Published in final edited form as:

J Speech Lang Hear Res. 2008 December ; 51(6): 1550–1568. doi:10.1044/1092-4388(2008/07-0038).

Typicality of Inanimate Category Exemplars in Aphasia Treatment: Further Evidence for Semantic Complexity

Swathi Kiran

Department of Communication Sciences and Disorders, University of Texas at Austin, Austin, TX
78712, USA

Abstract

Purpose—The typicality treatment approach on improving naming was investigated within two inanimate categories (*furniture* and *clothing*) using a single subject experimental design across participants and behaviors in five patients with aphasia.

Method—Participants received a semantic feature treatment to improve naming of either typical or atypical items within semantic categories, while generalization was tested to untrained items of the category. The order of typicality and category trained was counterbalanced across participants.

Results—Results indicated that two out of four patients trained on naming of atypical examples demonstrated generalization to naming untrained typical examples. One patient showed trends towards generalization but did not achieve criterion. Further, four out of four patients trained on typical examples demonstrated no generalized naming to untrained atypical examples within the category. Also, analysis of errors indicated an evolution of errors as a result of treatment, from those with no apparent relationship to the target to primarily semantic and phonemic paraphasias.

Conclusions—These results extend our previous findings (Kiran & Thompson, 2003) to patients with nonfluent aphasia and to inanimate categories such as *furniture* and *clothing*. Additionally, the results provide support for the claim that training atypical examples is a more efficient method to facilitating generalization to untrained items within a category than training typical examples (Kiran, 2007).

Naming therapies targeted at improving lexical retrieval in patients with aphasia have received extensive attention over recent years (Maher & Raymer, 2004; Nickels, 2002). Recently, an increasing number of studies have targeted treatment at the level of the naming impairment in individual patients. Naming deficits in aphasia can arise either from incorrect/incomplete activation of semantic or phonological nodes (Butterworth, 1989; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Foygel & Dell, 2000) or from a failure in the bi-directional link between them (Dell et al., 1997). Patients presenting with predominantly phonological errors may have a deficit in the phonological representation and often have concurrent deficits in real and nonword repetition (Caramazza, Papagno, & Rumel, 2000; Cuetos, Aguado, & Caramazza, 2000). Patients who demonstrate semantic errors devoid of co-existing semantic impairments may have difficulty accessing phonological representations from semantic representations (Caramazza & Hillis, 1990; Cuetos et al., 2000). Alternatively, presence of semantic errors may also suggest impairment at the semantic level (Hillis, Rapp, Romani, & Caramazza, 1990; Howard, 1984; McCleary & Hirst, 1986).

Consistent with the level of naming impairment, therapy tasks have either focused on facilitating access at the phonological or at the semantic level. In phonological treatments, tasks typically involve syllable judgment, rhyme judgment, word repetition and oral reading (Raymer, Thompson, Jacobs, & LeGrand, 1993; Wambaugh et al., 2001). In semantic treatments, tasks typically involve auditory and written word to picture matching tasks, answering yes/no questions about the target, spoken word categorization, relatedness judgment tasks and semantic attribute analysis (Boyle, 2004; Boyle & Coehlo, 1995; Davis & Pring, 1991; Howard, Patterson, Franklin, Orchid-Lisle, & Morton, 1985). In these studies, treatment has resulted in improvements on trained word; however, results of treatment studies examining generalization to untrained items have been mixed. Some studies have failed to show generalization to untrained items (Davis & Pring, 1991; Marshall, Pound, White-Thompson, & Pring, 1990; Pring, 1993). In contrast, other studies have been successful at facilitating generalization to untrained items (Boyle, 2004; Boyle & Coehlo, 1995; Drew & Thompson, 1999; Lowell, Beeson, & Holland, 1995), thereby illustrating that highlighting semantic attributes of trained items may be essential in facilitating generalization to items within a category (Drew & Thompson, 1999), and across semantic categories (Boyle, 2004; Boyle & Coehlo, 1995; Lowell et al., 1995).

In a previous study (Kiran & Thompson, 2003a) we employed a novel approach to facilitating lexical retrieval of trained and untrained items within a category in four patients with fluent aphasia. This study was based on a well-tested phenomenon in category representation in normal individuals, namely, typical examples of a category are processed faster and more accurately than atypical examples in a category. In a connectionist simulation examining relearning following damage within a computer network, however, Plaut (1996) showed that re-training atypical examples was more beneficial than training typical examples. The network was trained to recognize a set of artificial typical and atypical words (interpreted as comprehension), where typical words shared more of the semantic features of the category prototype (encoded as a set of binary values) than did atypical words. Once training was complete, the network was lesioned and retrained on either the typical items or the atypical ones. Plaut found that retraining atypical items resulted in improvements in recognition of typical items as well. However, training typical items improved performance only on trained items, while performance on atypical words deteriorated.

Plaut's simulation results were replicated by us during word retrieval in individuals with fluent aphasia (Kiran & Thompson, 2003a). Training spoken naming of atypical examples and their semantic features within two animate categories resulted in generalization to naming of intermediate and typical examples within each category. Training spoken naming of typical examples and their semantic features, however, did not result in generalization to the intermediate and atypical examples. These results presented a counterintuitive approach to facilitating lexical retrieval in patients with aphasia by manipulating exemplar typicality during treatment. We, therefore argued that atypical examples were more complex than typical examples within the category, and hence, generalization occurred from atypical examples to typical examples but not vice versa.

More recently, Stanczak, Waters and Caplan (2006) attempted to replicate the findings by Kiran and Thompson (2003a) in two patients with anomic aphasia. Stanczak et al. found that one of the two patients who was trained on atypical examples demonstrated generalization to untrained typical examples but this patient also showed marginally significant generalization from trained typical examples to untrained atypical examples. The second patient showed no learning of atypical examples of one category and no generalization from typical to atypical examples for the second category (Stanczak et al., 2006). While Stanczak's results generally support Kiran and Thompson's findings, they highlight the fact that not all participants with naming deficits respond to treatment the same way.

Our conceptualization of semantic complexity fits within the general framework of the Complexity Account of Treatment Efficacy hypothesis (Thompson, Shapiro, Kiran, & Sobecks, 2003). According to CATE, the basic principle of the complexity effect is that a subset relationship exists between the trained and untrained material in that greater generalization occurs when training items that encompass information relevant to untreated items (Thompson, 2007). Although this hypothesis is preliminary, evidence for the complexity effect comes from various strands of research including treatment for sentence production deficits in patients with agrammatic aphasia (Thompson & Shapiro, 2007), and in children with phonological deficits (Gierut, 2007).

The present study aimed to extend the examination of semantic complexity within animate categories to inanimate categories (*furniture* and *clothing*) as part of a broader effort to demonstrate that training atypical examples was a more efficient way to promote generalization within a category than training typical examples. A comprehensive theoretical account of semantic complexity is provided in Kiran (2007). Consequently, the applicability of this framework is elaborated within the context of the present experiment. It is hypothesized that representation of semantic attributes (or features) and lexical representations within a category is akin to a connectionist network consisting of nodes across two levels (semantic and phonological) that are linked through bidirectional connections (Dell et al., 1997). Each category (e.g., *furniture*) consists of exemplars represented at the basic level (e.g., *chair*, *dresser*, *hammock*) all of which comprise a set of *core features*, those that are required for category membership (e.g., *comes in different shapes/sizes*, *found in homes*). Apart from that, the category consists of a central prototype, or the idealized set of features (e.g., *heavy*, *set on floor*). Typical examples within the category possess more *prototypical features* (e.g., *heavy*, *set on floor*) and fewer distinctive features (e.g., *used outside*, *kids furniture*). Also typical examples have a number of shared/intercorrelated features with other typical examples (e.g., *made of wood* and *heavy* are shared by *sofa*, *dresser* and *table*). Therefore, it was hypothesized that these features carry less weight within the category as they are shared by a number of other typical examples (see Hampton, 1993; 1995).

