The Role of Semantic Complexity in Treatment of Naming Deficits: Training Semantic Categories in Fluent Aphasia by Controlling Exemplar Typicality

Swathi Kiran
University of Texas at Austin

Cynthia K. Thompson
Northwestern University
Evanston, IL

The effect of typicality of category exemplars on naming was investigated using a single subject experimental design across participants and behaviors in 4 patients with fluent aphasia. 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 indicated that patients trained on naming of atypical exemplars demonstrated generalization to naming of intermediate and typical items. However, patients trained on typical items demonstrated no generalized naming effect to intermediate or atypical examples. Furthermore, analysis of errors indicated an evolution of errors throughout training, from those with no apparent relationship to the target to primarily semantic and phonemic paraphasias. Performance on standardized language tests also showed changes as a function of treatment. Theoretical and clinical implications regarding the impact of considering semantic complexity on rehabilitation of naming deficits in aphasia are discussed.

KEY WORDS: complexity, naming, treatment, aphasia, typicality

Naming deficits are the most common form of language impairment among individuals with aphasia and have been reported across aphasia classification categories, although the nature of naming errors occurring in persons with different types of aphasia may vary (Butterworth, Howard, & McLaughlin, 1984; Goodglass, 1980, 1998). Typically, patients with fluent (Wernicke's) aphasia produce primarily unrelated or jargon words (neologisms), semantically or phonologically related errors, or are not able to retrieve the word at all (Ellis, Miller, & Sin, 1983; Hillis & Caramazza, 1995). Patients with fluent aphasia also often present with concomitant semantic deficits, including impairments in category knowledge (see Shelton & Caramazza, 1999, for a review) leading to the hypothesis that naming deficits in fluent aphasia may derive from impairments within the semantic system.

The source of naming deficits in fluent aphasia, however, is not altogether clear. According to interactive activation models of naming (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Stemberger, 1985), the type of errors seen in fluent aphasia can arise from incomplete/incorrect activation of semantic nodes or phonological nodes during naming attempts or a failure in the bidirectional link between them. Such failure can result in no discernable overlap between the produced word and the intended word, as in neologistic errors. Semantic and phonological errors result from activation of competing entries in the lexicon, which share features with the target, or the correct target may receive interference from other activated words.

Numerous researchers have examined recovery of naming in patients with aphasia when provided with treatment (Howard, Patterson, Franklin, Orchid-Lisle, & Morton 1985a, 1985b; Marshall, Pound, White-Thomson, & Pring, 1990; Nickels & Best, 1996), although few have focused on patients with fluent aphasia. Several studies have used semantically based treatment by means of auditory and written word to picture matching tasks, answering yes/no questions about the target, spoken word categorization, and relatedness judgment tasks (Boyle & Coehlo, 1995; Davis & Pring, 1991; Howard et al., 1985a). Other studies have compared the effects of semantic and phonological treatment on naming and, in general, have found that a combination of both treatments is most effective (Le Dorze, Boulay, Gaudreau, & Brassard, 1994; Howard et al., 1985b). Although most of these studies have reported improvement of trained items, few have found generalization to untrained items (Davis & Pring, 1991; Marshall et al., 1990; Pring, Hamilton, Harwood, & McBride, 1993), and still fewer have examined changes in error patterns resulting from treatment. That many of these studies are aimed at improving semantic access but do not emphasize the underlying aspects of lexical–semantic processing may be a possible reason for the limited generalization noted. Furthermore, the language material chosen to assess generalization has not always been related to the trained examples on important linguistic dimensions, thereby failing an important requirement for generalization (Thompson, 1988). Treatments based on models of lexical processing, which, for example, focus on the semantic features of items within a particular superordinate category (Drew & Thompson, 1999) or attempt to facilitate spreading activation of semantically related words (Boyle & Coehlo, 1995), have been more successful at facilitating generalization.

Generalization also may be enhanced by considering the hierarchical complexity of the stimuli selected for treatment. The complexity effect, that is, training items that are more complex to facilitate generalization to untrained simpler items now has robust evidence from treatment studies for agrammatic aphasia as well as for children with phonological deficits. As shown by Thompson and colleagues in several studies (e.g., Thompson et al., 1997; Thompson, Ballard, & Shapiro, 1998), training complex syntactic structures (e.g., object-cleft sentences) results in generalization to simpler structures (e.g., *wh-* questions) that are in a subset relation to trained structures in agrammatic aphasic patients. This observation led to the complexity account of treatment efficacy (CATE; Thompson, Shapiro, Kiran, & Sobecks, 2003). Likewise, Geirut and colleagues (for a review, see Geirut, 2001) have demonstrated that phonological treatment focused on complex sounds (either in terms of consistency, age of acquisition, or paired contrasts) results in greater generalization to untrained sounds in children with phonological deficits. Moving beyond language, there is supportive evidence of complexity in motor skill learning by adults, particularly as it relates to the conditions of practice in sports such as golf or tennis (Schmidt & Lee, 1999). The findings in this area demonstrate that practice of more difficult skills or levels of a skill results in greater learning of simpler motor skills.

In the present experiment, we apply the notion of complexity to semantic concepts with reference to naming deficits in patients with aphasia. Because existing naming treatment studies have not been completely successful in promoting generalization, the aim of the present experiment was to examine semantic complexity by controlling the typicality of category exemplars, with the postulate that training atypical (more complex) items would facilitate greater generalization to untrained items than training typical (less complex) category exemplars. The basis for such a hypothesis stems from Rosch's (1975) seminal work showing that typical examples (e.g., *robin*) have a different status within semantic categories (e.g., *bird*) than atypical examples (e.g., *ostrich*). Since then, there has been extensive experimental evidence supporting preferential processing (i.e., faster reaction times) for typical, as compared to atypical items (the typicality effect; Hampton, 1993, 1995; Kiran & Thompson, 2003; Larochelle & Pineu, 1994; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974; Storms, De Boek, & Ruts, 2000).

Moreover, Plaut (1996) has investigated differential processing of atypical and typical examples in a computer simulated network. The network was trained to recognize a set of artificial typical and atypical words, 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.

Plaut's (1996) findings, although not tested in humans, are particularly relevant to treatment of Wernicke's aphasia, in that patients with Wernicke's aphasia do not show a typicality effect, as do normals and patients with Broca's aphasia (Grober, Perecman, Kellar, & Brown, 1980; Grossman 1981; Kiran & Thompson, 2003). That is, Wernicke's patients do not show a difference in reaction times between typical and atypical examples on category verification tasks. We, therefore, extend Plaut's complexity model to individuals with fluent aphasia. Using a semantically based treatment, we emphasized the featural detail of typical versus atypical items in an attempt to facilitate improved naming in patients with fluent aphasia. The treatment provided was motivated by prototype/family resemblance models of typicality (Hampton, 1993, 1995; Rosch & Mervis, 1975; for a review of models of typicality, see Komatsu, 1992). According to these models, categories are represented by a set of features that are more or less salient for defining the category prototype. The more similar a particular item in a category is to the prototype, the more typical it is judged to be. The less similar it is to the prototype, the less typical the item and, therefore, fewer other examples in the category share its features. A multidimensional scaling of similarity judgments in a category (Rosch & Mervis, 1975) would place typical examples in the center of semantic space and atypical examples at its periphery.

