APHASIOLOGY
2009, iFirst Article, 1–19

Psychology Press
Taylor & Francis Group

Treatment for lexical retrieval using abstract and concrete words in persons with aphasia: Effect of complexity

Swathi Kiran and Chaleece Sandberg

University of Texas at Austin, TX, USA

Karen Abbott

Capitol School, Austin, Texas, TX, USA

Background: The significance of imageability and concreteness as factors for lexical tasks in aphasic individuals is under debate. No previous treatment studies have looked specifically at training abstract words compared to concrete for improved lexical retrieval in patients with chronic aphasia.

Aims: The goal of the present study was to determine the efficacy of a treatment for lexical retrieval that is based on models of lexical processing by utilising abstractness as a mode of complexity. It was hypothesised that training abstract words in a category will result in improvement of those words and generalisation to untrained target concrete words in the same category. However, training concrete words in a category will result in the retrieval of trained concrete words, but not generalisation to target abstract words.

Methods & Procedures: A single-participant experimental design across participants and behaviours was used to examine treatment and generalisation. Generative naming for three categories (church, hospital, courthouse) was tested during baseline and treatment. Each treatment session was carried out in five steps: (1) category sorting, (2) feature selection, (3) yes/no feature questions, (4) word recall, and (5) free generative naming.

Outcomes & Results: Although participant 1 demonstrated neither significant learning nor generalisation during abstract or concrete word training, participants 2, 3, and 4 showed significant learning during abstract word training and generalisation to untrained concrete words. Participants 3 and 4 were also trained on concrete words, on which they improved, but did not show generalisation to untrained abstract words.

Conclusions: The results of the present experiment support our hypothesis that training abstract words would result in greater learning and generalisation to untrained concrete words. They also tentatively support the idea that generalisation is facilitated by treatment focusing on more complex constructions (Kiran & Thompson, 2003; Thompson, Shapiro, Kiran, & Sobecks, 2003).

Keywords: Abstractness; Imageability; Training; Aphasia; Complexity; Generation.

Imageability refers to the ease in which a visual representation of a word can be accessed. Concreteness is the degree to which a word's referents can be perceived through the senses. Imageability and concreteness are highly correlated (Paivio, Yuille, & Madigan, 1968), although notable exceptions exist. For example, emotion

Address correspondence to: Swathi Kiran PhD, CCC-SLP, Communication Sciences and Disorders, The University of Texas at Austin, A1100, CMA 7.210, Austin, Texas 78712, USA. E-mail: s-kiran@mail.utexas.edu

http://www.psypress.com/aphasiology DOI: 10.1080/02687030802588866

words such as *anger* are rated high for imageability but low for concreteness; also, very low-frequency nouns such as *armadillo* are rated high for concreteness but low for imageability. Abstract words (e.g., *religion*) typically have low imageability and concreteness, while concrete words (e.g., *gavel*) share high imageability and concreteness. In behavioural studies, normal participants exhibit longer lexical decision times for abstract words than for concrete words (Bleasdale, 1987; de Groot, 1989; James, 1975), longer word association times for abstract than concrete words (de Groot, 1989), better recall of concrete word pairs and sentences than abstract word pairs and sentences (see Paivio, 1991 for a review), and increased ease of predication (generating semantic features) for concrete over abstract words (Jones, 1985). This behavioural preference for concrete over abstract words is referred to as a concreteness effect. Many theories have been proposed to account for the concreteness effect.

The dual coding theory (DCT) posits that there are two systems for encoding words into semantic memory: verbal or linguistic and nonverbal or sensory. Abstract words are encoded into the semantic system with only verbal information, whereas concrete words are encoded into the semantic system with both verbal and multimodal sensory information (Paivio, 1991). Schwanenflugel, Harnishfeger, and Stowe (1988) proposed the context availability theory (CAT) as an alternative to the DCT. In it they posit that the concreteness effect is due to the relative difficulty of retrieving the relevant contextual information (world knowledge) associated with the word(s) under consideration. Abstract words have less contextual information associated with them because they are weakly associated with many concepts, as opposed to concrete words, which are strongly associated with just a few concepts (Schwanenflugel et al., 1988).

Considering behavioural evidence from patients with aphasia provides further insights into how imageability and concreteness affect lexical retrieval. In general, patients with aphasia exhibit an exaggerated concreteness effect. Increased performance on words with higher imageability and/or higher concreteness has been well documented during single and paired word repetition (Martin, Saffran, & Dell, 1996), single word reading (Newton & Barry, 1997), word recognition (Crutch & Warrington, 2005), and reading comprehension (Barry & Gerhand, 2003). Nickels and Howard (1995) investigated various factors affecting naming performance in people with fluent and nonfluent aphasia, and found that imageability and concreteness significantly predicted naming performance for their participants. To explain reading deficits in deep dyslexia, the normal isolated centrally expressed (NICE) model proposed by Newton and Barry (1997) posits that lexicalisation (the process in which a semantic representation activates the phonological form) is driven by concreteness. Concepts with higher concreteness values will produce strong and specific activations of the word form, with less spreading activation to other word forms, resulting in fewer errors in lexical retrieval. Further, the NICE model proposes that abstract words are loosely associated with many different concepts, whereas concrete words are strongly associated with only a few concepts. In contrast to the extensive literature in behavioural psycholinguistics, the utility of considering imageability and concreteness has only been examined in the context for remediation of reading deficits (Kim & Beaudoin-Parsons, 2007) and not for treatment of lexical retrieval deficits in aphasia.

The present study was based on our previous treatment work extending the complexity account of treatment efficacy hypothesis (Thompson & Shapiro, 2007) within the semantic domain (Kiran, 2007). Specifically, Kiran (2007) argued that training the more complex atypical examples in a category results in generalisation to the less-complex typical examples because atypical items are less representative of their category than typical items. These findings have been borne out in three studies examining generalisation from atypical to typical examples and vice versa using animate categories (*birds*, *vegetables*) (Kiran & Thompson, 2003), inanimate categories (*clothing*, *furniture*) (Kiran, 2008), and in well-defined categories (*shapes*) (Kiran & Johnson, 2008). However, these three studies have only examined lexical retrieval of typical/atypical examples within a category in the context of picture naming. In the present study we examine the effect of abstractness in treatment within categories defined by their specific location (e.g., *church*, *courthouse*). Because abstract words cannot be pictured, we examine lexical retrieval of concrete and abstract words in a category generation task. Based on the NICE model, we hypothesised that abstract concepts are defined through their relationship with concrete concepts and other abstract concepts. Concrete concepts, on the other hand, are defined through visual characteristics, somatosensory experience, and their relationship with other concrete concepts. These differences in semantic representation make abstract concepts more complex than concrete words. As an example, in a contextual category such as "*church*" activation of an abstract word such as "*prayer*" entails activation of related concrete words such "*candle*", "*bible*", "*hymn*". In contrast, a concrete word such as "*candle*" will activate other related concrete words but not necessarily activate abstract words such as "*solace*" or "*penance*" (see Figure 1). Therefore, the current study aimed to determine the effectiveness of a treatment for lexical retrieval that is based on models of lexical processing by utilising abstractness as a mode of complexity. It was hypothesised that training abstract words in a category will result in generalisation to untrained target concrete words in the same category. However, training concrete words in a category will result in the retrieval of trained concrete words, but not generalisation to target abstract words.

