtypes were entered as clusters. All variables were equally weighted in the analysis and a Euclidean measure was used to compute distance. To determine distance between clusters, we used the single linkage distance (nearest neighbor) to capture the similarity between neighboring clusters. Therefore, this analysis begins with objects in its own cluster, the nodes representing the distance at which the two closest clusters join and progresses to a single cluster (agglomerative). Figure 8 illustrates the dendrograms for the two categories. Based on amalgamation coefficients, the final cluster solution for each category was determined to be a cut-off of linking distance of 25 as the clusters that were merged after a distance of 25 were quite distinct. Consequently, there are four clusters for *garage sale* (1. food/electronic/furniture, 2. kitchen/entertainment/miscellaneous, 3. home garden, and 4. typical examples). Likewise, there were four clusters for *camping* (1. games/entertainment/wildlife, 2. other/transportation, 3. supplies/tools/food/cooking/clothing/personal care and 4. production errors). These results indicate that across patients and independent of the typicality of trained stimuli, the evolution of responses produced during the course of treatment clustered along a semantic similarity dimension.

Standardized test results

Overall, the participants in this study showed improvements or maintained performance on the various standardized measures that were administered pre- and post-treatment. P1, P3, and P4 showed small improvements on WAB AQ scores. No patient showed improvements on the BNT, and two of six patients showed improvements on the WAB category scores. On the PALPA and PAPT, with the exception of written picture naming which improved for all but one participant, improvements on the individual subtests were mixed. Wilcoxon matched pair tests conducted on the pre-post scores did not reveal significance for any subtests.

Discussion

Our previous work (Kiran & Thompson, 2003; Kiran & Johnson, 2007; Kiran, 2008) has shown that training atypical examples is a more efficacious way to facilitate generalization within categories than training typical examples. In the present experiment, the effect of varying exemplar typicality within ad-hoc categories in six individuals with aphasia was examined. Recall that in the introduction, we hypothesized that strengthening access to semantic attributes and phonological representations for target atypical examples would facilitate access to these items as well as to corresponding semantic and phonological representations of untrained typical examples. In contrast, strengthening access to semantic attributes and phonological representations of typical items was predicted to improve only those items, no generalization to untrained atypical examples was expected. The results of the present experiment confirm our hypotheses and extend our previous findings of typicality treatment for lexical retrieval deficits in two ways. First, the present typicality treatment was effective in improving lexical access within a category generation task,

whereas our previous work was focused on improving confrontation picture naming. Second, these results illustrate that typicality within categories can be extended to ad-hoc/goal-derived categories, which are categories that have fairly loose category boundaries. Based on Table 4, it is apparent that when atypical members are trained, generation of untrained atypical items improves in three of the five patients (P1 - garage sale, P3 - garage sale, P6 - garage sale) and generation of target typical examples improves in five patients (P1 - garage sale, P2 - camping, P3 - garage sale, P4 - garage sale, and P6 - garage sale) and generation of untrained typical items improves in three patients (P1 - garage sale, P3 - garage sale, P4 - garage sale). In contrast, when training typical examples, there is no improvement in generation of untrained typical examples, only one patient showed an improvement (P2 - garage sale) on the untrained target atypical examples, and no patients improved on generation of untrained non-target atypical examples. These results suggest that there is benefit to only training atypical examples in the category as generalization extends to typical examples within the category. A second goal of the study was to closely examine the nature of response produced in treatment and if the evolution of responses was influenced by whether trained stimuli were typical or atypical items in the category. Results from this analysis suggested that training atypical examples resulted in more examples generated for the category compared to when typical examples were trained.

The results of the present study have clear theoretical and clinical significance. From a clinical standpoint, there have been several studies aimed at improving lexical access using picture naming as the treatment task by strengthening semantic representations of target items. The use of category fluency as a behavioral variable to improve lexical access has been relatively less common in treatments for patients with aphasia. In the present study, category fluency was considered to be an appropriate task for patients with relatively mild levels of naming impairment. These patients varied on their confrontation picture naming ability but all were impaired on their ability to retrieve items during a category fluency task. Further, word generation more closely resembles the word-finding required for conversation than confrontation naming, lending itself as a more suitable treatment task for real-world application. Notably, participants who were trained on atypical examples in a category generated more items for the category at the end of treatment as compared to participants who were trained on typical examples of a category. These results supplement our previous work (Kiran et al., 2009) showing that category generation of items can be facilitated through a semantically based treatment.