For the purpose of the present experiment, it was hypothesized that if indeed atypical examples are at the periphery of semantic categories, then training features associated with them would emphasize the variation of features within the category (e.g., *ostrich: runs, long legs; penguin: swims, eats fish*), as well as features of the prototype (e.g., *lays eggs, has beak*). Thus, features associated with the typical examples would be strengthened by atypical item training, and hence, access to typical items would be improved. Conversely, training semantic features of typical examples was not expected to result in generalization to intermediate or atypical items. Theoretically, typical examples entail little variation of semantic features within the category and, thus, training these examples was predicted to improve only items at the center that share similar features. We, therefore, hypothesized that within a category, atypical exemplars are more "complex" than typical ones, because collectively these items convey more diverse information about the category and its semantic features than typical items.

In the present experiment, we also examined the nature of naming errors occurring throughout treatment. We predicted that, if successful in improving naming, treatment would result in an evolution of errors reflective of enhanced access to both semantic and phonological targets during naming attempts. Prior to treatment, patients would be unable to access any specific information about target items, resulting in predominately neologistic errors, unrelated words, or no responses. Based on premises of interactive activation models, we predicted that semantically based treatment would facilitate improved access to semantic and phonological approximations of target words. Following treatment, therefore, we predicted a greater proportion of semantic and/or phonemic errors. Finally, performance on standardized language measures which examine aspects of lexical-semantic processing also was expected to improve.

Method

Participants

Four monolingual, English-speaking individuals with fluent aphasia, with pervasive naming deficits, participated in the study. The participants were selected from the Northwestern University Speech and Language Clinic and were recruited from stroke groups in the greater Chicago area. Several participant 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 9 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 degree (see Table 1). All participants also passed a pure-tone hearing screening at 40 db HL bilaterally at 500, 1000, and 2000 Hz and showed no visual impairment as measured by the Snellen chart. All participants had received varying amounts of traditional language treatment (with the exception of Participant 4), which was discontinued at least 3 months prior to the present study.

Table 1. Demographic and stroke related data for the 4 participants in the study.

	P1	P2	P3	P4
Age (in years)	64	63	72	75
Gender	Female	Male	Female	Female
Handedness	Right	Right	Right	Right
Occupation	Homemaker	Retired VP	Homemaker	Homemaker
Etiology	Left MCA	Left MCA	Left MCA	Left MCA
Months postonset	99	13	9	14

The diagnosis of fluent aphasia was determined by administration of the Western Aphasia Battery (WAB; Kertesz, 1982) and other standardized language measures. Results showed that all patients presented with fluent speech (range = 5–8), impaired comprehension (range = 5.8–6.8), and naming deficits (range = 20%–45% accuracy), and were able to comprehend written single words and phrases on the Reading Comprehension test of the WAB. All participants also showed impaired naming of high and low frequency items on the Boston Naming Test (BNT; Goodglass, Kaplan, & Wientraub, 1983; range = 7%–15% accuracy; see Table 2.)

Subtests of the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992) also were administered. Results showed that although some inconsistency was noted across participants on the spoken and written word to picture matching tasks (range = 57%–88%; range = 62%–90%, respectively), all participants demonstrated impairments in judging auditory and written word pairs (i.e., synonyms; range = 66%–70%, range = 60%–72%, respectively). Similarly, all participants had difficulties associating semantically related word pairs (high imageable pairs: range = 33%–60%, low imageable pairs: range = 0%–53%, respectively). These data indicated impairments in the semantic system for all participants. Performance on single word repetition (range = 58%–95%) and single word oral reading (range = 46%–79%) was varied, suggesting some phonological output lexicon impairment as well.

In addition to the aphasic participants, 30 normal young adults (range = 21–40 years) and 30 normal older adults (range = 41–75 years) participated in the various stimulus development tasks described below. Older participants were recruited from the Buehler Center on Aging Registry and from Northwestern University, and the young individuals were students and staff at Northwestern University. All participants had normal or corrected-to-normal vision, normal hearing, and at least a high school degree. Exclusionary criteria included history of neurological disorders such as stroke, transient ischemic attacks, Parkinson's disease, Alzheimer's disease, psychological illnesses, alcoholism, learning disability, seizures, and attention deficit disorders.

Stimuli

For purposes of the present experiment, norms for typicality of categories and their examples were developed prior to initiation of treatment, because previously published norms (Rosch, 1975; Uyeda & Mandler, 1980) are relevant only for young participants. The following sections describe the procedures used for stimulus development.

Development of Typicality Rankings

Ten normal young and 10 normal older participants were provided with a list of 12 superordinate category

Table 2. Individual histories and performance on the WAB (Kertesz, 1982), BNT (Goodglass et al., 1983), and PALPA (Kay et al., 1992) pre- and posttreatment.

	P1		P2		P3		P4	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
WAB								
AQ	43.4	54.4	50.9	51.50	70	79.7	46.4	58.00
Fluency score	5.00	8.00	7.00	7.00	8.00	8.00	7.00	8.00
Comprehension	5.80	7.50	5.85	6.95	6.80	8.05	5.70	7.60
Naming (%)	28.3	33.3	30.0	58.3	45.0	60.0	20.0	38.3
BNT (%)	15.0	15.0	13.0	28.3	8.3	36.7	6.7	11.7
PALPA								
Spoken word to picture matching (%)	56.7	90.0	87.5	95.0	70.0	85.0	87.5	100.0
Written word to picture matching (%)	61.7	85.0	90.0	90.0	62.5	82.5	82.5	90.0
Auditory word pair judgment (%)	68.3	78.3	68.3	86.7	66.0	80.0	70.0	88.3
Written word pair judgment (%)	68.3	70.0	71.7	86.7	68.3	80.0	60.0	78.3
High imageable-word association (%)	60.0	66.7	40.0	66.7	53.3	53.3	33.3	80.0
Low imageable-word association (%)	0.0	46.7	53.3	46.7	40.0	33.3	20.0	20.0
Written naming (%)	30.0	35.0	35.0	45.0	5.0	42.5	32.0	52.5
Single word oral reading (%)	79.2	70.8	45.8	58.3	79.2	95.8	54.2	70.8

Note. WAB = Western Aphasia Battery; AQ = Aphasia Quotient; BNT = Boston Naming Test; PALPA = Psycholinguistic Assessment of Language Processing in Aphasia.

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. Participants were instructed not to provide synonyms (e.g., *rabbit, hare*) or descriptive subordinate labels (e.g., *furniture: kitchen chair*).

Following completion of this task, responses from both the young and older participants were pooled, resulting in at least 50 examples for each category. A list with items for each superordinate category was then given to a new group of 20 participants (10 young and 10 older individuals). Two versions of the list were created, with the order of examples under each category randomized. Each version was given to half of the participants in each age group. Using instructions developed by Rosch (1975), participants were asked to rate on a 7-point scale, the extent to which each example represented their idea or image of the category term (typicality). A rating of 1 corresponded to the item being a very good example of the category; a rating of 7 indicated that item was considered a very poor example; a rating of 4 indicated a moderate fit. Participants were also required to mark *U* for examples that were unfamiliar to them (Malt & Smith, 1982). Once the participants completed the task, the average rating score, standard deviation, and median value for each example of each category were calculated across all participants.

Development of Treatment Categories and Their Examples

Several criteria were used to select categories and their examples to be used in treatment. First, categories were eliminated in which (a) more than 40% of the examples in a category were marked as being unfamiliar (e.g., *tools*), (b) an unequal distribution was noted where most examples were considered typical (e.g., *clothing*), and (c) atypical items overlapped in two categories (e.g., *foot* for *weapons* and *transportation*).