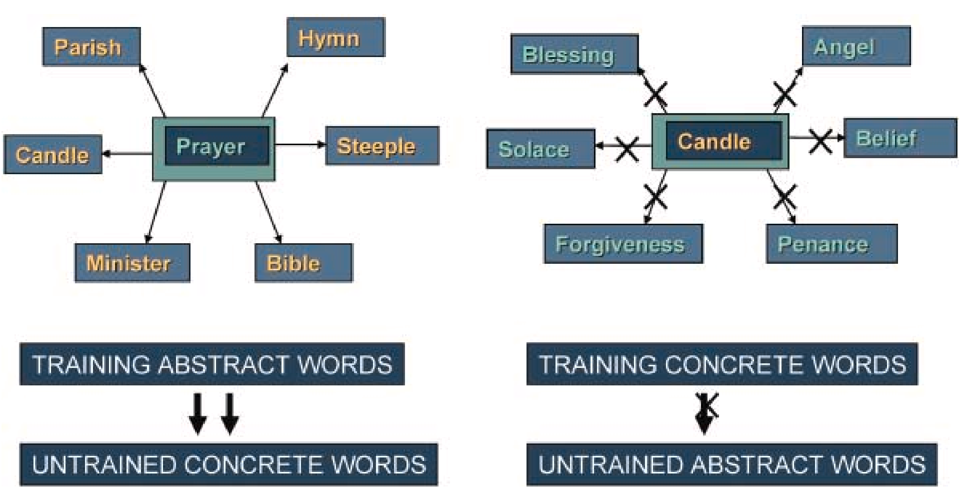

Figure 1. Schematic illustration of hypothetical representation of abstract and concrete words within a specific location context (e.g., *church*). Abstract words have stronger connections to other concrete words whereas concrete words are not strongly linked to other abstract words. Therefore, training abstract words will activate related concrete words whereas training concrete words will not activate related abstract words.

METHOD

Participants

Four monolingual English speakers with anomic aphasia and relatively high-level language skills participated in the study (see Table 1 for a complete description of participant demographics). The diagnosis of anomic aphasia was determined by administration of the *Western Aphasia Battery* (WAB; Kertesz, 1982). Results showed that all participants presented with fluent speech, naming deficits, and mildly impaired comprehension. All participants showed impaired naming on the *Boston Naming Test* (Kaplan, Goodglass, & Weintraub, 2001). Subtests of the *Psycholinguistic Assessment of Language Processing in Aphasia* (PALPA; Kay, Lesser, & Coltheart, 1992) and the *Pyramids and Palm Trees Test* (PAPT; Howard & Patterson, 1992) revealed that all participants presented with mild semantic impairments with performances ranging between 75% and 100% accuracy. Lower accuracy on low-imageability pairs compared to high-imageability pairs was observed on both synonym judgement tasks on the *PALPA* (see Table 2 for complete pre-testing results).

Stimuli

Development of categories. First, we identified eight location categories (*hospital*, *school*, *park*, *church*, *office*, *courthouse*, *restaurant*, and *museum*). A total of 14 normal young adults were provided with the list of eight locations and were asked to write down at least 15 words that they associated with each place. They were instructed to consider nouns, verbs, adjectives, high- and low-frequency words, and high- and low-imageability words. Then categories were eliminated in which (a) too few abstract words were obtained (*office*, *restaurant*), or (b) too many examples overlapped in two categories (*school*, *park*). Based on these criteria, four categories

TABLE 1
Patient demographic information

	P1	*P2*	*P3*	*P4*
Age (yrs)	54	57	39	77
Gender	male	female	female	male
Handedness	right	right	right	right
Occupation	retired auto body technician	retired office manager	software tester	retired lawyer
Aetiology	left MCA	left MCA (parietal, temporal, basal ganglia)	left temporal intra-parenchymal haemorrhage	left thalamic haemorrhage
Months post onset	19	43	8	32
Previous SLP therapy	6/04–1/05 outpatient rehab; 5/05–8/05 previous treatment study in this lab	1 year outpatient rehab; 1/06–4/06 previous treatment study in this lab	1/07–4/07 outpatient rehab; 5/07–8/07 previous treatment study in this lab	12/04–5/05 outpatient rehab; 6/05–12/05 previous treatment study in this lab
Aphasia diagnosis	anomic aphasia	anomic aphasia	anomic aphasia	anomic aphasia

Patient demographic information including age, gender, handedness, occupation, aetiology, time post-onset at start of therapy, previous SLP services, and aphasia diagnosis.

TABLE 2
Individual histories and test performance

	Participant 1		*Participant 2*		*Participant 3*		*Participant 4*	
TEST	*Pre-test*	*Post-test*	*Pre-test*	*Post-test*	*Pre-test*	*Post-test*	*Pre-test*	*Post-test*
Western Aphasia Battery								
Spontaneous Speech	9	9	10	9	9	9	8	9
Auditory Comprehension	8.8	9.0	9.4	8.4	10	10	9.6	9.9
Repetition	7.8	9.6	6.2	8.8	8.0	8.8	8.3	8.8
Naming	8.5	8.8	8.9	8.8	7.5	8.0	8.4	8.1
Aphasia Quotient	86.2	90.8	89.1	88.2	89	91.6	86.7	91.7
Boston Naming Test	81.7%	80.0%	71.7%	65.0%	23.3%	35.0%	66.7%	70.0%
PALPA								
Auditory lexical decision	81.3%	85.0%	88.1%	93.8%	94.4%	95.6%	81.9%	86.3%
Visual lexical decision task	95%	93.3%	95.8%	93.3%	94.2%	95.0%	96.7%	95.0%
Letter length reading	95.8%	100%	100%	100.0%	100.0%	100.0%	100.0%	100%
Spoken word–picture matching	92.5%	97.5%	97.5%	97.5%	97.5%	100.0%	95.0%	95.0%
Written word–picture matching	92.5%	95%	100%	100%	100.0%	100.0%	100.0%	95.0%
Auditory synonym judgements	78.3%	86.7%	73.3%	83.3%	86.7%	86.7%	86.7%	86.7%
high imageability	76.3%	86.7%	83.3%	90%	90.0%	93.3%	86.7%	90.0%
low imageability	80%	86.7%	80%	76.7%	83.3%	80.0%	86.7%	83.3%
Written synonym judgements	85%	88.3%	85%	88.3%	91.7%	90.0%	91.7%	93.3%
high imageability	90%	93.3%	93.3%	93.3%	96.7%	93.3%	96.7%	96.7%
low imageability	80%	83.3%	70%	83.3%	86.7%	86.7%	86.7%	90.0%
Spoken picture naming	87.5%	85%	95%	95%	95.0%	97.5%	97.5%	92.5%
Writing picture names	75%	65%	97.5%	100%	92.5%	92.5%	95.0%	95.0%
Reading picture names	100%	92.5%	97.5%	100%	100.0%	100.0%	100.0%	97.5%
Repeating picture names	100%	100%	97.5%	100%	100.0%	100.0%	92.5%	100.0%
Spelling picture names	85%	55%	92.5%	95%	95.0%	100.0%	95.0%	100.0%
Pyramids and Palm Trees								
Three pictures	98.1%	92.3%	92.3%	94.2%	90.4%	94.2%	90.4%	100.0%
Three written words	96.2%	84.6%	92.3%	92.3%	94.2%	94.2%	96.2%	100.0%

