Semantic Complexity in Treatment of Naming Deficits in Aphasia: Evidence From Well-Defined Categories

Swathi Kiran
Lauren Johnson
University of Texas at Austin

Purpose: Our previous work on manipulating typicality of category exemplars during treatment of naming deficits has shown that training atypical examples generalizes to untrained typical examples but not vice versa. In contrast to natural categories that consist of fuzzy boundaries, well-defined categories (e.g., *shapes*) have rigid category boundaries. Whether these categories illustrate typicality effects similar to natural categories is under debate. The present study addressed this question in the context of treatment for naming deficits in aphasia.
Methods: Using a single-subject experiment design, 3 participants with aphasia received a semantic feature treatment to improve naming of either typical or atypical items of shapes, while generalization was tested to untrained items of the category.

Results: For 2 of the 3 participants, training naming of atypical examples of shapes resulted in improved naming of untrained typical examples. Training typical examples in 1 participant did not improve naming of atypical examples. All 3 participants, however, showed weak acquisition trends.
Conclusions: Results of the present study show equivocal support for manipulating typicality as a treatment variable within well-defined categories. Instead, these results indicate that acquisition and generalization effects within well-defined categories such as shapes are overshadowed by their inherent abstractness.

Key Words: aphasia, semantic complexity, typicality, well-defined categories

Treatment for naming deficits in participants with aphasia is an issue of considerable research focus (Maher & Raymer, 2004; Nickels, 2002). Recent studies have investigated the effects of treatment targeted at the level of naming impairment. For instance, facilitating semantic access to improve lexical access has typically involved tasks such as matching auditory and written words to pictures, answering yes/no questions about the target, spoken word categorization, and relatedness judgment tasks (Boyle, 2004; Boyle & Coehlo, 1995; Davis & Pring, 1991; Drew & Thompson, 1999; Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1985a, 1985b; Wambaugh et al., 2001). Several of these studies have focused on facilitating access to trained items and generalization to untrained items.

For example, the semantic feature analysis approach (Boyle, 2004; Boyle & Coehlo, 1995) has been successful in facilitating improvements of trained items and in generalization to semantically related untrained items across categories. Our previous work examining the influence of exemplar typicality on naming accuracy also espouses the notion of generalization to untrained items (Kiran, in press; Kiran &

Thompson, 2003). This work stemmed from a connectionist simulation that examined relearning after damage within a computer network (Plaut, 1996). Plaut found that retraining a lesioned computer network on atypical items resulted in improvements in recognition of typical items as well. However, training typical items improved performance only on trained items, while performance on atypical words deteriorated. In two studies (Kiran, in press; Kiran & Thompson, 2003) examining animate and inanimate categories, we have demonstrated that training atypical examples resulted in improvements to untrained typical examples in participants with aphasia. In contrast, training typical examples did not result in generalization to untrained atypical examples. These trends were obtained for 9 participants with either fluent or nonfluent aphasia except for 1 participant with fluent aphasia receiving treatment for the category *clothing* (Kiran, in press).

In an attempt to replicate findings by Kiran and Thompson (2003), Stanczak, Waters, and Caplan (2006) recently examined the effect of typicality treatment in 2 participants with anomic aphasia. Stanczak et al. found that 1 participant who had semantic deficits and was trained on atypical examples

demonstrated generalization to untrained typical examples within the category *birds* as well as marginally significant generalization from trained typical examples to untrained atypical examples within the category *vegetables*. The second participant who did not show semantic deficits did not learn atypical examples within *vegetables* and did not show generalization from typical to atypical examples within *birds*. Stanczak et al. argued for the utility of deficit-oriented treatment in the individual with semantic deficits. Further, Stanczak et al. suggested that the generalization patterns observed from atypical to typical words may be due to large-scale changes in the semantic network after treatment for atypical examples.

The treatment studies reviewed above have used natural language categories that consist of living things (i.e., categories like *birds, vegetables*) or man-made items (e.g., *clothing, transportation*). These categories are generally considered to be fuzzy, that is, they have no clear boundaries separating members from nonmembers (McCloskey & Glucksberg, 1978). These categories are inherently different from well-defined categories that have a clear definition (e.g., *even numbers*: numbers that are divisible by 2; *planets*: large bodies in the solar system) and rigid category boundaries (e.g., *4 is either an even number or not*). Well-defined categories are such that most participants can list all members (e.g., *season, continent, measurement unit*). What sets well-defined categories apart from natural categories is the nature of their representation. According to Armstrong and colleagues (Armstrong, Gleitman, & Gleitman, 1983), well-defined categories consist of a conceptual core with specific attributes relevant to specific examples, which make it possible to define them fairly specifically (e.g., *uncle: male paternal figure*). Larochelle and colleagues (Larochelle, Richard, & Soulierres, 2000) have argued that well-defined categories may have more defining properties and fewer characteristic properties. It has been suggested recently that categorization in well-defined categories occurs in a rule-based fashion (Hampton, 1998; Keil, Carter Smith, Simons, & Levin, 1998). In contrast, for natural categories such as *fruit* and *bird,* it has been hypothesized that category membership depends on their similarity to a prototype (Smith & Medin, 1981).

The representation of typicality within well-defined categories is, however, under debate. In a seminal paper, Armstrong and colleagues (1983) examined categories thought to have clear defining properties (e.g., *female* and *shapes*) and found that participants in their study believed examples belonging to these well-defined categories met membership requirements to the same degree (i.e., all-or-none membership). However, when asked to rate the typicality of examples within well-defined categories, participants showed gradation in their ratings; that is, some examples were rated more typical than others (e.g., *mother* is considered more typical of the category *female* than *cowgirl*). Armstrong et al. further compared participants' performances on natural categories to their performance on well-defined categories and found that both natural and well-defined categories yielded graded responses in a typicality rating task. In addition, although mean response times were overall longer for all members of the well-defined categories, the authors found that response times during the category verification task were relatively longer for the more atypical members of both natural and well-defined categories. In contrast, Larochelle et al. (2000) found faster reaction times for typical exemplars of well-defined categories but argued that factors other than typicality may have influenced their results. Larochelle et al. found that the typicality effect disappeared when category dominance (i.e., the frequency with which the category label is produced) and familiarity were controlled.