Several additional criteria were used to eliminate problematic examples within categories. For instance, examples that at least 60% (12 of 20) of the participants marked as unfamiliar (*U*) were eliminated. Examples were also excluded if they (a) had a standard deviation of 2 or more, (b) conveyed the same meaning, (e.g., *zeppelin* and *blimp* for *transportation*), (c) were both atypical and unfamiliar (e.g., *persimmon* for *fruit*), (d) were homophones (e.g., *duck*), (e) lacked any salient features (e.g., *finch, kale*), and (f) were questionable category members (e.g., *seaweed* for *vegetables*). Once specific examples from each category were eliminated, if the number of remaining examples in the category was below 30 the entire category was eliminated. Based on these criteria, 10 categories were eliminated: *occupations, transportation, sports, fruits, musical instruments, weapons, clothing, animals, furniture,* and *tools*. Two remaining categories (*birds*, *vegetables*) were selected for treatment.

Twenty-four items within each category were selected by converting typicality ratings for each item into *z* scores. For each category, items with the highest *z* scores ($N = 8$) were selected as typical examples (range = –1.0 to –.50), and items with the lowest *z* scores ($N = 8$) were selected as atypical examples (range = 1.0 to .07). Items with *z* scores ranging from –.49 to .01 were selected as intermediate examples ($N = 8$). In general, the 24 examples selected from each category were low frequency words according to written word frequency norms (Frances & Kucera, 1982), with one exception (*Chicken* = 49). Low frequency examples were selected to eliminate the possibility of generalization as a function of frequency rather than typicality. Each of the three sets (typical, atypical, and intermediate) for both categories (*birds* and *vegetables*) was matched for written word frequency and number of syllables (see Appendix A for a list of treatment items). Corresponding photos of each item were printed on 4 × 6 in. cards. Additionally, examples from other superordinate categories (*fruits, animals,* and *musical instruments*) were selected to serve as distractors during treatment. In summary, there were two treatment categories with 24 examples each and three distracter categories with 12 examples each.

Development of Semantic Features for Treatment

Thirty features that were either physically (e.g., *is red in color, has feathers*), functionally (e.g., *is made into pie, is a predator*), characteristically (e.g., *is juicy, lays eggs*), or contextually (e.g., *found in a grocery store, lives near water*) related to items in each category were selected from published norms (Barr & Caplan, 1987) and from the Internet. Only features that 18 of 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 (e.g., *birds: lays eggs, has a beak*), and 15 features were relevant to at least two items, and were used to reinforce features of both typical and atypical examples (e.g., *penguin, swan: swims, lives near water*). Finally, 20 distractor features belonging to the categories *sports, transportation, animals, insects, flowers,* and *weapons* were selected and were evenly distributed in terms of the attribute types (e.g., physical, functional, contextual, characteristic).

Design

A single participant experimental design (i.e., a multiple baseline design across participants and behaviors;

Connell & Thompson, 1986; McReynolds & Kearns, 1983) was used to examine acquisition of trained items and generalization to untrained items within and across categories. In addition to varying the number of baseline probes preceding treatment, the order of categories trained and typicality of stimulus sets within each category were counterbalanced across participants. Participants 1 and 4 received five baseline probes prior to treatment, whereas Participants 2 and 3 received three. Although Participants 1 and 2 were trained on *birds* first, Participants 3 and 4 were trained first on *vegetables*. Further, for Participants 1 and 3, the eight typical items were treated first, while the remaining intermediate and atypical items of the category ($N = 16$) and all 24 items of the untrained category were tested for generalization. For these 2 participants, if naming accuracy for the trained typical items achieved criterion level (7 of 8 naming for two consecutive sessions) and no improvement was observed in naming of the untrained intermediate or atypical items, treatment was shifted to the intermediate set. If no generalization to naming of the atypical group was noted, while the accuracy of the trained intermediate items achieved criterion, treatment was finally shifted to the eight atypical items. Once all the items of the set ($N = 24$) were acquired, treatment was shifted to the typical examples of the second category and the same procedure followed (however, see results for Participant 1). The same protocol was followed for Participants 2 and 4, except in this case all atypical items were treated first while the remaining intermediate and typical items of the category ($N = 16$) and all 24 items of the untrained category were probed as in baseline. For these 2 participants, if naming accuracy for the trained atypical items achieved criterion level with no generalization, treatment was shifted to the intermediate subset and, following the same criteria, to the typical subset. For all participants, two true baseline probes were acquired for the untrained (second) semantic category, prior to its treatment.

Baseline Naming Procedures

Confrontation naming of all 48 items (24 examples from each category) was tested during baseline. Participants were shown each picture (presented in random order) and were instructed to name the bird or vegetable depicted. Responses were considered correct if they were clear and intelligible productions of the target item occurring within 20 s of stimulus presentation. Self-corrected responses, dialectical differences, and distortion/substitution of one vowel or consonant (e.g., *rovin/robin*) were allowed. All other responses including (a) superordinate labels (e.g., *bird/cardinal*); (b) circumlocutory responses; (c) unrelated, out of the category responses (e.g., *apple/lettuce*); (d) no responses or "I don't know"; (e) neologisms (i.e., less than 50% of the word resembling the target, such as *barnett/chicken*); (f) semantic paraphasias (e.g., *pelican/seagull*); and (g) phonemic paraphasias (e.g., *bravin/robin*) were counted as incorrect responses. Percent correct named, as well as the percentage of each error type relative to all errors, was calculated.

Treatment

All participants were treated concurrently, once a day for 2 hr, two times per week. During each treatment session, participants performed the following steps for each of the eight 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. During the category sorting task, the examiner randomized 60 pictures, of which 24 were from the target category and 12 were from each of three distractor categories. Once the patient demonstrated 100% accuracy on picture sorting for 10 consecutive treatment sessions, this step was eliminated for each target item and was performed once at the beginning of every treatment session. For specific instructions that were used in treatment see Appendix B.

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 24 items of the category in training were administered prior to every second treatment session. Naming probes for both the trained and untrained category were also conducted at the completion of treatment for each subset (e.g., typical, intermediate, atypical). The order of presentation of items was randomized during each probe presentation.

Responses to naming probes, coded in the same way as in baseline, served as the primary dependent measure in the study. Additionally, evolution of errors and performance on standardized language tests were examined. Treatment was discontinued when naming accuracy of 7 of 8 items was observed for two consecutive sessions or when a total of 20 treatment sessions (10 probe sessions) were completed. Generalized naming to the untrained examples was considered to have occurred when levels of performance changed by 40% over baseline levels.

Follow-Up Probes

Naming accuracy for both categories was again assessed between 6 and 10 weeks following completion of the study. Procedures were identical to those used during baseline testing.

Reliability

All baseline and treatment probe sessions were recorded on audiotape and 30% of the treatment sessions were recorded on videotape. Reliability on the dependent variable (naming responses) was calculated for 38% of the probe sessions, resulting in 90%–100% agreement. Reliability on the independent variable (i.e., presentation of the treatment protocol) was calculated by an independent observer viewing videotaped sessions. Point-to-point agreement ranged from 95%–100%.