Individual histories and performance on WAB (Kertesz, 1982), BNT (Kaplan et al., 2001), PALPA (Kay et al., 1992), and PAPT (Howard & Patterson, 1992) before (pre-) and after (post-) treatment.

(*church*, *hospital*, *museum*, and *courthouse*) were selected to develop stimuli for treatment.

Development of category examples. To select specific examples, words were excluded if (a) they were noun–verb ambiguous (e.g., *hope*, *cut*), or (b) they were two-word phrases (e.g., *blood pressure*, *waiting room*). A computer-based task was then administered to 10 normal young adult participants to classify the remaining category examples as concrete or abstract. The participants selected 1 to indicate the word was abstract, 3 to indicate the word was concrete, and 2 if they were not sure or if the word did not fit abstract or concrete. Abstract was defined as existing only in the mind or theoretical. Concrete was defined as existing in reality, perceptible by the senses. The category words appeared on the computer screen one at a time, blocked by category. In general, words selected for treatment had at least 70% classification agreement into abstract or concrete. This approach ensured that individual differences across participants did not influence the outcome of selection of abstract/concrete items. Five words (*evidence*, *government*, *judgement*, *angel*, and *prayer*) were included that had less than 70% agreement. At this point, the category *museum* was eliminated due to very low agreement scores.

A paper-and-pencil task was administered to the same 10 young adult participants to assess their familiarity with the examples from the categories (*church*, *courthouse*, *hospital*). Participants were instructed to rank each word on a scale of 1 to 5, with 1 indicating "not at all familiar" and 5 indicating "very familiar". The mean familiarity scores for all the words were between 3.8 and 5.0. Sets of abstract and concrete words were created for each category and matched for written word frequency, familiarity, and number of syllables based on the MRC Psycholinguistic Database (Coltheart, 1981) (see Appendix A). Individual typed cards were printed for all words.

Development of semantic features for treatment. Each category contained 45–50 semantic features. A total of 17 general features were created based on the dictionary definitions of concrete and abstract (e.g., *exists in nature*, *exists only in the mind*) and perceptual characteristics (e.g., *can hear it*, *can see it*). Five distractor features (e.g., *put on windows*) were obtained from a previous study (Kiran, 2008). The remaining 13–18 features for each category were based on the participants' input during the first session of each category training period.

Design

This study used a single-participant experimental design with the order of category and abstractness counterbalanced across the four participants (see Table 3). Criterion for switching treatment from one category to the next was set at either 8/10 accuracy on category generation on two consecutive treatment probes or the completion of 20 training sessions.

Baseline naming procedures. Generative naming for three categories/locations (*church*, *courthouse*, and *hospital*) was tested during baseline. Participants were instructed to name as many words associated with each category/location as they could in 1–2 minutes. Responses were divided into four categories: (1) target concrete words (e.g., *doctor*, *candle*), (2) target abstract words (e.g., *emergency*, *holy*),

TABLE 3
Order of treatment for each patient across category type and abstractness type

Participant	*# of baselines*	*Category trained*	*Typicality trained*	*Summary of generalisation patterns*
P1	3	1. Church	Abstract	Abstract ≠> Concrete*
		2. Hospital	Concrete	Concrete ≠> Abstract*
P2	3	1. Church	Abstract	Abstract => Concrete
		2. No treatment		
P3	5	1. Hospital	Abstract	Abstract => Concrete
		2. Courthouse	Concrete	Concrete ≠> Abstract
P4	5	1. Church	Concrete	Concrete ≠> Abstract
		2. Hospital	Abstract	Abstract => Concrete

Also shown is a summary of generalisation patterns observed for each patient. * Criterion not reached for acquisition during treatment indicating no learning and no generalisation.

(3) other concrete words (e.g., *sling*, *pew*), and (4) other abstract words (e.g., *diagnosis*, *religion*). Target abstract and concrete words were category exemplars that were normed based on procedures described above. They were marked as generated (1) or not generated (0) and were considered correct if they were clear and intelligible productions of the target words or semantically similar variations of the target word (e.g., *jurors* for *jury*) or a very close synonym (e.g., *physician* for *doctor*). Other abstract words and other concrete words were category exemplars that were spontaneously generated by each participant. These responses were only considered correct *other* words if they were intelligible productions of words that were appropriate for the category, including semantically dissimilar variations of a target word (e.g., *healthcare* counted as a correct other response, not as target word *health*). All other responses including (a) superordinate category labels, (b) circumlocutory responses, (c) nonspecific or unrelated responses, (d) neologisms, (e) phonemic paraphasias, and (f) repeated words or multiple forms of the same root word (e.g., *pray* and *prayer*; *forgiver* and *forgiving*) were counted as incorrect responses. Additionally, items that were all a specific type of thing were counted as one item (e.g., *Baptist*, *Methodist*, *Catholic*, and *Lutheran* all counted as *religions*) (see Table 4 for examples of errors). The criterion for initiation of treatment was $\leqslant 40\%$ accuracy at generating items with the amount of variability in baselines not exceeding 30%.

Treatment. Each treatment session was carried out in five steps: (1) category sorting, (2) feature selection, (3) yes/no feature questions, (4) word recall, and (5) free generative naming. Steps 1 and 5 occurred once per treatment session, whereas steps 2–4 were repeated for each trained target word (see Appendix B for specific instructions).