The theoretical implications of the present study span three domains of semantic representation and are explained below. One obvious theoretical implication is the extension of the typicality effect to ad-hoc categories that are not fixed entities in memory but are constructed on a more dynamic basis. As noted in the introduction, ad-hoc (or goal-derived) categories (e.g., *a grocery list*) do not have rigid features that constitute category membership; instead, category members follow a loosely combined thread of common features. Although complexity is more difficult to define in such categories, several studies show that typical examples (e.g., *tent*) are more illustrative of the central goal of the category (e.g., *things to take camping*) than atypical examples (e.g., *playing cards*). The treatment protocol used in this study took into account the fact that core (shared) features in goal-derived categories were restricted to the goal (e.g., *things needed at a camp*); nevertheless, features for typical examples were selected such that they consisted of shared features (e.g., *for nature protection, for sleeping*) whereas atypical examples also included distinctive features that were not shared by other category members (e.g., *things to do at a camp, for personal hygiene*). Consequently, training atypical examples in the category appeared to strengthen semantic features that are relevant for the range of examples within the category, whereas training typical examples only improved the core features in the category. Importantly, training atypical examples resulted in a greater number of overall

responses produced than training typical examples. Further, the cluster analysis for both categories indicated that changes in response types as a function of treatment were along a semantic similarity dimension. Therefore, words in the subcategories that were tokens of *electronics*, *food/drink* and *furniture* for *garage sale* evolved similarly as a function of treatment. Likewise, words denoting *supplies*, *tools*, *food/cooking* and *clothing* for *camping* evolved similarly as a function of treatment.

One way to interpret the results is within the now well-established framework of semantic feature theory (McRae et al., 1997; McRae, Cree, Seidenberg, & McNorgan, 2005) that word meaning is distributed across semantic features. McRae et al. (2005) found that concepts with many shared features are easier to respond to than concepts with fewer shared features. They further showed that the facilitative effect of the shared features was greater than any inhibitory effect from distinctive features. In another study, using computational simulation and behavioral data (Cree, McNorgan, & McRae, 2006), the same authors showed that distinctive features are activated more quickly when a concept is activated with a subsequent feature name and that distinctive features are better cues for retrieving a concept. Taken together, these studies suggest that both distinctive and shared features influence activation of concepts but in slightly different ways and inform the present study in providing a framework for the generalization patterns observed. Our future work is focused on disentangling the nature of semantic feature representation critical to the prediction of semantic complexity.

The present results also raise an interesting issue regarding semantic distance from the central goal. Although central tendency is not a strong predictor for goal-derived categories (Barsalou, 1985), patients with aphasia tend to be anchored to the center of the referential field for these types of categories (Hough, 1989; 1993). This observation may help to explain why training items on the periphery of an ad-hoc/goal-derived category also improves the items at the center of the semantic field. In other words, all examples selected for treatment fulfilled the goal of the category (e.g., *things to take camping*), however, some examples (typical) had features that were necessary to fulfill the goal whereas other examples (atypical) were less important in fulfilling the goal but nonetheless plausible. However distantly linked the atypical items are with the typical items, if the tendency is toward generating the typical items, the slightest overlap in features may trigger activation of typical items and other items that overlap with the goal.

Finally, the results are also consonant with the *situated conceptualization* hypothesis proposed by Barsalou (2003). Barsalou and colleagues (Barsalou, 2003; Yeh & Barsalou, 2006) suggest that concepts activate their relevant associated background situations and allow a perceptual experience of the person imagining the event/concept. In effect, the person simulates a perceptual, motor, and introspective experience whenever presented with a concept or an event. Inferences regarding concepts are generated in a situation-specific manner and when individuals are asked to retrieve instances for a concept, they likely imagine themselves in the specific situation (Vallee-Tourangeau, Anthony, & Austin, 1998; Yeh & Barsalou, 2006). Applying this notion to the present study, we can posit that when patients are being trained on a specific concept (e.g., *things to have in a garage sale*); they simulate themselves being in that specific situation and generate inferences about objects and events likely to be present in that situation (Barsalou, 2003). Therefore, training the atypical examples likely necessitates patients to imagine themselves within a context that includes images of a larger set of examples than when patients are trained on the typical examples, which only requires activation/imagination of items that directly fulfill the goal.