Results

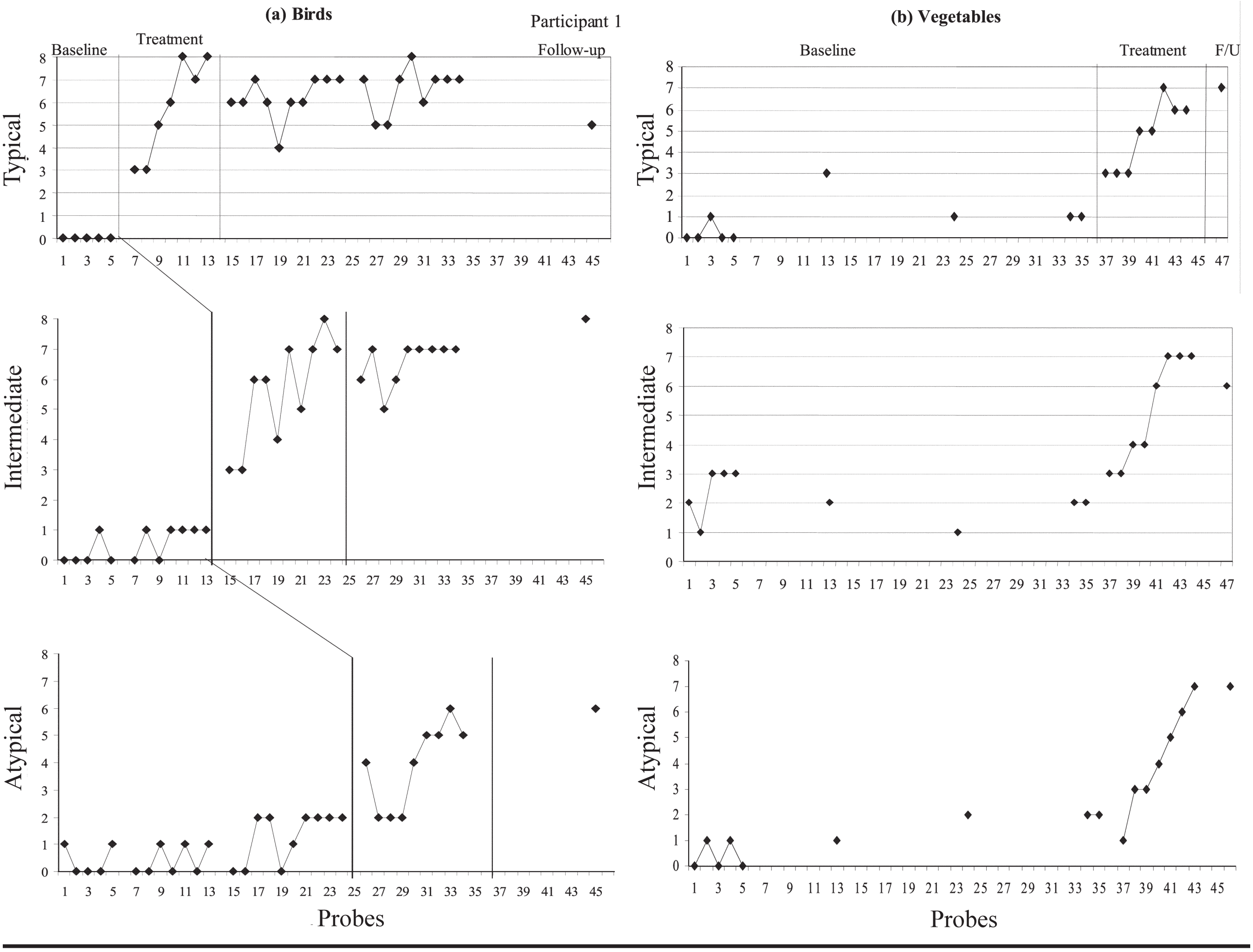
Naming Accuracy

Results are presented in Figures 1, 2, 3, and 4 in multiple baseline formats showing the number of items named correctly for each subset (typical, intermediate, 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 less than 2 points fluctuation across sessions), with the exception of Participant 3 who showed an increase in production of intermediate examples of *vegetables*.

Participant 1

Following baseline testing, treatment was initiated on typical items of *birds* for Participant 1, which resulted in acquisition of trained items, with training criterion met within 7 weeks. However, generalization to intermediate or atypical examples was not observed during this training. Only direct treatment of intermediate items resulted in their acquisition and, once again, this treatment had no effect on atypical items. Finally, when treatment was shifted to atypical examples, improvement was noted on the trained atypical items (see Figure 1a).

Figure 1. (a) Naming accuracy on typical, intermediate, and atypical items for the category *birds* and (b) naming accuracy on atypical, intermediate, and typical items for the category *vegetables* across baseline, treatment, and follow-up phases for Participant 1.



Administration of naming probes on the untrained category, *vegetables* indicated no changes throughout treatment focused on *birds* (see Figure 1b). Because this participant expressed frustration with treatment for the category *birds* (which required a total of 25 weeks to train) the second category (*vegetables*) was trained by first targeting atypical, rather than typical items as was originally planned. This alteration in the design also allowed for examination of differential responsiveness to typical or atypical training within the same participant. Results showed that when treatment was extended to atypical examples of *vegetables*, immediate improvement was noted and criterion was met in 8 weeks. More important, concomitant generalized naming of both untrained intermediate and typical items was observed during atypical exemplar training.