Atypical examples (e.g., *hammock*, *bean bag*), however, consist of core (e.g., *comes in different shapes/sizes*, *found in homes*) and *distinctive features* (*used outside*, *kids furniture*) that presumably carry more weight in their representation within the category. Also, as a group, features belonging to typical examples have a subset relationship with those of atypical examples. That is, atypical examples consist of a wider range of features (e.g., *found in home*, *decorative accessory*, *needs electricity*) that inherently include features relevant to typical examples. The evidence that atypical examples are processed slower than typical examples during category verification tasks (Kiran, Ntourou, & Eubanks, in press; Kiran & Thompson, 2003b; Rosch, 1975; Smith, Shoben, & Rips, 1975) further illustrate that atypical examples are more complex than typical examples (for a similar proposal equating processing time with complexity see Gennari & Poeppel, 2003).

The fundamental assumption of treatment is that strengthening access to semantic attributes results in facilitation of target semantic nodes at the semantic level which cascades downstream to the phonological representations, thereby strengthening phonological nodes as well. Also, enhanced access to target semantic representations facilitates semantically related neighbors, which consequently, results in facilitation of corresponding phonological representations. Because atypical examples and their features are presumed to represent a greater variation of semantic features, strengthening access to atypical examples also strengthens features relevant to typical examples, thereby facilitating phonological access to both typical and atypical examples. Conversely, typical examples and their features do not influence features relevant to atypical examples and therefore, phonological representations specific to typical examples

only will improve. Consequently, when typical examples are targeted in treatment, atypical examples are not accessed until directly targeted in treatment.

The present study examined inanimate categories as there is extensive evidence documenting the dissociation between animate and inanimate categories in their representation and processing subsequent to brain damage (Forde & Humphreys, 1999; Moore & Price, 1999). Further, typicality appears to be determined differentially across animate and inanimate categories in that inanimate categories show greater typicality effects than animate categories in normal individuals (e.g., *rug* is more likely to be judged a partial member of *furniture* than *tomato* is judged a partial member of *fruit*) (Diesendruck & Gelman, 1999; Estes, 2003).

Finally, another aspect of the present study was the inclusion of patients with nonfluent aphasia/apraxia in addition to patients with fluent aphasia. Whereas all five patients presented with breakdown in lexical retrieval at either the semantic level and/or the phonological level, two of these individuals presented with additional impairments downstream at the motor programming/planning problem as indicated by their apractic errors. The aim of the study was to examine the effect of a semantic based treatment on lexical access and understand whether the selective generalization patterns from atypical to typical examples were observed in these patients also. Finally, the nature of naming errors occurring throughout treatment was also examined. Within the theoretical framework described above, it was predicted that patients would be unable to access any specific information about target items, resulting in predominately neologistic errors, unrelated words, or no responses before initiation of treatment. The semantically based treatment was expected to facilitate improved access to semantic and phonological approximations of target words. Following treatment a greater proportion of semantic and/or phonemic errors was expected.

Methods

Participants

Five monolingual, English speaking individuals with aphasia recruited from local hospitals within the Austin area participated 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 seven 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.

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 50% accuracy (See Table 1). Another criterion for inclusion was performance lower than 85% on two or more subtests across the *Psycholinguistic Assessment of Language Processing in Aphasia* (PALPA) (Kay, Lesser, & Coltheart, 1992) and the *Pyramids and Palm Trees* (PAPT) (Howard & Patterson, 1992). Impairment in semantic processing was hypothesized to be integral to the success of treatment since the principle component of treatment focused on explicit manipulation of semantic information (i.e. semantic features)(see Table 2). Written naming was tested to examine if lexical retrieval impairments were limited to spoken output or across output modalities. Single word oral reading, single word repetition and written spelling were tested to measure phonological processing abilities.

The diagnosis of aphasia was determined by administration of the *Western Aphasia Battery* (WAB) (Kertesz, 1982). Results showed that participants 1-3 presented with language characteristics consistent with fluent aphasia whereas participants 4 and 5 presented with nonfluent aphasia and apraxia (See Table 1 for details). All participants except P4 were also administered the Cognitive Linguistic Quick Test (CLQT) (Helm-Estabrooks, 2001) which was acquired as part of another experimental protocol. P4 did not meet inclusionary criteria for the protocol. Scores on this task indicated all participants exhibited deficits in the memory and language domains, both of which contain a significant language component in the stimuli (See Table 1). Finally, the Apraxia Battery for Adults (ABA) (Dabul, 1979) was administered to P4 and P5 to assess the level of co-existing apraxia (see Table 3). Performance on this test indicated both participants presented with mild-moderate severity of apraxia, specifically on increasing word lengths and when utterance times for responses were measured.

To assist in development of norms for stimuli employed in the study, 20 young (range = 21 - 40 years) and 20 older individuals (range = 41 - 75 years) were recruited from Northwestern University and the Evanston community (Kiran, 2002). All participants had normal or corrected to normal vision, normal hearing, and had at least a high school degree. Exclusionary criteria included history of neurological disorders, psychological illnesses, alcoholism, learning disability, seizures and attention deficit disorders.

Stimuli

Development of typicality rankings—Ten young and 10 older participants were provided with a list of 12 superordinate category labels (vegetables, transportation, weapons, tools, clothing, furniture, sports, animals, fruits, birds, occupations and musical instruments) (Rosch, 1975; Uyeda & Mandler, 1980) and were asked to write down as many basic level examples as they could think of for each category. Following completion of this task, a list with items for each superordinate category was then given to another group of 20 participants (10 young and 10 older individuals). Using instructions developed by Rosch (1975), participants were asked to rate on a 7-point scale (1 indicating a good example, 7 indicating a poor example), the extent to which each example represented their idea or image of the category term (typicality). Mean average ratings and standard deviations were calculated for each example in the category.

Development of treatment categories and their examples—For the present experiment, two inanimate categories (*clothing*, *furniture*) were chosen from the above set based on three criteria; (a) the category contained at least 45 examples, (b) atypical items did not overlap across categories and (c) there was a relatively equal distribution of typical and atypical examples. Several additional criteria were utilized to eliminate problematic examples within categories. For instance, examples that at least 60% (12/20) of the participants marked as unfamiliar (*U*) were eliminated. Also eliminated were (a) those examples whose average typicality rating occurred with a standard deviation greater than 2, (b) alternate meanings for the same word (e.g., *pantyhose* and *stockings* for *clothing*), (c) examples that were both atypical and unfamiliar (e.g., *étagère* for *furniture*), (d) examples that lacked any salient features (e.g., *credenza*), and (e) examples that were questionable members (e.g., *plants* for *furniture*).

In order to normalize the average ratings across participants, z scores were calculated for the average ratings (across 20 participants) for each item within the two categories. For the category *furniture*, the z values were -1.37 to -0.42 (typical) and 1.12 to 0.41 (atypical). For *clothing*, the z scores ranged from -1.22 to -0.44 (typical) and from -0.01 to 0.05 (atypical) examples. Stimuli were controlled for written word frequency (Frances & Kucera, 1982), familiarity and imageability (MRC Psycholinguistic Database, Coltheart, 1981; http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.html), and number of syllables (see

appendix A for a list of stimuli). Separate 2 (typicality: typical, atypical) $\times$ 2 (category: *clothing*, *furniture*) ANOVAs, performed on the variables revealed nonsignificant effects for all variables.