Treatment probes. Throughout treatment, generative naming probes like those used in the baseline condition were presented every second treatment session to assess retrieval of the trained and untrained items. Generalised naming to the untrained examples was considered to have occurred when levels of performance changed by at least 4 points over baseline levels.

Reliability. All baseline sessions and treatment sessions were recorded on videotape. Reliability on the dependent variable for participants was calculated

TABLE 4
Responses

		Other abstract	*Other concrete*	*Incorrect*
Participant 1	CHURCH	god (33.3%)	pew (44.4%)	forgotten (100%) [semantic/phonemic paraphasia]
		religion (23.8%)	cross (22.2%)	
		Christ (14.3%)	preacher (11.1%)	
		heaven (9.5%)	choir (11.1%)	
		unfaithful (4.8%)	grail (11.1%)	
	HOSPITAL	sick (33.3%)	nurse (60%)	hearing (100%) [unrelated]
		aid (16.7%)	technician (10%)	
		well (16.7%)	nurse's aid (10%)	
		helpful (8.3%)	sling (10%)	
		cure (8.3%)	stroke (10%)	
Participant 2	CHURCH	communion (40.0%)	choir (16.4%)	
		god (13.3%)	piano (16.4%)	
		vows (13.3%)	pew(s) (14.5%)	
		Presbyterian (6.7%)	stained glass (12.7%)	
		love (6.7%)	cross (7.3%)	
Participant 3	COURTHOUSE	law (23.3%)	judge (11.1%)	court (30.8%) [superordinate]
		appeal (13.3%)	witness (9.5%)	courthouse (15.4%) [superordinate]
		innocence (10%)	criminals (7.9%)	12 people (7.7%) [circumlocution]
		defence (10%)	police (7.9%)	hammer (7.7%) [semantic paraphasia]
		verdict (10%)	defendants (7.9%)	offendant (7.7%) [neologism/phonemic paraphasia]
	HOSPITAL	sleeping (33.3%)	nurse (10.2%)	socks (16.7%) [unrelated]
		death (33.3%)	MRI (8.8%)	pictures (8.3%) [unrelated]
		reading (16.7%)	babies (8.0%)	white stuff (8.3%) [circumlocution]
		emotional (16.7%)	bed (7.3%)	neo (8.3%) [phonemic paraphasia]
			stethoscope (5.8%)	french fries (8.3%) [unrelated]

TABLE 4
(*Continued*)

		Other abstract	*Other concrete*	*Incorrect*
Participant 4	HOSPITAL	death (35.3%)	nurse (23.4%)	city hospital (9.1%) [superordinate]
		cleanliness (11.8%)	visitors (14.0%)	older hospital (9.1%) [superordinate]
		healing (5.9%)	rooms (4.7%)	medical hospital (9.1%) [superordinate]
		wellness (5.9%)	surgeons (4.7%)	presbyterian hospital (9.1%) [superordinate]
		diagnosis (5.9%)	paramedic (2.8%)	federal hospital (9.1%) [superordinate]
	CHURCH	god (50.0%)	funerals (16.2%)	preaching (28.6%) [form of target]
		Baptists (14.7%)	attendees (14.7%)	death services (28.6%) [circumlocution]
		Catholics (14.7%)	marriage (13.2%)	spear (14.3%) [phonemic/semantic paraphasia]
		Methodists (5.9%)	members (11.8%)	infant services (14.3%) [circumlocution]
		ordained (2.9%)	churchgoers (5.9%)	church house (14.3%) [superordinate/]

A breakdown of the most frequent other concrete and abstract responses and incorrect responses for each participant in each category. See text for description of error types.

for 75% of the probe sessions, resulting in 100% agreement. Reliability on the independent variable (i.e., presentation of the treatment protocol) was calculated for 50% of treatment sessions resulting in 100% agreement.

Data analysis. To calculate effects sizes (ES), the average baseline probe scores were subtracted from the average post-treatment scores and the result was divided by the standard deviation of the baseline scores (Beeson & Robey, 2006). Where post-treatment probe scores could not be obtained, the average of the final two treatment probe scores was used. Beeson and Robey (2008) recently updated their benchmarks for direct treatment: 6.5 = small ES, 8.0 = medium ES, 9.5 = large ES, and for generalisation of treatment, 2.0 = small ES, 5.0 = medium ES, 8.0 = large ES.

RESULTS

Participant 1

Participant 1 first received treatment for abstract words associated with *church*, which did not improve significantly (only 4/10 items correct, ES = 4.62; see Figure 2) within 10 weeks of training. No generalisation was observed to untrained target concrete words (ES = .29), other abstract words (ES = −.58), or other concrete words (ES = 1.15). Following one additional baseline session for the second category, treatment was shifted to concrete words related to *hospital*, which did not improve to criterion within 7 weeks of training (high 4/10, ES = 0). No generalisation was observed to untrained target abstract words (ES = 0), or untrained other concrete words (ES = −.87), however, other abstract words improved from 1 to 3 items (ES = 3.46). After treatment, Participant 1 had another stroke that resulted in reduced right-hand function, but no apparent increase in language problems. Thus, no follow-up probes were conducted for *hospital.*

Participant 2

Participant 2 received 6 weeks of treatment for abstract words associated with *church*, which showed improvement to 7/10 items (ES = 5.48; see Figure 3). Generalisation also occurred for untrained target concrete words (ES = 2.74), other abstract words (ES = 4.04), and other concrete words (ES = 1.20).

Participant 3

Because P3 indicated a strong aversion to the category *church* during baseline probes, this category was not considered for treatment. Instead, she was first trained on abstract words in the category *hospital*. She improved from 1/10 items to 10/10 items in only eight sessions, which was maintained post treatment (ES = 18.78; see Figure 4). On the untrained concrete items P1 improved from 3 to 9 items that maintained post treatment (ES = 4.47). However, a decrease was observed in the number of other abstract (ES = −.45) and other concrete items (ES = −.42). Next, P3 began training on concrete words in the category *courthouse*, which improved to 10/10 items in two weeks (ES = 3.10). It should be noted that these items showed some improvement (from 2/10 items to 4/10 items) during treatment for *hospital*, possibly due to cross-category generalisation during the abstract word training of the