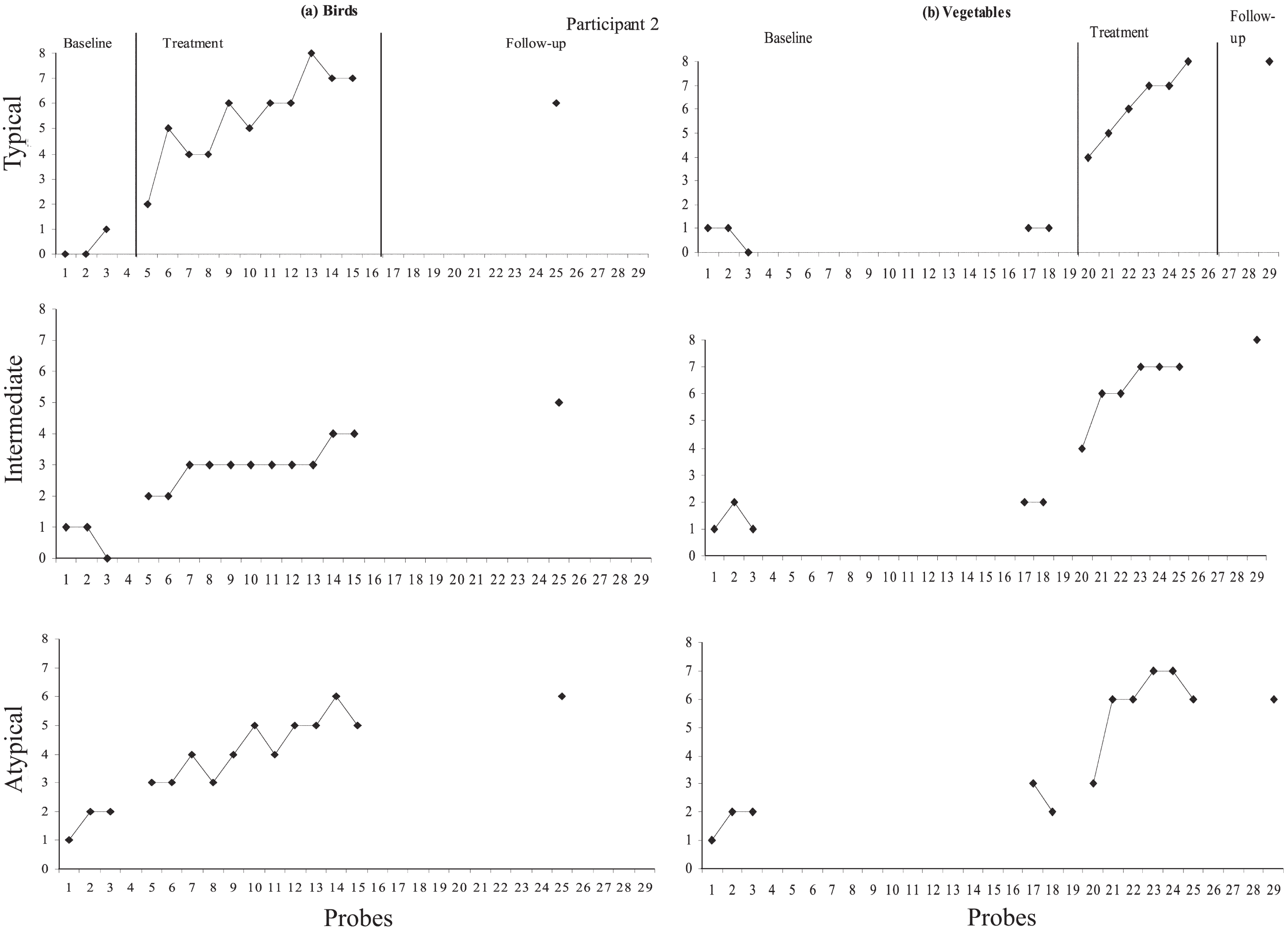
Participant 2

Participant 2 received treatment on atypical exemplars of *birds*. As can be seen in Figure 2a, this treatment resulted not only in improved production of trained items (criterion was reached in 11 weeks), but also generalization to both intermediate and typical items. Notably, no change in *vegetables* was noted during this training. However, when treatment was extended to *vegetables*, the training effect noted for *birds* was replicated. That is, treatment initiated on atypical *vegetables* items resulted in acquisition of both trained atypical items, and untrained intermediate and typical *vegetable* items within 6 weeks (see Figure 2b).

Participant 3

Participant 3 received treatment focused on typical *vegetables* items. The observed rising baseline occurring during the first three baseline probes disappeared on subsequent treatment probes of these items, eliminating any threat to internal validity imposed by unstable baselines. Like Participant 1, this treatment resulted in no generalization from typical items to intermediate or atypical items for Participant 3, even

Figure 2. (a) Naming accuracy on atypical, intermediate, and typical items for the category *birds* and (b) naming accuracy on atypical, intermediate, and typical items for the category *vegetables* across baseline, treatment, and follow-up phases for Participant 2.



though clear acquisition of trained items was noted. A similar effect was noted when treatment was shifted to intermediate examples of *vegetables*. Once again, an acquisition curve was noted for intermediate items, with no effect on atypical item naming. Only when treatment was directly applied to the atypical items was naming of these items improved (see Figure 3a). Because treatment for *vegetables* required a total of 28 weeks (approximately 7 months), treatment for the second category was not provided.

Participant 4

For Participant 4, treatment of atypical items resulted in an effect similar to that seen for Participant 2, who also received treatment focused on atypical items. For both categories, Participant 4 showed improved naming of trained items, reaching criterion in 6 weeks for *vegetables* and 9 for *birds*, while generalized naming to untrained intermediate and atypical items was also observed (see Figures 4a and 4b).

Follow-Up Probes

Results of follow-up probes conducted 6 to 10 weeks following treatment are also illustrated in Figures 1, 2, and 4. Participant 3 did not receive follow-up probes for health reasons. In general, on follow-up, participants demonstrated naming performance higher than initial baseline levels and near the mean performance of the last two probes during treatment, indicating maintenance of training effects.

Evolution of Errors

Errors produced during the first two baseline sessions and the last two treatment probes for each category were

Figure 3. (a) Naming accuracy on typical, intermediate, and atypical items for the category *vegetables* across baseline and treatment phase and (b) naming accuracy on typical, intermediate, and atypical items for the category *birds* during baseline and throughout *vegetable* training for Participant 3.