The selected pictures were presented to a group of 10 normal individuals who were required to name the pictures with 80% agreement (examples of acceptable alternatives included *picture/painting*). Color photos for each stimulus were downloaded from the internet (<http://images.google.com/>) and printed on 4 $\times$ 6 inch cards. Photos were screened for visual complexity; only photos with the target picture in the center on a contrast black or white background were selected. Examples from other categories (*fruits*, *body parts* and *musical instruments*) were selected to serve as distracters during treatment. Thus, there were two treatment categories with 30 examples each, and three distracter categories with 15 examples each.

Development of semantic features for treatment—In each category, thirty features that were either physical (descriptions regarding physical appearance, Ahn, 1998) (e.g., *has shelves, made of cotton*), functional (descriptions regarding use or applications, e.g., *used for sleeping, worn on special occasions*), characteristic (descriptions that were conveyed salient information about an example e.g., *needs power to work for furnace, decorative accessory for bandana*) or contextual (descriptions referring to a spatial location, e.g., *found in hallway, buy at clothing store*) were selected from published norms (Barr & Caplan, 1987) and by looking up specific information for each example in the internet. Only features that 18/20 young and elderly participants marked as being features of the category were selected. Fifteen of these features were applicable to all items in the category (core features, e.g., *clothing: buy at clothing store, furniture: buy at furniture store*). Fifteen others were relevant to both typical and atypical examples (prototypical features, e.g., *chair, hammock: used for sitting*) or that were specific to atypical examples (distinctive features, e.g., *furniture: decorative accessory*). The main difference between core, prototypical and atypical features is that core features are relevant to all examples whereas prototypical features are relevant to most typical examples and some atypical examples. In contrast, atypical features mostly consisted of distinctive features specific to one or more atypical examples. Finally, 20 distracter features belonging to the categories sports, transportation, animals, insects, flowers and weapons (e.g., *made of petals, found in a crime scene*) were selected and were evenly distributed across attribute types (e.g., physical, functional, contextual, characteristic).

Design

A single subject experimental design (Connell & Thompson, 1986; McReynolds & Kearns, 1983) was used to examine acquisition of trained items and generalization to untrained items within and across categories. The number of baseline sessions, the order of categories trained and typicality of stimulus sets within each category were counterbalanced across participants (see Table 4), and consequently, allowed for examination of differential responsiveness to typical or atypical training within the same participant. For all participants, one set of items ($N = 15$) within a category (either typical or atypical) was introduced into treatment, while the untrained items within the trained category ($N = 15$) and all examples of the untrained category ($N = 30$) remained in baseline. This way, items from the untrained category served as a control set allowing inspection for any unexpected changes. For all participants, two baseline probes were acquired for the untrained (second) semantic category prior to its treatment following Horner and Baer (1978). In the previous study (Kiran & Thompson, 2003a) treatment was shifted to the untrained items within the trained category if no generalization was observed. In the present study, however, this protocol was only followed for participant 1. For the remaining participants, treatment was only focused on one set of examples within a category to reduce any fatigue resulting from prolonged exposure to the same set of category examples. It should

be noted that Participant 2 performed at 70% accuracy during baseline naming of typical *clothing* hence, treatment was not provided for the second category. For participant 5, treatment was terminated after one category as she expressed fatigue following 24 weeks of treatment.

Baseline naming procedures

Confrontation naming of all 60 items (30 examples from each category) was tested during baseline. Participants were shown each picture (presented in random order) and were instructed to name the *clothing* or *furniture* depicted (e.g., *Please name this piece of clothing*). Responses were considered correct if they were self corrected responses, dialectal differences, distortion/omission/substitution of one vowel or consonant (e.g. *hemet/helmet*) of the target item. Feedback as to accuracy of response was not given during baseline, however, intermittent encouragement was provided. All other responses including were classified into (a) No response/I don't Know (IDK), (b) unrelated word or visual errors (e.g., *research/overalls*, *horse/toybox*), (c) neologisms (utterances with less than 50% phonetic overlap with the target) (e.g., *perchers/pajamas*), (d) perseverations, defined as four or more repetitions of the same phoneme string within a probe session, (e) circumlocutions (defined as multiword responses with relevant semantic information, e.g., *when he went to the moon/flightsuit*), (f) superordinate label (e.g., *furniture/bed*), (g) semantic paraphasias (e.g., *electric/furnace*, *endtable/nightstand*), (h) phonemic paraphasias (utterances with greater than 50% phonetic overlap with the target) (e.g., *shamas/pajamas*) and (i) mixed semantic/phonemic errors (e.g., *mug/earmuffs*). Percent correct named as well as the percentage of each error type relative to all errors was calculated.

Treatment

Participants were treated consecutively. Treatment was conducted two times per week for two hours. During each treatment session, participants performed the following steps for each of the 15 examples of the subset: a) naming the picture, b) sorting pictures by category, c) identifying semantic attributes applicable to the target example from a set of category features, and d) answering yes/no questions pertaining to the semantic features of the target item (see appendix B). Both orthographic and phonological information were provided for the trained items.

Treatment probes

Throughout treatment, naming probes like those used in the baseline condition were presented to assess naming of the trained and untrained items. Naming probes for all 30 items of the category in training were administered prior to every second treatment session. The order of presentation of items was randomized during each probe presentation. An apriori criterion for termination of treatment was set at 80% accuracy (12/15) for two consecutive sessions or a total of 20 treatment sessions (10 probe sessions). However, treatment was extended beyond this criterion for P1, P3 and P5 in order to examine if trends in the data were maintained. Generalization to naming of untrained examples was considered to have occurred when performance accuracy improved by 40% over the maximum baseline levels. This criterion has been used by us in previous studies (Kiran, 2005; Kiran & Thompson, 2003a) and in conjunction with effect size calculation, allows a uniform comparison of generalization effects across our treatment studies. Further, this criterion is especially useful during visual inspection of generalization data when there are positive trends but the slope and level of these trends are not sufficient to draw conclusions on whether generalization had occurred.

The probe protocol was modified for participants 4 and 5 when these participants did not demonstrate improvements on the trained items after the specified number of sessions (see results for details). For participant 4, after 9 treatment sessions, probes were modified to incorporate written responses as acceptable responses. Scoring protocol for written targets

followed our previous work in writing therapy (Kiran, 2005). Briefly, a response was counted as correct when (a) the letters were clear and legible, and (b) one letter was substituted (e.g., *blousd* for *blouse*), transposed (e.g., *betl* for *belt*) or omitted (e.g., *banana* for *bandana*). All other responses were scored as errors and coded using the criteria described above.

Participant 5 was allowed to write responses to target probes from the inception of baselines, although this modification had no apparent effect on facilitating lexical retrieval. Hence, after 10 probe sessions, this participant was provided with the initial phoneme of the target word (e.g., *bed*: /b/; *chandelier*: /sh/) for each probe item (trained and untrained). No feedback was provided regarding accuracy of word retrieval. It should be noted that these modifications do not confound the interpretation of the results of the study since (a) no modifications were made to the treatment protocol for either patient, (b) both trained and untrained items were subjected to the modified probe protocol, and (c) the period prior to the introduction of the modification served as extended baselines for assessment of performance.

