The Pennsylvania State University

The Graduate School

College of Medicine

COMPARISON OF SEMANTIC SPACE MODELS FOR NEUROIMAGING WITH ABSTRACT AND CONCRETE WORDS

A Thesis in

Neuroscience

by

Dominick DiMercurio II

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

August 2017

The thesis of Dominick DiMercurio II was reviewed and approved* by the following:

Chaleece W. Sandberg

Assistant Professor of Communication Sciences and Disorders

Thesis Advisor

Diane Williams

Professor of Communication Sciences and Disorders

Ping Li

Professor of Psychology, Linguistics, and Information Sciences

Kevin Alloway

Professor of Neural and Behavioral Sciences

Chair of the Neuroscience Graduate Program

*Signatures are on file in the Graduate School.

Abstract

Psycholinguists have long noted a distinction between abstract and concrete words, especially in measures of performance thereof in lexical tasks. In addition, neuroscientists discovered differences in the localization of abstract and concrete word processing and in behavioral responses regarding abstract and concrete words from people with brain injury or impairment. Various psycho- and neurolinguistic theories have been put forward to explain these differences, including the Different Representational Frameworks (DRF) hypothesis, which states that abstract words are represented more paradigmatically and concrete words are represented more syntagmatically. Meanwhile, computational models are becoming increasingly attractive in neuroscience; in particular, previous researchers (Mitchell et al., 2008) have investigated the ability of a type of semantic space model to predict neuroimaging data from corpus data. The present study expands the previous research by applying the methods to both abstract and concrete words within different semantic space models in a 2×2 design of word type and model type to test the DRF hypothesis. Additionally, a third model type inspired by latent semantic indexing was developed and included in the analysis. The present study found that 3 out of 3 model types predicted stimuli significantly better than chance with abstract words, and that 2 out of 3 model types predicted stimuli significantly worse than chance with concrete words; however, these prediction accuracies deviated only slightly from chance. Furthermore, analysis of variance (ANOVA) revealed no significant effect of word type, model type, or interaction between the two. The present thesis discusses possible limitations to the methodology, considerations of deviations from previous research, and suggested trajectories for future study. Improvements to the models are highly desired as semantic space modeling represents a promising avenue of research for improving neuroscientific understanding, language-processing technologies, and speech therapies.

Table of Contents

List of Figures

List of Tables

Acknowledgments

My thesis would not have been possible without the many incredible people who guided me up to and through my graduate studies. Here I highlight a subset of the paragons who were pivotal to my research, my education, or my life more generally.

First and foremost, I am delighted to acknowledge Dr. Chaleece Sandberg as my advisor. Chaleece is an intelligent, enthusiastic, and hard-working researcher who also manages to promote the seemingly rare – at least, for academia – notion of work-life balance. She welcomed me into her lab when I felt lost in terms of my career direction and has since encouraged me to be a persistent professional. Beyond shared research interests, we also share an interest in science fiction. Though we may disagree on our favorite *Star Trek* captains, our nerdy conversations have always been a welcome diversion from the monotony of everyday research.

I would also like to acknowledge the rest of my master's committee: Dr. Diane Williams and Dr. Ping Li. Diane has provided me with the resources of the department as well as with insights from her personal experiences in the field of semantics. Meanwhile, Ping has been a constant mentor to me at Penn State. Words cannot express my appreciation for all that Ping has provided me: the opportunity to rotate in his lab, the connections he has helped me make, and his expertise on subjects pertaining to computation, language, and the brain.

In addition to Ping, several other fantastic mentors allowed me the opportunity to rotate or shadow in their lab. Many thanks to Drs. Dezhe Jin, Patrick Drew, Kristina Neely, and Carina Curto for welcoming me into their labs. From them, I learned a lot during my first year as a neuroscience student about neuroimaging and electrophysiology as well as computational and mathematical methods. With this knowledge and these experiences, I was able to successfully complete my thesis.

My financial support included research and teaching assistantships from the Huck Institute of Life Sciences and from the Departments of Biology and of

Communication Sciences and Disorders. Under the guidance of Drs. Denise Woodworth, Carla Hass, and Chaleece Sandberg, I learned how to be an effective educator. I am grateful for these experiences because passing on knowledge to the next generation of scientists continues to be my favorite part of science.

Just as I hope to pass my knowledge onto others, I appreciate those who have passed theirs onto me. My thesis would not exist without the education that supports it. I would like to thank my instructors, especially Drs. Kevin Alloway, Patrick Drew, Yingwei Mao, Chaleece Sandberg, and Xiao Liu, who provided me with knowledge relevant to the field. Also pivotal to my education were my classmates. I would especially like to thank fellow students in my cohort of the neuroscience program, Jasmine Caulfield and Insung Song. The three of us braved homework assignments and candidacy exams together and managed to find time to celebrate afterward with bowling and ice cream, where we often had much-needed "venting sessions" to release stress from our research frustrations.

In addition to classmates, I also acknowledge labmates. In each lab where I rotated, I found a new home of colleagues who warmly welcomed me and showed me the ropes. One such colleague was Dr. Sumithra Sunrendralal, then a graduate student in the Jin group. Sumithra was the first person to help me adjust to grad school, and as an overwhelmed first-year student, I was very appreciative of her encouragement and wisdom. Another example of a helpful labmate is Ben Schloss, a graduate student in Ping's lab. I appreciate all the advice and assistance that Ben, even while abroad in Beijing, has been able to offer me with my research. I sincerely appreciate having worked alongside many talented graduate students.

Finally, my education before Penn State was just as pivotal to this thesis as is my graduate education. In my undergraduate institution, I had wonderful mentors like Dr. Michelle Starz-Gaiano, who inspired me with her love for biology, and Dr. Jodi Kelber-Kaye, who taught me how to value interdisciplinarity; both of whom, coincidentally, championed feminism and instilled in me a great respect for women in academia. And before then, my education included amazing high school teachers like Mrs. Connie Burley and Mr. Peter O'Connor who introduced me to biology and computer science, respectively, and inspired me to seek out science as a career. And well before then, it all started with my very first educators, my parents Dom and Donna DiMercurio, who I can never thank enough. My parents assisted me not only through encouragement for education but also in nearly every aspect of my life, as did the rest of my family. I am extremely thankful for all the love and support from my family and friends over the years. From that category are three final, but no less important, people who undoubtedly deserve an explicit mention: my sister Melissa DiMercurio, my brother Nicholas DiMercurio, and my boyfriend Brandon Alexander. We have withstood many life challenges together, and I acknowledge them for providing me with inspiration, motivation, and joy.