Results from the standardized assessments administered pre and post treatment indicate that participants showed modest improvements or maintained performance. P1, P3, and P4

showed small, but significant, improvements on WAB AQ scores. These findings are consistent with a previous study wherein we argued that patients who received a semantically-based naming treatment improved on standardized language tests of aphasia as compared to a non-treated patient group (Warfield & Kiran, 2008). In that study, we found that patients who received a semantically-based naming treatment significantly improved by an average of 8.6 points on the WAB AQ at the end of the maintenance period. In contrast, the non-treated group decreased an average of 5.7 points on the WAB AQ; however this decrease was not significant. Results on other standardized tests (BNT, PALPA and PAPT) were inconclusive. One possible reason for this lack of significance is that participants P1, P2, P3, and P6 were already at ceiling level on most subtests at pretesting.

Conclusion

In conclusion, this study has added to the corpus of research showing the effectiveness of utilizing complexity in treatments for word retrieval. We should note that although the experimental design is set up as a counterbalanced design, because of the length of treatment (20 weeks) and the relatively rigid structure of treatment, not all participants completed the entire treatment protocol. Therefore, there is an imbalance in the number of participants who received typical treatment versus atypical treatments. To mitigate this potential confound, we set a more stringent criteria for successful acquisition and generalization. Nevertheless, the results need to be interpreted in the context of a limited number of participants. The results of the present study can be considered as a preliminary proof of concept for the use of a typicality treatment within ad-hoc categories using a task such as category generation. We intend to continue examining the effects of typicality training on a larger and more diverse sample of aphasia patients as well as with different types of ad-hoc categories.

Acknowledgments

This research was supported by NIDCD # DC006359-03 and a New Century Research Scholars Grant from American Speech Language Hearing Foundation to the author. The author wishes to thank Shilpa Shamapant and Lisa Edmonds for their assistance during various stages of the project and Crystal Cavin and Jessica Guzman for their role in analyzing the naming errors. Finally, the author thanks the participants in the experiment for their patience and cooperation. Contact author: Swathi Kiran, Ph.D. CCC-SLP, Speech Language and Hearing Sciences, Sargent College of Health & Rehabilitation Sciences, Boston University 635 Commonwealth Ave. Boston, MA 02215

Appendix A: Target and distractor semantic features for the two categories

Category: Garage sale		Category: Camping	
Target	Entertainment	Target	Safety items
Target	Furniture	Target	Needed for cooking
Target	To sleep on	Target	For cleaning
Target	Kitchen or cooking Items	Target	For personal hygiene
Target	Art	Target	For comfort
Target	Bedroom Items	Target	For warmth
Target	Mens clothing	Target	Things needed at a camp
Target	Clothing	Target	Multipurpose
Target	Things that cover the body	Target	For repairs
Target	Electronics	Target	Things to do at a camp
Target	Items for travel	Target	To take out what you take with you

Target	Baby items	Target	For carrying/storing things
Target	Unwanted/ unneeded	Target	For lighting at night
Target	Outdated/outgrown	Target	Nature protection(not weather)
Target	Collectible	Target	Protection from weather
Target	Estate sale	Target	Clothing
Target	For kids	Target	For setting up camp
Target	Household items	Target	For sleeping
Target	Exercise/recreation	Target	Picnic goods
Target	Memorabilia	Target	Obtain at sporting store
Target	Communication	Target	Buy at drug store
Target	Ways to hear music	Target	Optional items
Target	For research	Distracter	Found in a crime scene
Target	things for reading	Distracter	Found under the soil
Target	things that provide light	Distracter	Carries disease
Target	Decorations	Distracter	Uses bullets
Target	Office or Work related items	Distracter	Kills for food
Target	Useful items	Distracter	Makes music

Appendix B: Treatment Protocol

One set of items ($N = 15$; e.g., typical or atypical) were treated at a time. For each item in the training set, the treatment protocol was as follows:

1. **Category Generation:** This step was performed only once at the beginning of each session. The patient was asked generate as many examples as possible for the category *things to sell in a garage sale*. The clinician wrote down all the responses produced and displayed them to the participant as feedback. Then, the clinician selected one item from the target list (typical or atypical) and said “Okay let’s pick this word (bicycle). Let’s go through the training steps and I’ll help you understand more about the features/details of (bicycle) and why it’s something you can have at a garage sale”.
2. **Category Sorting.** This step was performed once at the beginning of each session. The clinician placed written category cards on table: Garage sale and things that you see during Christmas time. (distracter category). The clinician presented the patient with the word cards and ask him (her) sort the cards according to their superordinate category by placing the cards on the category cards. If the patient categorized a word incorrectly he/she was given feedback “Are you sure this (bicycle) is found at a crime scene? It’s actually something you can sell at a garage sale”.
3. **Feature Generation/Selection.** The clinician placed the target word at the center of the table and ask the patient to generate as (at least 4-6) attributes regarding the target (e.g., bicycle) that makes it a good item to fit into the category (e.g., *garage sale*). For example, the participant could say, “children outgrow it, old model” etc. Then, the clinician presented the patient with the features of the target category and asked the patient to select the first 6 semantic features that are pertinent to bicycle. For example, for *bicycle*, the features that were practiced included five that were

