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

Swathi Kiran
University of Texas at Austin

Purpose: The typicality treatment approach on improving naming was investigated within 2 inanimate categories (*furniture* and *clothing*) using a single-subject experimental design across participants and behaviors in 5 patients with aphasia.
Method: Participants received a semantic feature treatment to improve naming of either typical or atypical items within semantic categories, whereas generalization was tested to untrained items of the category. The order of typicality and category trained was counterbalanced across participants.
Results: Results indicated that 2 out of 4 patients trained on naming of atypical examples demonstrated generalization to naming untrained typical examples. One patient showed trends toward generalization but did not achieve criterion. Furthermore, all 4 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.
Conclusion: These results extend our previous findings (S. Kiran & C. K. Thompson, 2003a) 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 of facilitating generalization to untrained items within a category than training typical examples (S. Kiran, 2007).

KEY WORDS: aphasia, treatment, typicality

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 bidirectional 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, & Ruml, 2000; Cuetos, Aguado, & Caramazza, 2000). Patients who demonstrate semantic errors devoid of coexisting 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 & Orchard-Lisle, 1984; McCleary & Hirst, 1986).

Consistent with the level of naming impairment, therapy tasks have focused on facilitating access at either the phonological or 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–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 improvement on trained words; 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, Harwood, & McBride, 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 4 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 retraining 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, whereas performance on atypical words deteriorated.

We replicated Plaut's simulation results 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 argued that atypical examples were more complex than typical examples within the category; 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 2 patients with anomic aphasia. Stanczak et al. found that 1 of the 2 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 the Stanczak et al. 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 (CATE) 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 are 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 make up 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; 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). Furthermore, 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 5 patients presented with breakdown in lexical retrieval at either the semantic level and/or the phonological level, 2 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 semantically based treatment on lexical access and to understand whether the selective generalization patterns from atypical to typical examples were also observed in these patients. Finally, the nature of naming errors occurring throughout treatment was also examined. Within the theoretical framework described previously, 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.

Method

Participants

Five monolingual, English-speaking individuals with aphasia recruited from local hospitals within the Austin, Texas, 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 7 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

Table 1. Demographic and stroke-related data for the 5 participants in the study.

Variable	P1	P2	P3	P4	P5
Age	55	77	63	47	50
Months postonset	10	7	9	8	7
Gender	Female	Female	Female	Male	Female
Years of education	14	14	12	15	12
			WAB		
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
			CLQT		
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

Note. Performance on the Western Aphasia Battery (WAB; Kertesz, 1982) and Cognitive Linguistic Quick Test (CLQT; Helm-Estabrooks, 2001) is reported. P = Participant; Dx = diagnosis; WNL = within normal limits; N/A = data not available.

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 test (PAPT; Howard & Patterson, 1992). Impairment in semantic processing was hypothesized to be integral to the success of treatment because the principal 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. Participant P4 did not meet inclusionary criteria for the protocol. Scores on this task indicated that 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 coexisting apraxia (see Table 3). Performance on this test indicated that both participants presented with mild to 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, Illinois, community (Kiran, 2002). All participants had normal or corrected-to-normal vision, had 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.

Table 2. Performance (in percentage points) on specific subtests of single word production and semantic processing on the Boston Naming Test (BNT), Psycholinguistic Assessment of Language Processing in Aphasia (PALPA), and Pyramids and Palm Trees (PAPT).

Test	Mean for non-brain-damaged individuals	P1		P2		P3		P4		P5	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
WAB AQ		56.7	69.5	72.1	77.1	62.2	73.1	46.4	50.9	37.0	49.5
BNT (*N* = 60)	91.0	25	42	17	25	15	22	13	38	0	8
PALPA											
Single Word Reading (*N* = 24)	100.0	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
PAPT											
Three Pictures (*N* = 52)	98.0	80	85	73	80	90	83	80	96	62	81
Three Words (*N* = 52)	98.0	77	87	72	75	62	75	100	94	73	92

Note. Changes for Western Aphasia Battery (WAB) are shown in terms of Aphasia Quotient (AQ). Mean performance for typically developing individuals is also provided in percentage points. *N* = number of items on the test.

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 that 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 = *good example,* 7 = *poor example*) the extent to which each example represented their idea or image of the category term. Mean average ratings and standard deviations were calculated for each example in the category.

Table 3. Performance on the Apraxia Battery for Adults prior to initiation and following completion of treatment for Participants 4 and 5.