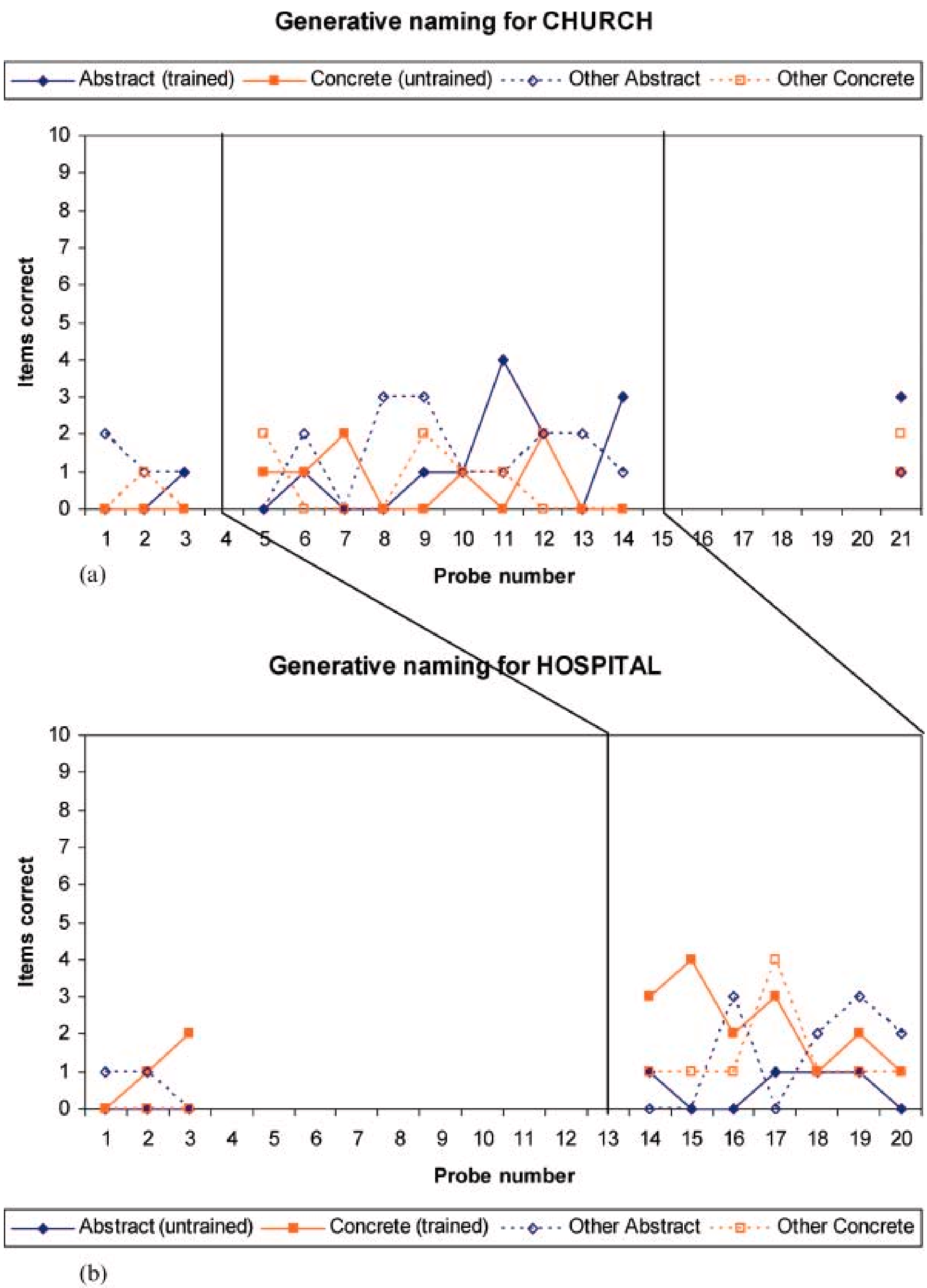

Figure 2. Generative naming accuracy for (a) abstract (trained), concrete (untrained), and other abstract and concrete (untrained) items for the category *church*, and (b) concrete (trained), abstract (untrained), and other abstract and concrete (untrained) items for the category *hospital* during baseline and treatment phases for Participant 1.

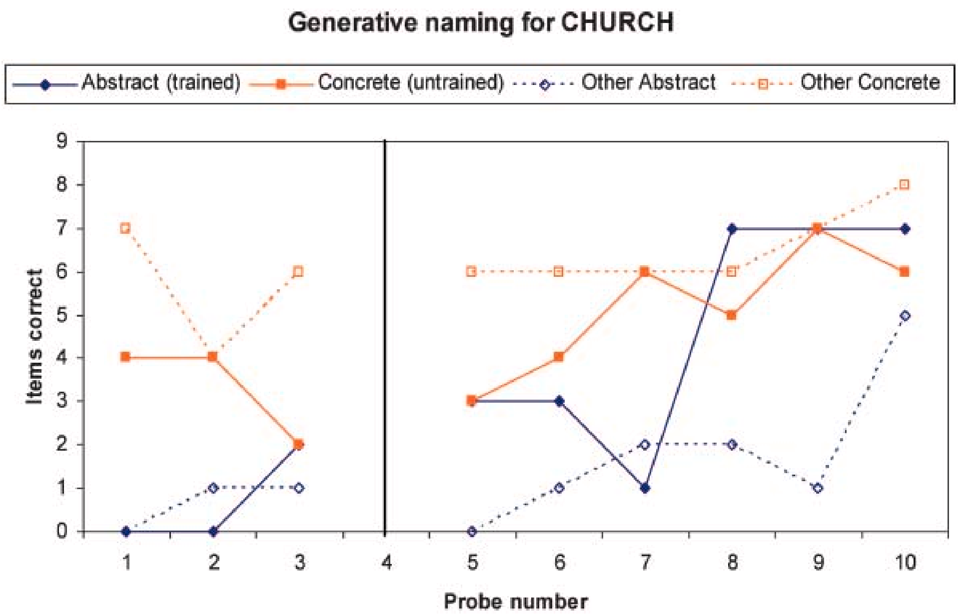

Figure 3. Generative naming accuracy for abstract (trained), concrete (untrained), and other abstract and concrete (untrained) items for the category *church* during baseline and treatment phases for Participant 2.

first category. No improvement was observed for untrained abstract items (ES = 1.47), other concrete items (ES = −2.25), and other abstract items (ES = .71).

Participant 4

To reduce bias towards the category *courthouse*, due to P4's experience as a lawyer, this category was not considered for treatment. P4 was first trained on concrete words in the category *church*, which improved from 2/10 to 8/10 and was maintained post treatment (ES = 7.84). No changes were observed on the untrained abstract examples; however, the effect sizes appear artificially inflated (ES = 3.42) due to a possible cross-category facilitation of these items during the abstract word training of the second category. No generalisation to other concrete (ES = −.59) or other abstract words (ES = −.24) was observed. Treatment was then shifted to abstract words in the category *hospital*, which showed some fluctuation but did not improve to criterion (ES = 1.6). Interestingly, P2 improved on untrained target concrete words from 4/10 to 7/10 (ES = 3.31) but no improvement was observed on other abstract words (ES = 0) or concrete words (ES = −.38).