In a recent study, we examined the effects of typicality during online category verification for three well-defined categories, *shapes, females,* and *body parts* (Kiran, Shamapant, & Delyria, 2006). Results demonstrated that for normal controls but not participants with aphasia, typical examples were processed faster than atypical examples, but this effect was significant only for shapes. Typicality effects were not observed for the other two well-defined categories tested, females and body parts. These results suggested that only some well-defined categories such as shapes demonstrated graded representations.

If the results obtained by Larochelle et al. (2000) and Kiran et al. (2006) are valid, then it appears that well-defined categories (other than shapes) do not show the expected typicality effects and consequently may not be graded. This notion violates the fundamental assumption driving the predictions of selective atypical to typical generalization that we have demonstrated in our previous work (Kiran, in press; Kiran & Thompson, 2003). Importantly, we have argued (Kiran, 2007) that longer processing times during category verification tasks for atypical examples reflect the nature of their representation in the periphery of the category and consequently make them more complex than typical examples. However, within the context of well-defined categories, because atypical examples and typical examples may be represented differently, generalization from atypical to typical examples is not expected to occur. Thus, it may not matter if typical or atypical examples are trained in such categories if one wishes to facilitate their naming by individuals with aphasia. On the other hand, if well-defined categories are represented no differently than natural fuzzy categories, and atypical examples take longer to respond to than typical examples, then selective generalization patterns would be predicted to occur.

The present study examined the effect of typicality training in participants with aphasia using well-defined categories. If indeed well-defined categories are represented similarly to other natural language categories, then their representation of semantic attributes and lexical representations is akin to a connectionist network consisting of nodes across two levels (semantic and phonological) that are linked through bidirectional connections. Within this network, attributes that define membership in the category (e.g., *shapes: can be drawn*) and shared prototypical features (e.g., *can be measured by a ruler*) carry less weight in their representation of the category, as most examples in the category possess these features. Typical examples (e.g., *octagon, cylinder*) comprise these features and therefore have less influence on other category examples. Atypical examples (e.g., *spade, asterisk*), however, consist of defining features but also consist of characteristic features (e.g., *curvy borders, symbolic*) that more

broadly represent the variation in the category. Secondly, features belonging to typical examples have a subset relationship to those of atypical examples (e.g., *three dimensions* is a feature often associated with atypical examples, whereas *two dimensions* is a feature associated with typical examples).

Based on this framework, it was hypothesized that a semantic-feature-based typicality treatment would improve access to semantic features and corresponding phonological representations for atypical examples as well as typical examples. Further, training atypical examples should improve access to untrained typical examples but not vice versa (i.e., no generalization from trained typical to untrained atypical examples). Stemming from the unresolved status of these categories in psycholinguistic research, an alternate hypothesis was proposed, wherein representation of examples in these categories was determined by factors other than typicality. Therefore, training atypical or typical examples would be inconsequential to the overall outcome of improving naming of items within the category.

Method

Development of Stimuli

To assist in development of norms for stimuli, 50 normal young and older individuals with no history of neurological impairment were recruited. Fifteen normal young individuals age 20 to 26 years (*M* = 23 years) and 15 normal older individuals age 40 to 84 years (*M* = 53 years) participated in category selection and typicality rating tasks. Additionally, 10 normal young individuals 19–29 years old (*M* = 23.6) and 10 normal older individuals 45–70 years old (*M* = 55.5) participated in the naming and semantic feature tasks.

Five potential categories (*body parts, females, shapes, colors,* and *males*) determined to be well-defined based on an initial survey conducted on 30 normal adults were selected as preliminary options to be used as categories in this study. In each of these categories, 35 potential stimuli were selected for use in the typicality rating section of this study. Identical to procedures used in Kiran and Thompson (2003), the same participants were asked to rate the extent to which each example represented their idea or image of the category term (typicality). Participants also marked U next to items that they were unfamiliar with (Malt & Smith, 1982).

Development of treatment categories and their examples. The average rating scores were converted into *z* scores to account for individual variability. For each category, items with the highest *z* scores were considered to be the more typical examples of the category, and the items with the lowest *z* scores were considered to be the more atypical examples. Several exclusionary criteria were employed to eliminate problematic examples. These included (a) if 60% or more of the participants marked an item as unfamiliar (e.g., *fuchsia* for *colors*), (b) if an item had a standard deviation of 2.5 or higher (e.g., *trident* for *shapes*), (c) if an item could potentially belong to another category (e.g., *heart* for *shapes*), (d) if 2 or more examples conveyed the same meaning (e.g., *prism/pyramid*), (e) if an item was not picturable (e.g., *gentleman* for *males*), and (f) if an item's membership in a category was questionable (e.g., *peach* for *color*). Based on

the above criteria, the entire categories of *colors, males,* and *body parts* were eliminated. Because none of the participants reported in the study received treatment for *females,* even though stimuli from both *females* and *shapes* were normed, only aspects of stimuli development relevant to *shapes* are discussed.

A picture (5 in. × 7 in.) representing each stimulus for *shapes* was found as clip art on the Internet and using Art Explosion software. To ensure name agreement, 20 normal individuals named each picture. There was no difference in naming accuracy between typical items (85%) and atypical items (81%), *t*(18) = 0.52, *p* = .60. Stimuli were controlled for written word frequency (Frances & Kucera, 1982), familiarity and imageability (MRC Psycholinguistic Database; see Coltheart, 1981), and number of syllables (see Appendix A for a list of stimuli). Finally, 20 examples from two superordinate categories (*body parts* and *musical instruments*) were selected to serve as distracters.