Variable	P4		P5	
	Pre	Post	Pre	Post
Diadochokinetic rate	Moderate	Mild	Moderate	Mild
Increasing word length (A)	Moderate	Mild	Moderate	None
Increasing word length (B)	Severe	Severe	Moderate	Moderate
Limb apraxia	Mild	WNL	Severe	Mild
Oral apraxia	Moderate	Mild	Severe	Moderate
Utterance time	Severe	Mild	Severe	Severe
Repeated trials	Moderate	Moderate	Moderate	Mild

Development of treatment categories and their examples. For the present experiment, two inanimate categories (*clothing, furniture*) were chosen from the 12-category 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 used to eliminate problematic examples within categories. For instance, examples that at least 60% (12 out of 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 *furniture* category, the *z* values were –1.37 to –0.42 (typical) and 1.12 to 0.41 (atypical).

For *clothing,* the *z* values were –1.22 to –0.44 (typical) and –0.01 to 0.05 (atypical). 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.htm), and number of syllables (see Appendix A for a list of stimuli). Separate 2 (typicality: typical, atypical) × 2 (category: *clothing, furniture*) analyses of variance (ANOVAs) performed on the variables revealed nonsignificant effects for all variables.

The selected pictures were presented to a group of 10 non-brain-damaged 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 in. × 6 in. (10.16 cm × 15.24 cm) 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, 30 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 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 on the Internet. Only features that 18 out of 20 young and older 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 were specific to atypical examples (distinctive features; e.g., *furniture: decorative accessory*). The main difference among 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 of *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 the 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, whereas 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, or 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 included responses were classified into one of the following categories: (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

Table 4. Number of baselines and counterbalanced order of category and typicality exposed in treatment.

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

Note. Also shown is a summary of generalization patterns observed for each participant.

[a]Participant 5 showed trends toward generalization (7%–40% accuracy).

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, end table/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 2 hr. 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 such as 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 a priori criterion for termination of treatment was set at 80% accuracy (12 out of 15) for two consecutive sessions or a total of 20 treatment sessions (10 probe sessions). However, treatment was extended beyond this criterion for Participants 1, 3, and 5 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. Furthermore, 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 improvement on the trained items after the specified number of sessions (see Results section for details). For Participant 4, after nine 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 were coded using the criteria described previously.

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 because (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 an extended baseline for assessment of performance.

Data Analysis

Effect sizes (ESs) were calculated comparing the mean of all data points in the treatment phase relative to the baseline mean divided by the standard deviation of the baseline (Busk & Serlin, 1992). On the basis of comparable naming treatment studies in aphasia, an ES of 4.0 was considered small, 7.0 was considered medium, and 10.0 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 online 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% to 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 were categorized into the corresponding subtypes by the author. Interrater 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; Edmonds & Kiran, 2006; Kiran & Thompson, 2003a) for the trained items.