Qualitative analysis of other responses and incorrect responses

As stated before, all responses that were not part of the normed stimuli but were generated spontaneously by each participant were recorded and categorised into other (abstract or concrete) responses and incorrect responses. We conducted a qualitative analysis to evaluate the nature of these responses. As shown in Table 3, P1 and P2 produced almost no incorrect responses. P3 and P4 produced a few incorrect responses

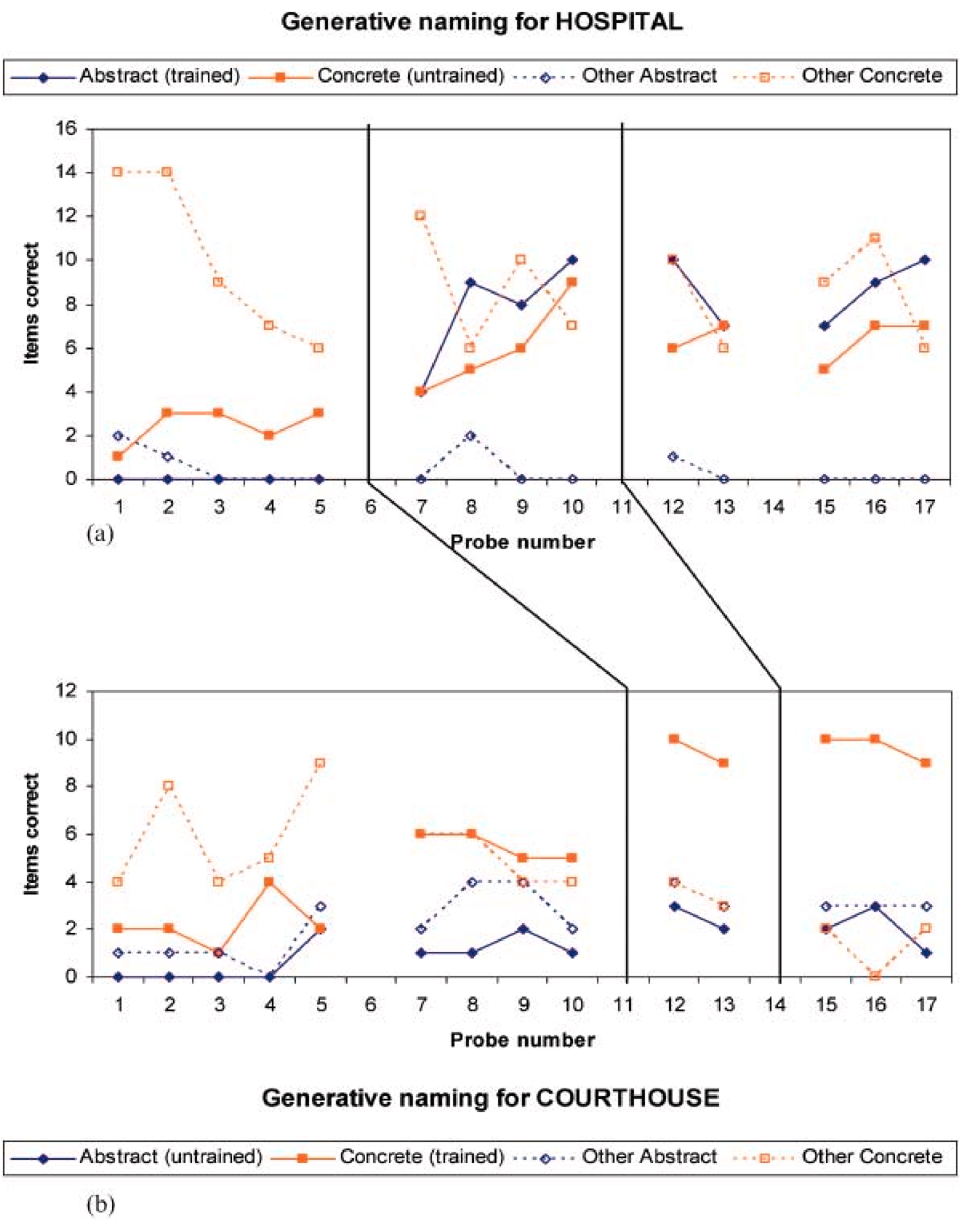

Figure 4. Generative naming accuracy for (a) abstract (trained), concrete (untrained), and other abstract and concrete (untrained) items for the category *hospital*, and (b) concrete (trained), abstract (untrained), and other abstract and concrete (untrained) items for the category *courthouse* during baseline and treatment phases for Participant 3.

and the majority of them were either circumlocutions or superordinate responses. Analysis of the participants' other responses revealed that certain words that were not target words were highly salient for a particular category. For example for the category *hospital*, *nurse* and *death* were highly salient for P1, P3, and P4. Additionally, each