pertinent to the example, (e.g., *outgrown/outdated*); five that belong to the category but not the example (e.g., *bedroom*); five that do not belong to the category (e.g., *lives in a forest*). The clinician selected six features and read the features aloud to the patient.

4. Yes/No Questions. The clinician then removed the target picture and the written phrases and instructed the patient “I’m going to ask you some questions about (bicycle) now. Please answer yes or no for each of these questions.” The clinician asked a total of 15 questions, up to five questions that are relevant to the target example, (e.g., *has pedals*); five that belong to the category but not the example (e.g., *is feminine*); five that do not belong to the category (e.g., *lives in a forest*).

References

- Barsalou L. Situated simulation in the human conceptual system. *Language and Cognitive Processes*. 2003; 18:513–562.
- Barsalou LW. Ad-hoc categories. *Memory and Cognition*. 1983; 11(3):211–227.
- Barsalou LW. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 1985; 11(1-4):629–654.
- Beeson PM, Robey RR. Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychological Review*. 2006; 16(4):161–169.
- Coltheart M. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*. 1981; 33A:497–505.
- Cree GS, McNorgan C, McRae K. Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *J Exp Psychol Learn Mem Cogn*. 2006; 32(4):643–658. [PubMed: 16822138]
- Everitt, BS.; Landau, S.; Leese, M. *Cluster Analysis*. 4 ed.. Arnold Publishers; London: 2001.
- Goodglass, H.; Kaplan, E.; Weintraub, S. *Boston Naming Test*. Lea & Febiger; Philadelphia: 1983.
- Hampton JA. Polymorphous concepts in semantic memory. *Journal Of Verbal Learning and Verbal Behavior*. 1979; 18(4):441–461.
- Helm-Estabrooks, N. *Cognitive Linguistic Quick Test*. Harcourt Assessment; London, England: 2001.
- Hough MS. Categorization in aphasia: Access and organization of goal-derived and common categories. *Aphasiology*. 1993; 7(4):335–357.
- Hough MS. Incidence of word finding deficits in normal aging. *Folia Phoniatri Logop*. 2007; 59(1):10–19. [PubMed: 17172782]
- Howard, D.; Patterson, K. *Pyramids and Palm Trees*. Harcourt Assessment; London, England: 1992.
- Kay, J.; Lesser, R.; Coltheart, M. *The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)*. Lawrence Erlbaum Associates; Hove, U. K: 1992.
- Kertesz, A. *The Western Aphasia Battery*. Grune and Stratton; Philadelphia: 1982.
- Kiran S. Semantic complexity in the treatment of naming deficits. *American Journal of Speech Language Pathology*. Feb.2007 16:1–12.
- Kiran S. Typicality of Inanimate Category Exemplars in Aphasia Treatment: Further Evidence for Semantic Complexity. *Journal of Speech Language and Hearing Research*. 2008; 51:1550–1568.
- Kiran S, Bassetto G. Evaluating the effectiveness of semantic based treatment for naming deficits in aphasia: what works? *Seminars in Speech and Language*. 2008; 29(1):71–82. [PubMed: 18348093]
- Kiran S, Johnson L. Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories. *American Journal of Speech Language Pathology*. 2008; 17:389–400. [PubMed: 18845698]
- Kiran S, Sandberg C, Abbott K. Treatment for lexical retrieval using abstract and concrete words in persons with aphasia: Effect of complexity. *Aphasiology*. 2009; 23(7-8):835–853. [PubMed: 19816590]