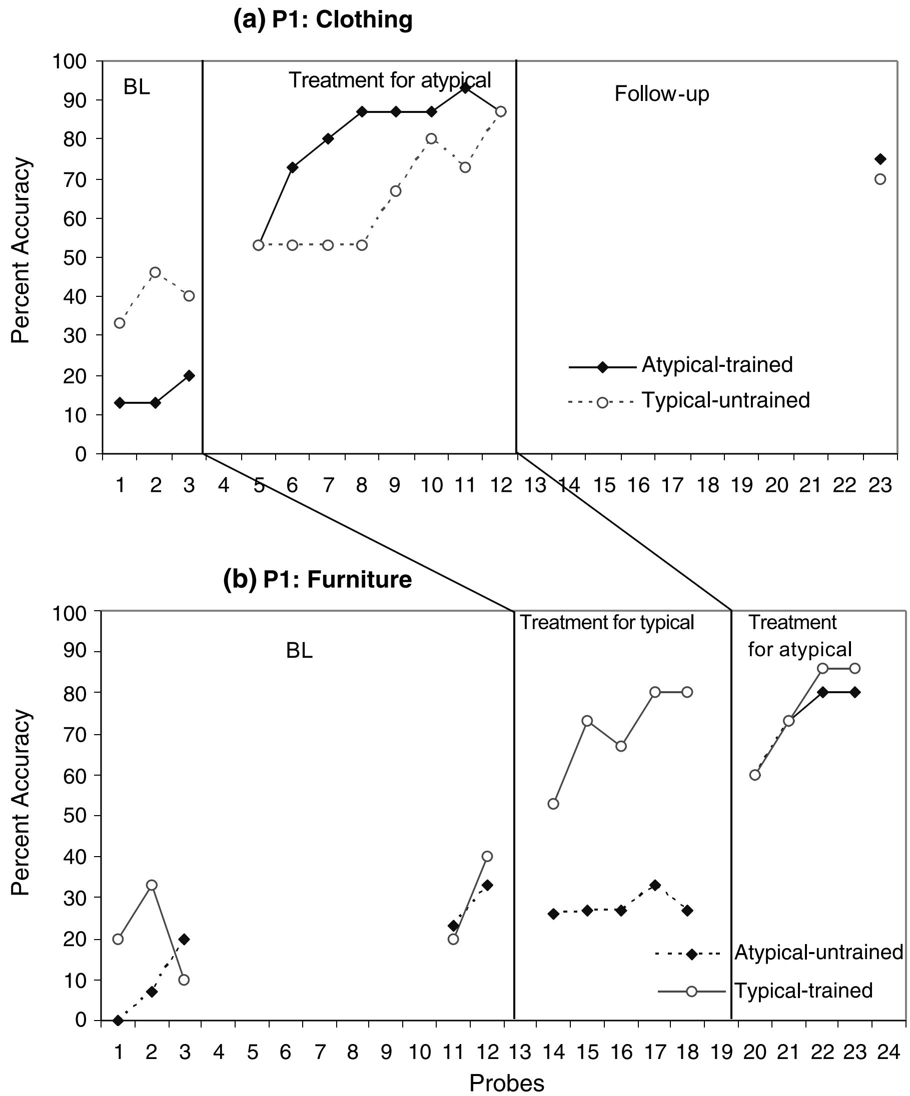
Participant 1. Following three initial baselines, Participant 1 first received treatment for atypical examples of *clothing,* which improved to criterion (high of 93% accuracy; ES = 16.21) within 8 weeks, as generalization to the untrained typical examples was observed to a high of 87% accuracy (ES = 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 5 weeks. Improvement on the untrained atypical examples was 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 = 0.5). 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 9 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). Although this participant 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

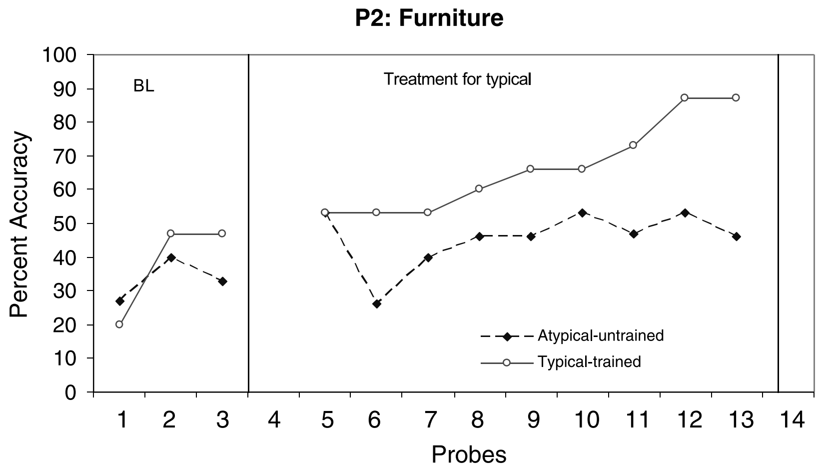
Figure 1. (a) Naming accuracy for atypical (trained) and typical (untrained) items for the *clothing* category and (b) naming accuracy for typical (trained) and atypical (untrained) when treatment was provided for typical examples for the *furniture* category for Participant 1. Treatment was subsequently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



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% to 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 were not conducted because of 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

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



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 (ESs 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, the data suggest 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.