Dedication

I dedicate my thesis to my mother and father,

Who helped me learn how to spell and count,

So that one day I could articulate and calculate.

Chapter 1

Introduction

A child can discern between concrete words (such as ‘house’) and abstract words (such as ‘property’);[1] however, providing a definition to these classifications remains more challenging. Concrete words, like object nouns and action verbs, take physical referents, which can be detected directly by one or more sensory modalities, whereas abstract words refer to concepts like emotions (Kousta et al., 2011; Newcombe et al., 2012; Vigliocco et al., 2013, 2014; Vinson et al., 2015) or numbers (Pobric et al., 2007, 2009; Lambon Ralph et al., 2009; Cappelletti et al., 2001; Jeffries et al., 2005) that are not as directly linked to the senses. The border between these two classifications, or word types, is often fuzzy, and there are some words like ‘alert’ that seem to be more of an intermediate[2] between abstract and concrete words. Nevertheless, when comparing distinctly different words on the concreteness scale,

[1]Elementary school teachers may ask their pupils to classify nouns as abstract or concrete. An example of a teacher providing this type of assignment can be inferred in the message where a fourth-grader asks “Grammar Girl” to explain concreteness (Fogarty, 2014).

[2]The word ‘alert’ has a subjective concreteness rating of 400 on a scale ranging from 100 to 700 (Coltheart, 1981). Note that 400 is precisely in the middle of both extrema.

important phenomena arise, as demonstrated by behavioral and brain research. For instance, the presence of double dissociations between abstract and concrete nouns (Coltheart, 2000; Jeffries et al., 2007; Martin & Saffran, 1992; Plaut & Shallice, 1993; Jeffries et al., 2009; Snowden et al., 1989) suggests that abstract and concrete word representations are distinct processes in the human brain. Perhaps the most crucial observation regarding this classification is the concreteness effect, where performance in a variety of lexical tasks is faster and more accurate for concrete words than for abstract words. Examples include better performances in word recall, word recognition, or reading and spelling accuracy (Paivio & Rowe, 1970; Begg, 1972; Walker & Hulme, 1999; Binney et al., 2016).

The concreteness effect has been studied extensively in both behavior and the brain, and many theories have been put forth to explain the phenomenon, but none of them have been definitively demonstrated to be true. The notion that abstract and concrete words are represented through distinct processes in the brain reflects neuroimaging data that suggests that anterior regions of the brain are more involved during abstract word processing, while more posterior regions of the brain are more involved during concrete word processing (Binder et al., 2005, 2009; Wang et al., 2010). Meanwhile, the emerging approaches of computational modeling within neuroscience offer new avenues to study these processes in more depth than they have been studied before. An example of such a model is a predictive type of semantic space model that makes use of both brain imaging and corpus data (Mitchell et al., 2008). It is prudent when studying an explanatory model's behavior to *a.)* modify the model to be consistent with empirical findings and *b.)* use the results to make inferences about neurolinguistic theory or predictions for future experiments based on the model. Models that are highly reliable at predicting, not just fitting, experiments are better able to explain the observations, which can inform scientific, technological, and clinical knowledge. The present thesis will investigate neuroimaging-predictive semantic space models, motivated by behavioral and brain research, and apply the model to a novel set of experiments to see if the model captures a hypothesized scheme of abstract and concrete word representation in the brain.

1.1 Psycholinguistic Theory

The concreteness effect is the tendency for a human's performance for a variety of lexical tasks to be significantly better (i.e., with a significantly lower reaction time or with a significantly higher accuracy rate) when the stimuli are concrete words as opposed to when they are abstract words. For instance, in a lexical decision task where participants are asked to discriminate between true words in their language and nonwords comprised of a non-meaningful string of letters (e.g. 'mfqkj'), concrete words tend to have an advantage in speed, a finding that has been reproduced multiple times (James, 1975; Whaley, 1978; Kroll & Merves, 1986), although the difference may sometimes disappear with low-frequency words

(James, 1975), and the effect may be sensitive to the order of presentation of words, as abstract words displayed after concrete words exhibited concreteness effect in reaction time while concrete words displayed after abstract words show no significant difference (Kroll & Merves, 1986).

In addition to concreteness, which is defined "in terms of reference to [sensory] experience" (Paivio et al., 1968), a related construct is imageability, the ability to visualize the referent of a word. A concrete word like 'house' is relatively easy to visualize whereas an abstract word like 'property' does not as readily evoke a specific image. While concreteness and imageability are not identical notions, there is such a strong correlation between these two metrics that most people refer to concreteness and imageability interchangeably (Reilly & Kean, 2007). As such, the concreteness effect can also be observed as an "imageability effect" of sorts. For instance, as imageability increases, so do recognition scores (Paivio & Rowe, 1970).

There are numerous examples of the concreteness effect in the literature, a few of which are highlighted in the present thesis. For example, both forward and backward recall are higher for concrete words than for abstract words (Begg, 1972; Walker & Hulme, 1999), a finding which is independent of speech or writing rate and roughly constant across serial positions (Walker & Hulme, 1999). The concreteness effect is also present in bilingualism and in persons with aphasia (PWA). When 15 neurotypical adults (NTA) and 1 PWA, all English-Spanish bilinguals, were studied in both naming-to-definition tasks and semantic priming tasks, a concreteness effect was present across tasks and languages (Kiran & Tuchtenhagen, 2005). The meanings of newly learned concrete words also tend, overall, to be remembered more accurately than the meaning of abstract words (Ter Doest & Semin, 2005; Romani et al., 2007). These examples are just a small sampling of experiments that have shown a connection between task performance and psychometric measures like concreteness and imageability. While reversals of the concreteness effect do exist (Kousta et al., 2011; Crutch et al., 2009), they are not the norm and tend to be the result of clever manipulations that attempt to coax out the difference between abstract and concrete word processing. Reversals of the concreteness effect may also occur in people with lesions to specific regions of the temporal lobe (Jeffries et al., 2009; Binney et al., 2016).

