AJSLP

Research Article

Abstract Semantic Associative Network Training: A Replication and Update of an Abstract Word Retrieval Therapy Program

Chaleece W. Sandberg[a] and Teresa Gray[b]

Purpose: We report on a study that replicates previous treatment studies using Abstract Semantic Associative Network Training (AbSANT), which was developed to help persons with aphasia improve their ability to retrieve abstract words, as well as thematically related concrete words. We hypothesized that previous results would be replicated; that is, when abstract words are trained using this protocol, improvement would be observed for both abstract and concrete words in the same context-category, but when concrete words are trained, no improvement for abstract words would be observed. We then frame the results of this study with the results of previous studies that used AbSANT to provide better evidence for the utility of this therapeutic technique. We also discuss proposed mechanisms of AbSANT.
Method: Four persons with aphasia completed one phase of concrete word training and one phase of abstract word training using the AbSANT protocol. Effect sizes were calculated for each word type for each phase. Effect sizes for this study are compared with the effect sizes from previous studies.
Results: As predicted, training abstract words resulted in both direct training and generalization effects, whereas training concrete words resulted in only direct training effects. The reported results are consistent across studies. Furthermore, when the data are compared across studies, there is a distinct pattern of the added benefit of training abstract words using AbSANT.
Conclusion: Treatment for word retrieval in aphasia is most often aimed at concrete words, despite the usefulness and pervasiveness of abstract words in everyday conversation. We show the utility of AbSANT as a means of improving not only abstract word retrieval but also concrete word retrieval and hope this evidence will help foster its application in clinical practice.

Anomia is the term used to describe the word retrieval difficulties observed in persons with aphasia (PWA). Aphasia[1] is an acquired language disorder that affects approximately 2.5 million Americans (Simmons-Mackie, 2018). While PWA can present with deficits in different modalities (reading, writing, verbal expression, auditory comprehension) and at different levels (letter/phoneme, word, sentence, discourse), the most prominent, ubiquitous, and lasting feature of aphasia is anomia. Word retrieval is often a target in therapy because it is a basic building block of conversation, and PWA invariably express frustration with word-finding difficulties.

Several word retrieval treatments exist and often focus on different aspects of the word retrieval process. For example, phonological cueing hierarchies are aimed at supporting those persons whose word retrieval deficits appear to affect phonological selection/encoding, evidenced by a preponderance of phonemic paraphasias or naming errors caused by phoneme substitutions (e.g., saying “gog” for “dog”). The word retrieval treatment that is most germane to the current study is semantic feature analysis (SFA), which was originally developed by Ylvisaker et al. (1987) for the treatment of semantic memory organization in persons with traumatic brain injury and was subsequently applied to word finding for PWA by Boyle and Coelho (1995). In SFA, the client is asked to generate semantic features for everyday objects. A semantic feature is a phrase that helps describe a concept. For example, a semantic feature for a dog is that it barks. In their seminal 1995 case study, Boyle

[a]Department of Communication Sciences and Disorders, Penn State University, University Park, PA
[b]Department of Speech, Language and Hearing Sciences, San Francisco State University, CA

Correspondence to Chaleece W. Sandberg: cws18@psu.edu

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Aimee Dietz

Received August 29, 2019
Revision received February 10, 2020
Accepted April 15, 2020
https://doi.org/10.1044/2020_AJSLP-19-00066

[1]Aphasia and dysphasia have both been used to describe the same syndrome. For consistency, this article will use the term “aphasia” throughout.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

and Coelho explain that the mechanism of SFA is based on spreading activation within the semantic system (Collins & Loftus, 1975). Briefly, when the semantic features are generated, the semantic network connected to the target word is activated, and the activation spreads to the target word, making it more easily accessible. In a recent systematic review of 21 studies of SFA, including 55 PWA, Efstratiadou et al. (2018) showed that SFA is effective, with 45 of the 55 participants showing improvement for trained items. However, treatment effects tended to be small, and only 40% of the participants showed generalization to untrained items. Perhaps more importantly, in its prescribed form, SFA can only be used with concrete, picturable concepts.

Why Train Abstract Words?

From an impairment-based perspective, the observation that abstract words are often more impaired than concrete words in PWA (e.g., Barry & Gerhand, 2003; Crutch & Warrington, 2005; Martin & Saffran, 1997, 1999; Martin et al., 1996; Newton & Barry, 1997; Nickels & Howard, 1995) makes them a natural target for therapy. Additionally, as described by Newton and Barry (1997) and outlined in detail below, the mechanism of the relative impairment for abstract words is subthreshold activation. If this is the case, then increasing the activation level for abstract words is a logical therapeutic strategy, as increasing activation in the semantic system has been an effective therapeutic technique to improve lexical access for concrete words, evidenced by the success of semantically based treatment protocols, such as SFA.

From a psychosocial-based perspective, abstract words are an especially ecologically valid target. First, abstract words are essential for natural daily conversation in which opinions and emotions are often expressed. In a recent review, Renvall et al. (2013) analyzed two language corpora and found that the most frequently used words had significantly lower imageability and concreteness ratings than words typically used in naming therapy (i.e., concrete words). The authors rightly questioned whether the conventional focus on concrete words is as functional as is standardly assumed. Targeting abstract words may be especially important in light of the fact that PWA have difficulty expressing opinions, feelings, and attitudes (Armstrong, 2005).

Finally, training abstract words (*diagnosis*) has been shown to promote generalization to concrete words (*doctor*) in the same context-category (*hospital*), while the reverse is not true (Kiran et al., 2009). Generalization, or the transfer of benefit from trained items to untrained items, is a priority in anomia therapy studies, as the average adult vocabulary size is far too large to tenably address each individual lexical entry. One process-based approach to promoting generalization is the Complexity Account of Treatment Efficacy (CATE), first proposed by Thompson et al. (2003) in the context of therapy for agrammatism. The authors found that training more complex syntactic structures (object relative clauses) promoted generalization to less complex, related syntactic structures (object clefts and *wh*-questions), but not vice versa. Kiran and Thompson (2003) extended this theory to anomia therapy in the context of semantic typicality, showing that training atypical items (*ostrich*) promoted generalization to typical items (*robin*) in the same taxonomic category, but not vice versa. What we will hereafter refer to as Abstract Semantic Associative Network Training (AbSANT) has extended the CATE to anomia therapy in the context of semantic concreteness, showing that training abstract words (*diagnosis*) promotes generalization to concrete words (*doctor*; Kiran et al., 2009; Sandberg & Kiran, 2014), but not vice versa (Kiran et al., 2009). In this case, abstract words serve as the more complex items, with concrete words as the related, less complex items.

As implied by the breadth of applications of the CATE, what makes related items more or less complex necessarily depends upon the goal of training and the type of stimulus. With regard to AbSANT, there are several dimensions of abstract and concrete words that help shape their differences in complexity, including differences in the quantity and quality of semantic features; differences in semantic diversity and inherent available context; and differences in strength, quantity, and quality of semantic connections. These characteristics of abstract and concrete words emerge from theories developed to help explain the concreteness effect (e.g., Newton & Barry, 1997; Paivio, 1991; Schwanenflugel & Shoben, 1983).

The Concreteness Effect

Concrete concepts (e.g., *pencil*) are easily imagined (i.e., imageability) and can be experienced by the senses (i.e., concreteness), while abstract concepts (e.g., *truth*) are those that cannot be experienced by the senses and for which a mental image is difficult to produce (Paivio et al., 1968). Thus, concrete words typically have high concreteness and high imageability, whereas abstract words typically have low concreteness and low imageability. The psycholinguistic factors of concreteness and imageability affect word processing in neurologically intact adults, such that individuals exhibit better performance (in the form of higher accuracy and lower reaction times) for concrete words than abstract words in a variety of lexical tasks. This well-studied phenomenon is called the concreteness effect (see Stoke, 1929, for an early description of the effect of concreteness on word recall). In some clinical cases, the concreteness effect has been shown to be reversed, with better performance for abstract than concrete words (e.g., Breedin et al., 1994; Warrington, 1975). This double dissociation suggests differing representation and/or processing mechanisms for abstract and concrete concepts, prompting several theories aimed at understanding these differences. While differing in their overarching approach to the organization of the semantic system, key aspects of a number of these theories are compatible and not only aid in the understanding of the concreteness effect but also help inform the basis for the current therapeutic approach as well as frame the interpretation of its effects. Relevant theories and their contributions are outlined below.

The Normal Isolated Centrally Expressed Model

Specific to word retrieval in aphasia, imageability and concreteness have been shown to influence performance, even when the items tested are concrete enough to be used in confrontation naming (Nickels & Howard, 1995). Furthermore, PWA appear to be especially vulnerable to the concreteness effect, exhibiting much better performance for concrete than abstract words (e.g., Alyahya et al., 2018; Bird et al., 2003; Sandberg & Kiran, 2013) and, in one case series study, showing a robust neural distinction between abstract and concrete words (Sandberg & Kiran, 2013). Newton and Barry (1997) presented a case study in which an individual with deep dyslexia showed a concreteness effect during production, but not during comprehension of written words, suggesting that the deficit was at the level of lexicalization, in which the appropriate word form is selected based on the activated semantic representation, that is, word retrieval. They proposed the Normal Isolated Centrally Expressed (NICE) model, which states that lexicalization is normally influenced by concreteness because the specific nature of concrete word representations promotes strong activation of the target word and activation spreads to only a few related concepts while abstract words have less specificity and, therefore, weaker activation and more spreading activation. Thus, the selection of the appropriate label for an abstract concept is complicated by a number of weakly activated alternatives. In an individual with impaired word production, there is an increased threshold of activation required for production, which makes abstract words even less likely to cross the threshold. The NICE model provides a mechanism of differential impairment for abstract versus concrete words in aphasia, as well as an aspect of differential complexity for abstract versus concrete words in the strength and quantity of semantic connections. In addition to supporting the notion that abstract words are more complex than concrete words, the NICE model also provides a mechanism for more spreading activation for abstract words, which aids in the interpretation of the effects of training abstract versus concrete words. This concept will be explored more in the discussion.