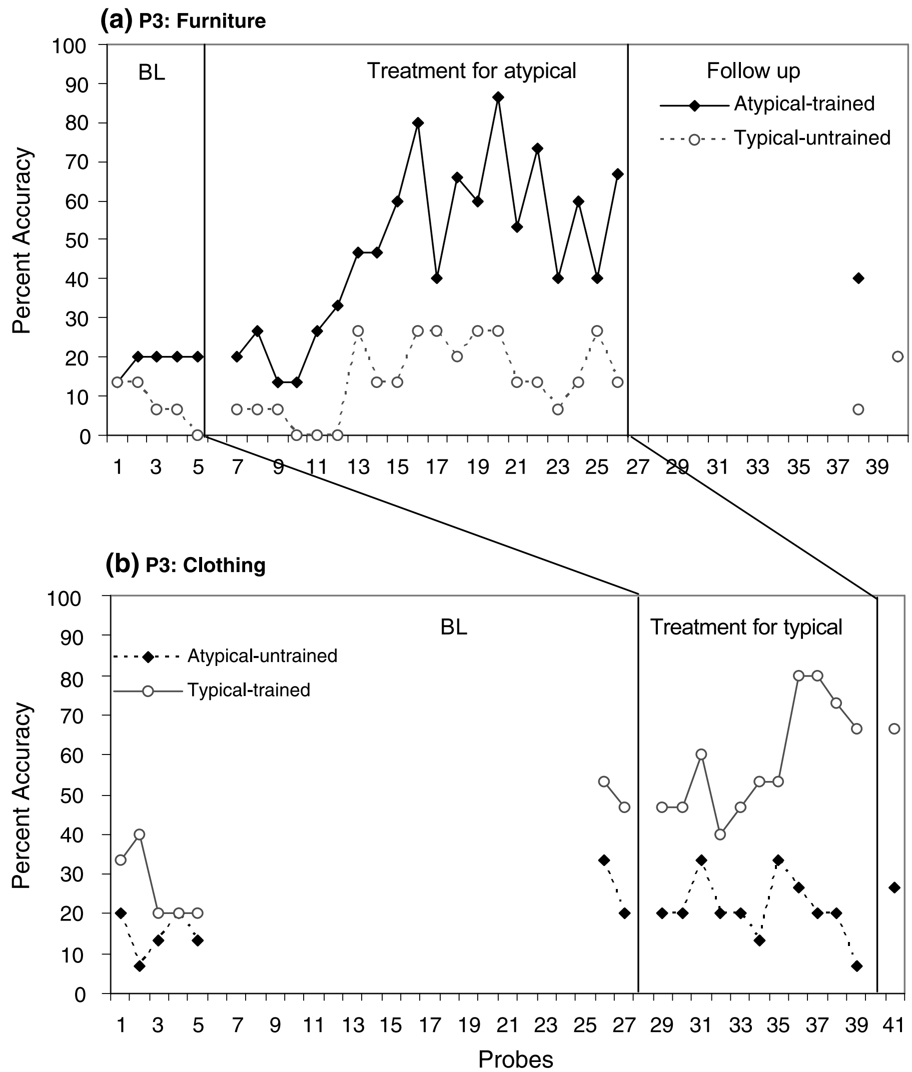
Evolution of Errors

For each participant, errors produced during baseline sessions and equal numbers of sessions at the end of treatment were compared across the two categories (see Table 5). All participants showed a 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 improvement on the standardized tests is reported here, but no interpretations are drawn regarding significant changes. To further understand the relationship between improvement 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 those reported by Hickin, Best, Herbert, Howard, and Osborne (2002). Specifically, improvement 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 improvement 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).