Data Analysis

Effect sizes (ES) were calculated comparing the mean of all data points in the treatment phase relative to the baseline mean divided by the standard deviation of baseline (Busk & Serlin, 1992). Based on comparable naming treatment studies in aphasia, an ES of 4.0 was considered small, 7.0 was considered medium and 10.1 was considered large (Beeson & Robey, 2006). McNemar tests were administered to inspect changes on the error analysis for each category for each participant. A nonparametric Spearman rank correlation was performed to examine the relationship between improvements in naming trained/untrained items and on standardized language tests.

Reliability

All the baseline and probe sessions were recorded on audiotape and 50% of the responses were also scored on-line by both the clinician and by an independent observer seated behind a one-way mirror. Point-to-point agreement was 95% across probe sessions. Daily scoring reliability checks by the independent observer were undertaken to ensure accurate presentation of the treatment protocol by the clinician. Point-to-point agreement ranged from 90-100%. Error analysis on the data was conducted by one independent scorer blind to purposes of the study. Twenty-five percent of the errors were randomly selected and categorized into the corresponding subtypes by the author. Inter-rater reliability was 100%.

Results

Naming accuracy

Results are presented in Figures 1, 2, 3, 4 and 5 in multiple baseline formats showing the percent accuracy (out of 15 items) for each subset (typical and atypical) within each category. Data are presented for baseline, treatment, and follow-up phases of the experiment. All participants demonstrated stable baselines (criterion of no more than 30% fluctuation across baselines) (Kiran & Thompson, 2003, Edmonds & Kiran, 2006) for the trained items.

Participant 1—Following three initial baselines, Participant 1 first received treatment for atypical examples of *clothing*, which improved to criterion (high of 93% accuracy; ES = 16.21) within eight weeks as generalization to the untrained typical examples was observed to a high of 87% accuracy (ES = 3.87). Following two baseline sessions for the second category, treatment was shifted to typical examples of *furniture* which also improved to criterion (80% accuracy; ES = 2.87) within five weeks. Improvements on the untrained atypical examples were not observed (high of 33% accuracy; ES = 0). For this participant treatment was then shifted to atypical examples of *furniture* which improved to criterion (80% accuracy; ES =

15.9), whereas performance on the previously trained typical examples was maintained ($ES = .5$). Follow up probes for the first treatment category conducted approximately 5 weeks after termination of treatment for that category revealed naming performance higher than initial baseline levels (see Figure 1).

Participant 2—Following three baselines, Participant 2 received treatment for typical examples of *furniture* which improved from 47% to a high 87% accuracy within nine weeks ($ES = 1.8$). Performance on the untrained atypical examples did not change appreciably from baseline levels (40% to 53% accuracy; $ES = 1.8$) indicating no generalization to these items (see Figure 2).

Participant 3—Following five baseline sessions, Participant 3 received treatment on atypical examples of *furniture* (See Figure 3). While this patient achieved accuracy at or above 80% on the trained items on two separate occasions ($ES = 9.7$), performance was not maintained for two consecutive sessions. Hence, treatment for this category was terminated after 20 weeks (66% accuracy on the final session). Generalization to untrained typical examples did not meet criterion ($ES = 1.13$). When treatment was shifted to typical examples of *clothing*, performance improved to a high of 80% accuracy in 9 weeks. Once again, however, performance declined after criterion was achieved and treatment was terminated after 11 weeks ($ES = 1.8$). No generalization to the untrained atypical examples of *clothing* was observed for this participant (high of 33% accuracy; $ES = -5.7$). Follow up probes were conducted 10 weeks and 18 weeks after termination of treatment for the first category which indicated a lack of maintenance of treatment effects for the trained items. Follow up probes on the second category were conducted approximately 8 weeks after treatment indicating maintenance of typical and atypical item accuracy at or above baseline levels.

Participant 4—Following five baseline sessions, participant 4 received treatment on typical examples of *clothing*. Performance on these items remained unchanged until a modification in the probe protocol was incorporated at 9 weeks to allow written responses as acceptable responses (see Figure 4). At this point, performance on the trained typical examples increased to criterion in 6 weeks ($ES = 3.6$), whereas performance on the untrained atypical examples increased from 20%-40% accuracy ($ES = 1.13$), below the criterion for generalization. Because the modification in treatment protocol was instituted midway during treatment and no baselines were obtained for written naming performance, the results of acquisition and generalization are interpreted with caution. For the second category (*furniture*), two baselines on written naming performance were obtained prior to initiation of treatment. Upon treatment of atypical examples of *furniture*, performance of trained items improved to criterion in 13 weeks ($ES = 8.35$), and generalization to untrained typical examples was also observed ($ES = 4.20$). Follow up probes for this participant was not conducted due to transportation and scheduling issues.

Participant 5—Following three baseline sessions, participant 5 received treatment for atypical examples of *furniture*. Despite allowing written responses for target probes, performance did not improve for trained items although a trend was noted. Hence, after 10 weeks of minimal acquisition of trained items, the participant was provided with only the initial phoneme for each probe item (typical and atypical) (see Figure 5). Performance was initially variable although eventually, participant 5 demonstrated 100% accuracy on naming of the trained items in 11 weeks (effects sizes are invalid due to the lack of variation during baseline). Performance on untrained typical examples remained unchanged during the initial phase of treatment although, when the initial phoneme cue was provided, limited improvement was observed to untrained typical examples (7% to 40%; $ES = 1.89$). Even though baselines with phonemic cued naming were not obtained, given the slow acquisition and maintenance of performance at follow up, we are reasonably confident that improvements observed on the

untrained typical examples were not due to other extraneous factors influencing performance outcome. Treatment was not continued for the second category per the participant's desire to terminate treatment.

Evolution of errors

For each participant, errors produced during baseline sessions and equal number of sessions at the end of treatment was compared across the two categories (see Table 5). All participants showed significant decrease in errors as indicated by McNemar tests. Participant 1 showed a reduction in the number of circumlocutions for *furniture* whereas the proportion of semantic and phonemic errors increased as a result of treatment. Participant 2 showed a reduction in the proportion of neologisms with a corresponding increase in the proportion of semantic errors. For participant 3, no clear trends emerged in the evolution of errors as the proportion of perseverations and unrelated words fluctuated across the two categories. Participant 4 showed a dramatic reduction in the proportion of perseverations, whereas the proportion of semantic and phonemic errors increased in both categories as a result of treatment. Finally, for participant 5, a slight reduction in the proportion of IDK/no responses and perseverations were replaced with an increase in semantic errors at the end of treatment.

Pre-post standardized language measures

All tests administered prior to initiation of treatment were readministered upon completion of treatment and are shown in Tables 2 and 3. Without a control group that did not receive treatment, it is difficult to ascertain if changes observed in the present study are due to repeated exposure to test items. Therefore, the improvements on the standardized tests are reported here but no interpretations are drawn regarding significant changes. To further understand the relationship between improvements on items in treatment and changes on standardized tests, improvements in naming of trained and untrained items following items were correlated with improvements on standardized tests reported in Table 2. For the purpose of this analysis, procedures followed are similar to that reported by Hickin et al. (2002). Specifically, improvements in naming trained items was calculated by subtracting the average of baseline performance for trained items in both categories (only one category for P2 and P5) from the average of the final treatment in both categories (only one category for P2 and P5). The same formula was applied for improvements in naming untrained items. As an example for P3, improvement in naming trained items was calculated as [AVERAGE Final probe (Atypical Furniture, Typical Clothing)] – [AVERAGE Baseline (Atypical Furniture, Typical Clothing)]. For the same patient, improvement in naming untrained items was calculated as [AVERAGE Final probe (Typical Furniture, Atypical Clothing)] – [AVERAGE Baseline (Typical Furniture, Atypical Clothing)]. Then, percent change on trained items and untrained items was correlated with percent change on the standardized tests (except WAB AQ) described in Table 2 using a nonparametric Spearman R test for ranks. The results show that improvement on trained items correlated with improvements on PALPA single word reading ($r_s = .90$; $N = 5$, $p < .05$), PALPA auditory synonym judgment ($r_s = .97$; $N = 5$, $p < .05$), and PAPT three pictures subtests ($r_s = .90$; $N = 5$, $p < .05$). Importantly, improvements on trained words correlated with improvements on untrained words ($r_s = .90$; $N = 5$, $p < .05$), confirming the findings of positive generalization reported. The remaining correlations between trained items and other subtests as well as between untrained items and the standardized tests were weak and did not approach significance.