Concreteness scores often correlate with a number of other psychometrics and properties of words. Concreteness was found to positively correlate strongly with imageability ($r = +.9$) and moderately with familiarity (Reilly, 2007). Concreteness was found to negatively correlate with measures of word frequency (i.e., Kucera-Francis Frequency and Brown Frequency) in slight but significant ways as well as to have a moderate significant negative correlation with number of phonemes ($r = -.45$), syllables ($r = -.44$), and morphemes ($r = -.45$) (Reilly & Kean, 2007). Low imageability (thus, generally more abstract) words were found to be significantly longer than high imageability words in terms of number of syllables and phonemes, were consistently more morphologically complex across syllable lengths, and showed more variation in the placement of stress (Reilly & Kean, 2007). Language of origin, at least in English, was also found to be significantly different between the classifications with abstract words in English having more Latinate etymologies and concrete words in English having more Germanic etymologies (Reilly & Kean, 2007). A factor analysis found concreteness, imagery, meaningfulness, and age of acquisition to be part of the same latent factor in the concreteness effect results from Whaley et al. (1978); however, it is worth pointing out that the concreteness effect can persist even after controlling for imageability (Barber et al., 2013). Age of acquisition is an important variable as words acquired earlier in life are processed faster, and distinct age effects have been discovered in action verbs and concrete nouns (Boulenger et al., 2007). Various parameters, such as age of acquisition, word length, word frequency, and even imageability, which often correlate with concreteness, have been put forth in attempts to explain the concreteness effect; however, the effect often remains after controlling for these parameters (Barber et al., 2013; Crutch et al., 2009).

Taking all of these psycholinguistic effects into account, the Dual Coding Theory (DCT) is the predominant explanation for the concreteness effect. DCT is the notion that semantic knowledge is stored both in verbal domains (e.g., the word itself) and nonverbal domains (e.g., the visual perception of the word). According to the theory, with concrete words, both verbal and nonverbal codes are utilized, whereas with abstract words, only the verbal code is employed (Paivio, 1991). The argument from DCT is that the dual coding of concrete words makes concrete words more resistant to degradation and easier to recall than abstract words (Paivio, 1991). The representation of the word in the verbal domain is called a logogen; meanwhile, the representation in the nonverbal domain is called the imagen. For instance, one could imagine the word "phone" by itself as a logogen, whereas the image of a phone or the ringing sound a phone can make would be the associated imagens (Paivio, 1991). An abstract word, however, may not have any strongly associated imagens to it, which would account for the greater difficulty in accessing, recognizing, or generating abstract words.

Not all researchers of the concreteness effect are convinced by DCT. In fact, the Context Availability Theory (CAT) is a leading alternative to DCT to explain the effect (Kemmerer, 2015; Binney et al., 2016). CAT suggests that concrete words are context-independent and abstract words are context-dependent, which allows concrete words to be understood more easily, with processing for abstract words being more complicated (Binney et al., 2016; Schwanenflugel et al., 1998). The idea behind CAT is that all words are amodal, but that some words (typically abstract words) can be more variable in their meaning. Saffran and Sholl (1999), and Hoffman et al. (2010) provide thought-provoking examples of context-dependent and -independent words. For instance, concrete words like "rose" or "spinach" refer to specific, almost singular meanings; however, the meanings of words like "phase" or "chance" are more dependent on context. "Phase" could appear in *phase of the moon* as well as it could appear in *phase of infant development.* “Chance” is even more variable in that it could refer to probability (*What are the chances?*), time (*If you get a chance...*), or risk (*I took a chance.*). The cognitive effort needed to disambiguate these different possible meanings could account for the increased reaction times or decreased accuracies seen in abstract words, as proposed by the CAT.

While the DCT and the CAT are the leading theories to explain the concreteness effect, they are by no means exhaustive. Others have proposed that the Grounded Cognition Model, which states that concepts are grounded in modality-specific representations (Allport, 1985; Thompson-Schill et al., 2006), may play a role in the concretness effect. Others still suggest that the emotional valence of words may be a significant confound in concreteness effect studies, presenting criticisms for the DCT (Kousta et al., 2011); however, there has been ample debate on whether the confound of emotion is as relevant as Kousta et al. suggest (Paivio, 2013; Vigliocco et al., 2013). Reversals of the concreteness effect, the reasons for the reversals, and their implications on neuropsychology have been offered and hotly debated in the field, yet no definitive answer has arisen. One possible explanation is that the two word types are in different representations, perhaps in a manner as suggested by the Different Representational Frameworks (DRF) hypothesis, which will be discussed in detail later in the present thesis. More information on the concreteness effect from a neuroscientific lens will illuminate how these words are processed in the brain and may give insight to the neural underpinnings of semantics as a whole.

1.2 Neurolinguistic Theory

Theories and observations in psychology (more specifically, in studies of behavior and cognition) often then feed theories and observations in neuroscience (more specifically, in studies of the anatomy and mechanisms of neurons or neural regions). The concreteness effect as observed in behavioral studies and the theories that attempt to explain it may motivate cognitive scientists and neuroscientists to understand more about the anatomy and mechanisms behind concreteness and semantic processing. Questions like *Where are the meanings of these words stored in the brain?* and *How are the meanings of these words represented in the brain?* are somewhat philosophical but can also be empirically tested through indirect means. Any theories put forth at first are bound to be overly simplistic; however, from these simpler theories, one can build a more sophisticated view that may have the ability to both explain and predict the observations not only of the brain but also of the behavior that results.