The Dual Coding Theory and Related Theories

The dual coding theory suggests that there are two systems for representing and processing information—a verbal system and a nonverbal, sensorimotor system. Abstract concepts are thought to rely more on the verbal system, whereas concrete concepts have the additive benefit of the support of both systems, which results in better performance for concrete than abstract concepts (Paivio, 1991). The differential reliance of abstract and concrete words on the verbal and sensorimotor systems aligns with the notion that abstract words have fewer semantic features than concrete words, as illustrated in the connectionist model developed by Plaut and Shallice (1991). Semantic features tend to be about the physical, functional, and contextual characteristics of concepts, which tap into sensorimotor systems. In practice, this would mean that it would be more difficult to create factual statements about abstract than concrete words, as demonstrated by Jones (1985) and which he described as ease of predication. Together, these theories lay the groundwork for understanding the differences in semantic features for abstract and concrete words. Taking into account these differences in semantic features was important during the development of the AbSANT protocol, as analysis of semantic features is a key ingredient in AbSANT and similar semantically based therapy approaches. While differences in semantic features may also support the differential complexity for abstract versus concrete words, their role in generalization appears to differ from other semantic treatments based on CATE (e.g., Kiran & Thompson, 2003), as will be addressed in detail in the discussion.

The Context Availability Theory and Related Theories

In a different vein, the context availability theory suggests that both abstract and concrete concepts rely on a single system but have "differential availability of context," such that the context for concrete concepts is more easily retrieved than for abstract concepts. The readily available context for concrete concepts is thought to be responsible for the better performance observed for these concepts based on the fact that the concreteness effect is diminished when context is provided (Schwanenflugel & Shoben, 1983). This lack of readily available context for abstract words is related to the finding that abstract words have more semantic diversity than concrete words (Hoffman et al., 2013). Semantic diversity is a measure of variability in meaning across contexts. For example, a quick search in the thesaurus for the word *chance* returns an adjective meaning *accidental*; three noun meanings of *possibility*, *luck*, and *risk*; and two verb meanings of *endanger* and *happen*. Together, context availability and semantic diversity provide perhaps the most salient contribution to the difference in complexity between abstract and concrete words, as these factors highlight how the process of identifying the appropriate meaning for abstract words requires additional search and inhibition mechanisms.

The Different Representational Frameworks Hypothesis

More recently, Crutch and Warrington (2005) presented a case of semantic refractory access aphasia, in which the activation of related semantic information causes extreme difficulty with lexical access due to increased competition. The authors found that comprehension performance was worse when abstract words were presented in an associative (exercise–healthy) versus similarity (look–peek) context and when concrete words were presented in a similarity (carrot–onion) versus associative (cow–barn) context. Furthermore, in a group of neurologically intact adults using an odd-one-out judgment task, Crutch et al. (2009) found better performance for abstract words in an associative versus similarity context and for concrete words in a similarity versus associative context. Based on this converging evidence, the authors proposed the different representational frameworks (DRF) hypothesis that abstract and concrete concepts have qualitatively different organizational structures within the semantic system, such that abstract concepts are organized associatively, whereas concrete concepts are organized by semantic similarity. This hypothesis further refines the notion proposed in the NICE model regarding the differences in semantic connectivity of abstract and concrete words and may be especially relevant for the discussion of the effects of the current therapeutic approach, which places both abstract and concrete words into an associative context.

The Current Study

While the direct training and generalization effects of AbSANT have been shown in a number of participants (Kiran et al., 2009; Sandberg & Kiran, 2014), support for the absence of a generalization effect to abstract words when concrete words are trained is less robust (Kiran et al., 2009). Thus, the goal of the current study is to replicate that of Kiran et al. (2009) to determine whether this pattern holds. In the study of Kiran et al., four persons with anomic aphasia were recruited. All four PWA completed one phase of abstract word training, and three of the four also completed one phase of concrete word training. Word type (abstract vs. concrete) and category (hospital, courthouse, church) were counterbalanced across participants. P1 showed neither direct training nor generalization effects during abstract or concrete word training. During abstract word training, P2 and P3 showed both direct training (improvement of trained abstract words) and generalization (improvement of untrained related concrete words) effects, whereas P4 only showed generalization effects. During concrete word training, P3 and P4 showed direct training (improvement of trained concrete words) but not generalization (improvement of untrained related abstract words) effects. P2 did not complete a concrete word training phase.

To the best of our knowledge, only a handful of studies have addressed the treatment of abstract words since the publication of the seminal Kiran et al. (2009) study. In 2014, Sandberg and Kiran tested AbSANT in 12 persons with varying types of aphasia. All 12 PWA participated in one phase of abstract word training. The authors separated the PWA into groups, depending on their treatment outcome. Eight of the participants were deemed "generalizers" because they showed improvement on both the trained abstract words (*justice*) and the untrained related concrete words (*jury*). Two of the participants were deemed "nongeneralizers" because they showed improvement only on the trained abstract words. The remaining two participants were deemed "nonresponders" because they did not show any improvement.

In a case study, McCarthy et al. (2017) found that when a person with deep-phonological aphasia was trained to repeat abstract nouns within semantically cohesive adjective–noun phrases (*social exclusion*), repetition not only improved for the trained adjective–noun phrases but also generalized to the abstract nouns in isolation. While this treatment focused on abstract words, the treatment was aimed at repetition, not word retrieval per se, and was not designed to test generalization to concrete words.

In another case study, Renvall and Nickels (2019) trained a person with anomic aphasia to retrieve abstract emotive adjectives (*fantastic*) using repetition in the presence of a picture. They found that retrieval of the trained items improved, which is promising for improving the ability for PWA to express themselves. However, they found no generalization to untreated adjectives or pictures or to other tasks, such as connected speech.

Based on the Kiran et al. (2009) and Sandberg and Kiran (2014) studies, we hypothesize that training abstract words will result in direct training effects for the trained abstract words and generalization effects to related concrete words. Based on the study of Kiran et al., we hypothesize that training concrete words will result in direct training effects for the trained concrete words, but no generalization effects to related abstract words.

Method

Participants

Four native English speakers with aphasia participated in the study (see Table 1 for a complete description of participant demographics). All participants were right-handed and in the chronic stage of recovery from stroke, which was operationalized as 6 months poststroke. Three participants provided informed consent in accordance with the institutional review board at San Francisco State University, and one participant provided informed consent in accordance with the institutional review board at Penn State University.

Assessment

All participants completed a battery of standardized tests. The Western Aphasia Battery–Revised (Kertesz, 2006) was used to determine the type and severity of aphasia. Two participants fit the classification of conduction aphasia, one fit the Broca's classification, and one fit the anomic classification. Three participants exhibited moderate aphasia and moderate apraxia, and one exhibited mild aphasia with no apraxia. All participants showed impaired naming on the Boston Naming Test (Kaplan et al., 2001). The Pyramids and Palm Trees Test, Three-Picture Version (Howard & Patterson, 1992), tested nonverbal semantic reasoning. Three of the four participants scored within normal limits, and one scored slightly below the other three, revealing a mild semantic impairment. Subtests of the Psycholinguistic Assessment of Language Processing in Aphasia (Kay et al., 1992) were used to examine word recognition (lexical decision) and semantic processing (synonym judgment) abilities for both abstract and concrete words. For three participants, the average accuracy across the five Psycholinguistic Assessment of Language Processing in Aphasia subtests was higher for concrete than abstract words, in line with the pattern predicted by the concreteness effect. Overall, the scores obtained on these language tests indicate

Table 1. Demographic and assessment data for all participants.

Variable	P1		P2		P3		P4	
Age	70		51		47		51	
Sex	Male		Male		Female		Male	
Education	16		16		16		15	
Months poststroke	132		48		156		21	
Lesion region	LMCA		LMCA		LMCA		LMCA	
	Pre-Tx	**Post-Tx**	**Pre-Tx**	**Post-Tx**	**Pre-Tx**	**Post-Tx**	**Pre-Tx**	**Post-Tx**
Western Aphasia Battery								
Aphasia Quotient	63.9	62.1	59.8	62.9	93.5	93.5	51.1	55.4
Aphasia Type	Conduction	Conduction	Broca's	Broca's	Anomic	Anomic	Conduction	Conduction
Boston Naming Test	46.67%*	50.00%*	50.00%*	65.00%*	81.67%	93.33%	5.00%*	18.33%*
Psycholinguistic Assessment of Language Processing in Aphasia								
Auditory Lexical Decision								
High imageability	100.00%	100.00%	95.00%*	97.50%	100.00%	100.00%	100.00%	100.00%
Low imageability	97.50%	90.00%*	80.00%*	97.50%	100.00%	95.00%	67.50%*	77.50%*
Visual Lexical Decision								
High imageability	90.00%*	96.67%	96.67%	100.00%	100.00%	93.33%	96.67%	96.67%
Low imageability	93.33%	96.67%	90.00%*	96.67%	96.67%	96.67%	66.67%*	90.00%*
Auditory Synonym Judgment								
High imageability	86.67%	83.33%	93.33%	93.33%	100.00%	100.00%	63.33%	83.33%
Low imageability	60.00%	60.00%	53.33%	63.33%	93.33%	93.33%	73.33%	63.33%
Written Synonym Judgment								
High imageability	86.67%	83.33%	86.67%	76.67%	100.00%	100.00%	76.67%	86.67%
Low imageability	60.00%	66.67%	60.00%	76.67%	90.00%	93.33%	80.00%	83.33%
Word Semantic Association								
High imageability	66.67%*	66.67%*	60.00%*	80.00%	93.33%	93.33%	66.67%*	86.67%
Low imageability	80.00%	6.67%*	20.00%*	40.00%*	80.00%	93.33%	40.00%*	46.67%*
Pyramids and Palm Trees Test								
3 Pictures	92.31%	90.38%	88.46%*	80.77%*	92.31%	98.08%	92.31%	92.31%
Cognitive Linguistic Quick Test								
Composite Severity	Mild	Mild	Moderate	Moderate	Mild	WNL	Mild	Mild
Test of Nonverbal Intelligence	Average	Average	Average	Below average	Above average	Above average	Average	Average

Note. For subtests that have norms, scores that fall below the cutoff for unimpaired performance are marked with an asterisk. Cutoff values for the Boston Naming Test were obtained from Nicholas et al. (1989). Cutoff values for the Pyramids and Palm Trees Test were obtained from Howard and Patterson (1992). Cutoff values for the Psycholinguistic Assessment of Language Processing in Aphasia were extrapolated from the norms provided in the test by Kay et al. (1992). However, no norms are available for the Synonym Judgment tasks. LMCA = left middle cerebral artery; Tx = treatment; WNL = within normal limits.

that these participants have difficulty with naming, but have relatively spared semantic processing, and are able to make decisions about the meanings of both abstract and concrete words, although performance for abstract words is generally lower than for concrete words. This language profile suggests that these participants are able to complete the tasks in the therapy protocol.