- 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):773–787.
- McRae K, de Sa VR, Seidenberg MS. On the nature and scope of featural representations in word meaning. *Journal of Experimental Psychology General*. 1997; 126:99–130. [PubMed: 9163932]
- McRae K, Cree GS, Seidenberg MS, McNorgan C. Semantic feature production norms for a large set of living and nonliving things. *Behav Res Methods*. 2005; 37(4):547–559. [PubMed: 16629288]
- Pexman PM, Holyk GG, Monfils M-H. Number-of-features effects and semantic processing. *Memory and Cognition*. 2003; 31(6):842–855.
- Posner MI, Keele SW. On the genesis of abstract ideas. *Journal of Experimental Psychology*. 1968; 77:353–363. [PubMed: 5665566]
- Rosch, E. On the internal structure of perceptual and semantic categories. In: Moore, TE., editor. *Cognitive development and the acquisition of language*. Academic Press; New York: 1973.
- Rosch E. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception & Performance*. 1975; 1(4):303–322.
- Sandberg C, Sebastian R, Kiran S. Processing of typicality within well defined categories and ad-hoc categories in patients with aphasia. submitted. Manuscript submitted for publication.
- Thompson CK. Complexity in language learning and treatment. *Am J Speech Lang Pathol*. 2007; 16(1):3–5. [PubMed: 17329670]
- Vallee-Tourangeau F, Anthony SH, Austin NG. Strategies for generating multiple instances of common and ad-hoc categories. *Memory*. 1998; 6(5):555–592. [PubMed: 10197163]
- Vigliocco G, Vinson DP, Damian MF, Levelt W. Semantic distance effects on object and action naming. *Cognition*. 2002; 85(3):B61–69. [PubMed: 12169413]
- Warfield, E.; Kiran, S. Impact of Model-Based Semantic Treatment on Standardized Tests of Aphasia; Paper presented at the American Speech Language and Hearing Association Convention; 2008;
- Yeh W, Barsalou LW. The situated nature of concepts. *Am J Psychol*. 2006; 119(3):349–384. [PubMed: 17061691]

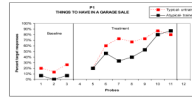


Figure 1. Percent of target responses produced for atypical (trained) and typical (untrained) items for the category garage sale. Treatment was not provided for the second category.

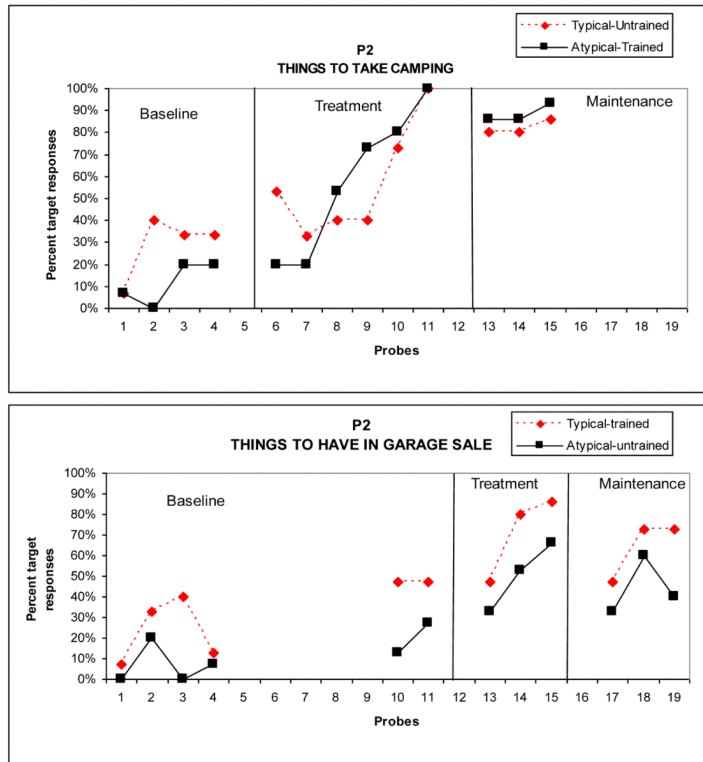


Figure 2. (a) Percent of target responses produced for atypical (trained) and typical (untrained) items for the category camping, and (b) percent of target responses produced for typical (trained) and atypical (untrained) when treatment was provided for category garage sale for Participant 2.

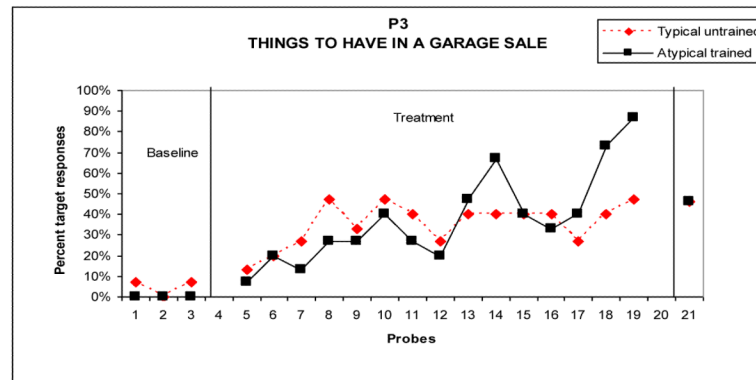


Figure 3. Percent of target responses produced for atypical (trained) and typical (untrained) items for the category garage sale for participant 3. Treatment was not provided for the second category.