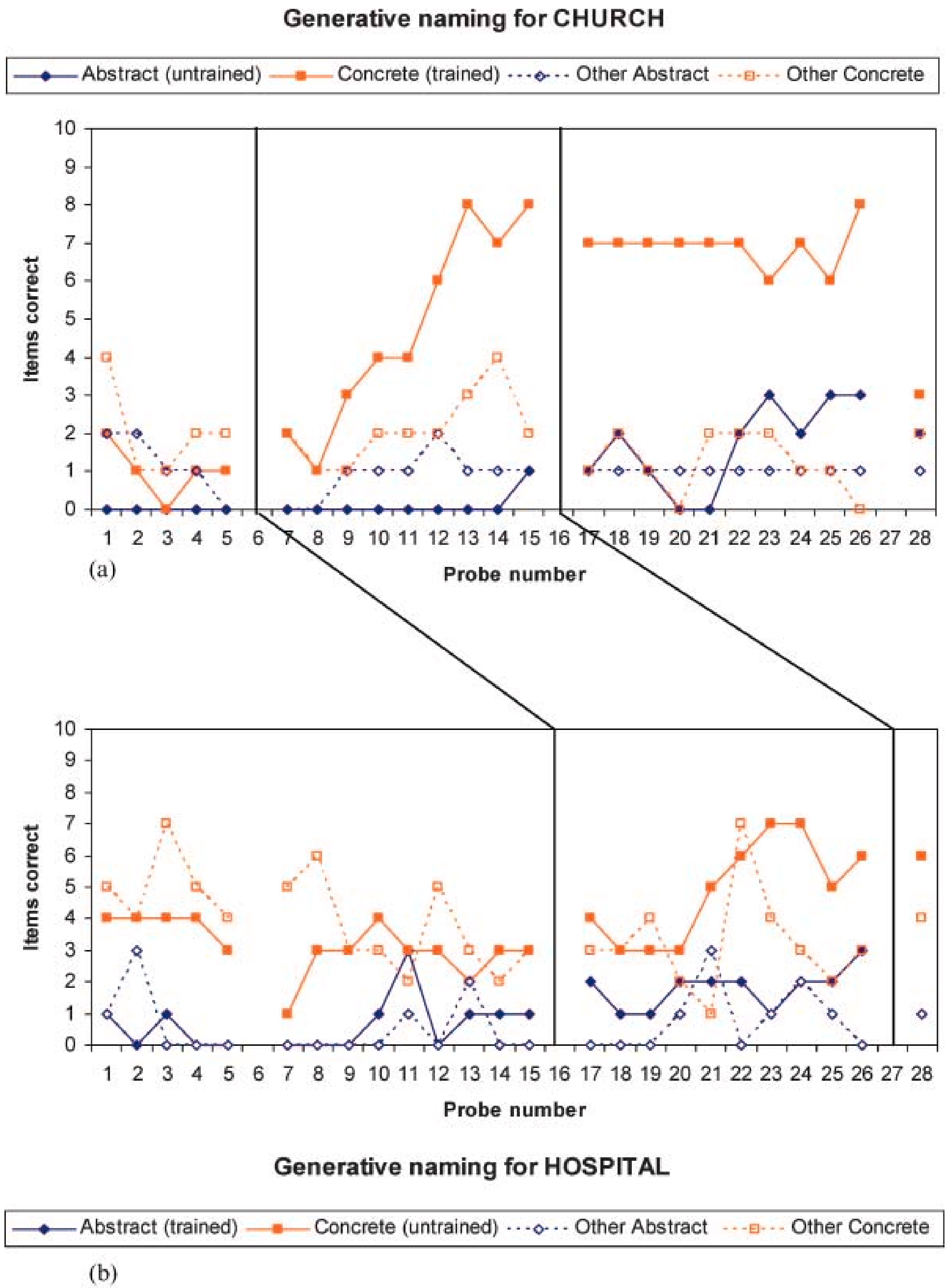

Figure 5. Generative naming accuracy for (a) concrete (trained), abstract (untrained), and other abstract and concrete (untrained) items for the category *church*, and (b) abstract (trained), concrete (untrained), and other abstract and concrete (untrained) items for the category *hospital* during baseline and treatment phases for Participant 4.

participant tended to repeat certain words during consecutive probe sessions (e.g., P1: *God* – 33%; P2: *communion* – 40%). Other, less-salient and arguably less-typical items of a category were usually given less frequently (e.g., P4: *spear* for church).

Standardised test results

Overall, all participants showed improvements or maintained performance on the various standardised measures that were administered pre and post treatment. P1, P3, and P4 showed small improvements on *WAB* AQ scores, whereas only P3 and P4 showed improvements on the BNT. Likewise, all participants showed small improvements on the PALPA and PAPT but no consistent patterns were observed for any specific participant.

DISCUSSION

The purpose of this study was to examine the hypothesis that training abstract examples resulted in greater generalisation within a contextual category than training concrete examples of the same category. The results showed that P1 did not demonstrate significant learning or generalisation during abstract word training for category *church* or concrete word training for category *hospital*. For P2, when trained on abstract words for category *church*, she improved on those items and showed generalisation to naming untrained target concrete words. Likewise, P3 showed learning for abstract words and generalisation to untrained concrete words for the category *hospital*. In contrast, she showed learning of only concrete words for the category *courthouse*; no generalisation was observed to the untrained abstract words. Finally, P4 who was initially trained on concrete words for the category *church* showed improvements on those items but did not generalise to the untrained abstract items. Interestingly, when trained on abstract items for the category *hospital*, he did not improve on the trained items but showed generalisation to the untrained concrete items (see Table 2).

With the exception of P2, no other participant exhibited generalisation to *other* concrete or abstract words in the trained category, regardless of whether abstract or concrete words were being trained. In fact, *other* abstract and concrete responses actually decreased for P3 and P4 in both trained categories, but *total* responses for each category increased. It seems that participants shifted their generative naming responses from a majority of *other* responses to a majority of *target* responses as treatment progressed. Taken together, these results provide preliminary support for the extension of the complexity hypothesis into the domain of the representation of abstractness/imageability. Specifically, training the complex abstract items resulted in generalisation to the less-complex concrete items. In contrast, training concrete items did not promote generalisation to the more-complex abstract items. Therefore, as hypothesised in the introduction, strengthening semantic features of abstract words facilitated activation of associated abstract and concrete words. In contrast, strengthening semantic features of concrete words only activated those items; semantic features of abstract words and their corresponding representations remained unchanged. In other words, strengthening abstract words during treatment results in similar activation of concrete items not specifically targeted in treatment. Therefore, concrete items inherently receive additional practice when training abstract items. Alternatively, it may be that the semantic-based treatment (i.e., category sorting task and yes/no questions) resulted in strengthening the entire category regardless of whether the words were abstract or concrete in nature. Such a premise may explain P4's data, as he showed more improvements on the untrained concrete words than the trained abstract words of *hospital*. Therefore, the precise mechanisms underlying these selective generalisation patterns cannot be concluded in this study and require further elaboration. Further data in future studies would permit a deeper inspection of the assumptions concerning the way in which concrete and abstract lexical items are acquired. Nonetheless, the present data indicate that the complexity effect seen in syntactic treatment for agrammatic aphasia (Thompson et al., 2003) and semantic treatment for naming deficits utilising typicality as a mode of complexity (Kiran, 2007; Kiran & Thompson, 2003) can be extended to a semantic treatment for naming deficits that utilises abstractness as a mode of complexity.

The results of the present study should be considered preliminary for several reasons. First, the variation in acquisition and generalisation patterns observed across patients indicates that not all participants responded to treatment the same way. For instance, P1 did not show improvements on the trained items or generalisation to untrained items. This participant, however, showed subtle changes during treatment: he performed the card-sorting task more quickly, and read and understood the semantic features more accurately. Likewise, P4 reached treatment correct criterion only after 10 weeks for the first category and never quite reached criterion for the second category. In contrast, P3 reached and surpassed the treatment criterion in a very short period for both categories.

Second, this was the first study to investigate the effects of semantic feature training on generative naming, whereas previous studies with semantic feature training examined its efficacy for picture naming. Without the visual prompts involved in picture-naming tasks, generative naming is a much harder task; it likely recruits additional processing mechanisms than just lexical retrieval including executive functioning and short-term memory. Even though the precise influence of these mechanisms on the treatment outcomes cannot be conclusively interpreted with the present data, they offer a useful next step to determine the functionality of training single word retrieval.

Third, at least two of the participants (P3 and P4) showed some form of cross-category generalisation after abstract stimuli treatment. Upon training of concrete items of *church*, P4 did not improve on abstract items, but improved slightly on these items when treatment for abstract items of *hospital* commenced. A similar improvement was observed for P3 for concrete items of *courthouse* during treatment for abstract words in *hospital*. While these results can be interpreted as further support for the beneficial effects of abstract item training, such a conclusion would be tentative at best and would require systematic examination in future studies.