Discussion

This experiment was undertaken for two reasons: (a) to establish the effectiveness of the typicality treatment approach in facilitating lexical retrieval and generalization in individuals with naming deficits, and (b) to investigate the relevance of semantic complexity as a treatment

variable to understanding language recovery patterns in aphasia. A previous study (Kiran & Thompson, 2003a) showed 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 inanimate categories across five individuals with fluent/nonfluent aphasia was examined. Recall that in the introduction, we hypothesized that strengthening access to semantic attributes and phonological representations for target atypical examples will facilitate access to these items as well as to corresponding 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. Results revealed that reinforcing semantic features associated with atypical examples resulted in generalization to untrained typical examples in participant 1 when trained on atypical examples of *clothing* and participant 4 when trained on atypical examples of *furniture*. It should be noted that P5 showed a trend toward generalization (from 7% to 40% accuracy) but did not reach criterion. In contrast, training typical examples did not result in generalization to untrained atypical examples in participants 1 and 2 when trained on typical examples of *furniture*, and participants 3 and 4 when trained on typical examples of *clothing*. Since the experimental design allowed the examination of differential responsiveness to typical or atypical training within the same participant, the present results provide evidence for the beneficial effects of training atypical examples (instead of typical examples) within a category. These results are consistent with a growing body of evidence suggesting that semantic based treatment that emphasizes the explicit analysis of semantic attribute information is a successful approach for improving naming skills and facilitating generalization (Boyle, 2004; Boyle & Coehlo, 1995; Coehlo et al., 2000; Drew & Thompson, 1999; Lowell et al., 1995).

Second, the present results indicate that the effect of manipulating typicality as a treatment variable to examine semantic complexity in animate categories (Kiran & Thompson, 2003) extends to inanimate categories as well. The mechanism underlying the selective generalization patterns from trained atypical examples to untrained typical examples is likely the same across animate and inanimate categories. Specifically, training atypical examples highlight the featural variation within the categories and consequently improve typical examples. In contrast, training the features associated with typical examples has no influence on atypical examples of the category. While the present study examined complexity within the lexical-semantic domain, these results further reinforce the applicability of complexity as a viable treatment approach across various aspects of language impairment (Thompson, 2007).

Furthermore, the beneficial effects of semantic based treatment extended beyond improvements in naming of trained and untrained examples within each category to changes in errors. These results fit within the theoretical framework discussed in the introduction where strengthening semantic representations facilitates access to specific target nodes at the semantic and phonological level. With treatment, activation of random nodes or diffuse multiple nodes (manifested as no responses, perseverations and circumlocutions) are replaced by specific nodes that are semantically or phonologically related to the target. This claim is consistent with theoretical models of naming impairment which suggest that nonwords, formal paraphasias and no responses tend to occur in patients with more severe naming deficits whereas semantic and mixed errors arise independent of naming severity (Dell et al., 1997; Schwartz & Brecher, 2000). In addition, with recovery, both severe and less severe patients exhibit an increase in semantic errors. The present data provide empirical clinical evidence regarding the effect of treatment in promoting a similar transition from nonspecific information to specific semantic and phonemic information about the target and are resonant with other studies examining evolution of errors over time (Basso, Corno, & Marangolo, 1996; Edmonds & Kiran, 2006; Jokel, Rochon, & Leonard, 2004; Kiran & Thompson, 2003a; Raymer et al., 2000).

To further understand the relationship, if any, between improvements observed in treatment and changes on standardized tests, a nonparametric correlation analysis revealed that improvement on naming of trained items during therapy was associated with improvements on single word reading and two measures of semantic processing, (a) auditory synonym judgment on the PALPA and (b) three pictures test on PAPT. One possible explanation for this finding comes from the fact that specific steps employed in treatment such as judging the relationship between pictures (Step 1: category sorting) reading written word cards (Step 3: feature selection), and judging auditorily presented semantic features (Step 3: Yes/No questions) resulted not only in improvements on naming of trained items but also translate to improvements in oral reading and semantic processing. It should be noted however, that the scope of such broad changes should theoretically also extend to improvements on the untrained categories. No improvements in the untrained category was observed until directly trained. One possible explanation for this finding is that assessment of the untrained categories was only done through picture naming which may be a more difficult task for patients than oral reading or semantic triplet judgment. Any explanations proposed at this point, however, are purely speculative and would require future systematic examinations.

A limitation in the interpretation of the results is that the predictions for selective generalization patterns are not equally borne out across all participants. Other related work has also revealed such variable patterns (Kiran, 2005). The present results are unambiguous for three participants (P1, P2 and P4) whereas as results for P3 and P5 are not completely aligned with the predictions. What sets apart these patients from P1 and P4 who show strong generalization patterns is unclear and requires further examination. Despite P3 being similar to participants 1 and 2 in their language profile, this participant did not demonstrate generalization from the trained atypical examples to the untrained typical examples for the category *clothing*. Also, this participant did not demonstrate the expected trends in evolution of errors. P5 showed a severe language deficit prior to inception of treatment and also showed limited benefit from the semantic based treatment. Moreover, this participant's treatment outcome may also be influenced by the provision of phonemic cues. Both participants 3 and 5 performed in the severe range on all tests of CLQT, especially on the executive function portion. Several recent studies have shown that patients who present with concurrent cognitive deficits show lesser benefits of language therapy than patients without cognitive impairments (Goldenberg, Dettmers, Grothe, & Spatt, 1994; Murray, Ballard, & Karcher, 2004). Further, Helm-Estabrooks (Helm-Estabrooks, 2002) suggests that individual differences in response to language therapy may be attributable to various aspects of a patient's cognitive abilities. While it may be premature to assume a relationship between performance in treatment and on the CLQT, the present study at least underscores the value of obtaining a complete cognitive-linguistic profile for each participant.

Recall that one aim of the study was to examine the effect of treatment in nonfluent participants who presented with co-existing apraxia. Participants 4 and 5's impaired performance on the semantic processing and naming tasks qualified them for the present treatment program. Nevertheless, the treatment protocol was ineffective in facilitating acquisition of trained items until it was altered to accommodate for their concurrent apractic impairments. Therefore, it appears that the typicality treatment in its pure form was less effective for nonfluent patients than fluent patients. That these participants demonstrated changes on standardized semantic processing tests similar to the three fluent patient counterparts do indicate that assigned modifications in treatment were beneficial to these participants in terms of their lexical retrieval abilities. These results suggest the need for future work that examines why some individuals are sensitive to the complexity effect, while others are not.