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

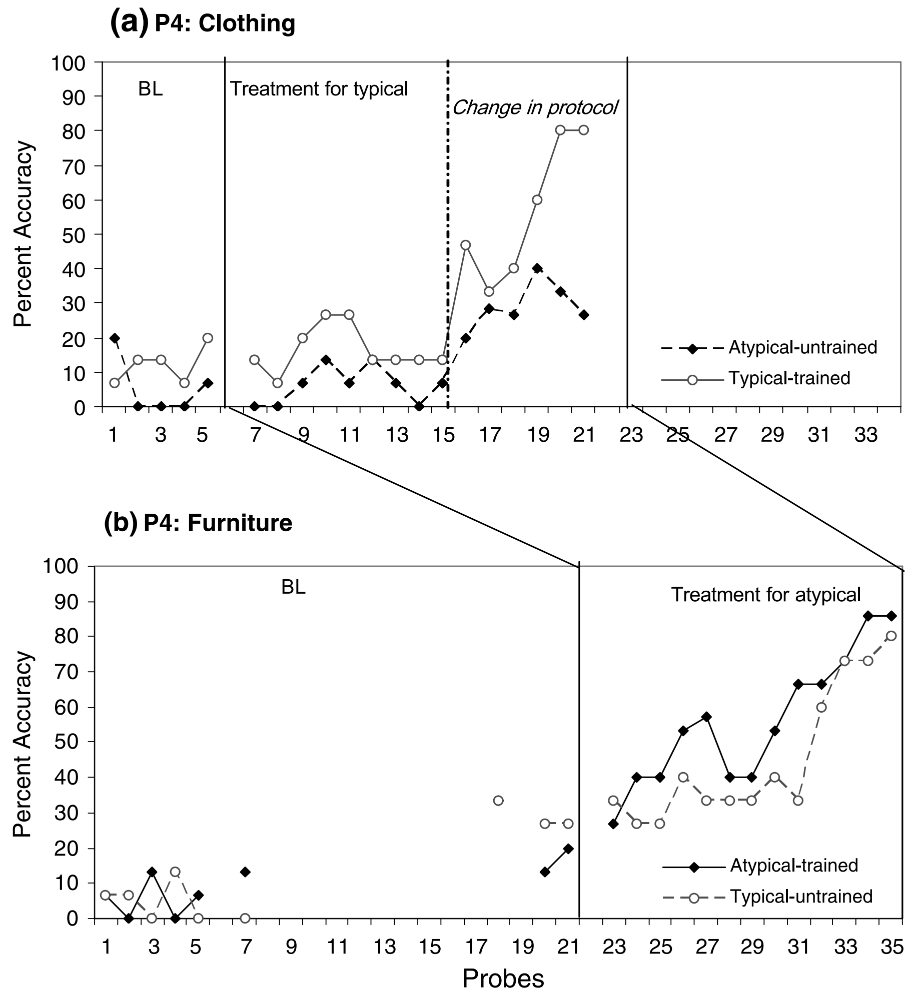


For the same patient, improvement in naming untrained items was calculated as Average Final Probe (Typical Furniture, Atypical Clothing) – Average Baseline (Typical Furniture, Atypical Clothing). Then, percent change on trained items and untrained items was correlated with percent change on the standardized tests (except Western Aphasia Battery: Aphasia Quotient [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

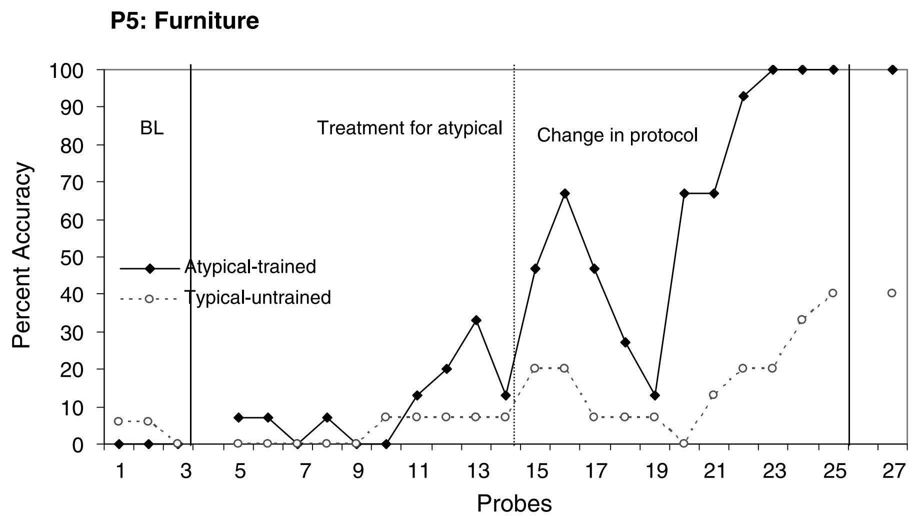
This experiment was undertaken for two reasons: (a) to establish the effectiveness of the typicality treatment approach in facilitating lexical retrieval and generalization in individuals with naming deficits and (b) to investigate the relevance of semantic complexity as a treatment variable to understanding language recovery patterns in aphasia. A previous study (Kiran & Thompson,

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



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 5 participants 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 Participant 5 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*. Because the experimental design allowed the examination of differential responsiveness to typical or atypical training within the same participant, the 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 semantically based treatment that emphasizes the explicit analysis of semantic attribute information is a successful approach for improving naming skills and facilitating

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



generalization (Boyle, 2004; Boyle & Coehlo, 1995; Coehlo, McHugh, & Boyle, 2000; Drew & Thompson, 1999; Lowell et al., 1995).