Participants also completed the Cognitive Linguistic Quick Test Plus (Helm-Estabrooks, 2001) to examine attention, memory, and executive functions in addition to language functions. Three participants obtained a mild composite severity rating, and one participant was within normal limits. The Test of Nonverbal Intelligence, Fourth Edition (Brown et al., 2010) measures intelligence without the influence of language. Three participants exhibited average performance, and one participant exhibited above-average performance. The results of these cognitive tests suggest that these participants do not have any major nonlinguistic cognitive deficits that would interfere with completing the tasks in the therapy protocol.

Treatment

Design

A single-subject multiple baseline design was implemented. Each participant underwent four phases: baseline, concrete word treatment, abstract word treatment, and posttreatment. Treatment category was counterbalanced across treatment type and across participants.

Stimuli

The stimuli from Kiran et al. (2009) and Sandberg and Kiran (2014) were used in this study.

Categories. As described in detail in the study of Kiran et al. (2009), categories were developed by asking healthy young adults to list words associated with eight different locations (hospital, school, park, church, office, courthouse, restaurant, and museum). From these, three were chosen, which had low overlap with each other and elicited a large number of both abstract and concrete words—hospital, church, and courthouse.

Words. As described in detail by Sandberg and Kiran (2014), the original stimuli from Kiran et al. (2009) were expanded and refined by checking the words generated for each category against association norms (Kiss et al., 1973; Nelson et al., 1998). Furthermore, the word lists within and across categories were balanced on psycholinguistic variables, such that the abstract and concrete words significantly differed on imageability and concreteness but did not significantly differ on frequency and familiarity (Frances & Kucera, 1983; Gilhooly & Logie, 1980; Paivio et al., 1968). The result was three categories (*church*, *hospital*, and *courthouse*), each with 10 target abstract words (*belief*) and 10 target concrete words (*Bible*). See Appendix A for a complete list of the stimuli.

Features. As described by Kiran et al. (2009) and Sandberg and Kiran (2014), 45 semantic features for each category were used to train the target words in that category. Fifteen of the 45 semantic features were based on dictionary definitions of abstract and concrete (*can be touched*) and were referred to as generic features. Fifteen of the 45 features belonged to words in an unrelated category (e.g., *has feathers* would be a feature applicable to any item in the category *birds*, but not to any target item in the categories *hospital*, *courthouse*, or *church*) and were referred to as distractor features. Fifteen of the 45 features were generated by each participant during the first training session, which was called a brainstorming session. In the brainstorming session, the clinician prompted the participant to explain the meaning and associations of each target word in as much detail as possible. After the brainstorm session, the research team met to create features from the descriptions generated by the participant. Care was taken to ensure that none of the features contained target words. Appendix A contains the predetermined semantic features and some examples of participant-derived features.

Probes

The dependent variable for treatment effects was performance on generative naming probes. For each context-category, participants were asked to generate as many abstract and concrete words as possible in 2 min. The definitions of abstract and concrete were provided, and the participants were encouraged to imagine themselves in the context (e.g., *courthouse*) and think of as many things, people, ideas, and feelings associated with that context as possible. This scaffolding was faded as appropriate. Percent accuracy was calculated based on how many of the 10 abstract and 10 concrete predetermined target words (see Appendix A) were provided by the participant. Other responses that were provided were analyzed separately. A detailed administration and scoring protocol for the probes is located in Appendix B.

Each participant received five baseline probes prior to beginning the first phase of training, a probe at the beginning of every other training session (such that there were always two training sessions between each probe), five posttreatment probes after completing the second phase of training, and two maintenance probes administered on consecutive days, approximately 1 month after finishing the last posttreatment probe. Exceptions included P3, who received four baseline probes and did not complete maintenance probes due to scheduling issues, and P4, whose maintenance probes occurred approximately 3.5 months after finishing the last posttreatment probe due to scheduling issues. Additionally, P1 and P2 completed several "interim" probes between training phases, as their schedules did not allow them to continue with the second training phase immediately after the first training phase.

Treatment Protocol

The treatment used in this study is the AbSANT program developed by Sandberg and Kiran (Kiran et al., 2009; Sandberg & Kiran, 2014). For the current study, we converted the original paper-and-pencil version of the task into a computerized version using Qualtrics survey software (Qualtrics, 2020). Screenshots of the task are located in Appendix B with the training steps.

Each participant completed two phases of therapy. Each phase consisted of approximately 10 weeks of therapy, with two sessions each week, for up to 20 sessions. Each session was approximately 2 hr long, for a total of 4 hr of therapy per week. Each participant was first trained on concrete words in either the category *hospital* or *courthouse*. Category assignment was counterbalanced across participants (see Table 2). After reaching criterion for the first phase of training (i.e., concrete word training), each participant was trained on abstract words in the remaining category (*hospital* or *courthouse*). The criteria for ending a training phase was 80% accuracy for two probes in a row or 20 training sessions, whichever came first. In summary, the actual total number of treatment sessions for each participant was 20 sessions in each phase for P1 and P2, 16 sessions in Phase 1 and 20 sessions in Phase 2 for P3, and 20 sessions in Phase 1 and 16 sessions in Phase 2 for P4.

Throughout both phases of training, the category *church* served as a control category for which target items were seen, but not trained (these items were used in the category sorting step explained below), which we will call the "exposed control category" hereafter. All three categories were probed throughout all phases of the study. For P3, the first phase was completed remotely using Zoom (2019) due to difficulties arranging in-person sessions during that time. Zoom is a secure videoconferencing platform. Because part of the therapy is designed to be used on the computer and the other part consists of verbal exchanges, the use of videoconferencing for this phase did not change the administration of the protocol.

Training Steps

Each participant completed the same training steps in each treatment session, except for the brainstorming session, of each phase. The training protocol is provided in detail in Appendix A but is briefly summarized here.

Category sorting. The first step was category sorting in which the participant sorted 40 words into their respective categories. Twenty of the words were in the trained

Table 2. Counterbalancing of context-categories across participants.

	Phase 1 training		Phase 2 training		Control
ID	Context-category	Word type	Context-category	Word type	Context-category
P1	hospital	concrete	courthouse	abstract	church
P2	courthouse	concrete	hospital	abstract	church
P3	hospital	concrete	courthouse	abstract	church
P4	courthouse	concrete	hospital	abstract	church

context-category (either *hospital* or *courthouse*—only 10 of which were trained in subsequent steps), and 20 were in the context-category *church* (i.e., the exposed control category), which were only seen in this step. In both categories, there were 10 target abstract words and 10 target concrete words. Participants sorted the words independently. However, the clinician could provide support, such as reading each word, as needed, fading support as appropriate. If a word was placed in the incorrect category, the clinician would provide feedback regarding the correct response as detailed in Appendix B.

Feature selection. After category sorting, one of the trained target words was presented, along with the 45 semantic features, one at a time, in random order. The participant was asked to choose the first six features that helped describe the target word. If the participant had difficulty reading a feature, the clinician would read it aloud to the participant and then fade this support as appropriate. While the selection of features was highly subjective, the clinician would often ask the participant to explain their choice, especially if the choice was not obvious to the clinician. Occasionally, this discussion and reanalysis of the meaning of the word within the specific context of the trained context-category would result in the choice of a different feature that was a better fit. Anecdotally, this discussion was considered to be a worthwhile and even anticipated exchange by participants.[2]

After the participant chose the six features, the clinician asked the participant to read the target word and each feature in sentence form aloud. For example, if the target word was *diagnosis* and the feature was “exists only in the mind,” then the clinician would ask the participant to say, “Diagnosis exists only in the mind.” Scaffolding was provided based on participants’ needs. For example, for those with apraxia, the clinician would provide cues to help with pronunciation and accept approximations.[3]

Yes/no questions. After feature selection, the clinician would remove the stimuli from view and ask the participant 15 yes/no questions about the target word, using the 45 features. Five questions had an expected “yes” response, five questions had an expected “no” response for the word being trained but could be a possibility for another word in the context-category, and five questions were from the distractor features and thus had a definite expectation of a “no” response. During this step, the clinician was careful not to state the target word in order to reduce auditory repetition of the word because the next step required the participant to recall the word. The questions were pseudorandomized to avoid too many of the same responses in a row in order to limit the predictability of the expected response. If the participant provided an unexpected response, the clinician would ask the participant to explain their choice. As during the feature selection step, if the explanation was considered sound by the clinician, then it was accepted; if not, the clinician would provide an alternative viewpoint for the participant to consider. If the participant and clinician could not agree on the best response to the question, then a different question was asked in its place. Again, this rarely occurred, and any discussion during this step was considered positive by both the participants and the clinicians.

Synonym, word type, word recall. After all 15 yes/no questions had been asked, without stating the target word, the clinician asked the participant if the target word was abstract or concrete, then asked the participant to provide a synonym for the target word, and finally, to recall the target word. After this step, the clinician would move on to the next target word.

Free generative naming. Near the end of the session, with approximately 5–15 min remaining, the clinician would ask the participant to name as many words in the trained context-category as possible. This step was distinct from the generative naming task used to probe treatment effects in two important ways. First, there was no time limit (other than the end of the session). This allowed the participant to work at their own pace to generate as many words as possible. Second, the clinician was allowed to provide feedback and scaffolding during this step. This allowed the participant successful practice with the generative naming task.

Effect Size Calculation

Treatment effect size (ES) was calculated based on the study of Beeson and Robey (2006, 2008), who used a modified Cohen’s *d*. The formula for the ES calculation is the mean of the posttreatment performance minus the mean of the baseline performance, divided by the standard

[2]We did not track the selection of nonobvious features and subsequent discussions, but tracking the effect of this aspect of the protocol on treatment outcomes in future work would be informative.
[3]This therapy was not developed to address apraxia; therefore, we did not specifically measure the effect of apraxia on treatment outcomes or the effect of this protocol on apraxia outcomes. Future work would benefit from including this information.

deviation of the baseline performance. ESs were also calculated for the maintenance probes using the same formula, substituting the maintenance probes for the posttreatment probes. The interpretation of the magnitude of the ES for direct training effects and generalization effects was also based on the study of Beeson and Robey (2008), which suggested 6.5, 8, and 9.5 as small, medium, and large thresholds, respectively, for direct training effects and 2, 5, and 8 as small, medium, and large thresholds, respectively, for generalization effects.