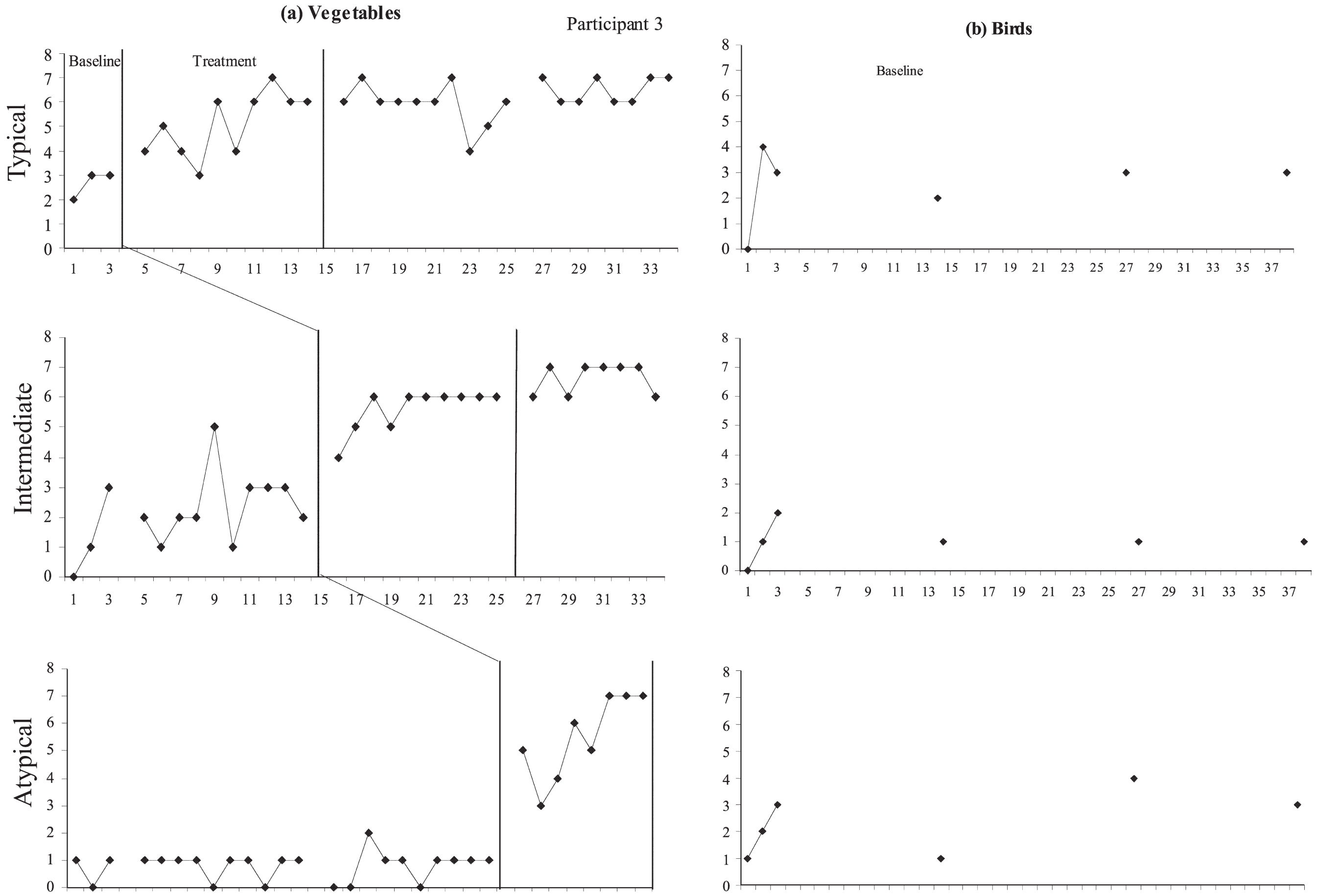
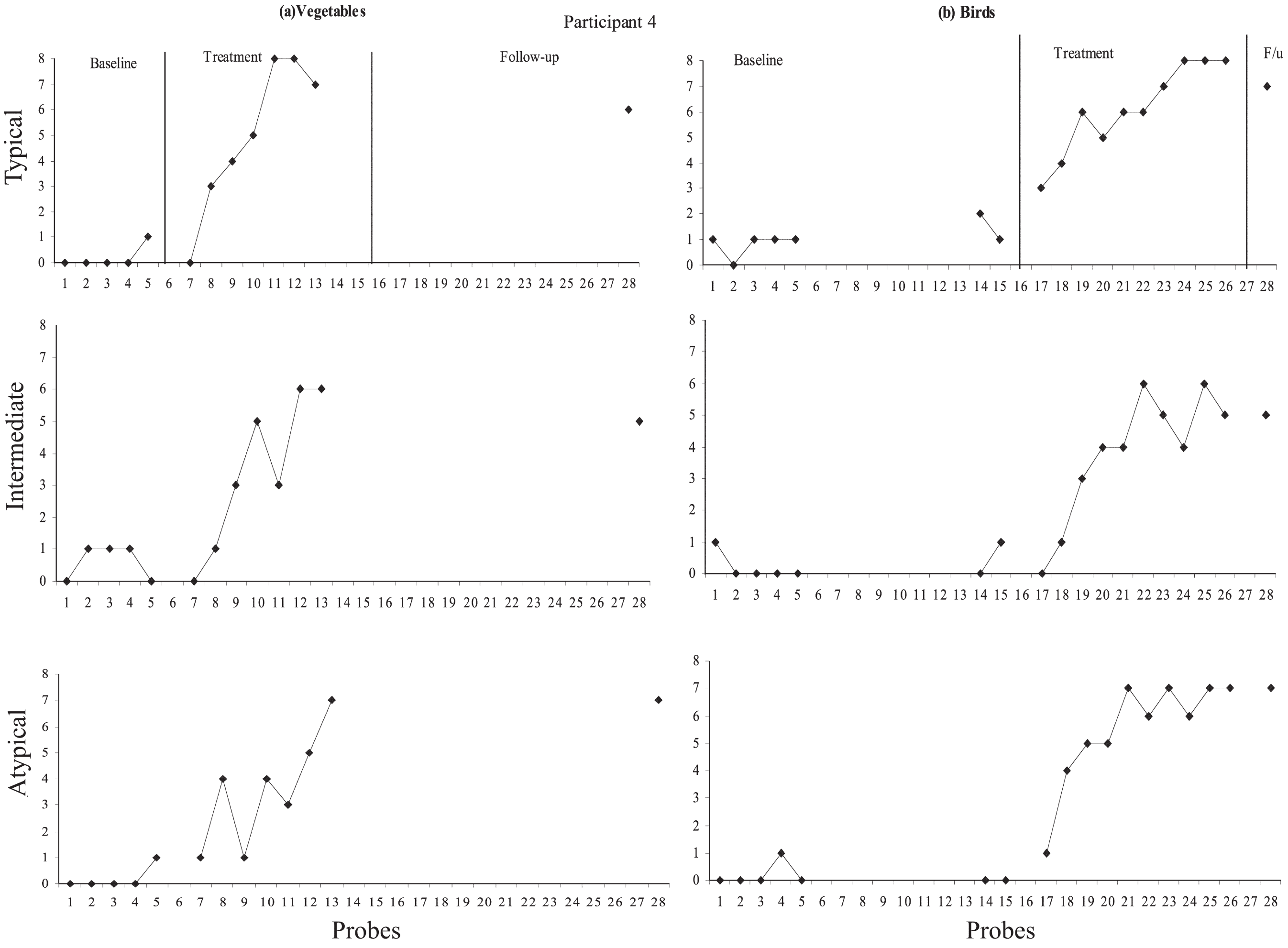


Figure 4. (a) Naming accuracy on atypical, intermediate, and typical items for the category *vegetables* and (b) naming accuracy on atypical, intermediate, and typical items for the category *birds* across baseline, treatment, and follow-up phases for Participant 4.



compared for each participant. Errors were collapsed into general (superordinate, circumlocutions, neologisms, and no responses) and specific (semantic and phonemic errors). The proportions of errors by type are included in Table 3.

Chi-square analyses, examining change in error types, showed significant effects for all participants. For Participant 1, changes were significant following treatment of both *birds,* $\chi^2(1, N = 55) = 24.5$, $p < .0001$, and *vegetables,* $\chi^2(1, N = 49) = 3.6$, $p < .05$; for Participant 2, significant changes were noted following treatment of *birds,* $\chi^2(1, N = 56) = 8.95$, $p < .01$, and *vegetables,* $\chi^2(1, N = 44) = 4.3$, $p < .05$; for Participant 3, changes were significant for *vegetables,* $\chi^2(1, N = 30) = 4.39$, $p < .05$; and for Participant 4, changes were significant following treatment of *vegetables,* $\chi^2(1, N = 55) = 22.9$, $p < .001$, and *birds,* $\chi^2(1, N = 52) = 27.8$, $p < .001$. For all participants, decreases in the proportion of general responses and increases in specific errors were noted as both categories were trained. The order of treatment (whether typical or atypical items were trained first) had no effect on the nature of errors produced.

Pre–Post Standardized Language Measures

In general, all 4 participants demonstrated improvements on the standardized language tests conducted following completion of treatment (see Table 2). Improvements were noted on the auditory comprehension subtests of WAB, while small or no improvements were noted on naming subtests on the WAB and BNT. All 4 participants demonstrated improvements on the PALPA subtests, where test scores improved by an average of 10 points after treatment compared to pretreatment (P1: $M = 14.8\%$, $SD = .18$; P2: $M = 10.4\%$, $SD = .10$; P3: $M = 13.5\%$, $SD = .13$; P4: $M = 17.6\%$, $SD = .13$).

Table 3. Evolution of errors, reported in total number of errors produced (in italics) and percentage of specific errors to total errors.