Second, the results indicate that the effect of manipulating typicality as a treatment variable to examine semantic complexity in animate categories (Kiran & Thompson, 2003a) 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 highlights the featural variation within the categories and consequently improves typical examples. In contrast, training the features associated with typical examples has no influence on atypical examples of the category. Although 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 semantically based treatment extend beyond improvement 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 the evolution of errors over time (Basso, Corno, & Marangolo, 1996; Edmonds & Kiran, 2006; Jokel, Rochon, & Leonard, 2004; Kiran & Thompson, 2003a; Raymer, Maher, Foundas, Gonzalez Rothi, & Heilman, 2000).

To further understand the relationship, if any, between improvement 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) the 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 2: feature selection), and judging auditorily presented semantic features (Step 3: yes/no questions) resulted not only in improvement on naming of trained items but also translated to improvement in oral reading and semantic processing. It should be noted, however, that the scope of such broad changes should theoretically also extend to improvement

Table 5. Evolution of errors reported in raw numbers and in proportion to total errors.

	Clothing				Furniture			
	Pre (raw)	Pre (%)	Post (raw)	Post (%)	Pre (raw)	Pre (%)	Post (raw)	Post (%)
P1								
Total errors**	44		9		51		10	**
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
P2								
Total errors**	—			—	58		29	
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
P3								
Total errors**	121		80		128		97	
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
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
P4								
Total errors**	137		83		142		44	
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

(Continued on the following page)

Table 5 *Continued.* Evolution of errors reported in raw numbers and in proportion to total errors.

	Clothing				Furniture			
	Pre (raw)	Pre (%)	Post (raw)	Post (%)	Pre (raw)	Pre (%)	Post (raw)	Post (%)
				P5				
Total errors**					88		28	
No response/IDK					76	86.4	21	75.0
Unrelated word					5	5.7	0	0.0
Neologism					0	0.0	0	0.0
Perseveration					7	8.0	0	0.0
Circumlocution					0	0.0	1	3.6
Superordinate					0	0.0	0	0.0
Semantic					**0**	**0.0**	**6**	**21.4**
Phonemic					0	0.0	0	0.0
Mixed					0	0.0	0	0.0

Note. Changes over 10% are highlighted in bold. IDK = I don't know.

**$p < .05$.

on the untrained categories. No improvement 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 Participants 1, 2, and 4, whereas results for Participants 3 and 5 are not completely aligned with the predictions. What sets these participants apart from Participants 1 and 4 (who show strong generalization patterns) is unclear and requires further examination. Despite Participant 3 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 *clothing* category. Also, this participant did not demonstrate the expected trends in evolution of errors. Participant 5 showed a severe language deficit prior to inception of treatment and also showed limited benefit from the semantically based treatment. Moreover, this participant's treatment outcome may also have been influenced by the provision of phonemic cues. 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). Furthermore, Helm-Estabrooks (2002) suggests that individual differences in response to language therapy may be attributable to various aspects of a patient's cognitive abilities. Although it may be premature to assume a relationship between performance in treatment and on the CLQT, the present study 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 coexisting apraxia. The impaired performance of Participants 4 and 5 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. The fact that these participants demonstrated changes on standardized semantic processing tests similar to the three fluent participant counterparts does 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, whereas others are not.

Conclusion

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 improvement on trained typical examples, but no generalization to untrained atypical examples was observed. These findings have important clinical implications. Given the current health care environment, which restricts treatment to a limited number of sessions, a naming therapy that promotes optimal generalization patterns is ultimately more beneficial than one that does not facilitate generalization. The present data show that the typicality-based naming treatment is applicable for some patients with fluent and nonfluent aphasia.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant DC006359-03 and a New Century Research Scholars Grant from the American Speech-Language-Hearing Foundation to the author. Thanks to 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, thanks to the participants in the experiment for their patience and cooperation.

References

Ahn, W.-K. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69,* 135–178.

Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory and Cognition, 15*(5), 397–418.

Basso, A., Corno, M., & Marangolo, P. (1996). Evolution of oral and written confrontation naming errors in aphasia. A retrospective study on vascular patients. *Journal of Clinical and Experimental Neuropsychology, 18,* 77–87.

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

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology, 13,* 236–249.