For P1 and P2, the ES calculation used the "interim" probes as the posttreatment probes for the first training phase and as the baseline probes for the second training phase. For P3 and P4, the ES calculation used the last three probes of the first training phase as the baseline for the second training phase and the first three probes of the second training phase as the posttreatment probes for the first training phase. As a supplement to ESs, similar to the study of Sandberg and Kiran (2014), paired *t* tests were carried out on the change from the baseline average to the posttest average and from the posttest average to the maintenance average to examine the significance of the training effect across participants.

Reliability / Treatment Fidelity

Reliability was performed on an average of 49% of the probes (range across participants: 42%–53%), resulting in an average of 97% agreement (range across participants: 96%–98%). In addition to percent agreement, Cohen's kappa was used to test interrater reliability across observations for each participant. The average agreement, based on guidelines from McHugh (2012), was strong (k = .86, SD = .10), ranging from moderate to almost perfect (.71–.91) across participants. Treatment fidelity was performed on an average of 32% of the treatment sessions (range across participants: 25%–38%), resulting in an average of 98% fidelity (range across participants: 92%–100%). See Appendix C for an example of the scoresheet used for treatment fidelity.

Results

In general, participants showed improvement for directly trained items, regardless of whether concrete or abstract words were trained. However, participants only showed generalization to other target words in the same context-category when abstract words were trained. Table 3 details the ESs for every condition for every participant.

Participant 1

The graph of P1's performance is shown in Figure 1. P1 was first trained on concrete words in the context-category *hospital*. They showed a small direct training effect for the target concrete words (ES = 7.67), but no generalization effect for the related target abstract words (ES = 0.89). During the second phase of training, P1 was trained on abstract words in the context-category *courthouse*. Although a small positive change was noted, neither the direct training ES for the target abstract words (ES = 2.24) nor the generalization ES for related target concrete words (ES = 1.83) passed the small threshold (6.5 and 2, respectively). For the exposed control category, P1 showed no effect for the exposed abstract words (ES = 1.79) but showed a small effect for the exposed concrete words (ES = 2.91). At the maintenance probes 1 month later, the ES for the trained abstract words dropped to 1.79, the ES for the generalized concrete words dropped to −0.73, and the ES for the trained concrete words dropped to 3.83.

Participant 2

The graph of P2's performance is shown in Figure 2. P2 was first trained on concrete words in the context-category *courthouse*. They showed no direct training effect for the target concrete words (ES = −0.57), nor a generalization effect for the related target abstract words (ES = 1.11). During the second phase of training, P2 was trained on abstract words in the context-category *hospital*. Like P1, the direct training effect for target abstract words (ES = 2.67) did not pass the small threshold (6.5); however, unlike P1, there was a small generalization effect for related target concrete words (ES = 4.44). For the exposed control category, P2 showed no effects for the exposed abstract (ES = −0.96) or concrete (ES = 0.15) words. At the maintenance probes 1 month later, the ES for the trained abstract words dropped to 1.11, the ES for the generalized concrete words increased to 5.56, and the ES for the trained concrete words increased to 0.53.

Participant 3

The graph of P3's performance is shown in Figure 3. P3 was first trained on concrete words in the context-category *hospital*. They showed a large direct training effect for the target concrete words (ES = 10.10), but no generalization effect for the related target abstract words (ES = −0.09). During the second phase of training, P3 was trained on abstract words in the context-category *courthouse*. They showed a large direct training effect for target abstract words (ES = 12.70) and a small generalization effect for related target concrete words (ES = 2.31). For the exposed control category, P3 showed no effect for the exposed abstract (ES = 0.24) or concrete (ES = 1.10) words. We were unable to collect maintenance probes for P3. Notably, the use of videoconferencing for P3 during Phase 1 does not appear to have negatively affected treatment outcomes, since P3 achieved a large direct training effect during this phase.

Participant 4

The graph of P4's performance is shown in Figure 4. P4 was first trained on concrete words in the context-category *courthouse*. They showed a large direct training effect for the target concrete words (ES = 10.73), but no

Table 3. Effect sizes for each condition for each participant.

	Concrete word training		Abstract word training		Exposed control	
ID	Abstract	Concrete	Abstract	Concrete	Abstract	Concrete
P1	0.89	7.67^	2.24	1.83	1.79	2.91*
P2	1.11	–0.57	2.67	4.44*	–0.96	0.15
P3	–0.09	10.10^^^	12.7^^^	2.31*	0.24	1.10
P4	0.00	10.73^^^	15.95^^^	3.18*	2.78*	1.55
Average	0.48	6.98^	8.39^^	2.94*	0.96	1.43

Note. Carets and asterisks indicate strength of direct training and generalization effect sizes, respectively, based on Beeson and Robey (2008). Direct training: ^^^ = large, ^^ = medium, ^ = small. Generalization: *** = large, ** = medium, * = small.

generalization effect for the related target abstract words (ES = 0.00). During the second phase of training, P4 was trained on abstract words in the context-category *hospital*. They showed a large direct training effect for target abstract words (ES = 15.95) and a small generalization effect for related target concrete words (ES = 3.18). For the exposed control category, P4 showed a small effect for the exposed abstract words (ES = 2.78) and no effect for the exposed concrete words (ES = 1.55). For P4, the maintenance probes were not able to be given until approximately 3.5 after the posttreatment probes. At this time, the ES for the trained abstract words dropped to 10.34, the ES for the generalized concrete words dropped to −0.29, and the ES for the trained concrete words increased to 12.97.

Group Results

While an average ES across four participants is not powerful, it is included here to show the group trend. The average direct training ES for trained target concrete words was small (6.98), and the average generalization ES for untrained target abstract words in the same context-category did not pass the small threshold (0.48). The average direct training ES for trained target abstract words was medium (8.39), and the average generalization ES for untrained target concrete words in the same context-category was small (2.94). The other conditions did not show any appreciable change.

The paired t tests on the change from baseline to posttest and from posttest to maintenance resulted in only one change that was significant at an alpha level of $p < .05$. Across participants, untrained concrete words that were related to trained abstract words significantly improved during abstract word training, that is, generalization, $t(6.71, 3)$, $p = .007$. This suggests that the generalization effect may be the strongest effect observed in this group of participants for this therapy. It is worth noting here that power is reduced with the small sample size. Additionally, while the lack of significance of the paired t tests of the change from posttreatment to maintenance probes suggests maintenance of treatment effects, these results may be an artifact of reduced power.

Discussion

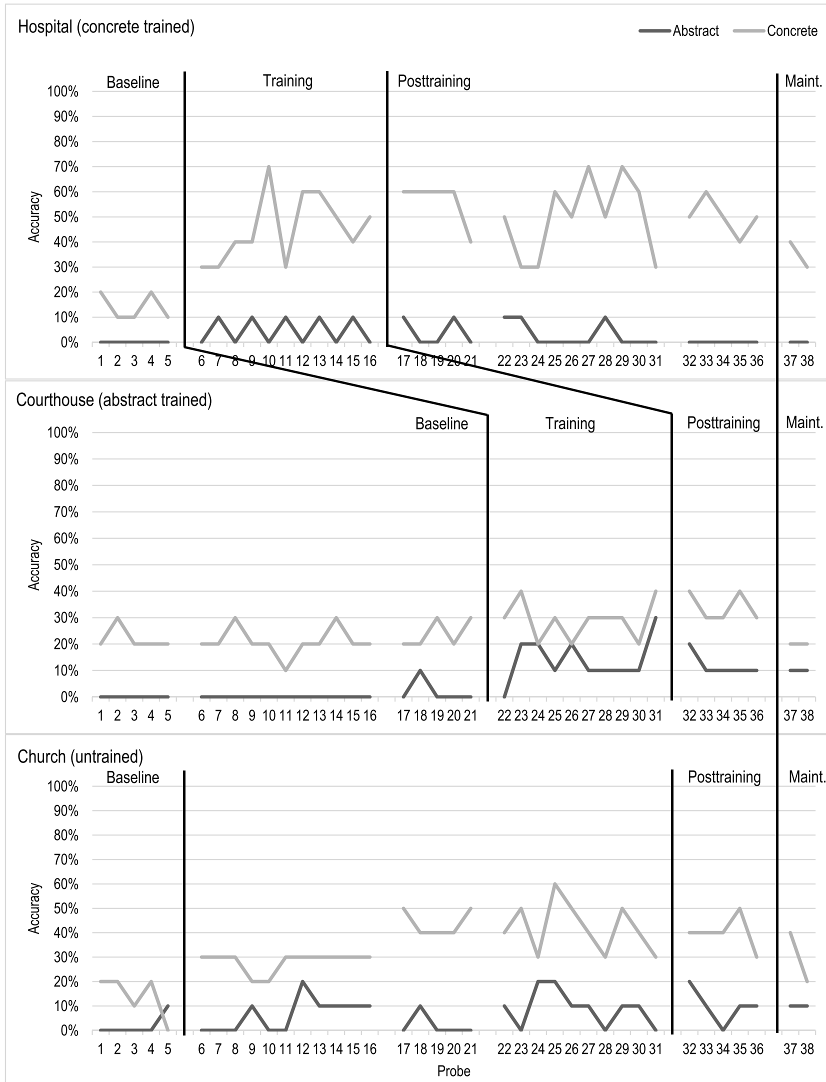
The Effect of Training Abstract Words

Our first hypothesis was that when abstract words were trained, both direct training and generalization would be observed. On average, the group showed a medium direct training effect for abstract words and a small generalization effect for related concrete words, supporting this hypothesis. This is similar to the average effects in the study of Sandberg and Kiran (2014), which equated to a medium direct training effect for abstract words and a small generalization effect for concrete words, and slightly better than the average effects in the study of Kiran et al. (2009), which equated to a small direct training effect and a small generalization effect (see Table 4).

At the individual level, the direct training effects appear to be driven by two participants, P3 and P4. These participants showed large direct training effects for abstract words. Although P1 and P2 did not reach the threshold for a small direct training ES for abstract words, they did show positive change, with ESs of 2.24 and 2.67, respectively. This is a similar ratio to that of Sandberg and Kiran (2014), which showed large direct training effects for six of the 12 participants and small positive changes that did not reach the small direct training threshold for four additional participants, and slightly better than that of Kiran et al. (2009), which showed a large direct training effect for one participant and small positive changes that did not reach the small direct training threshold for two participants.