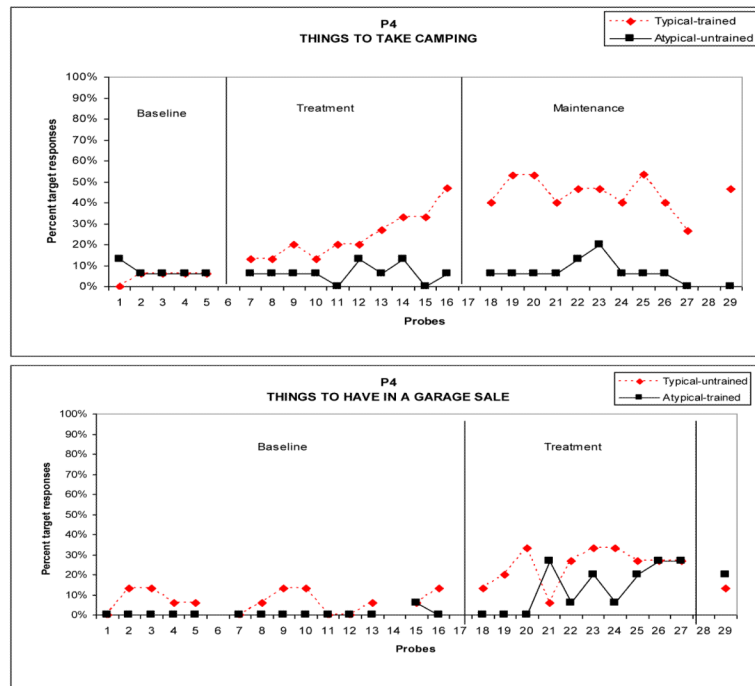


Figure 4. (a) Percent of target responses produced for typical (trained) and atypical (untrained) items for the category camping, and (b) percent of target responses produced for atypical (trained) and typical (untrained) when treatment was provided for category garage sale for Participant 4.

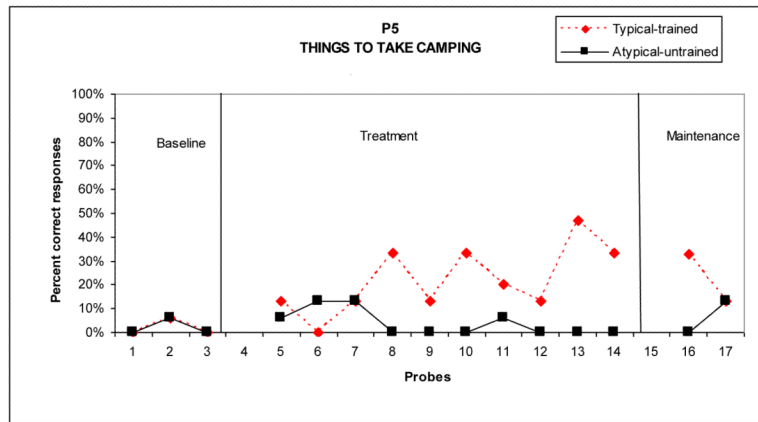


Figure 5. Percent of target responses produced for atypical (trained) and typical (untrained) items for the category garage sale for participant 5. Treatment was not provided for the second category.

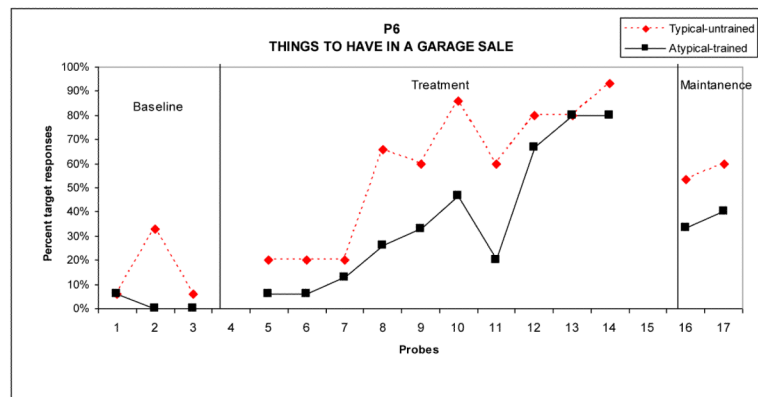


Figure 6. Percent of target responses produced for atypical (trained) and typical (untrained) items for the category garage sale for participant 6. Treatment was not provided for the second category.