To conclude, the present study illustrates a semantic-based treatment approach based on the complexity account of treatment efficacy by utilising abstractness as a mode of complexity. These results hold considerable promise to facilitate generative naming in patients with aphasia. We intend to continue this question by examining the effects of abstract training on a larger and more diverse sample of aphasia patients.

Manuscript received 7 August 2008
Manuscript accepted 30 October 2008
First published online day/month/year

REFERENCES

Barry, C., & Gerhand, S. (2003). Both concreteness and age-of-acquisition affect reading accuracy but only concreteness affects comprehension in a deep dyslexic patient. *Brain and Language*, *84*(1), 84–104.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychological Review*, *16*(4), 161–169.

Beeson, P. M., & Robey, R. R. (2008). *Meta-analysis of aphasia treatment outcomes: Examining the evidence*. Paper presented at the Clinical Aphasiology Conference, Jackson Hole, Wyoming, USA.

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 582–594.

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

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*(Pt 3), 615–627.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824–845.

Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Bury St. Edmunds, UK: Thames Valley Test Company.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology (Human Perception and Performance)*, *104*(2), 130–136.

Jones, G. V. (1985). Deep dyslexia, imageability, & ease of predication. *Brain and Language*, *24*(1), 1–19.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

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

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

Kim, M., & Beaudoin-Parsons, D. (2007). Training phonological reading in deep alexia: Does it improve reading words with low imageability? *Clinical Linguistics and Phonetics*, *21*(5), 321–351.

Kiran, S. (2007). Semantic complexity in the treatment of naming deficits. *American Journal of Speech Language Pathology*, *16*(Feb), 1–12.

Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia: Further evidence for semantic complexity. *Journal of Speech, Language and Hearing Research*, *51*, 1550–1568.

Kiran, S., & Johnson, L. (2008). Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories. *American Journal of Speech Language Pathology*, *17*, 389–400.

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*(4), 773–787.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory – verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52*(1), 83–113.

Newton, P. K., & Barry, C. (1997). Concreteness effects in word production but not word comprehension in deep dyslexia. *Cognitive Neuropsychology*, *14*, 481–509.

Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, *33*(10), 1281–1303.

Paivio, A. (1991). Dual Coding Theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499.

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

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech Language and Hearing Research*, *46*(3), 591–607.

APPENDIX A

Target words and average values on various psycholinguistic variables derived from MRC Psycholinguistic Database. CNC = concreteness; IMG = imageability; KF FREQ = Kucera-Francis written frequency; SYLL = syllable; FAM = familiarity.

HOSPITAL			
Abstract words		Average ratings	
1. emergency	*2. quiet*	CNC	354.17
3. hygiene	*4. sensitive*	IMG	432.29
5. science	*6. recovery*	KF FRQ	50.20
7. health	*8. insurance*	SYLL	2.60
9. compassion	*10. sterile*	FAM	545.14
Concrete words		Average ratings	
1. examination	*2. ambulance*	CNC	524.78
3. doctor	*4. blood*	IMG	563.22
5. injury	*6. birth*	KF FRQ	52.40
7. disease	*8. surgery*	SYLL	2.50
9. medicine	*10. patient*	FAM	552.56
COURTHOUSE			
Abstract words		Average ratings	
1. truth	*2. protection*	CNC	298.00
3. guilt	*4. evidence*	IMG	372.13
5. justice	*6. legal*	KF FRQ	83.40
7. equality	*8. judgement*	SYLL	2.20
9. logic	*10. freedom*	FAM	535.63
Concrete words		Average ratings	
1. trial	*2. lawyer*	CNC	520.60
3. prosecutor	*4. prison*	IMG	535.1
5. jury	*6. bench*	KF FRQ	81.8
7. prisoner	*8. flag*	SYLL	2.4
9. constitution	*10. government*	FAM	502.1
CHURCH			
Abstract words		Average ratings	
1. grace	*2. blessing*	CNC	328.20
3. prayer	*4. forgiveness*	IMG	447.40
5. angel	*6. holy*	KF FRQ	23.90
7. baptism	*8. penance*	SYLL	2.10
9. belief	*10. solace*	FAM	498.40
Concrete words		Average ratings	
1. hymn	*2. chapel*	CNC	552.20
3. wedding	*4. bell*	IMG	558.50
5. minister	*6. organ*	KF FRQ	24.90
7. candle	*8. parish*	SYLL	1.90
9. Bible	*10. steeple*	FAM	484.40

APPENDIX B: TREATMENT PROTOCOL

Step 1: Category sorting

At the beginning of each treatment session, the participant was asked to sort all 60 target words (10 abstract and 10 concrete from each category) into their respective categories by placing the word cards into piles labelled *church*, *courthouse*, and *hospital*. The order of the piles was randomly assigned by the clinician in each session. Participant was provided with feedback regarding accuracy.

Step 2: Feature selection

After the word cards had all been correctly sorted into the three categories, the 10 target words that were being trained (either abstract or concrete from a particular category) were retained. The participant was then given the 45 feature cards for the category (17 general features, 13 features specific to the target words in the category, 15 distractor features), and a trained target word (randomly selected from the 10) was presented for feature selection. The participant was instructed go through the stack of features and select the first six features that applied to the word. After all six features were selected and placed under the word card, the participant read each of the features aloud to the clinician and explained why each feature was applicable to the target word. At this point the clinician either agreed with the selection of features or engaged the participant in a discussion as to why a particular feature did or did not apply to the target word.

Step 3: Yes/No feature questions

After six features were selected and reviewed for a trained target word, the features and the word were removed from the participant's view and the clinician asked the participant 15 yes/no feature questions. For example, if the trained target word was *doctor*, five questions related to the trained target word, five questions did not relate to the trained target word but to other target words in the category, and five questions were completely unrelated. Care was taken not to mention the trained target word unless it was obvious that the participant had forgotten which concept the questions were referring to.

Step 4: Word recall

After the yes/no questions were complete the clinician asked the participant to say the trained target word. The participant was given feedback regarding accuracy but not given the target word if incorrect. After this the clinician randomly selected another trained target item and steps 2–4 were repeated.

Step 5: Generative naming

At the end of each session the participant was asked to name as many words as s/he could think of in the trained category. This task was unlike the generative naming task used for the probes in that the clinician gave the participant feedback on each response and there was no time limit. The clinician was also free to give prompts such as "Think of all the words we worked on today. Remember all the characteristics that we talked about for each one," and "Good, you came up with X words that belong in the category X. There were X more that we worked on today, can you think of those?"