Development of semantic features for treatment. The same participants who named the pictures were also involved in development of semantic features for treatment. Each participant was asked to produce at least 10 semantic features for *shapes.* Of the 50 total features that were generated, 10 were core/defining features of the category, those that were equally prevalent in both typical and atypical examples. The remaining features were characteristic features either more relevant to typical examples (*N* = 15) or to atypical examples (*N* = 15). Further, the 50 features were equally distributed across easily recognizable visual features, geometric physical features, and functional features. Finally, 15 distracter features were also selected (see Appendix B for a list of features).

Participants

Three monolingual, English-speaking individuals with aphasia (age range = 54–75 years) participated in the treatment experiment. The participants were recruited from local hospitals within the Austin, Texas, area. Several initial selection criteria were met, including (a) a single left hemisphere stroke in the distribution of the middle cerebral artery confirmed by a computed tomography/magnetic resonance imaging scan, (b) onset of stroke at least 7 months before participation in the study, (c) premorbid right-handedness as determined by a self-rating questionnaire, and (d) at least a high school diploma (see Table 1). All participants also passed an audiometric hearing screening at 40 db HL bilaterally at 500, 1000, and 2000 Hz, and showed no visual impairment as measured by the Snellen chart. All participants had received some traditional language treatment during the initial months following their stroke but were not involved in concurrent therapy during the study.

The diagnosis of aphasia was determined by administration of the Western Aphasia Battery (WAB; Kertesz, 1982). All 3 participants were classified as having anomic aphasia. Further, performance on the Boston Naming Test (BNT; Goodglass, Kaplan, & Weintraub, 1983) was at or below 81% accuracy (see Table 1). All participants demonstrated less than 85% accuracy on two or more subtests from the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992) and the

TABLE 1. Participants' demographic data.

Participant	Sex	Age	Occupation	Site of lesion	MPO	Type of aphasia	WAB AQ
1	M	54	Body mechanic	Left MCA	11	Anomic	82.5
2	M	75	Lawyer	Left thalamus	7	Anomic	84.3
3	F	58	Insurance agent	Left parietal, temporal, left BG	36	Anomic	87.3

Note. M = male; F = female; MPO = months postonset; WAB AQ = Western Aphasia Battery Aphasia Quotient; MCA = middle cerebral artery; BG = basal ganglia.

Pyramids and Palm Trees (PAPT; Howard & Patterson, 1992; see Table 2). Based on our previous work (Kiran, Ntourou, & Eubanks, 2007; Kiran & Thompson, 2003), it was reasoned that participants demonstrating impairments on semantic processing tasks were most likely to benefit from the semantic feature-based naming treatment. Participant 1 (P1) performed below criterion on the written word–picture matching, word pair judgment (auditory and written word) subtests on the PALPA, and the three-word subtest on PAPT. Participants 2 (P2) and 3 (P3) performed below criterion on auditory and written word pair judgment on the PALPA. Written naming was tested to examine whether lexical retrieval impairment was limited to spoken output or across output modalities. Compared to the other 2 participants, P1 was relatively impaired on the written naming subtest of the PALPA. Overall, even though P1 had less difficulty naming items on the BNT, this participant was included in the treatment because he presented with marked semantic impairments and was unable to name any of the experimental stimuli. In contrast, both P2 and P3 were impaired on naming pictures on the BNT but showed relatively milder semantic processing impairments.

Design

A single-subject experimental design with multiple baselines across behaviors (Connell & Thompson, 1986; Thompson, 2006) was used to examine acquisition of trained items and generalization to untrained items. For all participants, one set of items ($N = 10$) within a category (either typical or atypical) was introduced into treatment, while the untrained items within the trained category ($N = 10$) remained in baseline. The order of typicality entered into treatment was counterbalanced across participants. Thus, P1 and P3 received treatment for atypical examples of *shapes*, whereas P2 received treatment for typical examples of *shapes*. If naming accuracy on the trained items achieved criterion level of 8/10 accuracy for two consecutive sessions and no improvement was observed on the untrained items, treatment was then shifted to the untrained set. Naming accuracy on the previously trained

TABLE 2. Pretreatment and posttreatment performance on the WAB, BNT, PALPA, and PAPT.

	Participant 1		Participant 2		Participant 3	
	Pre	Post	Pre	Post	Pre	Post
WAB						
I. Spontaneous Speech	15	18	17	19	18	20
II. Auditory Verbal Comprehension	8.1	8.65	9	10	9.15	9.45
III. Repetition	9.8	9.8	7.8	7.9	8.2	6.2
IV. Naming	8.3	7.9	8.3	7.3	8.3	8.9
Aphasia Quotient	82.5	88.6	84.3	88.4	87.3	89.1
BNT (%)	81	80.0	**26.7**	**56.7**	**55.0**	**71.0**
PALPA (%)						
Letter Length Reading	**70.8**	**95.8**	100.0	100.0	100.0	100.0
Spoken Word–Picture Matching	97	92.5	95.0	100.0	97.5	97.5
Written Word–Picture Matching	**82.5**	**92.5**	95.0	95.0	100.0	100.0
Auditory Synonym Judgments	83.5	78.3	**66.7**	**93.3**	81.7	73.3
Written Synonym Judgments	81.6	85.0	81.7	85.0	81.7	85.0
Spoken Picture Naming	97.5	87.5	**82.5**	**92.5**	90.0	95.0
Writing Picture Naming	67.5	75.0	90.0	95.0	95.0	97.5
Reading Picture Naming	97.5	100.0	97.5	97.5	97.5	97.5
Repeating Picture Naming	**87.5**	**100.0**	82.5	90.0	97.5	97.5
PAPT (%)						
3 pictures	94.2	98.1	94.2	94.2	92.3	92.3
3 words	**80.2**	**96.2**	88.5	94.2	88.5	92.3

Note. Changes exceeding 10% are highlighted in bold. Pre = pretreatment; Post = posttreatment; WAB = Western Aphasia Battery (Kertesz, 1982); BNT = Boston Naming Test (Goodglass et al., 1983); PALPA = Psycholinguistic Assessment of Language Processing in Aphasia (Kay et al., 1992); PAPT = Pyramids and Palm Trees (Howard & Patterson, 1992).

set was monitored for maintenance of treatment effects. If, however, improvement on the untrained items exceeded 40% change over maximum baseline levels, generalization was considered to have occurred and treatment was not initiated on that set of stimuli (Kiran, in press; Kiran & Thompson, 2003).