One classic way of studying neural anatomy and mechanisms is through lesion studies. With a uniquely human behavior like language, one faces the conundrum that no experimental lesions on a human brain would be ethical and no experimental lesions on a nonhuman animal brain would be fully informative. Instead, in order to perform a lesion study for language, one must look for an “experiment” performed by nature; that is to say, one must find someone who has already inherited or acquired brain injury in the region of interest. These lesions, unfortunately, will not be specific, but will, nevertheless, be informative. For instance, many PWA, who typically have lesions in perisylvian cortex exhibit concreteness effects in their deficits (Coltheart, 2000; Jeffries et al., 2007; Martin & Saffran, 1992; Plaut & Shallice, 1993); however, several cases of people with semantic dementia (PWSD), which typically involves degradation of inferior temporal poles and amygdalae, may have reversals of the concreteness effect (Jeffries et al., 2009; Snowden et al., 1989), which suggests that abstract and concrete words are doubly dissociated.[3] It is worth noting, however, that not all PWSD have more deficits in concrete words than abstract words, and there is some debate as to whether the reversals discovered are misleading. In either case, it is reasonable to conclude, given the limitations of natural lesion studies, that studying brain injury or impairment is not sufficient to understand completely abstract and concrete word processing. For this reason, one may turn to other methods, such as neuroimaging, as a supplementary way of obtaining this knowledge.

Beginning in the 1990s, functional magnetic resonance imaging (fMRI) has become an increasingly attractive technique (Bandettini, 2012) for observing human participants as they engage in a variety of cognitive tasks, such as in the studies of the concreteness effect. In particular, blood-oxygen-level dependent (BOLD)

[3]A single dissociation is when some brain impairment leads to deficits in some behavior A but not some behavior B. A double dissociation occurs when the reverse is also found in another individual (i.e., where the individual experiences deficits in B but not in A). Double dissociations are important discoveries in neuroscience as they tend to suggest that those two behaviors are possibly represented in distinct manners or locations in the brain.

signals from fMRI are widely employed as an indirect measure of local neural activity as these signals can be obtained noninvasively with better temporal and spatial resolution than other common neuroimaging techniques like positron emission tomography (Kemmerer, 2015). While BOLD signal may not necessarily accompany an increase in neural activity (Tootell et al., 1998; Shmuel et al., 2002; Muller & Kleinschmidt, 2004), the present study will refer to significant increases in BOLD signals or their latent components as "activation" in line with other literature, but do note that this activity is vascular rather than neural. BOLD signal results from a detection of changes in the magnetic field that occur when oxygenated hemoglobin becomes deoxygenated, which is followed by a rapid replenishment of oxygenated hemoglobin in what is called a hemodynamic response function. The main foundation for fMRI is the principle of neurovascular coupling, the idea that regions that have more neural activity and processing will require more blood supply to support a higher metabolic demand (Arbib, 2015), an assumption which is a general trend that is sometimes violated. Even though fMRI results in an indirect measure, the fact that there are patterns of activity that are physiological, systematic, and quantifiable allows for the development of theories and models that are empirically based; nevertheless, one should remain cautious when interpreting fMRI data so as not to assume that a one-to-one relationship exists between the BOLD signal and neural activity, let alone cognitive representations.

People who are engaging with abstract and concrete words often show distinct profiles of activation in many brain imaging studies. Meta-analyses of these studies (Binder et al., 2005, 2009; Wang et al., 2010) have been able to uncover multiple patterns of activation in the brain during the semantic processing of abstract and concrete words. While certain regions, like the left lateral temporal lobe, appear to be engaged nearly equally with both abstract and concrete words (Binder et al., 2005), some regions show a potential preference for one of the classifications over the other after performing a subtraction analysis. For concrete words, these regions are bilateral angular gyrus, dorsal prefrontal cortex, posterior cingulate gyrus, precuneus, fusiform gyrus, and parahippocampal gyrus (Binder et al., 2005; Wang et al., 2010). For abstract words, these regions are inferior frontal gyrus and middle temporal gyrus (Binder et al., 2009; Wang et al., 2010). If both hemispheres are not equally involved, the brain regions involved in word processing are typically left-lateralized.

The presence of the concreteness effect, concrete-abstract double dissociations, and distinct profiles of fMRI activation during concrete and abstract word processing suggests that concrete and abstract word representations are distinct processes in the human brain. As mentioned previously, one attempt to understand these semantic processes from a psycholinguistic perspective is the DRF hypothesis, which states that abstract and concrete words are represented in systems with qualitatively different properties (Crutch & Warrington, 2005). According to the DRF hypothesis, abstract words are represented in a more associative manner (a concrete word pair like "'apron" and "kitchen" or an abstract word pair like "faith" and "prayer" would be examples) than in a manner of similarity (take, for instances, the pair "cheetah" and "tiger" or the pair "practice" and "exercise"); meanwhile, the hypothesis purports that concrete words are represented more strongly in the reverse manner (Crutch et al., 2009). To uncover these semantic representations in the brain and untangle the evidence for these different hypotheses, we need to explore characteristics like concreteness at the cognitive and behavioral levels through both theoretical and experimental lenses. To that end, empirically-based computational models have arisen as a promising avenue for neurolinguistics research.

1.3 Semantic Space Models

Computational modeling is important, especially in brain research, because the brain is an extremely complicated organ with 100 billion neurons connected through 100 trillion synapses through extremely complicated configurations. Thus, understanding the interactions of neural regions and the behaviors that they control (such as language) cannot come directly from dissection and observation as can be done with simpler organs and their behaviors. Computers give researchers the ability to perform a large number of calculations, such as when studying the massive amount of brain imaging data,[4] and can supplement observations and experiments

[4]The base unit of fMRI data is the voxel, the three-dimensional version of a pixel. An fMRI scan can have hundreds of thousands of voxels. The sheer size of this data often necessitates sophisticated computational techniques.

by providing testable and quantifiable predictions. Computational and statistical methods can give neuroscientists a clearer understanding of the brain (as they do to scientists in other fields for phenomena beyond the brain), and so a computational model may have benefits for the discipline, and the methods developed may help other disciplines as well.