In the current study, three of the four participants showed at least a small generalization effect for related concrete words, with P2 nearly reaching the medium ES threshold. This is similar to the results of both Kiran et al. (2009) and Sandberg and Kiran (2014), in which three of four and eight of 12 participants, respectively, showed generalization to related concrete words. While the majority of participants in these studies show both direct training effects and generalization, there are a small number of participants who do not show these effects. It is unclear exactly what patient-specific factors may contribute to differences in direct training and generalization patterns. More work is needed that systematically examines the role

Figure 1. Generative naming accuracy for abstract (dark gray) and concrete (light gray) words in each context-category for each phase of treatment for P1.



of both linguistic and nonlinguistic factors in treatment outcomes.

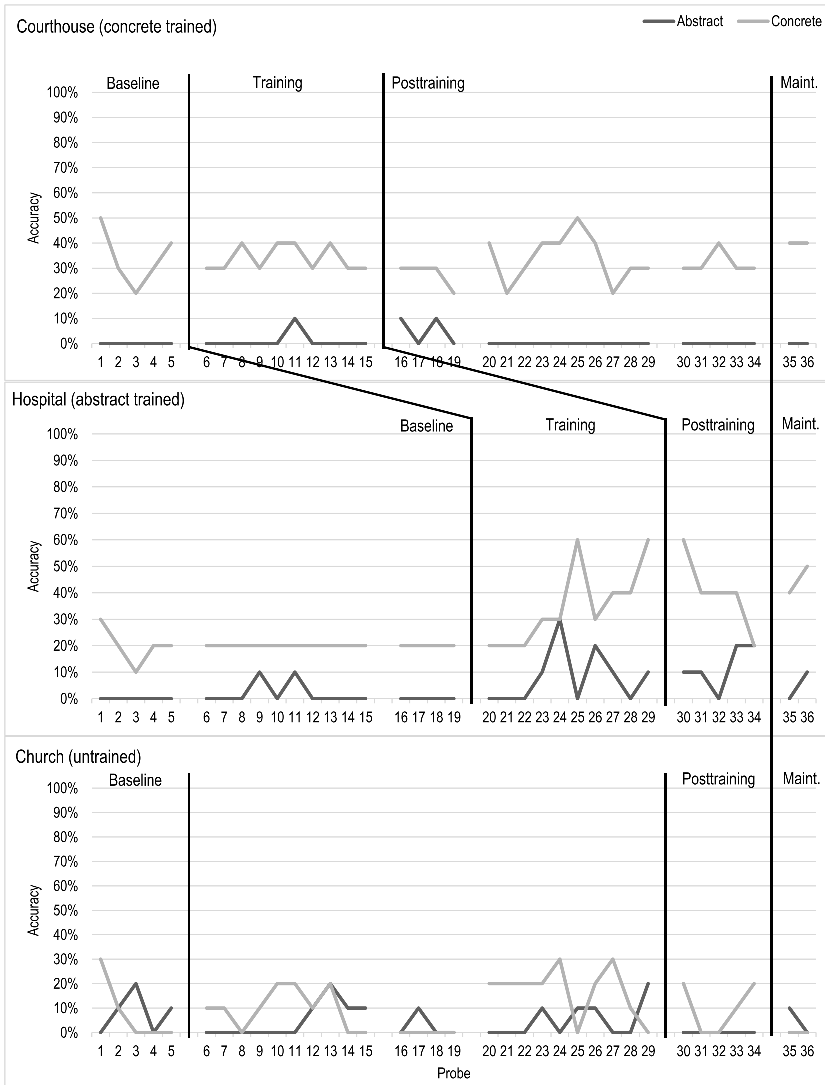
The Effect of Training Concrete Words

Our second hypothesis was that when concrete words were trained, there would be direct training, but not generalization effects. On average, the group showed a small direct training effect for concrete words and no generalization effect for related abstract words, supporting our second hypothesis. This is similar to average effects in the study of Kiran et al. (2009), which equated to positive changes for concrete words that did not reach the threshold for the small direct training effect and no generalization effect for related abstract words.

At the individual level, P3 and P4 showed large direct training effects for concrete words, while P1 showed a small direct training effect and P2 showed no effect. In the study of Kiran et al. (2009), only three of the four participants completed concrete word training. Of those three, one showed a large direct training effect, one showed a small positive change that did not reach the small direct training effect threshold, and one showed no effect.

One difference between the study of Kiran et al. (2009) and the current study is that concrete words were trained in the second phase for two of the three participants who received concrete word training in the study of Kiran et al., while concrete words were trained in the first phase of the current study. It is possible the order of training affected the results. However, the similarity in the patterns of treatment effects

Figure 2. Generative naming accuracy for abstract (dark gray) and concrete (light gray) words in each context-category for each phase of treatment for P2.



for both studies suggests no role of treatment order. That being said, more work is needed to systematically examine effects of the order of training abstract versus concrete words.

Maintenance

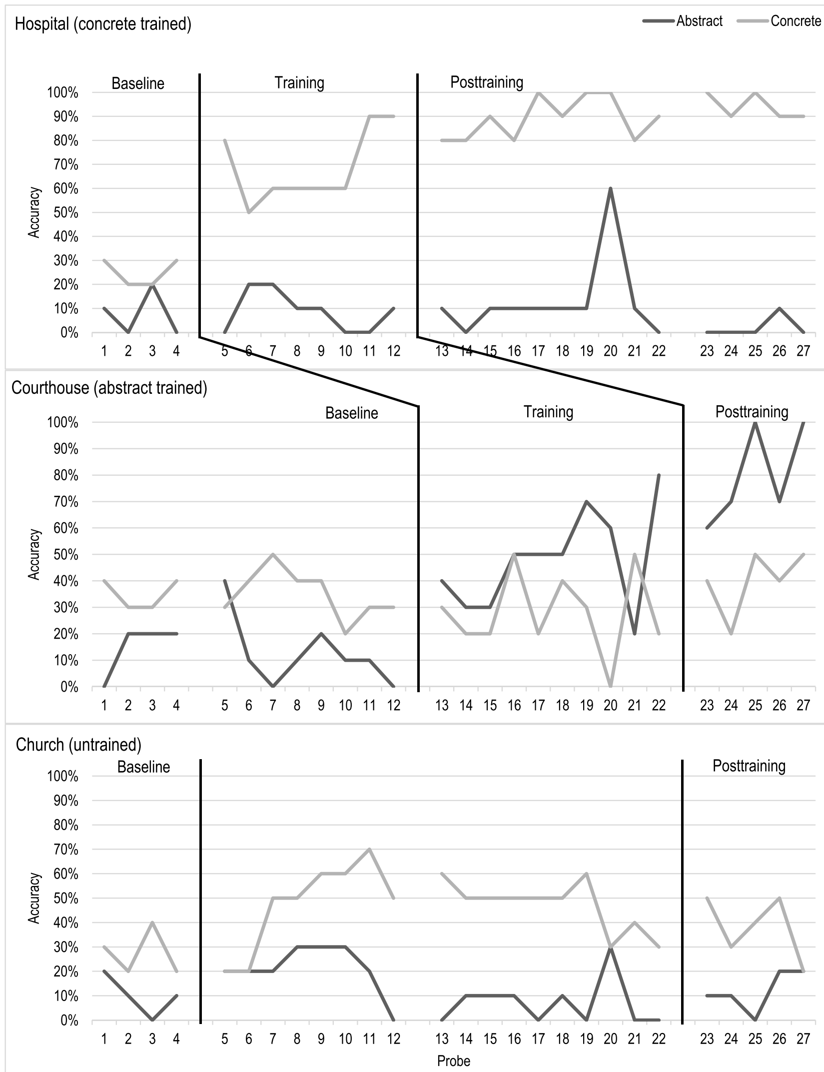
For P1 and P2 (P3 did not complete maintenance probes), none of the ESs at 1 month posttreatment reached the small threshold. However, for items that improved, performance remained above baseline, except for generalization to concrete words for P1, and some improved, such as generalized concrete words and directly trained concrete words for P2. This metric has been used previously to determine maintenance for a similar therapeutic approach (Boyle, 2010).

P4 provided an interesting case study of long-term maintenance, as he was not probed again until 4 months after treatment ended. Interestingly, while P4 showed a drop in ES for the directly trained abstract words, the ES was still large. For directly trained concrete words, the ES actually increased but was large to begin with. Conversely, for generalized concrete words, the ES decreased to lower than the small threshold. With the current data, it is not possible to determine the basis for this difference in change among directly trained abstract words, generalized concrete words, and directly trained concrete words. More work is needed examining long-term maintenance effects of both direct training and generalization effects of AbSANT.

Converging Evidence

To date, 20 PWA across three studies conducted in four different cities across the United States have received

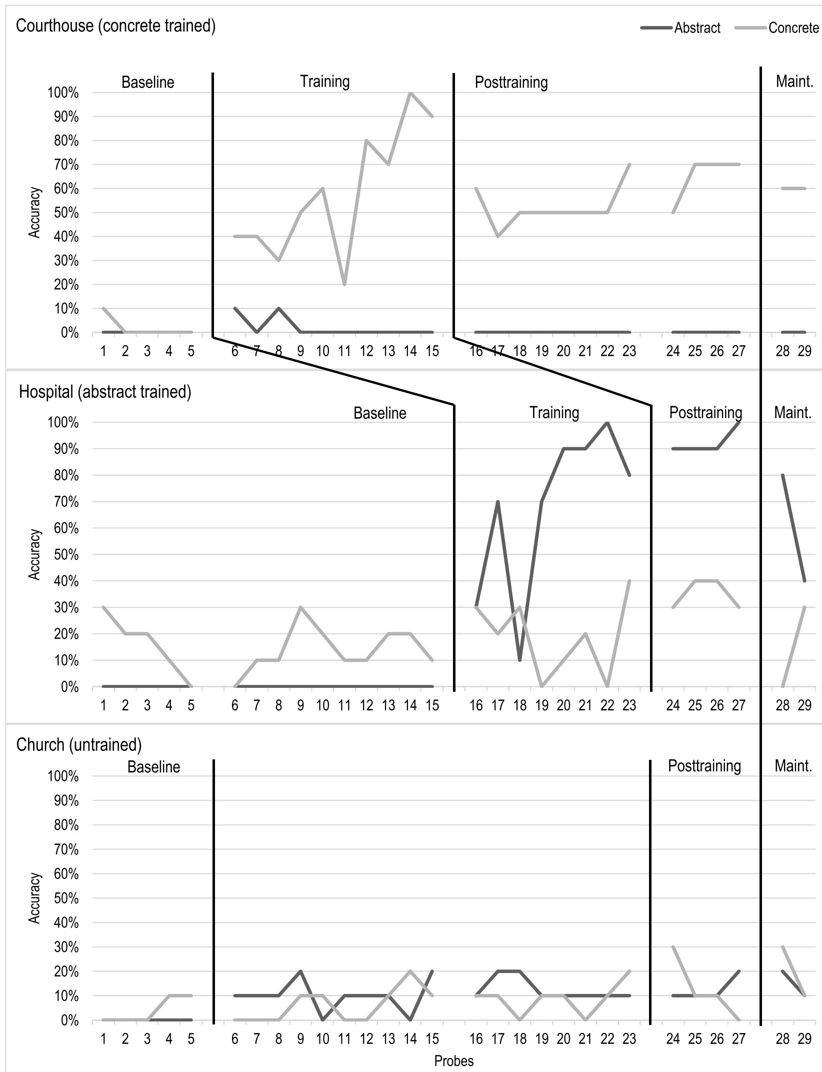
Figure 3. Generative naming accuracy for abstract (dark gray) and concrete (light gray) words in each context-category for each phase of treatment for P3.