Conclusions

The present data indicate that the typicality based semantic treatment resulted in acquisition of trained atypical examples and generalization to untrained typical examples for some patients with aphasia. In contrast, the treatment facilitated improvements on trained typical examples but no generalization to untrained atypical examples was observed. These findings have important clinical implications. Given the current healthcare environment which restricts treatment to a limited number of sessions, a naming therapy that promotes optimal generalization patterns is ultimately more beneficial than those that do not facilitate generalization. The present data shows that the typicality based naming treatment is applicable for some patients with fluent and nonfluent aphasia.

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 Joyce Harris, Shilpa Shamapant and Lisa Edmonds for their assistance during various stages of the project, and Sarah Key DeLyria, Karen Abbott, Kimberly Mellen, Heather Wagner and Maitreyi Vishwanathan for their role in data collection and error analysis. Finally, the author thanks the participants in the experiment for their patience and cooperation. Contact author: Swathi Kiran, Ph.D. CCC-SLP, A1100, 1 University Station, Department of Communication Sciences and Disorders, University of Texas, Austin, Texas 78712. Email: s-kiran@mail.utexas.edu

Appendix

Appendix A

Stimuli used in treatment

	Clothing		Furniture	
	Typical	Atypical	Typical	Atypical
	Pants	Tie	Chair	Trunk
	Jeans	Rainwear	Sofa	Mirror
	Blouse	Gloves	Bed	ToyBox
	Sweater	Belt	Curio	Drapes
	Skirt	Flightsuit	Desk	Chandelier
	Suit	Cape	Dresser	Umbrella stand
	Shorts	Suspenders	Coffee Table	Porch Swing
	Jacket	Hood	Loveseat	Rug
	Overalls	Earmuffs	BookCase	Picture
	Vest	Helmet	EndTable	WasteBasket
	Sweatsuit	Garter	NightStand	Hammock
	Underwear	Apron	Recliner	Pillow
	Pajamas	Bib	Lamp	Furnace
	Socks	Bandana	Cabinet	Towelrack
	Shirt	Tights	FootStool	BeanBag
Average typicality	1.7 (.47)	4.1 (.73)	1.7 (.53)	5.4 (.45)
Average WWF	11.6 (10)	6.5 (10)	24.2 (38)	20.2 (10)
Average Familiarity	558 (45)	487 (82)	546 (79)	549 (77)
Number of syllables	1.6 (.82)	1.6 (.72)	2.0 (.88)	2.0 (.82)
Average imageability	591 (31)	536 (55)	572 (42)	564 (51)

Clothing		Furniture	
Typical	Atypical	Typical	Atypical

WWF: Written word frequency. Standard deviations are reported in parenthesis.

Appendix B: Treatment protocol

The treatment protocol for each target item was as follows:

1. Category Sorting

The examiner placed written category cards (*clothing/furniture, fruits, musical instruments*) on the table in random order. The examiner then randomized the 60 pictures and presented them one at a time for the participant to sort by superordinate category, by placing each picture on its written category card. If incorrect, the examiner placed the picture under the accurate category label.

2. Picture naming

The participant was presented with the picture and was asked to name it. Participants were provided with the verbal label if the patient was unable to retrieve the name. Irrespective of accuracy, the participant was guided through the next steps.

3. Feature Selection

The examiner placed the target picture (e.g., *sweater*) in the center of the table and provided the participant with approximately 40 written semantic feature cards belonging to the target category. The participant was then required to select the first six features that were pertinent to the target example. For example, for *sweater*: *made of fabric*, *keeps warm* were acceptable semantic features, while *decorative accessory*, and *worn on feet* were features that were not applicable. Once the participant selected six features, he/she was required to read aloud the selected features.

4. Yes/No Questions

The experimenter then asked the patient 15 questions about the target example (e.g., *sweater*), which included five acceptable semantic features, five unacceptable semantic features from the same category and five semantic features from a different category. The patient had to respond yes or no. Feedback regarding accuracy was provided.

5. Picture naming

Same procedure as Step 2.

References

- Ahn WK. Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition* 1998;69:135–178. [PubMed: 9894403]
- Ashcraft MH. Property dominance and typicality effects in property statement verification. *Journal of Verbal Learning and Verbal Behavior* 1978;17(2):155–164.

- Barr RA, Caplan LJ. Category representations and their implications for category structure. *Memory and Cognition* 1987;15(5):397–418.
- Basso A, Corno M, Marangolo P. Evolution of oral and written confrontation naming errors in aphasia. A retrospective study on vascular patients. *Journal of Clinical and Experimental Neuropsychology* 1996;18(1):77–87. [PubMed: 8926299]
- Beeson PM, Robey RR. Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychological Review* 2006;16(4):161–169.
- Boyle M. Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech Speech-Language Pathology* 2004;13(3):236–249.
- Boyle M, Coehlo C. Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology* 1995;4:94–98.
- Butterworth, B. Lexical access in speech production. In: Wilson, WM., editor. *Lexical representation and process*. Cambridge, MA: MIT Press; 1989. p. 108–235.
- Caramazza A, Hillis AE. Where do semantic errors come from? *Cortex* 1990;26(1):95–122. [PubMed: 2354648]
- Caramazza A, Papagno C, Rumel W. The selective impairment of phonological processing in speech production. *Brain and Language* 2000;75(3):428–450. [PubMed: 11112296]
- Connell PJ, Thompson CK. Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders* 1986;51(3):214–225. [PubMed: 3525988]
- Coltheart M. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* 1981;33A:497–505.
- Cuetos F, Aguado G, Caramazza A. Dissociation of semantic and phonological errors in naming. *Brain and Language* 2000;75(3):451–460. [PubMed: 11112297]
- Dabul, B. *Apraxia Battery for Adults*. Tigard, OR: C.C. Publications; 1979.
- Davis A, Pring T. Therapy for word-finding deficits: More on the effects of semantic and phonological approaches to treatment with dysphasic patients. *Neuropsychological Rehabilitation* 1991;1(2):135–145.
- Dell GS, Schwartz MF, Martin NM, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 1997;104:801–838. [PubMed: 9337631]
- Diesendruck G, Gelman SA. Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review* 1999;6(2):338–346. [PubMed: 12199220]
- 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.
- Edmonds LA, Kiran S. Effect of semantic based treatment on cross linguistic generalization in bilingual aphasia. *Journal of Speech, Language and Hearing Research* 2006;49(4):729–748.
- Estes Z. Domain differences in the structure of artifactual and natural categories. *Memory and Cognition* 2003;31(2):199–214.
- Forde EM, Humphreys GW. Category specific recognition impairments: A review of important case studies and influential theories. *Aphasiology* 1999;13(1):169–193.
- Foygel D, Dell GS. Models of impaired lexical access in speech production. *Journal of Memory and Language* 2000;43(2):182–216.
- Frances, N.; Kucera, H. *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin; 1982.
- Gennari S, Poeppel D. Processing correlates of lexical semantic complexity. *Cognition* 2003;89(1)
- Gierut J. Phonological complexity and language learnability. *American Journal of Speech Language Pathology* 2007;16:6–17. [PubMed: 17329671]
- Goldenberg G, Dettmers H, Grothe C, Spatt J. Influence of linguistic and nonlinguistic capacities on spontaneous recovery of aphasia and on success of language therapy. *Aphasiology* 1994;8:443–456.
- Goodglass, H.; Kaplan, E.; Weintraub, S. *Boston Naming Test*. Philadelphia: Lea & Febiger; 1983.
- Helm-Estabrooks, N. *Cognitive Linguistic Quick Test*. London, England: Harcourt Assessment; 2001.
- Helm-Estabrooks N. Cognition and aphasia: a discussion and a study. *Journal of Communication Disorders* 2002;35(2):171–186. [PubMed: 12036150]