One group of computational models of interest in the present study is a class of models called semantic space models, which includes a variety of related yet diverse models (Lowe, 2001; de Gemmis et al., 2015). In order to understand semantic space models, one must first examine some concepts about language, semantic spaces, and mathematical modeling. First, consider a particular language such as the English language. One can conceptualize a language, reductionistically, as having a collection of words W, such that each word w in a language is an element of that collection. Next, consider a semantic space V to be a special type of vector space where each vector $\vec{v}$ is built from numeric values that represent some property of w as it relates to some "aspect of language" so to speak. That aspect of language could be the contexts that contain a specific word or term, in which case that semantic space is more specifically a word space, or it could be the documents where the word or term occurs, in which case that semantic space is more specifically a document space (de Gemmis et al., 2015). Each vector has a size or dimensionality D, often with $D = 300$ considered the "magic number" as it approximates the peak where some models perform at optimal performance (Landauer & Dumais, 1997).

Semantic space models also involve the notions of transformation function A, set of basis elements B, similarity measure S, and map M as formalized by Lowe (2001). Every ith basis element b_i in B is some constituent of the language, such as words, lemmas, articles, or documents. The function A is such that the ith component of vector $\vec{v}$ for any word w in W can be composed as $v_i = A(b_i, w)$ (Lowe, 2001). For instance if the set B is of documents, one could compose document vectors in a document space using a function A that counts the occurrence of word w for each element b_i in B. With activation signals from fMRI, one could also imagine creating activation vectors from "activation bases" with each basis roughly approximating a hypothetical lemma but in an activation space as opposed to the hypothetical lemma space; perhaps both are part of a larger category of vector spaces, the "brain space" so to speak. In addition to bases and their transformations into a vector space, semantic space models often also take advantage of similarity measures; measures like Euclidean distance and cosine similarity have both been employed previously (Lowe, 2001). Finally, a map M is applied to transform one semantic space into another. In the literature, a semantic space model has been formalized as a 4-tuple of A, B, S, and M (Lowe, 2001); however, such formalism is a tangential consideration in the context of the present thesis, which focuses more specifically on the comparison of two semantic spaces (such as an activation space built from fMRI data and a document space built from corpus data) using an M derived from statistical regression techniques.

Multiple types of semantic models exist, such as latent semantic analysis (LSA) (Landauer & Dumais, 1997), random indexing (RI) (Sahlgren, 2005, 2006), and hyperdimensional analogue to language (HAL) models (Lund & Burgess, 1996; de Gemmis et al., 2015). These models are also sometimes called distributional models, because they rely on the distributional hypothesis, which states the meaning of a word depends on its distribution as it appears in the language (de Gemmis et al., 2015). LSA, or sometimes called latent semantic indexing (LSI), uses singular value decomposition to reduce corpus-derived semantic vectors into latent components, which then can be used to judge similarity between texts (Landauer & Dumais, 1997). Meanwhile, RI uses an incremental algorithm called Random Projections to reduce dimensions as in LSA but with less computational cost (Sahlgren, 2005, 2006). Meanwhile still, HAL uses windows of text called n-grams to create co-occurrence matrices for words in order to map words onto a semantic space and compare how similar the words are or in which categories they cluster (Lund & Burgess, 1996). A major application of semantic space models is as a bottom-up approach to content-based recommender systems (further elaborated in Chapter 4), as opposed to a top-down approaches (e.g. BabelNet or DBpedia), which require external knowledge to determine the representation of texts (de Gemmis et al., 2015).

1.4 Predictive Models

The specific type of semantic space model considered in the present study was introduced in Mitchell et al. (2008) which uses a regression of corpus and neuroimaging data to create learned scaled parameters for semantic vectors in order to predict the activation profile of the brain for each given word. Mitchell and colleagues refer to their model as a semantic space model or a "predictive model" but for the sake of precision, the present thesis will refer to this type of model as a neuroimaging-predictive document space model. In addition, this thesis will also investigate a neuroimaging-predictive dependency space model, which is similar to the model of Mitchell and colleagues but involves dependency vectors as introduced by Fyshe et al. (2013). Mitchell and colleagues have shown that their document space model predicts activation space significantly better than chance, a finding that has since been reproduced across various dimension sizes and in document, dependency, and concatenated model types (Schloss & Li, 2016).

The corpus data can be turned into semantic vectors either through calculating the frequency of words co-occurring in the same document (called "document vectors") or through calculating the frequency of words appearing in a syntactic relationship parsed by dependencies (called "dependency vectors"). Document vectors can be characterized as paradigmatic, topic-based, or associatively represented;

meanwhile, dependency vectors can be characterized as syntagmatic, type-based, or similarly represented (Fyshe et al., 2013).

Fyshe et al. (2013) find that clustering models using the dependency vectors perform more accurately than when using document vectors in concrete words, but that this disparity in performance is lost when the concrete words are mixed with abstract words. Schloss & Li (2016) reveal a seemingly reverse finding that document vectors perform more accurately for concrete words than dependency vectors, and they also find that dependency vectors exhibit destructive interference at high dimensionalities. It is worth noting that the Fyshe study only investigates how well the model works on a behavioral judgment task without any input from experimental brain data, whereas the Schloss study uses neuroimaging-predictive semantic space models. Also, while the Fyshe study does look at abstract words after mixing them with concrete words, neither of these studies look at abstract words on their own, which is unfortunately missing from the literature. Both the seemingly contradictory findings and the absence of abstract words studied on their own motivates a further investigation into the performance of neuroimaging-predictive semantic space models with relation to abstract and concrete words.

Computational modeling is an important method for better understanding topics of neuroscience. While these models do not necessarily demonstrate what exactly happens in the brain, they do demonstrate what is possible given the evidence that already exists. If multiple models seem to converge on the same idea, then that suggests more strongly that a hypothesis about a phenomenon is true. Models allow researchers to further test those hypotheses because models will make specific, quantifiable predictions about what should happen according to the hypothesis and previous evidence. If a model's predictions are correct, then more credence can be lent to the hypothesis underlying the model. In the best paradigm, experimentation and modeling exist in a cyclical nature where experiments provide evidence, which in turn motivates models, which in turn provide predictions, which in turn motivate experiments. This view of a long-term symbiosis between models and experiments shows that neither type of study is superior to the other, and in fact, the two are interdependent on the success of the other type of study. The goal is to build a model that both is motivated by and will motivate experiments.