	P1		P2		P3		P4	
Error type	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**	**Pre**	**Post**
Birds total	45	10	43	13			45	7
Superordinate (%)	24.4	10	2.33	0			44.4	0
Circumlocution (%)	13.3	0	0	0			6.67	0
Unrelated (%)	0	0	0	0			15.6	0
No responses (%)	42.2	10	2.33	7.69			0	28.6
Neologisms (%)	17.8	20	69.8	23.1			31.1	0
Semantic (%)	2.2	0	18.6	30.8			0	14.3
Phonemic (%)	0	60	6.9	38.5			2.22	57.1
Vegetables total	39	10	38	6	23	7	46	9
Superordinate (%)	0	0	0	0	0	0	45.7	0
Circumlocution (%)	15.4	0	0	0	0	0	2.17	0
Unrelated (%)	2.56	0	0	0	8.7	14.3	17.4	0
No responses (%)	10.3	10	47.4	0	60.9	14.3	2.16	0
Neologisms (%)	25.6	10	13.2	16.7	0	0	19.6	11.1
Semantic (%)	35.9	0	21.1	33	30.4	71.4	13.5	44.4
Phonemic (%)	10.3	80	18.4	50	0	0	0	44.4

Discussion

Results of this experiment demonstrate that training atypical items within a category and their semantic features results in generalization to naming of intermediate and typical examples of the category selected for training. This effect was replicated on five separate occasions across 3 participants—for Participants 2 and 4 for both categories trained and for Participant 1 for the second category (*vegetables*). In addition, the lack of generalization resulting from applying treatment to typical items within a category was replicated on two occasions across two participations—for Participant 1 for *birds* and Participant 3 for *vegetables*. These data provide strong evidence that training atypical exemplars is superior to training typical ones for facilitating generalization to untrained items. The finding is further strengthened by considering the data from Participant 1, who received initial treatment focused on both typical (for *birds*) items and atypical (for *vegetables*) items, showing a complete lack of generalization under typical item training and robust generalization under atypical item training. These findings have important implications for treatment of naming deficits in patients with fluent aphasia. First, they show that semantically based treatment, emphasizing underlying aspects of semantic representation and processing is a successful approach for training naming, as has been shown in other studies examining semantic featural approaches (Boyle & Coehlo, 1995; Drew & Thompson, 1999). Indeed, this method results in stronger training effects than when semantic cueing or picture–word matching techniques are used as in, for example, Davis and Pring (1991) and Howard et al. (1985b).

Secondly, our data indicate that the complexity effect found when controlling syntactic complexity in treatment of sentence deficits for agrammatism as noted by Thompson and colleagues (Thompson et al., 1997, 1998, 2003) extends to the lexical–semantic domain in treating aphasic naming deficits. That is, training more complex items, which encompass variables relevant to simpler items, facilitates greater access to untrained items than training simple items. This effect likely results because, as shown by Plaut (1996) in connectionist simulation, exposure to items sharing some features of the prototype as well as disparate features results in activation of both typical and atypical entries, whereas exposure to items with features similar to a semantic prototype results in a high probability of activating only a limited set of items with comparable features. Similarly, our findings with aphasic patients showed that training atypical items (which were quite dissimilar to the category prototype with regard to semantic features) highlighted the featural variation within categories, whereas training the featural detail of typical items repeatedly emphasized only a few features that are common among typical items. Recall that for each category, 30 features were selected, of which 15 were defining features relevant to all typical items within the category (e.g., *has a beak,* for *birds)*. In addition, characteristic features were selected, some of which were more applicable to the typical items (e.g., *small in size, lives in trees*), while others covered a wider range relevant to the atypical items (e.g., *has long neck, lives near water* for *birds*). The main difference, then, between training typical examples and atypical examples concerned the variation of semantic features that were encountered in treatment. Training typical items repeatedly emphasized only a few features that were common among typical items; training atypical items (which were dissimilar to the category prototype) highlighted the featural variation within the category.

We recognize, however, that complexity in the semantic domain may be manifested differently than what is observed in the syntactic domain. That is, for sentence level deficits, grammatical representations and relations among elements are denoted through syntactic trees with a discernable hierarchical structure. Semantic representations (specifically, semantic categories), however, are considered to be represented in multidimensional vector spaces depending on the degree of featural overlap (e.g., Hampton, 1993; Rosch & Mervis, 1975; Smith et al., 1974), although Jackendoff (1976, 1983) and others have also characterized lexical–semantic features in terms of tree structures. However, even if the representations involved in the two domains differ substantively, the complexity effect appears to emerge when the items selected for training encompass information relevant to untreated ones (i.e., they are in a subset relation; Thompson et al., 2003). Here, typical items are composed of a subset of all features encountered within a semantic category. We, therefore, conclude, as did Thompson et al. (2003), that complexity is an overarching principle of recovery and generalization.

It is also noteworthy that the patients who received treatment on atypical items acquired all 24 items of the category much faster than the patients trained on typical examples (e.g., Participant 2 required 6 weeks to acquire all *vegetable* items compared to the 28 weeks required by Participant 3). Therefore, the present results suggest that training atypical examples is a more time efficient method for facilitating naming of category exemplars than training typical examples.

Another important aspect of the present results concerns the evolution of errors noted throughout the study. As noted above, although the number of errors was minimal by the end of treatment, all participants demonstrated significant changes from pre- to posttreatment in the type of errors produced. Prior to treatment, predominantly general errors (i.e., superordinate labels, neologisms, and no responses) were produced, indicating failure to access specific semantic and/or phonological detail for target items. Following treatment, accurate naming of trained items and generalized naming to untrained items illustrated situations in which access to both semantic and lexical/phonological representations was successful. The evolution of general errors at the beginning of treatment to semantic and phonemic errors at posttreatment suggest that although treatment resulted in a greater excitatory influence at the semantic/phonemic level during naming attempts, some type of interference occurred in the process of correctly selecting the target word from activated entries. These data, then, provide further evidence of the treatment effect, suggesting that the present treatment was successful in enhancing spreading activation to semantically related targets within the category, but that it was not entirely successful in ameliorating interference from multiple activated entries at either the semantic or phonological levels in naming attempts.

Finally, treatment resulted in improvements on standardized language measures that were conducted prior to and after treatment. Although improvements were negligible on the BNT and the WAB, improvements were evident on the auditory comprehension and semantic processing subtests of PALPA. It could be argued that lack of generalization to items on BNT and across categories minimizes the clinical significance of the present treatment. However, it should be emphasized that in order to obtain maximal experimental control, semantic features were carefully selected to be applicable only to the categories trained and, therefore, were not applicable to untrained categories. One process all patients underwent during treatment was making explicit judgments about semantic features that were both imageable (e.g., *does it have wings?*) and nonimageable (e.g., *is it a predator?*). The effect of such practice was evident on the semantic processing subtests on the PALPA, as participants improved by at least 10% on the semantic processing subtests.

We also point out that two important predictors for language recovery, namely, severity of aphasia and months postonset of the stroke (Kertesz, 1984), were controlled in this experiment and, therefore, had no influence on the generalization patterns observed. For instance, Participant 3 had the least severe deficit (WAB AQ = 70) prior to initiation of treatment and was inducted into treatment only 9 months following her stroke. This patient received treatment for typical examples and did not generalize to naming of intermediate or atypical examples. In contrast, Participant 4 presented with a moderately severe aphasia prior to treatment (AQ = 46.4) and reportedly was denied individual treatment prior to the present experiment, because her prognosis for improvement was considered poor. When trained on atypical examples, however, this patient generalized to naming of intermediate and typical examples and also demonstrated remarkable improvements on all other measures of language assessed through the course of treatment. Finally, Participant 1, who presented with moderately severe aphasia (AQ = 43.4) and was 11 years postonset of stroke, did not generalize to intermediate and atypical examples when trained on typical examples, but generalized to intermediate and typical examples when trained on atypical examples.

Conclusion

The current findings provide support for a semantically based treatment, focused on the featural detail of category items, for training naming in patients with fluent aphasia. The strong generalization effects observed in the present study also indicate that the items selected for treatment within categories are important to consider, in that training atypical items within semantic categories results in generalization to untrained items, whereas training typical items does not. These data suggest that the complexity account of treatment efficacy advanced by Thompson et al. (2003) extends to the semantic domain. That is, like treatment for sentence production deficits in patients with agrammatic aphasia (Thompson et al., 2003) and that for children with developmental phonological deficits (e.g., Geirut, 2001), the most effective approach for training naming seems to be to train more complex material first. These findings challenge the long-standing clinical notion that treatment must begin from simpler tasks and proceed to the more difficult ones. Instead, evidence such as that from the present study illustrates the facilitative effects of training more complex items that encompass variables relevant to simpler items.