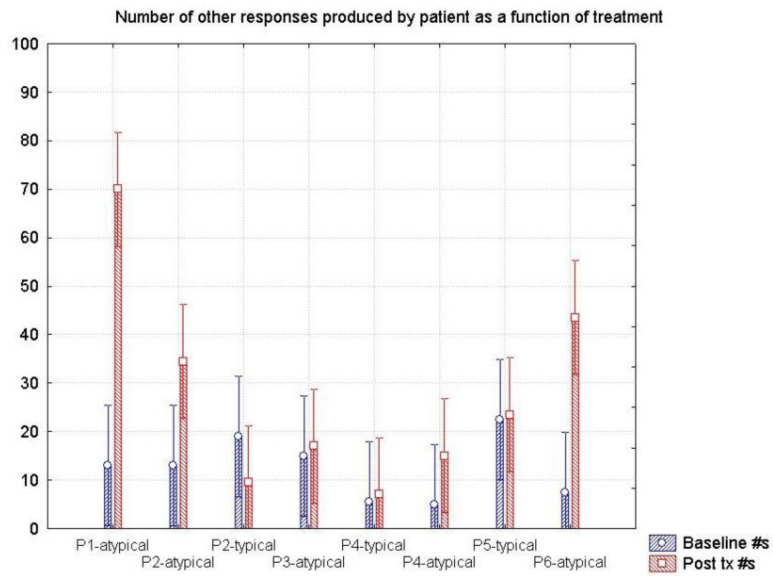


Figure 7.

Proportion of responses produced tabulated as “other” that were not target atypical or typical examples. Blue bars indicate the averaged responses generated during initial two probes and red bars indicate the average responses generated in the final two treatment probes.

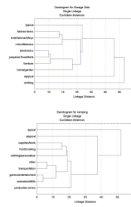


Figure 8. Dendrograms for two categories (camping, garage sale) are displayed as horizontal tree clusters. The x axis indicates linkage distance measured as Euclidean distances and the y axis indicates the cases/clusters considered for each category. See text for further details.

Table 1
Demographic and stroke-related data for the five participants in the study.

	P1	P2	P3	P4	P5	P6
Age (yrs)	76	39	76	69	84	64
Gender	F	F	M	M	F	M
Handedness	Right	Right	Right	Right	Right	Right
Occupation	Teacher	Software engineer	Retired Clerk	Business	Clerk	Attorney
Etiology	L CVA	Left TP hemorrhage	L CVA	Left TP CVA	Left TP CVA	L CVA
MPO	30	6	108	10	9	96
Aphasia AQ	79	82	84.3	72.1	70.9	84.8
Aphasia DX	Anomic	Anomic	Anomic	Conduction	Conduction	Anomic

Table 2
Performance on the Western Aphasia Battery (Kertesz, 1982), Boston Naming Test (Goodglass et al., 1983) and PALPA (Kay et al., 1992) and PAPT (Howard & Patterson, 1992). Changes for WAB are shown in terms of Aphasia Quotient.