- Hillis AE, Rapp BC, Romani C, Caramazza A. Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology* 1990;7(3):191–243.
- Hickin J, Best W, Herbert R, Howard D, Osborne F. Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology* 2002;16(1011):981–999.
- Horner DR, Baer DM. Multiple probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis* 1978;11(1):189–196. [PubMed: 16795582]
- Howard D, Orchard-Lisle V. On the origin of semantic errors in naming: evidence from case of global aphasia. *Cognitive Neuropsychology* 1984;1(2):163–190.
- Howard, D.; Patterson, K. *Pyramids and Palm Trees*. London, England: Harcourt Assessment; 1992.
- Howard D, Patterson K, Franklin S, Orchard-Lisle V, Morton J. The facilitation of picture naming in aphasia. *Cognitive Neuropsychology* 1985;2:49–80.
- Jokel R, Rochon E, Leonard C. Testing predictions of the interactive activation model in recovery from aphasia after treatment. *Brain and Cognition* 2004;54(3):251–253. [PubMed: 15050786]
- Katz RC, Wertz RT. The efficacy of computer-provided reading treatment for chronic aphasic adults. *Journal of Speech and Hearing Research* 1997;40(3):493–507.
- Kay, J.; Lesser, R.; Coltheart, M. *The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)*. Hove, U. K: Lawrence Erlbaum Associates; 1992.
- Kertesz, A. *The Western Aphasia Battery*. Philadelphia: Grune and Stratton; 1982.
- Kiran S. Training phoneme to grapheme conversion for patients with written and oral production deficits: A model-based approach. *Aphasiology* 2005;19(1):53–76.
- Kiran S. Semantic complexity in the treatment of naming deficits. *American Journal of Speech Language Pathology* 2007;16(Feb):18–29. [PubMed: 17329672]
- 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* 2003a;46(4):773–787.
- Kiran S, Thompson CK. Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language* 2003b;85(3):441–450. [PubMed: 12744956]
- Kiran S, Ntourou K, Eubanks M. Effects of typicality on category verification in inanimate categories in aphasia. *Aphasiology*. in press
- Kiran, SD. *Dissertation Abstracts International: Section B: the Sciences & Engineering*. Vol. 62. US: Univ Microfilms International; 2002. Effect of exemplar typicality on naming deficits in fluent aphasia; p. 5076
- Lowell S, Beeson PM, Holland AL. The efficacy of a semantic cueing procedure on naming performance of adults with aphasia. *American Journal of Speech Language Pathology* 1995;4(4):109–114.
- Maher L, Raymer A. Management of anomia. *Topics in Stroke Rehabilitation* 2004;11(1):10–21. [PubMed: 14872396]
- Marshall J, Pound C, White-Thompson M, Pring T. The use of picture/word matching tasks to assist word retrieval in aphasic patients. *Aphasiology* 1990;4(2):167–184.
- McCleary C, Hirst W. Semantic classification in aphasia: A study of basic, superordinate, and function relations. *Brain and Language* 1986;27(2):199–209. [PubMed: 3955339]
- McCloskey ME. The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior* 1980;19:485–502.
- McReynolds, LV.; Kearns, KP. *Single subject experimental designs in communicative disorders*. Baltimore: University Park Press; 1983.
- Moore CJ, Price CJ. A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 1999;122:943–962. [PubMed: 10355678]
- Murray LL, Ballard K, Karcher L. Linguistic specific treatment: Just for Broca's aphasia? *Aphasiology* 2004;18:785–809.
- Nickels L. Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology* 2002;16(1011):935–979.
- Plaut DC. Relearning after damage in connectionist networks: toward a theory of rehabilitation. *Brain and Language* 1996;52(1):25–82. [PubMed: 8741976]

- Pring TH, H A, Harwood A, Macbride L. Generalization of naming after picture/word matching tasks: only items appearing in therapy benefit. *Aphasiology* 1993;7(4):383–394.
- Raymer AM, Maher LM, Foundas AL, Gonzalez Rothi LJ, Heilman KM. Analysis of lexical recovery in an individual with acute anomia. *Aphasiology* 2000;14(9):901–910.
- Raymer AM, Thompson CK, Jacobs B, LeGrand HR. Phonological treatment of naming deficits in aphasia: Model-based generalization analysis. *Aphasiology* 1993;7:27–53.
- Rosch E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 1975;104(3):192–233.
- Smith EE, Shoben EJ, Rips LJ. Structure and process in semantic memory: A featural model of semantic association. *Psychological Review* 1975;81:214–241.
- Stanczak L, Waters G, Caplan D. Typicality-based learning and generalization in aphasia: Two case studies of anomia treatment. *Aphasiology* 2006;20(24):374–383.
- Schwartz MF, Brecher A. A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language* 2000;73(1):62–91. [PubMed: 10872638]
- Thompson CK. Complexity in language learning and treatment. *American Journal of Speech Language Pathology* 2007;16:3–5. [PubMed: 17329670]
- Thompson CK, Shapiro LP. Syntactic complexity in treatment of sentence production deficits. *American Journal of Speech Language Pathology* 2007;16:30–42. [PubMed: 17329673]
- Thompson CK, Shapiro LP, Kiran S, Sobecks J. The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research* 2003;46(3):591–607.
- Uyeda KM, Mandler F. Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation* 1980;12(6):587–595.
- Wambaugh J, Linebaugh CW, Doyle P, Martinez AL, Kalinyak-Fliszar M, Spencer K. Effects of two cueing treatments on lexical retrieval in aphasic patients with different levels of deficit. *Aphasiology* 2001;15(1011):933–950.

Figure 1.

(a) Naming accuracy for atypical (trained) and typical (untrained) items for the category *clothing*, and (b) naming accuracy for typical (trained) and atypical (untrained) when treatment was provided for typical examples for the category *furniture* for Participant 1. Treatment was subsequently shifted to atypical examples while maintenance of the previously trained typical examples was observed.

Figure 2.
 (a) Naming accuracy for typical (trained) and atypical items (untrained) for the category *furniture* during baseline and treatment phases for Participant 2.

Figure 3.

(a) Naming accuracy for atypical (trained) and typical (untrained) items for the category *furniture*, and (b) naming accuracy for typical (trained) and atypical (untrained) items for the category *clothing* across baseline, treatment and follow-up phases for Participant 3.

Figure 4.

(a) Naming accuracy for typical (trained) and atypical (untrained) items for the category *clothing*, and (b) naming accuracy for atypical (trained) and typical (untrained) items for the category *furniture* across baseline and treatment phases for Participant 4. The hashed line indicates a change in probe protocol.

(a) P5: Furniture

Figure 5.

(a) Naming accuracy for atypical (trained) and typical (untrained) items for the category *furniture* across baseline, treatment and follow up phases for Participant 5. The hashed line indicates a change in probe protocol.