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

Busk, P. L., & Serlin, R. (1992). Meta analysis for single case research. In T. R. K. Levin & J. R. Levin (Eds.), Single case research design and analysis: New directions for psychology and education (pp. 23–45). Hillsdale, NJ: Erlbaum.

Butterworth, B. (1989). Lexical access in speech production. In W. M. Wilson (Ed.), *Lexical representation and process* (pp. 108–235). Cambridge, MA: MIT Press.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex, 26,* 95–122.

Caramazza, A., Papagno, C., & Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language, 75,* 428–450.

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

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

Connell, P. J., & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders, 51,* 214–225.

Cuetos, F., Aguado, G., & Caramazza, A. (2000). Dissociation of semantic and phonological errors in naming. *Brain and Language, 75,* 451–460.

Dabul, B. (1979). *Apraxia Battery for Adults*. Tigard, OR: C.C. Publications.

Davis, A., & Pring, T. (1991). Therapy for word-finding deficits: More on the effects of semantic and phonological approaches to treatment with dysphasic patients. *Neuropsychological Rehabilitation, 1,* 135–145.

Dell, G. S., Schwartz, M. F., Martin, N. M., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104,* 801–838.

Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review, 6,* 338–346.

Drew, R. L., & Thompson, C. K. (1999). Model-based semantic treatment for naming deficits in aphasia. *Journal of Speech, Language, and Hearing Research, 42,* 972–989.

Edmonds, L. A., & Kiran, S. (2006). Effect of semantic based treatment on cross linguistic generalization in bilingual aphasia. *Journal of Speech, Language, and Hearing Research, 49,* 729–748.

Estes, Z. (2003). Domain differences in the structure of artifactual and natural categories. *Memory and Cognition, 31,* 199–214.

Forde, E. M., & Humphreys, G. W. (1999). Category specific recognition impairments: A review of important case studies and influential theories. *Aphasiology, 13,* 169–193.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43,* 182–216.

Frances, N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition, 89,* 27–41.

Gierut, J. (2007). Phonological complexity and language learnability. *American Journal of Speech-Language Pathology, 16,* 6–17.

Goldenberg, G., Dettmers, H., Grothe, C., & Spatt, J. (1994). Influence of linguistic and nonlinguistic capacities on spontaneous recovery of aphasia and on success of language therapy. *Aphasiology, 8,* 443–456.

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

Hampton, J. (1993). Prototype models of concept representation. In J. Van Mechelen, J. A. Hampton, & P. T. R. S. Michlanski (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). London: Academic Press.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34,* 686–708.

Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test*. London, England: Harcourt Assessment.

Helm-Estabrooks, N. (2002). Cognition and aphasia: A discussion and a study. *Journal of Communication Disorders, 35,* 171–186.

Hillis, A. E., Rapp, B. C., Romani, C., & Caramazza, A. (1990). Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology, 7,* 191–243.

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology, 16*(10–11), 981–999.

Horner, D. R., & Baer, D. M. (1978). Multiple probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis, 11,* 189–196.

Howard, D., & Orchard-Lisle, V. (1984). On the origin of semantic errors in naming: Evidence from case of global aphasia. *Cognitive Neuropyschology, 1,* 163–190.

Howard, D., & Patterson, K. (1992). *Pyramids and Palm Trees*. London, England: Harcourt Assessment.

Howard, D., Patterson, K., Franklin, S., Orchid–Lisle, V., & Morton, J. (1985). The facilitation of picture naming in aphasia. *Cognitive Neuropsychology, 2,* 49–80.

Jokel, R., Rochon, E., & Leonard, C. (2004). Testing predictions of the interactive activation model in recovery from aphasia after treatment. *Brain and Cognition, 54,* 251–253.

Kay, J., Lesser, R., & Coltheart, M. (1992). *The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)*. Hove, United Kingdom: Erlbaum.

Kertesz, A. (1982). *The Western Aphasia Battery*. Philadelphia: Grune and Stratton.

Kiran, S. (2005). Training phoneme to grapheme conversion for patients with written and oral production deficits: A model-based approach. *Aphasiology, 19,* 53–76.

Kiran, S. (2007). Semantic complexity in the treatment of naming deficits. *American Journal of Speech-Language Pathology, 16,* 18–29.

Kiran, S., & Thompson, C. K. (2003a). 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, 46,* 773–787.

Kiran, S., & Thompson, C. K. (2003b). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language, 85,* 441–450.