Baseline Naming Procedures

Confrontation naming of 40 items (20 examples from *shapes* and *females*) was tested during baseline. Each participant was shown a picture (presented in random order) and was instructed to name the shapes or females depicted. Feedback as to accuracy of response was not given, but intermittent encouragement was provided.

Analysis of Errors

During baseline and treatment sessions, responses were considered correct if they were self-corrected responses, dialectical differences, or distortion/omission/substitution of one vowel or consonant (e.g. *octanogon/octagon*) of the target item. All other responses were classified into (a) no response/"I don't know," (b) unrelated words (e.g., *streets/asterisk*), (c) neologisms (utterances with less than 50% phonetic overlap with the target; e.g., *scholl/ribbon*), (d) circumlocution (e.g., *not four equal sides, one side that's too long/rectangle*), (e) semantic paraphasias (e.g., *egg/oval*), or (f) phonemic paraphasias (utterances with greater than 50% phonetic overlap with the target; e.g., *octacle/octagon*). There were no perseverations, mixed errors, or superordinate labels in the data collected; hence these error types were not coded. Percentage correct named as well as the percentage of each error type relative to all errors was calculated. High scores on the category *females* by all subjects precluded its use in the experiment.

Treatment

Participants were treated consecutively. Treatment was conducted two times per week for approximately 2 hr each session. P1 received a total of 15 weeks of treatment, P2 received 18 weeks, and P3 received 8 weeks. Each participant was trained to name either 10 typical words or 10 atypical words following procedures in Kiran and Thompson (2003). Treatment steps for each word within the category were as follows:

1. Naming the picture (e.g., *cylinder* for *shapes*)

2. Sorting pictures of the target category with two distracter categories (*body parts, musical instruments*)

3. Selection of six written semantic features relevant to the target item from the set of semantic features and verbal discussion of the features with the clinician

4. Answering 15 yes/no questions of which 5 belonged to the target example (e.g., *has volume*), 5 belonged to the category but not the target example (e.g., *has different degree of angles*), and 5 did not belong to the target category (e.g., *sounds a siren*)

5. Naming the target (e.g., *cylinder*)

During the category-sorting task, the examiner randomized 40 pictures, of which 20 were from the target category and 20 were from two distracter categories (10 *musical instruments* and 10 *body parts*). This step was performed once at the beginning of every treatment session. Participants were also allowed to write the target word being practiced during step 1 to facilitate written feedback. Hence, both orthographic and phonological information was provided for the trained items.

Treatment Probes

Throughout treatment, naming probes similar to those used in the baseline condition were presented to assess naming of the trained and untrained items. Naming probes for all 20 items of the category in training were administered before every second treatment session. 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.

Data Analysis

The extent to which change from baseline to treatment phase was statistically reliable was determined through a time series analysis using the C statistic (Tryon, 1982). Since the C statistic suffers from the drawback of being too lenient, effect sizes (ES) were also calculated comparing the mean of all data points in the treatment phase relative to the baseline mean divided by the standard deviation of baseline (Busk & Serlin, 1992). Based on comparable studies in aphasia, an ES of 4.0 was considered small, 7.0 was considered medium, and 10.1 was considered large (Beeson & Robey, 2006).

Reliability

All the baseline and probe sessions were audio-recorded. Online reliability by an independent observer seated behind a one-way mirror was obtained for 86% of P1's probe sessions, 50% of P2's probe sessions, and 87% of P3's probe sessions. Point-to-point agreement was 95% across probe sessions. Daily scoring reliability checks by the independent observers were undertaken to ensure accurate presentation of the treatment protocol by the clinician. Point-to-point agreement was 100%. An independent scorer blind to purposes of the study performed the error analysis.

Results
Naming Accuracy

Results are presented in Figures 1, 2, and 3 in multiple baseline formats showing percentage correct responses for typical and atypical subsets for *shapes* for baseline and treatment phases of the experiment.

Participant 1 (atypical shapes training). P1 improved to 70% accuracy with treatment but did not reach criterion ($C = 0.11$, $z = 0.43$, $p = .31$, ES = 1.6). At the same time, naming of untrained typical examples increased from 10% to 60% ($C = 0.50$, $z = 2.2$, $p < .01$, ES = 6.11). This participant's performance was variable through the course of treatment,