Acknowledgments

Parts of this work comprised the first author's doctoral dissertation at Northwestern University. The authors wish to thank Drs. Douglas Medin, Karla McGregor, and James Booth for their valuable contributions as committee members. The authors also thank Dr. David Plaut for his helpful comments on the manuscript. This research was supported in part by the National Institute of Health Grant R01-DC01948 (Cynthia K. Thompson) and the Cognitive Science Program at Northwestern University.

References

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

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.

Butterworth, B., Howard, D., & Mclaughlin, P. (1984). The semantic deficit in aphasia: The relationship between semantic errors in auditory comprehension and picture naming. *Neuropsychologia, 22,* 409–426.

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

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

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 92,* 283–321.

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

Drew, R. L., & Thompson, C. K. (1999). Model based semantic treatment for naming deficits in aphasia. *Journal of Speech, Language, and Hearing Research, 42,* 972–990.

Ellis, A. W., Miller, D., & Sin, G. (1983). Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition, 15,* 111–144.

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

Geirut, J. A. (2001). Complexity in phonological treatment: Clinical factors. *Language, Speech, and Hearing Services in Schools, 32,* 229–241.

Goodglass, H. (1980). Disorders of naming following brain injury. *American Scientist, 68,* 647–655.

Goodglass, H. (1998). Stages of lexical retrieval. *Aphasiology, 12,* 287–298.

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

Grober, E., Perecman, E., Kellar, L., & Brown, J. (1980). Lexical knowledge in anterior and posterior aphasics. *Brain and Language, 10,* 318–330.

Grossman, M. (1981). A bird is a bird: Making references within and without superordinate categories. *Brain and Language, 12,* 313–331.

Hampton, J. A. (1993). Prototype models of concept representation. In J. A. Van Mechelen, R. S. Hampton, T. Michlanski, & P. Theuns (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.

Hillis, A., & Caramazza, A. (1995). The compositionality of lexical semantic representation: Clues from semantic errors in object naming. *Memory, 3,* 333–358.

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

Howard, D., Patterson, K., Franklin, S., Orchid-Lisle, V., & Morton, J. (1985b). The treatment of word retrieval deficits in aphasia: A comparison of two therapy methods. *Brain, 108,* 817–829.

Jackendoff, R. (1976). Toward an explanatory semantic representation. *Linguistic Inquiry*, 7, 89–150.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

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

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

Kertesz, A. (1984). Neurobiological aspects of recovery from aphasia in stroke. *International Rehabilitation Medicine, 6*(3), 122–127.

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

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112,* 500–526.

Larochelle, S., & Pineu, H. (1994). Determinants of response time in the semantic verification task. *Journal of Memory and Language, 33,* 796–823.

Le Dorze, G., Boulay, N., Gaudreau, J., & Brassard, C. (1994). The contrasting effects of a semantic versus a formal semantic technique for the facilitation of naming in a case of anomia. *Aphasiology, 8,* 127–141.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10,* 69–75.

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

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

Nickels, L., & Best, W. (1996). Therapy for naming disorders (Part I): Principles, puzzles and progresses. *Aphasiology, 10,* 21–47.

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

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

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

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–604.

Schmidt, A. M. & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis.* Champaign, IL: Human Kinetics.

Shelton, J. R., & Caramazza, A. (1999). Deficits in lexical and semantic processing: Implications for models of normal language. *Psychonomic Bulletin & Review, 6*(1), 5–27.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model of semantic association. *Psychological Review, 81,* 214–241.

Stemberger, J. P. (1985). An interactive model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Hillsdale, NJ: Erlbaum.

Storms, G., De Boek, P., & Ruts, W. (2000). Prototype and exemplar based information in natural language categories. *Journal of Memory and Language, 42,* 51–73.

Thompson, C. K. (1988). Generalization in the treatment of aphasia. In L. V. McReynolds & J. Spradlin (Eds.), *Generalization strategies in the treatment of communication disorders* (pp. 82–115). Philadelphia: B. C. Decker.

Thompson, C. K., Ballard, K., & Shapiro, L. (1998). Role of syntactic complexity in training wh- movements structures in agrammatic aphasia: Optimal order for promoting generalization. *Journal of International Neuropsychological Society, 4,* 661–674.

Thompson, C. K., Shapiro, L. P., Ballard, K. J., Jacobs, B., Schneider, S. L., & Tait, M. (1997). Training and generalized production of wh and NP movement structures in agrammatic aphasia. *Journal of Speech and Hearing Research, 40,* 228–244.

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.

Received April 4, 2002

Accepted November 13, 2002

DOI: 10.1044/1092-4388(2003/048)

Contact author: Swathi Kiran, PhD CCC-SLP, Rm 7.206, Department of Communication Sciences and Disorders, University of Texas, Austin, Texas 78712.
Email: s-kiran@mail.utexas.edu

Appendix A. Stimuli used for treatment.

Typical	Intermediate	Atypical
	Birds	
Bluebird	Eagle	Pheasant
Bluejay	Seagull	Peacock
Cardinal	Pigeon	Pelican
Robin	Owl	Chicken
Crow	Cockatoo	Ostrich
Woodpecker	Falcon	Turkey
Hummingbird	Swan	Penguin
Parrot	Vulture	Flamingo
Avg. frequency = 1.2	Avg. frequency = 4.2	Avg. frequency = 8.5
	Vegetables	
Carrot	Green beans	Scallions
Broccoli	Onion	Parsley
Cauliflower	Gourd	Artichoke
Celery	Mushroom	Kidney beans
Cucumber	Eggplant	Pumpkin
Lettuce	Lima beans	Rhubarb
Radish	Yam	Olive
Spinach	Tomato	Garlic
Avg. frequency = 2.6	Avg. frequency = 5.7	Avg. frequency = 1.5

Appendix B. Treatment protocol for each target item.

1. *Picture naming.* Initially, the participant was presented with the picture and was asked to name it. Irrespective of whether the picture was named correctly or not, the experimenter explained that he/she would now be aided in understanding more about the item.
2. *Category sorting.* The examiner placed written category cards (*birds/vegetables, animals, fruits, musical instruments*) on the table in random order. The examiner then randomized the 60 pictures and presented them 1 at a time for the participant to sort by superordinate category, by placing each picture on its written category card. If incorrect, the picture was placed under the accurate category label by the examiner.
3. *Feature selection.* For this task, an icon board with separate slots for the target picture and six semantic features was used. The examiner placed the target picture (e.g., *chicken*) in the center of the icon board and provided the participant with 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 *chicken: lays eggs, is food* were acceptable semantic features, whereas *flies distance,* and *swims* were features that were not applicable. Once six features were selected, the participant was required to read aloud the selected features.
4. *Yes/no questions.* The participant was asked questions about the target example and was required to answer yes or no in response. The experimenter then asked the patient 15 questions about the target example (e.g., *chicken*), which included five acceptable semantic features (e.g., *does it have wings?*), five unacceptable semantic features from the same category (e.g., *can it fly?*), and five semantic features from a different category (e.g., *is it made of metal?*).
5. *Picture naming.* Same procedure as Step 1.