Kiran, S., Ntourou, K., & Eubanks, M. (2007). Effects of typicality on category verification in inanimate categories in aphasia. *Aphasiology, 21,* 844–867.

Kiran, S. D. (2002). Effect of exemplar typicality on naming deficits in fluent aphasia. Doctoral dissertation, Northwestern University. Retrieved October 27, 2008, from www.proquest.com/division/cs-support.shtml.

Lowell, S., Beeson, P. M., & Holland, A. L. (1995). The efficacy of a semantic cueing procedure on naming performance of adults with aphasia. *American Journal of Speech-Language Pathology, 4,* 109–114.

Maher, L., & Raymer, A. (2004). Management of anomia. *Topics in Stroke Rehabilitation, 11,* 10–21.

Marshall, J., Pound, C., White-Thompson, M., & Pring, T. (1990). The use of picture/word matching tasks to assist word retrieval in aphasic patients. *Aphasiology, 4,* 167–184.

McCleary, C., & Hirst, W. (1986). Semantic classification in aphasia: A study of basic, superordinate, and function relations. *Brain and Language, 27*(2), 199–209.

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

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122,* 943–962.

Murray, L. L., Ballard, K., & Karcher, L. (2004). Linguistic specific treatment: Just for Broca's aphasia? *Aphasiology, 18,* 785–809.

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

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language, 52,* 25–82.

Pring, T. H., A. Harwood, A., & Macbride, L. (1993). Generalization of naming after picture/word matching tasks: Only items appearing in therapy benefit. *Aphasiology, 7,* 383–394.

Raymer, A. M., Maher, L. M., Foundas, A. L., Gonzalez Rothi, L. J., & Heilman, K. M. (2000). Analysis of lexical recovery in an individual with acute anomia. *Aphasiology, 14,* 901–910.

Raymer, A. M., Thompson, C. K., Jacobs, B., & LeGrand, H. R. (1993). Phonological treatment of naming deficits in aphasia: Model-based generalization analysis. *Aphasiology, 7,* 27–53.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1975). Structure and process in semantic memory: A featural model of semantic association. *Psychological Review, 81,* 214–241.

Stanczak, L., Waters, G., & Caplan, D. (2006). Typicality-based learning and generalization in aphasia: Two case studies of anomia treatment. *Aphasiology, 20*(2–4), 374–383.

Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language, 73,* 62–91.

Thompson, C. K. (2007). Complexity in language learning and treatment. *American Journal of Speech-Language Pathology, 16,* 3–5.

Thompson, C. K., & Shapiro, L. P. (2007). Syntactic complexity in treatment of sentence production deficits. *American Journal of Speech-Language Pathology, 16,* 30–42.

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). 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, 46,* 591–607.

Uyeda, K. M., & Mandler, F. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation, 12,* 587–595.

Wambaugh, J., Linebaugh, C. W., Doyle, P., Martinez, A. L., Kalinyak-Fliszar, M., & Spencer, K. (2001). Effects of two cueing treatments on lexical retrieval in aphasic patients with different levels of deficit. *Aphasiology, 15*(10/11), 933–950.

Received February 14, 2007

Accepted April 3, 2008

DOI: 10.1044/1092-4388(2008/07-0038)

Contact author: Swathi Kiran, Department of Communication Sciences and Disorders, University of Texas, 1 University Station, Austin, Texas 78712.
E-mail: s-kiran@mail.utexas.edu.

Appendix A. Stimuli used in treatment.

	Clothing		Furniture	
Variable	*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	End table	Wastebasket
	Sweatsuit	Garter	Nightstand	Hammock
	Underwear	Apron	Recliner	Pillow
	Pajamas	Bib	Lamp	Furnace
	Socks	Bandana	Cabinet	Towel rack
	Shirt	Tights	Footstool	Beanbag
Average typicality *M* (*SD*)	1.7 (0.47)	4.1 (0.73)	1.7 (0.53)	5.4 (0.45)
Average WWF *M* (*SD*)	11.6 (10)	6.5 (10)	24.2 (38)	20.2 (10)
Average familiarity *M* (*SD*)	558 (45)	487 (82)	546 (79)	549 (77)
Number of syllables *M* (*SD*)	1.6 (0.82)	1.6 (0.72)	2.0 (0.88)	2.0 (0.82)
Average imageability *M* (*SD*)	591 (31)	536 (55)	572 (42)	564 (51)

Note. WWF = written word frequency.

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 participant 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.