1.5 Experimental Overview

Psycho- and neurolinguistic theory has begun to unravel the differences between abstract and concrete word processing, but many questions still remain. While we see a concreteness effect in NTA as well as double dissociations between abstract and concrete words in PWA versus PWSD, theories like DCT and CAT that underlie these phenomena still continue to be tested and debated. The world of computational modeling may contribute to that debate, especially through the use of predictive models, which can predict brain activation based on words significantly better than chance (Mitchell et al., 2008). It may be possible to utilize semantic space models to capture interesting brain phenomenon, as it uses both brain imaging data as well as semantic vectors that derive from corpus data as a behavioral variable. Nevertheless, previous studies into concrete and abstract words with semantic space models revealed conflicting results, which opens up further questions about the nature of these representations. The competing ideas behind the concreteness effect, as well as the conflicting results from semantic space models, motivate further study.

The present study uses fMRI data previously collected from the Adult Neuroplasticity Laboratory led by Dr. Chaleece Sandberg at Penn State University. These fMRI studies on NTA involving both abstract and concrete words, which are further elaborated in the methods, have never been studied with these computational models before. In addition, the present study obtains semantic vectors from a database provided by Dr. Alona Fyshe, a computer scientist at the University of Victoria. The words included in this database (referred in the present thesis as the "Fyshe database") have a substantial overlap with the words tested in the fMRI data by the Adult Neuroplasticity Laboratory. The resources made available by both Dr. Sandberg and Dr. Fyshe have made this study possible, which allows for a combination of neuroimaging and corpus data that has not been previously combined or analyzed in this manner.

The DRF hypothesis, where abstract words are more associatively represented and concrete words are more similarly represented, is supported by behavioral studies such as the one performed by Crutch et al. (2009). Meanwhile, Fyshe et al. (2013) seem to reiterate support for the DRF hypothesis *in silico* as their computer models seem better able to perform clustering tasks with abstract words using a paradigmatic proxy and with concrete words using a syntagmatic proxy. The combination of neural representations and vectors from the Fyshe database appear to show an inconsistent pattern, where concrete words appear to perform better in the dependency space model for a behavioral prediction task but better in the document space model for neuroimaging prediction task (Fyshe et al., 2013; Schloss & Li, 2016). This inconsistent pattern, in part, motivates the present study.

In accordance with the DRF hypothesis and the findings by Fyshe et al. (2013), the hypothesis underlying the present study is that the document vectors, as a proxy for association, will perform more accurately for abstract words, thought to be represented more on the basis of association, than they do for concrete words, thought to be represented more on the basis of similarity. The present study also hypothesizes that the reverse pattern will be found with dependency vectors as it is possible that the dependency vectors serve more closely as a proxy for similarity due to their more syntagmatic design. It is currently unknown exactly how specific words or types of words are represented in the brain. However, when these words are accessed during an fMRI scan, the mechanisms by which words are processed are reflected as a particular pattern of fMRI activation. The present study assumes that the models will perform significantly better than chance in reproduction of previous findings and also assumes that the closer the model representation is to these mechanisms, the better the model will be able to pick out the "reflection" that the brain representation leaves in the fMRI data, and thus the more accurately the model will predict specific words.

The current study aims to test the hypothesis of an interaction between model- and word-types under the aforementioned assumptions of reproducing better-than-chance findings and of using paradigmatic and syntagmatic proxies to distinguish between different word representations as reflected in fMRI. The results of the present study may have implications in understanding semantic processing and their related computational models. Furthermore, research in this area will not only help to settle longstanding debates in neurolinguistics but may also lead to information for and improvements in technologies or therapies related to language.

Chapter 2

Methods

The present study used fMRI data from neurotypical adult (NTA) human participants who were engaged in a concreteness judgment task in the scanner. The 49 concrete and 49 abstract words used in the task were selected *a priori* for the study, and their semantic vectors were retrieved from the Fyshe database. The blood-oxygen-level dependent (BOLD) response of the participants when interacting with each word was processed for input into semantic space models. Four semantic space models were developed to cover each combination of word type (i.e., abstract and concrete) and model type (i.e., document vector and dependency vector). An additional model type was also developed as another consideration. The models were used to predict activation vectors for words in the fMRI data that were encountered by the model, and the accuracy of the four models were calculated using a leave-out-two paradigm. Finally, a statistical comparison of the accuracies with analysis of variance (ANOVA) was performed, and the significance of word type, model type, and the interaction thereof was used to make interpretations about model reliability and underlying neural representation.

2.1 Participants

All participants ($n = 11$, 6 male) were NTA with no history of neurodegenerative disease or psychiatric illness. All participants were right-handed and had at least a high school education, with ages varying from 47 to 74 years old ($\mu = 62.1$). A further breakdown of sex and age demographics is provided in Table 2.1. All participants scored within normal limits on the Cognitive Linguistic Quick Test and gave informed consent in accordance with procedures approved by the Penn State University Institutional Review Board.

Participant	**01**	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**
Sex	F	M	M	F	F	M	M	M	F	F	M
Age	64	66	47	55	72	74	72	57	59	61	56

Table 2.1: Breakdown of demographics. This table presents the sex and age of each participant, deidentified. F = Female; M = Male. Ages reported in years.

2.2 Concreteness Judgment Task

Each individual in the study participated in a concreteness judgment task in two functional magnetic resonance imaging (fMRI) scans approximately 10 weeks apart.