	P1		P2		P3		P4		P5		P6	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
<i>Western Aphasia Battery</i>												
<i>Spontaneous Speech</i>	19	20	9	9	16	12	16	16	16	19	15	16
<i>Auditory Comprehension</i>	8.3	8.35	10	10	9.15	7.6	8.05	7.05	7.25	6.6	10	9.8
<i>Repetition</i>	3.9	6.6	7.2	8	8.9	8.4	5.7	5.6	5.6	8.2	8.2	9.5
<i>Naming</i>	8.3	7.9	5.8	7.5	8.1	8.2	6.3	5.8	6.6	5.7	9.2	9.5
<i>WAB category fluency score</i>	12	9	4	10	6	7	7	5	12	6	16	16
<i>Aphasia Quotient</i>	79	85.7	82	89	84.3	85.6	72.1	68.9	70.9	73.1	84.8	89.6
<i>BNT</i>	43.3%	58.3%	21.7%	23.3%	68.3%	66.7%	18.3%	15.0%	26.7%	25.0%	56.7%	53.3%
<i>PALPA</i>												
<i>Auditory Lexical Decision (%)</i>	88.1%	90.6%	98.8%	94.4%	93.8%	85.0%	88.8%	92.5%	73.8%	80.6%	96.9%	96.3%
<i>Letter Length Reading (%)</i>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	50.0%	54.2%	87.5%	100.0%	70.8%	75.0%
<i>Visual Lexical Decision Task (%)</i>	93.3%	91.7%	93.3%	94.2%	90.0%	91.7%	73.3%	77.5%	90.0%	78.3%	98.3%	99.2%
<i>Spoken Word-Picture Matching (%)</i>	100.0%	100.0%	100.0%	97.5%	100.0%	100.0%	95.0%	92.5%	90.0%	95.0%	100.0%	100.0%
<i>Written Word-Picture Matching (%)</i>	100.0%	100.0%	100.0%	100.0%	100.0%	97.5%	97.5%	97.5%	95.0%	92.5%	100.0%	100.0%
<i>Auditory Synonym Judgments (%)</i>	88.3%	90.0%	85.0%	86.7%	95.0%	86.7%	60.0%	61.7%	65.0%	65.0%	90.0%	90.0%
<i>Written Synonym Judgments (%)</i>	100.0%	96.7%	86.7%	91.7%	95.0%	85.0%	56.7%	58.3%	73.3%	71.7%	91.7%	95.0%
<i>Spoken Picture Naming (%)</i>	100.0%	100.0%	92.5%	95.0%	95.0%	92.5%	62.5%	60.0%	62.5%	70.0%	90.0%	95.0%
<i>Writing Picture Names</i>	80.0%	85.0%	85.0%	92.5%	92.5%	100.0%	7.5%	132.5%	62.5%	92.5%	92.5%	95.0%
<i>Reading Picture Names</i>	97.5%	100.0%	97.5%	100.0%	100.0%	95.0%	37.5%	52.5%	92.5%	0.0%	85.0%	95.0%
<i>Spelling Picture Names</i>	92.5%	92.5%	95.0%	95.0%	85.0%	85.0%	10.0%	0	82.5%	60.0%	95.0%	97.5%

Table 3
Performance on the Cognitive Linguistic Quick Test (Helm-Estabrooks, 2001) is reported. WNL = Within Normal Limits

CLQT	P1		P2		P3		P4		P5		P6	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
<i>Attention</i>		WNL	WNL	WNL	WNL	WNL	Mod	Mod	WNL	WNL	WNL	WNL
<i>Memory</i>		WNL	WNL	Mild	WNL	WNL	Severe	Severe	Mild	Mild	WNL	WNL
<i>Executive Functions</i>		WNL	WNL	WNL	Mild	WNL	Mod	Mod	WNL	WNL	WNL	WNL
<i>Language</i>		Mild	Mild	Mild	Mild	WNL	Severe	Severe	Mod	Mod	WNL	WNL
<i>Visuospatial skills</i>		WNL	WNL	WNL	WNL	WNL	Mild	Mild	WNL	WNL	WNL	WNL

Table 4
Number of baselines and counterbalanced order of category and typicality exposed in treatment. Also shown is a summary of acquisition (bench mark = 6.5) and generalization effect sizes (benchmark = 2.0) observed for trained items and untrained items.

<i>Number of baselines</i>	<i>Category Trained</i>	<i>Trained item acquisition</i>	<i>ES-target atypical</i>	<i>ES-other atypical</i>	<i>Untrained item generalization</i>	<i>ES-target typical</i>	<i>ES-other typical</i>
P1	3	1. Things at garage sale	20.5	17.6	Typical	9.8	21.4
P2	4	1. Things to take camping	7.6	0.6	Typical	3.6	0.5
P3	3	1. Things at garage sale	11	4.6	Typical	10.4	4
P4	14	2. Things at garage sale	14.3	1.7	Typical	2.4	2.6
P6	3	1. Things at garage sale	10	8.1	Typical	2.6	1.7

<i>Number of baselines</i>	<i>Category Trained</i>	<i>Trained item acquisition</i>	<i>ES-target typical</i>	<i>ES-other typical</i>	<i>Untrained item generalization</i>	<i>ES-target atypical</i>	<i>ES-other atypical</i>
P2	6	2. Things at garage sale	1.9	-1	Atypical	3	0.1
P4	5	1. Things to take camping	14.7	0*	Atypical	-0.05	0*
P5	3	1. Things to take camping	6.06	1.2	Atypical	1.2	0.9

* In these cases, effect sizes could not be calculated due the standard deviation in the baseline equaling zero.