FIGURE 1. Naming accuracy in percentage points for atypical (trained) and typical items (untrained) for the category *shapes* during baseline and treatment phases for Participant 1.
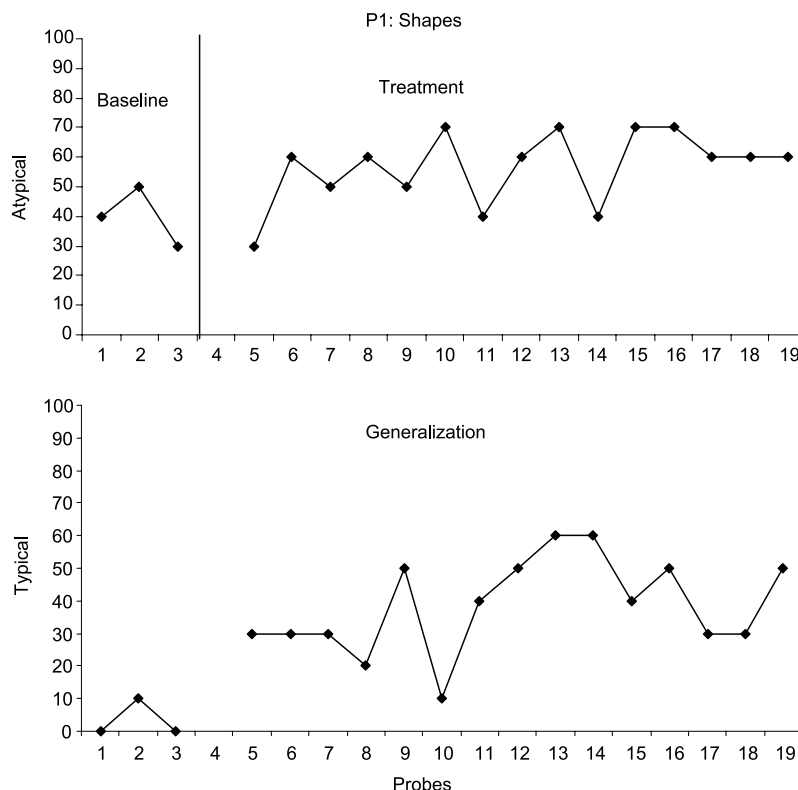
however (see Figure 1); performance on the untrained typical examples was more consistent than performance on the trained atypical examples.

Participant 2 (typical shapes training). Performance on those items improved from a high of 30% to a high of 60% (which was not significant; $C = 0.02$, $z = 0.08$, $p = .53$, $ES = 1.67$). There was no change in performance of the untrained atypical examples. Hence, treatment was then shifted to atypical examples, which then improved to criterion ($C = 0.90$, $z = 4.0$, $p < .001$, $ES = 31.8$). Even though treatment was no longer administered on the previously trained typical examples, performance on these items was maintained up to the final session.

Participant 3 (atypical shapes replication). P3 was administered three baselines, following which treatment was initiated on atypical examples of *shapes*. This participant demonstrated an increasing baseline but stabilized at 50%. With treatment, accuracy on the trained items improved to 100% ($C = 0.85$, $z = 3.118$, $p < .001$, $ES = 2.6$). At the same time, naming accuracy on the untrained typical examples improved from 20% to a high of 60% ($C = 0.70$, $z = 2.55$, $p = .01$, $ES = 11.3$).

Evolution of Errors

For each participant, errors produced during the three baseline sessions and equal numbers of sessions at the end

of treatment were compared (see Table 3 for proportions of errors by category). Chi-square tests revealed significant changes for P2, $\chi^2(3, N = 82) = 7.8$, $p < .05$, and P3, $\chi^2(5, N = 59) = 13.84$, $p < .01$. P1 showed a reduction in the proportion of no responses and semantic errors after treatment, although these effects were not significant, $\chi^2(3, N = 79) = 3.4$, $p > .05$. A modest increase in phonemic errors at the end of treatment was also noted. P2 showed a slight increase in the proportion of circumlocutions and no responses and a decrease in the proportion of semantic errors. Finally, P3 showed an increase in the number of no responses and phonemic errors and a decrease in the number of semantic errors. The few unrelated words, neologisms, and circumlocutions disappeared by the end of treatment for this participant.

Pre-Post Standardized Language Measures

All tests administered before initiation of treatment were readministered on completion of treatment and are shown in Table 2. Improvements of 5 points or more on the WAB Aphasia Quotient were considered clinically significant (Katz & Wertz, 1997). Improvements in naming on the BNT exceeded 10% for P2 (27%–57% accuracy) and P3 (55%–70% accuracy), whereas negligible changes were observed for P1 (81%–80% accuracy). All participants showed varying improvements on specific subtests of the PALPA and the PAPT that are designed to examine semantic processing

FIGURE 2. Naming accuracy in percentage points for typical (trained) and atypical (untrained) items for the category *shapes* for Participant 2. Treatment is shifted to the untrained atypical examples during the second phase of treatment. Performance on the previously trained typical examples is assessed in the form of maintenance probes.
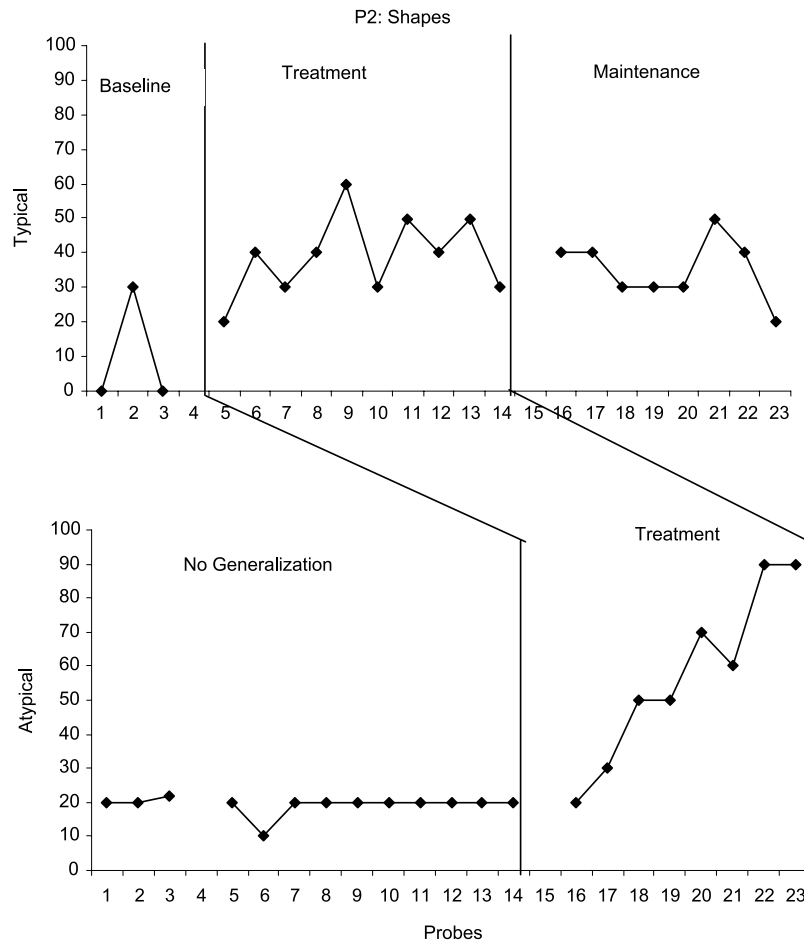
(see Table 2). For instance, P1 showed improvements exceeding 10% only on the written word–picture matching subtest of the PALPA and the three written words test on the PAPT. P2 improved on the auditory synonym judgment, spoken picture naming, and three words subtests. None of the changes demonstrated by P3 exceeded the preset criterion of 10%.