The task in each scan was identical; as a result, there are two instances of each word tested per participant. During the task, participants were presented with words (e.g. ‘admission’) or nonwords (e.g. ‘dmsdf’) and had to make a decision of whether the words were abstract or concrete or if the nonwords were strings of vowels or consonants by indicating with their finger via button presses. Stimuli were presented via E-Prime software (Schneider et al., 2002) in a random manner and were counterbalanced across both participants and sessions. The stimuli included 60 abstract and 60 concrete words in two categories (i.e., hospital and courthouse) as well as 60 nonwords, each appearing twice, which served as a control. Accuracy and reaction time for each stimulus was recorded, and signals for words where a participant responded incorrectly, regardless of which session had the inaccurate response, were excluded from the analysis.

During the same sessions that participants participated in the concreteness judgment task, they also engaged in a resting-state scan for 6 minutes (van Dijk et al., 2013), during which they focused on a white dot on black background while attempting to remain still. The data from the resting-state scan was not analyzed in the present study. The fMRI data originates from a study performed by the Adult Neuroplasticity Lab (Sandberg, in preparation).

The visual stimuli were presented on a screen behind the scanner, which was reflected by a mirror fitted to the head coil. When necessary, MRI-safe corrective lenses were used to correct the participant’s vision. All scans occurred at the Penn

State University Social, Life, and Engineering Sciences Imaging Center in a 3 Tesla Siemens Magnetom Primsa Fit MRI scanner.

High-resolution T1-weighted images have the following acquisition parameters: 160 sagittal slices, 1 mm^3 voxels, reconstruction matrix = 256, flip angle = 9°, phase encoding = AP, TR = 1650 ms, TE = 2.03 ms. BOLD-sensitive functional images have the following acquisition parameters: 42 axial slices (3 mm thick), 3 mm^3 voxels, reconstruction matrix = 80, flip angle = 90°, phase encoding = AP, TR = 2500 ms, TE = 25 ms (Sandberg, in preparation).

2.3 Selection of Word List

The 120 word stimuli used in the concreteness judgment were crosschecked with a list of words from a database of semantic vectors, provided publicly on a Carnegie Mellon University webpage (Fyshe et al., 2013). From the 120 words, several concrete words ($n = 11$, 6 hospital) and one abstract word ('nauseous', hospital) were excluded because there were no semantic vectors for these words provided in the database. To balance abstract and concrete word lists for better reliability of comparison, we also discarded additional abstract words ($n = 10$, 8 courthouse). The method used to discard words was systematic and determined *a priori*: Words with the lowest mean accuracy rates were removed in order to maximize the amount of data included for the model as all inaccurate events were discarded. The final word list contained 49 concrete and 49 abstract words to study. Mentions of "98 stimuli" hereafter in this thesis will refer to the final word list of 49 abstract words and 49 concrete words after removing words for which there were no semantic vectors or low-accuracy abstract words. More details on the words included or excluded can be found in Appendix A.

In order to ensure that the two word lists varied primarily in concreteness and not other known confounds, information about the words from each word list was collected, calculated, and compared. Such information included word frequencies like the Brown Frequency (BFRQ), the Kucera-Francis Frequency (KFFRQ), and the Thorndike-Lorge Frequency (TLFRQ), which were collected from the MRC Psycholinguistic Database (Coltheart, 1981). Word frequencies are the number of occurrences of words in a corpus. In specific, BFRQ comes from the Brown Corpus, a set of texts from native American English speakers edited and printed in 1961, excluding second editions or reprints (Brown, 2007); the KFFRQ comes from the norms given in Kucera et al. (1967), an automated count from texts in American English; and the TLFRQ comes from Lorge magazine count from Thorndike and Lorge (1944), a manual count from magazines written in American English.

Beyond word frequency, additional psychometrics such as age of acquisition (AOA), concreteness (CNC), familiarity (FAM), imageability (IMG), number of letters (NLET), and number of syllables (NSYL) were also collected from the MRC Psycholinguistic Database (Coltheart, 1981). Two-tailed t-tests were performed to compare abstract and concrete words in terms of participants' accuracy rates

	BFRQ	KFFRQ	TLFRQ
Concrete	6.00	28.22	232.77
Abstract	10.52	58.34	266.48
p-value	.236	.019	.684

Table 2.2: Comparison of word frequencies. The present table compares the available data for various word frequency counts across word classification. Kucera-Francis Frequency (KFFRQ) differs between abstract and concrete words at an uncorrected significance level of .05; meanwhile, Brown Frequency (BFRQ) and Thorndike-Lorge Frequency (TLFRQ) do not.

(ACC), participants' reaction time (RT), and available psychometrics for each word of the 98 stimuli, as reported in Table 2.3. ACC, RT, CNC, and IMG were all found to be significantly different across word classification; meanwhile, BFRQ, FAM, NLET, NSYL, and TLFRQ were not. While the p-values for AOA and KFFRQ fall below the uncorrected significance level, these values are nonsignificant after applying a Bonferroni correction with $n = 11$. Thus, the differences in AOA or KFFRQ between abstract and concrete words can be considered non-significant.

	ACC	RT	AOA	CNC	FAM	IMG	NLET	NSYL
Concrete	0.95	1175.83	336.43	570.52	519.96	575.93	6.94	2.06
Abstract	0.89	1485.43	333.65	425.60	540.56	393.76	7.00	2.31
p-value	$<.001$	$<.001$	.015	$<.001$	.091	$<.001$	.874	.217

Table 2.3: Comparison of psychometrics. The present table compares the available data for various psychometrics across word classification. Accuracy (ACC), Reaction Time (RT), Age of Acquisition (AOA), Concreteness (CNC), and Imageability (IMG) all differ between abstract and concrete words at an uncorrected significance level of .05; meanwhile, Familiarity (FAM), Number of Letters (NLET), and Number of Syllables (NSYL) do not.

The analysis of the selected abstract and concrete word lists was performed to ensure that the two word lists truly differed in terms of concreteness and the closely related concept of imageability but not in terms of other major confounds such as word frequency, length, or familiarity. In addition, the statistical analysis confirms that a concreteness effect in participant accuracy and reaction time exists between these words, even after removing the 10 lowest accuracy abstract words.