AbSANT. Comparing the average ES for each of the three studies, training abstract words results in average direct training ES ranging from 7.62 (small) to 8.39 (medium) and average generalization effects to concrete words ranging from 2.20 (small) to 2.94 (small). Together, these studies show that AbSANT is effective because it results in measurable improvement of the directly trained items, and it is efficient because it also results in generalization to related, but untrained items. However, until this study, there was only scant evidence (based on three cases) that training concrete words was not as efficient as training abstract words. With the current study, seven PWA have now completed both abstract and concrete word training using the same protocol. Across both studies, the average direct training ES for concrete words ranges from 3.65 (null) to 6.98 (small), and the average generalization ES for untrained abstract words ranges from 0.48 (null) to 1.63 (null). Thus, the combined evidence suggests not only that training abstract words results in both direct training and generalization effects while training concrete words only results in a direct training effect but also that training abstract words results in larger direct training effects than training concrete words.

It is worth noting here that there are factors that can influence observed differences in direct training ES for abstract versus concrete words. First, differences in baseline variability could affect the magnitude of the ES because baseline variability is the denominator of the ES equation. However, in the current study, there is no difference in the average baseline standard deviation between abstract and concrete words ($M_{abstract}$ = 0.60, $M_{concrete}$ = 0.77, p = .32). Second, there may simply be more "room for growth" for

Figure 4. Generative naming accuracy for abstract (dark gray) and concrete (light gray) words in each context-category for each phase of treatment for P4.



abstract words. However, in the current study, when the gain is calculated based on the difference between the average baseline accuracy and 100% accuracy (see Lambon Ralph et al., 2010, for a description and calculation of the proportion of the potential maximal gain), there is a similar proportion of gain for abstract and concrete words ($M_{\text{abstract}} = 0.48$, $M_{\text{concrete}} = 0.41$, $p = .72$). Finally, there may be "regression to the mean" as described by Howard et al. (2015), in which items that are below the mean appear to improve simply because later performance brings them more in line with the mean. However, if abstract word performance were simply approaching the mean, then untrained abstract words (both those in the same category as the trained concrete words and those in the untrained category) would improve, but this is not the case.

Mechanisms of Generalization

AbSANT was originally conceived based on the theory that training more complex items results in benefit to both the directly trained items and untrained items that are related but less complex (i.e., the CATE; Thompson et al., 2003). Prior to developing AbSANT, the CATE had successfully been applied to semantics using typicality as a mode of complexity. In the case of typicality, atypical items (*penguin*) are considered to be more complex than typical items (*robin*) because they have semantic features that are typical to the category (*lays eggs*) and additional semantic features that make them atypical exemplars of the category (e.g., *swims* rather than *flies*). Generalization from an atypical bird like *penguin* to a typical bird like *robin* is thought to occur because of the overlap in typical semantic

Table 4. Comparison of average effect sizes across studies.

Study	Concrete word training		Abstract word training	
	Abstract	Concrete	Abstract	Concrete
Kiran et al. (2009)				
Average	1.63	3.65	7.62^	2.70*
Sandberg & Kiran (2014)				
Average			8.28^^	2.20*
Current study				
Average	0.48	6.98^	8.39^^	2.94*

Note. Carets and asterisks indicate strength of direct training and generalization effect sizes, respectively, based on Beeson and Robey (2008). Direct training: ^^^ = large, ^^ = medium, ^ = small. Generalization: *** = large, ** = medium, * = small.

features (*lays eggs*) that exist between the two related items.

In the case of concreteness, the semantic features do not play the same type of role in complexity as those in the case of typicality. Abstract and concrete words do not share very many semantic features because they do not co-occur in natural (taxonomic) categories like birds. And, unlike in typicality, the semantic features for abstract words do not subsume the semantic features of concrete words. In fact, abstract words are thought to have fewer semantic features than concrete words overall. Thus, the featural overlap is an unlikely mechanism for generalization from abstract to concrete words, and the complexity account for abstract words may need to be expanded to fully understand what supports the transfer of benefit from abstract to concrete words.

In the study of Boyle and Coelho (1995), the authors suggest that training semantic features promotes spreading activation from the features to the concept, making that concept more active within the semantic network. Once a concept rises above a certain activation threshold (which may be higher in aphasia; Newton & Barry, 1997), its label is more easily retrieved. This interpretation of the effects of SFA is based on the spreading activation theory described by Collins and Loftus (1975), which states that semantic knowledge can be represented as a network in which concepts are the nodes and are linked to each other based on relatedness. When a concept is activated, activation spreads to related concepts. This theory assumes that properties of concepts (i.e., semantic features) help form the links in the network between concepts. Importantly, this theory also allows for the links between concepts to be based on associations, super- or subordinate relationships, or any other relationship between concepts. Although the conclusions of both studies (Boyle & Coelho, 1995; Collins & Loftus, 1975) are based on concrete words, we can extend this reasoning to abstract concepts. Let us also assume that the semantic network behaves as described in the NICE model, in which concrete concepts have strong and specific representations, with limited spreading activation to closely related nodes, and abstract concepts having weaker and less specific representations with more spreading activation to loosely related nodes (Newton & Barry, 1997). This aligns with the notion that abstract words have higher semantic diversity than concrete words (Hoffman et al., 2013). In this scenario, it can be argued that training the semantic features of an abstract concept increases the level of activation of that concept above the threshold required for lexical retrieval, and then activation spreads along the links from that concept to a wide variety of related concepts, including concrete concepts, helping to raise these concepts above the activation threshold required for lexical retrieval. Training the semantic features of concrete concepts, however, may only promote activation of that concept and a few closely related concepts that share semantic features.

Why might activation spread differently from an abstract concept than from a concrete concept and what consequences does this have for generalization? First, it may be appropriate to think of the semantic network as a directed graph, which means that the spread of activation would not have to be equally bidirectional between concept nodes. In this case, activation could spread from an abstract concept node to a concrete concept node, but not vice versa. For example, in the Florida Association Norms (Nelson et al., 1998), the word *abduct* primes the word *child*, but (thankfully) not vice versa. Thus, the link from *abduct* to *child* is stronger than the link from *child* to *abduct*. This hypothesis is supported by the NICE model, which posits that the more concrete a concept, the less spreading of activation to related concepts, and the more abstract a concept, the more spreading activation to related concepts.

This would suggest a very limited role for spreading activation in the generalization from trained concrete words to untrained words, either abstract or concrete. On the surface, such a suggestion may seem contrary to the results of Kiran and colleagues (Kiran, 2007, 2008; Kiran & Johnson, 2008; Kiran & Thompson, 2003), who reported generalization from trained atypical concrete to untrained typical concrete words within a given taxonomic category. Kiran hypothesized that the generalization occurred via overlaps in semantic features. Perhaps a more parsimonious view would be that when concrete words are trained, the spread of activation is reinforced in links between concepts that share semantic features. In the Kiran studies (Kiran, 2007, 2008; Kiran & Johnson, 2008; Kiran & Thompson, 2003), concrete concepts were trained within taxonomic categories, which may enhance the overlap of semantic features among concrete concepts. In the current study, it was not possible to use taxonomic categories, because abstract words are not easily categorized taxonomically and it is unlikely that a natural category exists that contains both abstract and concrete concepts. Thus, context-categories, such as *courthouse*, were devised that would satisfy the CATE requirement that the trained, complex items and the untrained, less complex items be related. Because context-categories are more thematically based, the concrete concepts in *courthouse* may have fewer semantic features in common with other words (either abstract or concrete) in the category and may therefore not enjoy the same amount of spreading activation as those concrete concepts when they are trained within taxonomic categories.

For abstract concepts, the reverse may be true. The NICE model posits that the more abstract a concept, the more spreading of activation to more loosely related concepts. This would suggest a larger role for spreading activation in the generalization from trained abstract words to untrained abstract and concrete words. If abstract words are trained within a context-category, such as *courthouse*, the associative links from abstract concepts to other abstract and concrete words in the category may be enhanced (just as taxonomic categories may enhance links between concrete words based on semantic features). This is in line with the DRF hypothesis, which proposes that abstract words are organized within the semantic network associatively, while concrete words are organized taxonomically (Crutch & Warrington, 2005). In practice, this would mean that links from abstract concepts to other abstract and concrete concepts in the semantic network would be stronger for concepts linked by association than those linked by similarity, while the reverse would be true for concrete concepts. Together, the spreading activation theory, the NICE model, and the DRF help to explain the pattern of direct training and generalization effects observed with AbSANT (Kiran et al., 2009; Sandberg & Kiran, 2014).

Conclusion

The retrieval of abstract words is an essential component of natural conversation. In aphasia, performance for abstract words is often more impaired than for concrete words, increasing the negative impact on successful communication. Paradoxically, few resources exist for training abstract word retrieval. This article replicated previous evidence for AbSANT. Importantly, when the present results are combined with previous results, there is a clear pattern of an advantage for training abstract words using this protocol. Specifically, when abstract words are trained, both direct training effects and generalization effects to untrained, related concrete words are observed. However, when concrete words are trained, no generalization occurs to the untrained, related abstract words, and the direct training effects are lower in magnitude. Together, these results suggest that AbSANT is an effective and efficient technique to improve both abstract and concrete word retrieval in aphasia.