Discussion

This experiment was undertaken to examine the effectiveness of the typicality treatment approach in facilitating lexical retrieval and generalization in individuals with naming deficits using well-defined categories. Three main findings emerged from the study. First, well-defined categories show generalization from atypical examples to typical examples, leading to the conclusion that some well-defined categories such as *shapes* are indeed graded. Second, the present results are less robust than our previous treatment studies both in terms of acquisition of treated items and

generalization to untrained items. Also, improvements on standardized tests and evolution of error patterns observed in the present study were also not as convincing as observed in our previous work (Kiran, in press; Kiran & Thompson, 2003). Finally, the results of the present study show that the complexity approach appeared to work with some, but not all, similar participants with aphasia. These results will be discussed in further detail.

Selective Generalization in Well-Defined Categories

Results of the present study revealed that training atypical examples resulted in generalization to untrained typical examples in P1 and P3. In contrast, training typical examples did not result in improvements on the untrained atypical examples in P2. It should be noted that this participant also showed no appreciable changes to the previously trained typical examples in subsequent atypical training. It is interesting to note that effect sizes for the trained items were minimal (P1 = 1.6; P2 = 1.6; P3 = 2.6), with the exception

FIGURE 3. Naming accuracy in percentage points for atypical (trained) and typical (untrained) items for the category *shapes* during baseline and treatment phases for Participant 3.
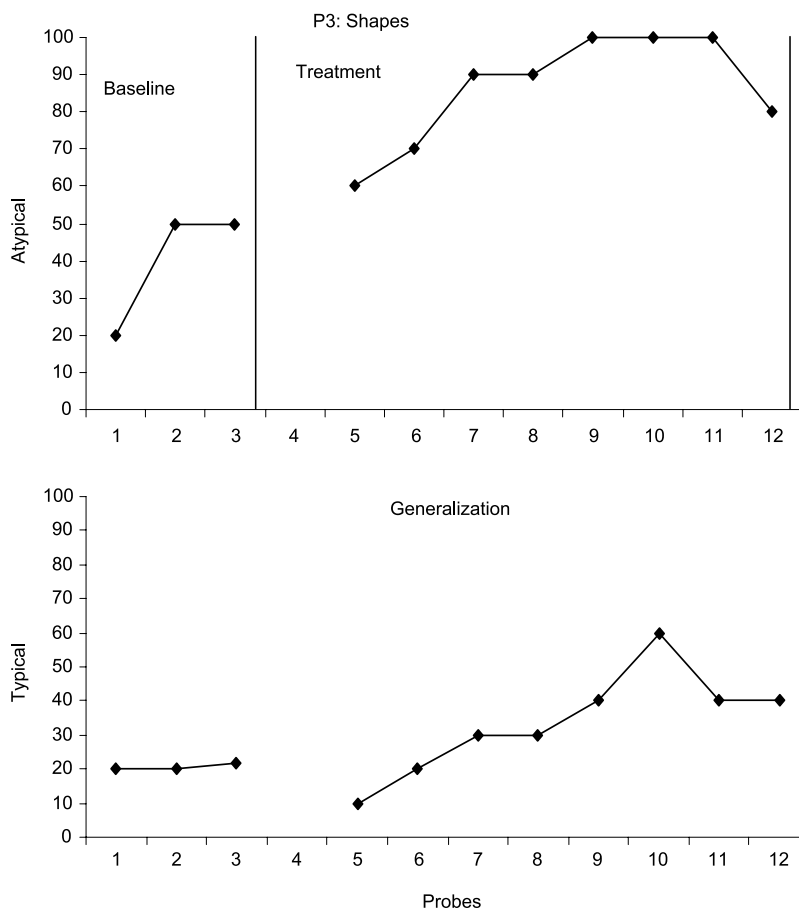
of atypical training for P2 (ES = 31.8). In contrast, the effect size for the untrained generalization items was 6.11 for P1 and 11.3 for P3. The present results, therefore, provide limited support for the beneficial effects of training the more complex atypical examples as opposed to training less complex typical examples to facilitate generalization to untrained items within well-defined categories. The notion of complexity

as a manipulable variable in treatment is supported by increasing evidence from various domains. For instance, the application of complexity has been demonstrated within the syntactic domain in treatment for sentence production deficits in participants with agrammatic aphasia (Thompson & Shapiro, 2007) and within the phonological domain in children with phonological deficits (Gierut, 2007).

TABLE 3. Evolution of errors reported in total number of errors produced (in italics) followed by proportion of error type within the total number of errors.