Table 1

Demographic and stroke related data for the five participants in the study. Performance on the Western Aphasia Battery (Kertesz, 1982) and Cognitive Linguistic Quick Test (Helm-Estabrooks, 2001) is reported. WNL = Within normal limits, N/A: Not available

	P1	P2	P3	P4	P5
Age	55	77	63	47	50
Months post onset	10	7	9	8	7
Gender	Female	Female	Female	Male	Female
Years of education	14	14	12	15	12
<i>Western Aphasia Battery</i>					
Aphasia Dx	Conduction	Conduction	Conduction	Broca/apraxia	Broca/ -apraxia
Fluency	6	9	8	4	4
Comprehension	8.95	7.85	7.2	6.5	5.7
Repetition	3.3	3.7	6.3	3.8	2.4
Naming	5.1	7.7	3.6	3.8	1.4
Aphasia Quotient	56.7	72.5	62.2	46.4	37
<i>CLQT</i>					
Attention	WNL	Mild	Mild	N/A	Mild
Memory	Severe	Moderate	Severe	N/A	Severe
Executive Function	WNL	Moderate	Severe	N/A	Severe
Language	Severe	Moderate	Severe	N/A	Severe
Visuospatial skills	WNL	Mild	Mild	N/A	WNL

Table 2
Performance (in percentage points) on specific subtests of single word production and semantic processing on the BNT (Goodglass et al., 1983), PALPA (Kay et al., 1992) and PAPT (Howard & Patterson, 1992). Changes for WAB are shown in terms of Aphasia Quotient. Mean performance for normal individuals is also provided in percentage points. N/A: Not available

	Normal mean		P1		P2		P3		P4		P5	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Western Aphasia Battery (AQ)												
Boston Naming Test (N = 60)	91	56.7	69.5	72.1	77.1	62.2	73.1	46.4	50.9	37	49.5	
		25	42	17	25	15	22	13	38	0	8	
Psycholinguistic Assessment of Language Processing in Aphasia												
Single word reading (N = 24)	100	88	96	92	88	92	96	25	42	0	33	
Written naming (N = 40)	97.5	85	98	60	65	18	55	32	52	0	3	
Single word repetition (N = 40)	99.2	75	100	78	93	98	93	83	72	98	100	
Spoken word to picture matching (N = 40)	98.2	93	100	95	98	68	93	85	100	65	90	
Written word to picture matching (N = 40)	98.6	95	93	93	93	83	88	97	100	73	88	
Auditory word synonym judgment (N = 60)		78	95	68	65	65	62	66	88	0	78	
Written word synonym judgment (N = 60)		82	88	67	78	0	95	75	68	63	77	
Written Spelling	98.7	88	85	50	100	68	65	25	N/A	0	0	
Pyramids and Palm Trees												
Three pictures (N = 52)	98	80	85	73	80	90	83	80	96	62	81	
Three words (N = 52)	98	77	87	72	75	62	75	100	94	73	92	

Performance on the Apraxia Battery for Adults (ABA, Dabul, 1979) prior to initiation and following completion of treatment for P4 and P5. WNL = Within normal limits

Kiran

	P4		P5	
	Pre	Post	Pre	Post
<i>Diadochokinetic Rate</i>	Moderate	Mild	Moderate	Mild
<i>Increasing word length (A)</i>	Moderate	Mild	Moderate	None
<i>Increasing word length (B)</i>	Severe	Severe	Moderate	Moderate
<i>Limb apraxia</i>	Mild	WNL	Severe	Mild
<i>Oral apraxia</i>	Moderate	Mild	Severe	Moderate
<i>Utterance time</i>	Severe	Mild	Severe	Severe
<i>Repeated trials</i>	Moderate	Moderate	Moderate	Mild

Table 4
Number of baselines and counterbalanced order of category and typicality exposed in treatment. Also shown is a summary of generalization patterns observed for each patient

Participant	# of baselines	Category trained	Typicality trained	Generalization patterns observed
P1	3	1. Clothing 2. Furniture	Atypical Typical	Atypical => Typical Typical ≠ Atypical
P2	3	1. Furniture 2. No treatment	Typical	Typical ≠ Atypical
P3	5	1. Furniture 2. Clothing	Atypical Typical	Atypical ≠ Typical Typical ≠ Atypical
P4	5	1. Clothing 2. Furniture	Typical Atypical	Typical ≠ Atypical Atypical => Typical
P5	3	1. Furniture 2. No treatment	Atypical	**

** Note that P5 showed trends towards generalization (from 7% to 40% accuracy).

Table 5

Evolution of errors reported in raw numbers and in proportion of total errors. Changes over 10% are highlighted in bold. **indicates values significant at $p < .05$.

	Clothing				Furniture			
	Pre (raw)	Pre (%)	Post (raw)	Post (%)	Pre (raw)	Pre (%)	Post (raw)	Post (%)
<i>Total errors**</i>	<i>44</i>		<i>9</i>		<i>51</i>		<i>10</i>	**
P1								
No response/IDK	7	15.9	0	0.0	4	7.8	1	10.0
Unrelated word	9	20.5	1	11.1	7	13.7	0	0.0
Neologism	0	0.0	0	0.0	1	2.0	0	0.0
Perseveration	0	0.0	0	0.0	0	0.0	0	0.0
Circumlocution	20	45.5	4	44.4	29	56.9	1	10.0
Superordinate	0	0.0	0	0.0	0	0.0	0	0.0
Semantic	6	13.6	1	11.1	5	9.8	6	60.0
Phonemic	2	4.5	3	33.3	4	7.8	2	20.0
Mixed	0	0.0	0	0.0	1	2.0	0	0.0
<i>Total errors**</i>					<i>58</i>		<i>29</i>	
P2								
No response/IDK					9	15.5	6	20.7
Unrelated word					9	15.5	1	3.4
Neologism					18	31.0	2	6.9
Perseveration					0	0.0	0	0.0
Circumlocution					5	8.6	2	6.9
Superordinate					0	0.0	0	0.0
Semantic					5	8.6	9	31.0
Phonemic					11	19.0	8	27.6
Mixed					1	1.7	1	3.4
<i>Total errors**</i>	<i>121</i>		<i>80</i>		<i>128</i>		<i>97</i>	
P3								
No response/IDK	4	3.3	2	2.5	6	4.7	19	19.6
Unrelated word	29	24.0	18	22.5	26	20.3	10	10.3
Neologism	5	4.1	6	7.5	7	5.5	3	3.1
Perseveration	34	28.1	31	38.8	59	46.1	40	41.2
Circumlocution	12	9.9	4	5.0	15	11.7	12	12.4

	Clothing				Furniture			
	Pre (raw)	Pre (%)	Post (raw)	Post (%)	Pre (raw)	Pre (%)	Post (raw)	Post (%)
Superordinate	0	0.0	0	0.0	0	0.0	1	1.0
Semantic	32	26.4	12	15.0	11	8.6	6	6.2
Phonemic	5	4.1	7	8.8	2	1.6	5	5.2
Mixed	0	0.0	0	0.0	2	1.6	1	1.0
<i>P4</i>	<i>137</i>		<i>83</i>		<i>142</i>		<i>44</i>	
<i>Total errors**</i>								
No response/IDK	1	0.7	1	1.2	0	0.0	5	11.4
Unrelated word	19	13.9	18	21.7	5	3.5	4	9.1
Neologism	12	8.8	11	13.3	10	7.0	8	18.2
Perseveration	95	69.3	4	4.8	122	85.9	4	9.1
Circumlocution	0	0.0	2	2.4	0	0.0	0	0.0
Superordinate	2	1.5	0	0.0	1	0.7	0	0.0
Semantic	5	3.6	26	31.3	2	1.4	17	38.6
Phonemic	3	2.2	16	19.3	2	1.4	6	13.6
Mixed	0	0.0	5	6.0	0	0.0	0	0.0
<i>P5</i>								
<i>Total errors**</i>								
No response/IDK					88		28	
Unrelated word					76	86.4	21	75.0
Neologism					5	5.7	0	0.0
Perseveration					0	0.0	0	0.0
Circumlocution					7	8.0	0	0.0
Superordinate					0	0.0	1	3.6
Semantic					0	0.0	0	0.0
Phonemic					0	0.0	6	21.4
Mixed					0	0.0	0	0.0