Although we are far from a complete understanding of the way abstract and concrete words are organized and processed within the semantic system, this article represents a valuable step forward in understanding not only how abstract and concrete words are organized and processed within the semantic system but also the impact this organization and processing has for training. Continued study of intact semantic systems, impaired semantic systems, and semantic training is important for making progress both in better understanding the semantic system and in identifying ways to improve the rehabilitation of word retrieval deficits in clinical populations.

Acknowledgments

We would like to thank the persons with aphasia who participated in this study and their family members for their support. We would also like to thank the students in our labs who helped with data collection and reliability.

References

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018). The behavioural patterns and neural correlates of concrete and abstract verb processing in aphasia: A novel verb semantic battery. *NeuroImage: Clinical, 17,* 811–825. https://doi.org/10.1016/j.nicl.2017.12.009

Armstrong, E. (2005). Expressing opinions and feelings in aphasia: Linguistic options. *Aphasiology, 19*(3–5), 285–295. https://doi.org/10.1080/02687030444000750

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. https://doi.org/10.1016/S0093-934X(02)00522-9

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*(4), 161–169. https://doi.org/10.1007/s11065-006-9013-7

Beeson, P. M., & Robey, R. R. (2008). *Meta-analysis of aphasia treatment outcomes: Examining the evidence* [Paper presentation]. Clinical Aphasiology Conference, Jackson Hole, WY, United States.

Bird, H., Howard, D., & Franklin, S. (2003). Verbs and nouns: The importance of being imageable. *Journal of Neurolinguistics, 16*(2–3), 113–149. https://doi.org/10.1016/S0911-6044(02)00016-7

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation, 17*(6), 411–422. https://doi.org/10.1310/tsr1706-411

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*(4), 94–98. https://doi.org/10.1044/1058-0360.0404.94

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11*(6), 617–660. https://doi.org/10.1080/02643299408251987

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence–Fourth Edition*. Pearson.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407
Crutch, S. J., Connell, S., & Warrington, E. K. (2009). The different representational frameworks underpinning abstract and concrete knowledge: Evidence from odd-one-out judgements. *The Quarterly Journal of Experimental Psychology, 62*(7), 1377–1390. https://doi.org/10.1080/17470210802483834
Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain, 128*(Pt. 3), 615–627. https://doi.org/10.1093/brain/awh349
Efstratiadou, E. A., Papathanasiou, I., Holland, R., Archonti, A., & Hilari, K. (2018). A systematic review of semantic feature analysis therapy studies for aphasia. *Journal of Speech, Language, and Hearing Research, 61*(5), 1261–1278. https://doi.org/10.1044/2018_JSLHR-L-16-0330
Frances, K. J., & Kucera, H. (1983). *Frequency analysis of English usage: Lexicon and grammar*. Houghton Mifflin.
Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12,* 395–427. https://doi.org/10.3758/BF03201693
Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test*. The Psychological Corporation.
Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x
Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology, 29*(5), 526–562. https://doi.org/10.1080/02687038.2014.985884
Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees Test*. Thames Valley Test Company.
Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language, 24*(1), 1–19. https://link.springer.com/article/10.3758%2FBF03201693
Kaplan, E., Goodglass, H., Weintraub, S., Segal, O., & Loon-Vervoorn, A. V. (2001). *Boston Naming Test*. Pro-Ed.
Kay, J., Lesser, R. P., & Coltheart, M. (1992). *The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)*. Erlbaum.
Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)*. Harcourt Assessment. https://doi.org/10.1037/t15168-000
Kiran, S. (2007). Complexity in the treatment of naming deficits. *American Journal of Speech-Language Pathology, 16*(1), 18–29. https://doi.org/10.1044/1058-0360(2007/004)
Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research, 51*(6), 1550–1568. https://doi.org/10.1044/1092-4388(2008/07-0038)
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*(4), 389–400. https://doi.org/10.1044/1058-0360(2008/06-0085)
Kiran, S., Sandberg, C., & Abbott, K. (2009). Treatment for lexical retrieval using abstract and concrete words in persons with aphasia: Effect of complexity. *Aphasiology, 23*(7), 835–853. https://doi.org/10.1080/02687030802588866
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*(3), 608–622. https://doi.org/10.1044/1092-4388(2003/048)
Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies*. Edinburgh University Press.
Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: Both language and cognitive status are key predictors. *Neuropsychological Rehabilitation, 20*(2), 289–305. https://doi.org/10.1080/09602010903237875
Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*(5), 641–682. https://doi.org/10.1080/026432997381402
Martin, N., & Saffran, E. M. (1999). Effects of word processing and short-term memory deficits on verbal learning: Evidence from aphasia. *International Journal of Psychology, 34*(5–6), 339–346. https://doi.org/10.1080/002075999399666
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. https://doi.org/10.1080/02687038.2016.1208803
Mccarthy, L. M., Kalinyak-Fliszar, M., Kohen, F., & Martin, N. (2017). Effects of semantic context on access to words of low imageability in deep-phonological dysphasia: A treatment case study. *Aphasiology, 31*(5), 542–562. https://doi.org/10.1080/02687038.2016.1208803
Mchugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://w3.usf.edu/FreeAssociation/
Newton, P. K., & Barry, C. (1997). Concreteness effects in word production but not word comprehension in deep dyslexia. *Cognitive Neuropsychology, 14,* 481–509. https://doi.org/10.1080/026432997381457
Nicholas, L. E., Brookshire, R. H., Maclennan, D. L., Schumacher, J. G., & Porrazzo, S. A. (1989). *The Boston Naming Test: Revised administration and scoring procedures and normative information for non-brain-damaged adults* [Paper presentation]. Clinical Aphasiology Conference, Cape Cod, MA, United States. http://aphasiology.pitt.edu/id/eprint/69
Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia, 33*(10), 1281–1303. https://doi.org/10.1016/0028-3932(95)00102-9
Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287. https://doi.org/10.1037/h0084295
Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25. https://doi.org/10.1037/h0025327
Plaut, D. C., & Shallice, T. (1991). *Effects of word abstractness in a connectionist model of deep dyslexia* [Paper presentation]. 13th Annual Conference of the Cognitive Science Society, Chicago, IL, United States.
Qualtrics. (2020). Qualtrics survey software.
Renvall, K., & Nickels, L. (2019). Using treatment to improve the production of emotive adjectives in aphasia: A single-case study. *Aphasiology, 33*(11), 1348–1371. https://doi.org/10.1080/02687038.2019.1643001

Renvall, K., Nickels, L., & Davidson, B. (2013). Functionally relevant items in the treatment of aphasia (Part I): Challenges for current practice. *Aphasiology, 27*(6), 636–650. https://doi.org/10.1080/02687038.2013.786804

Sandberg, C. W., & Kiran, S. (2013). Analysis of abstract and concrete word processing in persons with aphasia and age-matched neurologically healthy adults using FMRI. *Neurocase, 20*(4), 361–388. https://doi.org/10.1080/13554794.2013.770881

Sandberg, C. W., & Kiran, S. (2014). How justice can affect jury: Training abstract words promotes generalisation to concrete words in patients with aphasia. *Neuropsychological Rehabilitation, 24*(5), 738–769. https://doi.org/10.1080/09602011.2014.899504

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Simmons-Mackie, N. (2018). *Aphasia in North America*. AphasiaAccess. https://www.aphasiaaccess.org/white-papers/

Stoke, S. M. (1929). Memory for onomatopes. *The Pedagogical Seminary and Journal of Genetic Psychology, 36*(4), 594–596. https://doi.org/10.1080/08856559.1929.10532218

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. https://doi.org/10.1044/1092-4388(2003/047)

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology, 27*(4), 635–657. https://doi.org/10.1080/14640747508400525

Ylvisaker, M., Szekeres, S. F., Henry, K., Sullivan, D. M., & Wheeler, P. (1987). Topics in cognitive rehabilitation therapy. In M. Ylvisaker & E. M. R. Gobble (Eds.), *Community re-entry for head injured adults* (pp. 137–220). Little, Brown.

Zoom. (2019). Zoom Video Communications.

Appendix A (p. 1 of 2)

Treatment Stimuli

Target words

Hospital	***Courthouse***	***Church***
Abstract	Abstract	Abstract
admission	guilt	angel
care	justice	baptism
condition	law	belief
diagnosis	oath	blessing
emergency	pardon	forgiveness
health	perjury	grace
mortality	plead	holy
recovery	proof	penance
sterile	sue	prayer
treatment	truth	solace
Concrete	Concrete	Concrete
ambulance	bench	bell
bandage	trial	Bible
blood	flag	candle
chart	gavel	chapel
doctor	judge	hymn
medication	jury	minister
nurse	vlawyer	organ
patient	prison	parish
stethoscope	record	steeple
syringe	robe	wedding

(*table continues*)

Appendix A (p. 2 of 2)

Treatment Stimuli

Features

Distractors	***Generic abstract***	***Generic concrete***
builds a nest	exists only in the mind	exists only in the mind
has feathers	exists outside the mind	exists outside the mind
lives in trees	can be seen	can be seen
found in a garden	can be heard	can be heard
is furry	can be touched	can be touched
has a tail	can be tasted	can be tasted
used to cut paper	can be perceived	is an object
can fly	is an object	is a person
comes out of a cocoon	is an idea	is an idea
spins a web	is alive	is symbolic
is put in a salad	is a feeling or emotion	is located in a courthouse/hospital
lives in the water	has a physical presence	can be worn
grows in soil	has a different meaning for different people	is generally considered positive
has six feet	is generally considered positive	is generally considered negative
used to plow the fields	is generally considered negative	
Examples of participant-derived features for abstract hospital		
can make you sad	can be good or bad	
requires a quick response	is protective	
provides information	makes you feel good	
requires payment	needed for different things	
associated with waiting	requires working	
can be provided by spouse	can happen quickly	
makes you glad to have support	is a part of life	

Appendix B (p. 1 of 4)

Testing and Treatment Protocols

Generative Naming Probe Protocol

Five baseline probes should be given prior to the start of therapy. After baseline, a treatment probe should be given every other treatment session at the beginning of the session, so that there are two treatments between each probe. After treatment, conduct five posttreatment probes. One month after the last posttreatment probe, conduct two maintenance probes within 2 days.

Say, *List as many words as you can think of that are associated with the category X* (hospital, courthouse). *List both concrete words, such as things or people and abstract words, such as ideas and feelings. For example, for the category* school *you could say* teacher *or you could say* knowledge (if the category is *school*, use a different example).