	Participant 1		Participant 2		Participant 3	
	Pre	Post	Pre	Post	Pre	Post
Shapes	*48*	*31*	*54*	*28*	*43*	*16*
No response/"I don't know" (%)	16.6	12.9	**1.8**	**14.2**	**13.9**	**50.0**
Unrelated word (%)	0	0	5.5	0	2.3	0
Neologism (%)	0	0	0	0	2.3	0
Circumlocution (%)	2.0	3.2	68.5	71.43	9.3	0
Semantic (%)	**77.0**	**67.7**	**24.0**	**14.29**	**65.12**	**31.25**
Phonemic (%)	**4.1**	**16.1**	0	0	**6.98**	**18.75**

Note. Changes over 10% are highlighted in bold.

Recall that the main assumption at the outset of the study was that gradedness within the category was essential to facilitate the selective generalization patterns from atypical examples to typical examples. Specifically, in order for atypical examples to facilitate access to typical examples (but not vice versa), representation of and access to atypical examples had to be different from typical examples. Data from our online category verification study (Kiran et al., 2006) suggest that this is the case for *shapes* but not for *females* and *body parts*. That is, atypical examples were responded to more slowly and less accurately than typical examples for *shapes* but not for *females* and *body parts*. The results of the present study suggest that within the category of *shapes,* selective generalization from atypical to typical (but not from typical to atypical) does occur.

Certainly, a comparison of treatment effects between *females* and *shapes* would have conclusively resolved the issue of differential levels of gradedness in well-defined categories. None of the 3 participants reported here were trained on the category *females* because accuracy levels during baseline were high. Two participants who withdrew from this treatment were provided with treatment on *females.* One received treatment for 10 weeks, at which point he chose to terminate. During the time of his participation, he improved on naming of atypical examples to 50% accuracy, although no generalization to the untrained typical examples was observed yet. While it is impossible to conclude whether acquisition and generalization patterns would have continued had his treatment been extended, the results raise the issue of whether certain well-defined categories are more graded than others and whether these differences translate to selective generalization patterns after the typicality treatment.

Treatment in Well-Defined Categories Compared to Natural Categories

Although the results of the present study support the benefit of training atypical examples over typical examples within a category, treatment effects observed in this study were not as robust as those reported in our previous studies for perhaps a number of reasons. First, the 3 patients showed variable baselines prior to initiation of treatment, a feature that was not true in previous work. Second, patterns of evolution of errors in the present study also differed from our previous studies. That is, we have previously observed that participants typically produce no responses and unrelated words before treatment and that these patterns eventually evolve into semantic and phonemic errors by the end of treatment (Edmonds & Kiran, 2006; Kiran, in press; Kiran & Thompson, 2003). This finding is consistent with the predictions of a computational model proposed by Dell and colleagues (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). These authors have argued that nonwords, formal paraphasias, and no responses tend to occur in participants with more severe naming deficits, whereas semantic and mixed errors arise independently of naming severity (Dell et al., 1997; Schwartz & Brecher, 2000). In addition, with recovery, participants with a range of severity of deficits exhibit an increase in semantic errors. Likewise, participants in our

previous studies have demonstrated a shift toward semantic errors as their naming abilities improved. In the present study, however, initial errors were predominantly semantic errors and circumlocutions that in due course either reduced or resolved to no responses. Therefore, it is reasonable to assume that the patterns of evolution in the present study reflect subtle shifts toward a meta-awareness regarding the inability to access lexical items, a likely endpoint in the spectrum of failed lexical retrieval.

While recognizing the caveat that the differences in treatment effects using natural and well-defined categories may be extremely subtle, there are two possible explanations for why treatment effects in well-defined categories are not as robust as other natural language categories. As noted earlier, it could be argued that well-defined categories are less well-graded than are natural language categories because they have more defining properties and few characteristic properties compared to natural categories (Larochelle et al., 2000). In our experiment, however, there were 10 core/defining features and 40 characteristic features utilized in treatment. Thus, with respect to the semantic features used in treatment, every effort was made to represent the variation of attributes similar to procedures that were employed in previous treatment experiments using natural animate and inanimate categories.

An alternate, possibly more plausible, explanation may be that the category *shapes* is more abstract and less functional than previous categories entered into treatment (e.g., *vegetables/clothing*). Indeed, this proposition would be compatible with the finding that 2 of the 3 participants did not achieve the preset criterion of 80% accuracy across two consecutive sessions on the trained items (P1 trained on atypical *shapes,* P2 trained on typical *shapes*). Second, all 3 individuals went on to participate in subsequent treatment programs and all showed positive generalization from trained items to untrained items for the specific stimuli entered into treatment. Finally, even though we did not analyze accuracy on the semantic feature judgment task during treatment, reports from the daily treatment notes revealed that participants struggled with comprehension of specific features such as *multidimensional, has obtuse angles, has perpendicular lines,* and *has pointed corners.* Oftentimes, participants attempted personal associations with a specific example (e.g., *arch: sign for McDonald's*). In a series of studies, Marshall and colleagues have examined the effectiveness of personalized cues to facilitate word retrieval in participants with aphasia. For instance, Freed, Celery, and Marshall (2004) compared the effectiveness of personalized semantic cues with phonological cues and found that personalized cues were more effective in facilitating word retrieval and in the long-term retention of target words. The authors argued that for each participant, personalized cues were tied to individualized events in memory, hence making them more meaningful than the phonological cues. Taken together with the results of the present study, Freed et al.'s findings suggest that training well-defined categories such as *shapes* may not be productive in facilitating improvements in overall lexical retrieval, and the extent of improvements observed in treatment depend on the relevance of the treatment stimuli to real-life situations.

Differential Responsiveness to Complexity Treatment

We have suggested that several factors likely influence generalization patterns in response to the typicality treatment. Whereas the nature of the typicality was one factor and the focus of the present study, individual differences among the participants in treatment also may play a role. All 3 participants demonstrated anomic aphasia as measured by the WAB Aphasia Quotient. P1 also demonstrated semantic impairments as measured by the PAPT and PALPA; this participant clearly generalized from atypical to typical examples. In contrast, P2 presented with milder semantic impairments and did not show generalization from typical to atypical examples or from trained atypical to the previously trained typical examples.

One possible explanation may be that a semantic-based treatment has maximal benefit when the deficit corresponds to the treatment provided. This explanation was also suggested by Stanczak et al. (2006). Even though Stanczak et al. employed a slightly different treatment protocol (i.e., the feature verification task was a yes/no task), the results of Participant RB in their study are consistent with the present data. Given that participants in these studies showed differential responsiveness to the typicality treatment, such heterogeneity might be the result simply of individual differences. This issue raises the question of whether administering group studies with larger numbers of participants might yield different results. While detecting subtle differences in generalization patterns across participant groups in the presence of substantial intersubject variability appears less likely, future experiments that systematically compare treatment effects using different categories with the same individual can potentially provide greater insight into the influence of participant variability on treatment outcomes.

Conclusions

Results of this study provide equivocal support for manipulating typicality as a treatment variable within well-defined categories such as *shapes*. This is because even though participants show generalization patterns from atypical to typical examples that are consistent with the complexity hypothesis, the variable baselines and weak acquisition effects do not suggest any benefit from training these categories and their atypical (or typical) examples. From a clinical standpoint, the present results appear to indicate that acquisition and generalization effects within well-defined categories such as shapes are overshadowed by their inherent abstractness, rendering those difficult categories to train and of questionable utility in terms of real-world usage. In effect, although this form of training is developing a substantial track record, clinicians are reminded to be sensitive to other issues such as usage and individual variability that may affect results.

Acknowledgments

References

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13,* 263–308.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychological Review, 16*(4), 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*(4), 94–98.

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

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. Part III: Using flexibility to design or modify experiments. *Journal of Speech & Hearing Disorders, 51*(3), 214–225.

Davis, A., & Pring, T. (1991). Therapy for word-finding deficits: More on the effects of semantic and phonological approaches to treatment with dysphasic participants. *Neuropsychological Rehabilitation, 1*(2), 135–145.

Dell, G. S., Schwartz, M. F., Martin, N. M., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and non-aphasic 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,* 972–989.

Edmonds, L., & 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.

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

Freed, D., Celery, K., & Marshall, R. (2004). Effectiveness of personalized and phonological cueing on long-term naming performance by aphasic subjects: A clinical investigation. *Aphasiology, 18,* 743–757.

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

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

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65*(2-3), 137–165.

Howard, D., & Patterson, K. (1992). *Pyramids and palm trees.* London: Harcourt Assessment.

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

Howard, D., Patterson, K., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985b). Treatment of word retrieval deficits in aphasia. *Brain, 108,* 817–829.

Katz, R. C., & Wertz, R. T. (1997). The efficacy of computer-provided reading treatment for chronic aphasic adults. *Journal of Speech and Hearing Research, 40,* 493–507.

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

Keil, F. C., Carter Smith, W., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition, 65*(2-3), 103–135.

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

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

Kiran, S. (in press). Typicality of inanimate category exemplars in aphasia: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research.*

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

Kiran, S., Shamapant, S., & Delyria, S. (2006). Typicality within well-defined categories in aphasia. *Brain and Language, 99,* 149–151.

Kiran, S., & Thompson, C. K. (2003). 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.

Larochelle, S., Richard, S., & Soulierres, I. (2000). What some effects might not be: The time to verify membership in "well-defined" categories. *The Quarterly Journal of Experimental Psychology, 53A*(4), 929–961.

Maher, L., & Raymer, A. (2004). Management of anomia. *Topics in Stroke Rehabilitation, 11*(1), 10–21.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory and Cognition, 10*(1), 69–75.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory and Cognition, 6,* 462–472.

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

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language, 52*(1), 25–82.

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*(1), 62–91.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

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.

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders, 39*(4), 266–291.

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

Tryon, W. W. (1982). A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis, 15,* 423–429.

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 participants with different levels of deficit. *Aphasiology, 15,* 933–950.

Contact author: Swathi Kiran, 1100, 1 University Station, Department of Communication Sciences and Disorders, University of Texas, Austin, TX 78712. E-mail: s-kiran@mail.utexas.edu.

	Typical	Atypical
	Octagon	Arch
	Cone	Bell
	Diamond	Spade
	Cylinder	Semicircle
	Rectangle	Clover
	Cube	Prism
	Oval	Spiral
	Hexagon	Crescent
	Parallelogram	Asterisk
	Pentagon	Ribbon
Mean frequency	7.6	6.46
Mean familiarity	306	297

Appendix B

List of Semantic Features Used for *Shapes*

Functional	Geometric	Visual	Distracter
Ways for an architect to design/used in architecture	Has acute angles	Have more than one angle	Builds nests
Can be a building shape	Has obtuse angles	Have more than one side	Has 2 wheels
Ways to cut food	Has right angles	Two dimensional	Has feathers
Suits for decks of cards	Have area	Three dimensional	Grows underground
Religious symbols	Have volume	Multiple dimensions	Has a beak
Symbolic	Have arcs	Is a geometrical figure	Has webbed feet
Have different applications in different contexts	Have planes	Multiple straight lines	Grows on a vine
Can be found in nature	Can be measured by a ruler	Curvy edges	Has an engine
Can be drawn by humans	Its area can be calculated	Has rounded edges	Uses gasoline
Are the same across languages	Its volume can be calculated	Multiple curves	Lays eggs
Can be mechanically reproduced	Has different degrees of angles	Easily recognizable	Has handlebars
Used to define characteristics of many objects	All angles are the same	Is recognized as an outline or solid figure	Has a peel
Used in traffic and constructions	Has 4 right angles	No points	Usually eat it raw
Used in calculus/algebra	Has parallel sides	Half of another shape	Has petals
Used in statistics	Has perpendicular lines	Part of another shape	Found in a crime scene
On a keyboard	Can be equilateral	Has line segments	
Seen in picture books	Has pointed corners		
Most children know this shape			