Write down every word the participant provides. If the participant says a word that matches or comes close to one of the stimuli, write it down in the “response” column. If the word is unlike any of the given stimuli, write it down in the “other responses” column. Mark a “1” for each correct response and a “0” for each incorrect response on the response sheet. See the scoring protocol to assess the accuracy of a response.

If the participant is having difficulty, prompt with something like, *Imagine you are in a hospital/courthouse*. *What do you see? What are the ideas or feelings that go along with being there?* If the participant starts to provide a story-like response, prompt for single words. Throughout, provide only general encouragement (e.g., *You’re doing fine*), but **do not give specific feedback regarding accuracy of responses**.

Stop after 2 min.

Probe Scoring Protocol

A response is counted as correct (1) when:

- The response is clear and intelligible and is the target or a very close synonym (e.g., physician for doctor; attorney for lawyer; medication for medicine).
- The subject self-corrects and produces the intended target.
- The target is accurate but intelligibility is reduced due to minor distortions and/or prosodic/stress differences.
- The subject initially produces close phonological approximations of the target and then achieves the target.

- The response is a variation of the target, as long as the meaning is not changed (e.g., jurors for jury, weddings for wedding, guilty for guilt).
- The response is a dialectal difference (e.g., tomata for tomato).
- At a request for clarification by the experimenter, the subject is able to produce the target accurately.
- The response contains a substitution, deletion, or addition of one phoneme (e.g., menicine for medicine).
- For other (i.e., nontarget) responses, the same above rules apply, and additionally, the word must appear on the "acceptable other response" list.

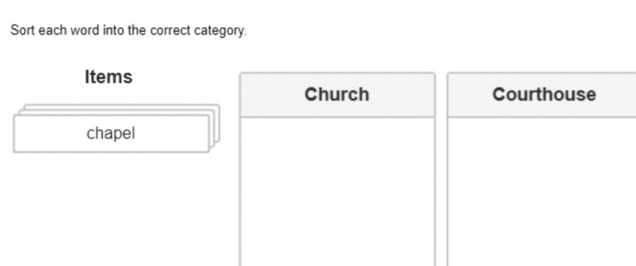
A response is counted as incorrect (0) when:

- The response is a neologism, meaning that less than 50% of the word resembles the target (e.g., rodifer for doctor).
- The response is a semantic paraphasia (e.g., harp for bell). This applies to the target items. If the word fits in the category, it may be counted as an "other acceptable response."
- The response has the same root word as the target, but a different meaning (e.g., forgiver for forgiveness).
- The response is a circumlocution (e.g., "12 people" for jury).
- The response is a phonemic paraphasia consisting of the substitution, deletion, or addition of two or more phonemes (e.g., stidal for steeple, minser for minister).
- The category is given as the response (e.g., children's hospital, mental hospital cannot be counted as other acceptable responses for the category hospital).
- An unrelated word out of the category (e.g., dentist for the category church) or an unrelated description ("I have one" for lawyer) is provided.
- No response or "I don't know."
- For other (i.e., nontarget) responses, the same above rules apply, and additionally, if the word does not appear on the "acceptable other response" list.

Treatment Protocol

One set of items (*N* = 10; abstract or concrete) will be treated at a time. Complete Steps 1 and 5 only once per session. Complete Steps 2–4 for each word. **In the first treatment session (i.e., the brainstorm session), Steps 2 and 3 will be replaced by a discussion aimed at generating features for the remainder of therapy.** The amount of support provided to participants should be adjusted based on the participants' ability and progression through treatment.

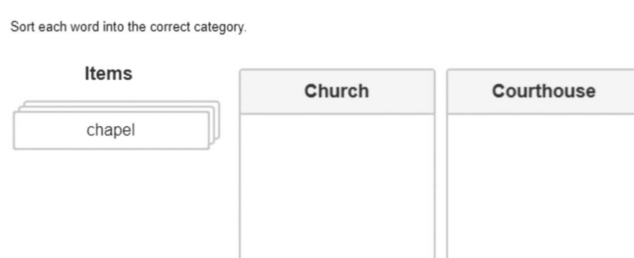
1. Category Sorting. This step will be performed only once at the beginning of each session. Click on the Qualtrics link for our participant's category sorting task. The program will present 40 words in random order in a "stack" on the left side of the screen. Twenty words (10 abstract, 10 concrete) are from the target category, and 20 are from the distractor category, *church*. The "bins" for each category will be on the right of the screen. Ask the participant to drag each word to the correct category.



If the participant has trouble reading the words, read the words for them during the first couple of sessions, then gradually allow the participant to read them on their own. If the participant is upset that they've categorized a word incorrectly, especially if a case can be made that it fits into both categories, say something like *"Yes, a robe can be in both categories, but it fits better in the category courthouse because a judge always wears a robe, but in a church, the minister doesn't always wear a robe, it depends what religion the church is for."*

a. **During the first treatment session**, at the end of this step, tell the participant something like *"Ok, now we're going to work on the characteristics/features of some of the words from the category X. One of the characteristics/features of words is whether they are abstract or concrete. Abstract words are thoughts, ideas, or feelings that can't be touched, seen, tasted, etc., like the word* 'knowledge.' *Concrete words are things or objects that can be seen, touched, etc., like the word* 'chair.' *Do you have any questions?"* Proceed to ask the participant generate features for each of the trained items. You will incorporate some of these participant-derived features into the feature list that will be used for the remainder of therapy. Say something like *"Let's start with the word EMERGENCY."* Place the written word in front of the participant and say something like *"How would you describe an EMERGENCY?"* Write down everything the participant says. Allow approximately 5 min for each word (remember that you have 10 words to get through in less than 2 hr). If the participant has difficulty, prompt with something like *"What is something that happens in an EMERGENCY?"* or *"What are some ideas/feelings associated with an EMERGENCY?"*

2. Feature Selection. This step will be performed for each word in each session, except the first session. Before starting with the first word in the session, tell the participant something like *"Ok, now we're going to work on the characteristics/features of some of the words from the category X. One of the characteristics/features of words is whether they are abstract or concrete. Abstract words are thoughts, ideas, or feelings that can't be touched, seen, tasted, etc., like the word* 'knowledge.' *Concrete words are things or objects that can be seen, touched, etc., like the word* 'chair.' *Do you have any questions?"* Click on the Qualtrics link for your participant's feature selection task. Select the word that you are working on. The features will be "stacked" in a random order on the left of the screen and the "bins" for the target word and the rejection pile will be on the right. Ask the participant to drag the first six features that apply to the target word into its "bin." If they appear to be struggling, go through each of the features with the participant. For example, *"Is emergency an object? No? Then we'll reject it." "Is emergency an idea? Yes? Then let's keep it."*



If the participant finds it difficult to comprehend a feature, provide an explanation. If the participant picks a feature that is not obviously well-fitting of the word, ask them to tell you why that feature applies to that word. If they can't, try to get them to think about why it would or would not fit by providing an argument for and against that feature and asking them which argument they agree with. Once six features have been selected, have the participant read aloud the features that have been selected (e.g., *"Emergency is an idea. Emergency is generally considered negative."*). You may read the features aloud, especially during the first few sessions, but try to progress toward the participant reading by themselves.

3. Yes/No Questions: Remove the Qualtrics survey from sight and tell the participant *"Now I'm going to ask you some questions about the word X. Please answer yes or no for each of these questions."* Ask the participant 15 questions for each target example: five that are acceptable semantic features, five that are unacceptable semantic features, and five that are distractors. For the example word *emergency*, *"(a) Is it an idea? (b) Does it have a practical use? (c) Does it have shelves?"* **Take care to avoid saying the target word during this step.** If the participant answers incorrectly, make a note of their response and then ask them to think about it and make sure they are certain of their response. The question may be reworded or a short explanation provided to assist in comprehension. For example, if you ask *"Can it be perceived?"* and the participant does not seem to understand or says *"no,"* then ask *"Can you perceive an emergency?"* Record the final response. Try to cycle through features so each is asked approximately the same amount of times per word.

4. Type, Synonym, Recall: After the participant has answered all of the yes/no questions for that word, ask if it is an abstract word or a concrete word. Reiterate the definitions of abstract and concrete, if necessary. Provide feedback and the correct response. Then ask for a synonym or another word that means the same thing *"Can you think of a synonym/ another word that means the same thing?"* If they say the target word prior to providing a synonym, confirm that they are correct. If they say a different target, correct them and then ask again for a synonym. If they can't think of one, provide a multiple choice or fill in the blanks, for example, *"What means the same thing as the word we're working on, crisis or decision?* or *The Cuban missle ____."* Finally, ask them for the target word "W*hat word are we working on?"*

5. Free Generative Naming: This step will be performed only once at the end of each session with only the category that the participant is working on. This step has no time limit, but one may be enforced if needed. Instruct the participant, "*List as many words as you can think of that are associated with the category X. You may list either concrete or abstract words. Think of all of the words we worked on today as well as others that belong to this category.*" Inform the participant about the accuracy of their responses and provide specific feedback on how to improve their responses. For example, *"Yes,* emergency *definitely goes with hospital."* Or: *"No, this word doesn't really belong in a hospital. It's more related to X." If the participant is struggling, you can prompt them to cluster, "You said emergency, now what is associated with an emergency?" and you can give them hints if requested. Try to make the hints something they can recall from the session: "Remember there was a word we were working on and you said that a synonym would be crisis."*

Appendix C

Example of Treatment Fidelity Scoresheet

Treatment Fidelity | Participant ID#: | Scorer:

Refer to the treatment protocol for the details of each step. Remember that support for each of these steps may be introduced or withdrawn depending on the patient's severity level; the important thing is to make sure that each step is completed. Be sure to fill in the date of each session. Mark complete steps with a plus sign and incomplete steps with a minus sign. For steps 2–4, write each word that was trained and tally the times each step was completed for each word. Please make notes of any variation to the treatment protocol (e.g., what was missing or if something was out of order). See example in the first row. Add rows as needed.

Date	Step 1: Category Sorting (only once per session)	Step 2: Feature Selection (choose 6 for each word)	Step 3: Y/N Questions (15 for each word)	Step 4: Recall, Synonym, Type (for each word)	Step 5: Free Generative Naming (only once per session)	Total steps completed
01/01/01	+	Guilt – Justice + Law + Oath + Total = 3/4	Guilt + Justice + Law + Oath + Total = 4/4	Guilt + Justice + Law + Oath – Total = 3/4	+	12/14

Notes: Only 3 features were chosen for the word guilt during step 2; only completed the recall and type for the word oath during step 4