2.4 Processing of Functional MRI Data

The fMRI data were subjected to a preprocessing pipeline that included slice timing correction, functional volume realignment, coregistration of the structural and functional volumes, segmentation of the structural image, and normalization to the MNI template. The data were not smoothed across voxels because it was important to keep raw data for the semantic space model, which works at the voxel-level, in order to boost any potential signal at the voxel-level that the model could use to perform reliably and not lose that information to smoothing. A high pass filter with a cutoff at 128 seconds was used to filter out low-frequency fluctuations, such as physiological factors like heart rate, technical artefacts like scanner drift, and other possible sources of noise.

A general linear model was performed for each stimulus, using 10 regressors for both scans and a constant for either scan resulting in a total of 22 beta values. The 10 regressors in a scan were 4 word-related measures (stimulus itself, abstract words sans stimulus, concrete words sans stimulus, and control nonwords) and 6 motion-related measures (x, y, z, roll, pitch, and yaw). The abstract or concrete word condition excluded the stimulus itself; these conditions also excluded stimuli where the participant responded inaccurately in either session. Only for words where the participant responded correctly both sessions were general linear models considered, and this entire process was repeated for each participant. More information on the inaccurate words that were excluded from the model can be found in Appendix A.

To determine significant beta values for the model for each word, a masking procedure was used with the beta values for the stimulus itself, averaged over both sessions. The mask used a contrast of considered words subtracted by control nonwords, or more specifically control nonword condition multiplied by 3 subtracted from the sum of stimulus condition, the abstract condition, and the concrete condition. Only the voxels where the contrast was significant (at a .001 significance level, uncorrected) were included in the mask. The resulting data per participant per word was converted into an activation vector that could be used as input for the semantic space model. The reason for the mask was to ignore nonsignificant voxels, which were most likely to be solely a source of noise, and to reduce computational time required to perform the regressions.

All fMRI data were modified using a combination of a MATLAB script provided online (Humphries, 2005), and MATLAB scripts developed by the author, versions of which are provided in Appendices B and C.

2.5 Development of Semantic Space Models

The semantic space models used the input of the activation vectors resulting from the fMRI data as described earlier and the first 300 dimensions of either document vectors or dependency vectors as retrieved from the database (Fyshe et al., 2013). A leave-two-out paradigm, like in previous research (Mitchell et al., 2008; Schloss & Li, 2016), was employed such that two words from the accurate subset of 49 words per classification were removed from each trial, resulting in a maximum of 1176 possible configurations of input for each model for each participant. The leave-out-two paradigm allows the model to be trained on multiple words while being blind to the two left-out words, between which the model will attempt to discriminate. A multiple regression was performed on the activation and semantic vectors for those $n-2$ words in each of the $n(n-1)/2$ configurations, where n is the number of words after excluding inaccurate trials ($n \leq 49$). Each multiple regression resulted in learned scaled parameters for each semantic vector dimension as it contributes to activation in a particular voxel. This was repeated across all 11 participants in 4 different modeling schemes.

The four modeling schemes are the result of a 2×2 design of concreteness×model-type: abstract×document (abs-doc), abstract×dependency (abs-dep), concrete×document (con-doc), concrete×dependency (con-dep). This 2×2 design allows for an analysis of word type effect, model type effect, and interaction between the two types. For each modeling scheme and for each participant in every word pair configurations, predictions were made for the two left-out words by taking the dot product of the learned scaled parameters from the multiple regression and the semantic vectors of the left-out words. The predicted activation vectors were then either successfully or unsuccessfully matched with the actual activation data through the cosine similarity measure. If the correct matching had a higher cosine similarity than the incorrect matching, then the trial was counted as a success. The number of successes were then tallied and divided by $n(n-1)/2$ total trials to obtain an accuracy rate. In this way, 4 accuracy rates per participant can be reported.

Performing the model involved use of custom MATLAB scripts, some of which were adapted from scripts kindly provided by Ben Schloss.

2.6 Development of Cosine Similarity Vector

An additional type of semantic vector was created as a derivative of other semantic vectors. To do this, 300 semantic vectors from the Fyshe database were selected at random. Then the cosine of the angle between the document vector ($D = 300$) of each word under study and the document vector ($D = 300$) of the random words was taken to create a matrix. Each element of a word's cosine similarity semantic vector was the cosine of the angle of that word and one of the selected 300 words. The rationale behind developing this additional model type was to see if these vectors were a better proxy for similarity than the dependency vectors are.

Dependency vectors, although with some syntagmatic considerations in their design, are ultimately based on text co-occurrences, thus making them very association-like. The cosine similarity model was inspired by latent semantic analysis, which uses almost identical methods to judge similarity of words, and thus the thinking is that the cosine similarity vectors will be a better proxy for similarity. Nevertheless, this model type was not part of the original hypothesis and is only included as an additional consideration.

2.7 Comparison of Semantic Space Models

The number of successes/failures for each classification×model-type was compared with a chi-squared test to see if they differed significantly from the expected numbers of successes/failures (4629/4628 for concrete[1] and 5708/5708 for abstract) which were based on 50% expectation for a random choice between two options. A significance level of .05 was used to determine if any of these models were significantly different than chance. In addition, a two-way ANOVA was performed on the accuracy rates across model type (i.e., document space model and dependency space model) and word classification (i.e., abstract and concrete). Statistical effects of type, classification, and their interaction are reported at a significance level of .05 with a correction for Tukey's Honestly Significant Difference.

[1]The odd number of trials was biased in favor of getting the extra success so that a decree of "significantly better than chance" would be more conservative.

Chapter 3

Results

3.1 Performance Breakdown

Each model proceeded within a leave-out-two paradigm per participant with a certain subset of words (either abstract or concrete) using a certain type of semantic vector (document, dependency, or cosine similarity). Each model was repeated in $K = \frac{n(n-1)}{2}$ leave-out-two trials where n is the number of words for a given participant and word classification; these trials represent all possible combinations of two words in each set of accurate words for a given participant and word classification. Information about the number of trials can be found in Table A.6. A complete breakdown of the accuracy of these models can be seen in Table 3.1. The average across participants for each word classification and model type is 51% (Abstract Document), 53% (Abstract Dependency), 52% (Abstract Cosine), 47% (Concrete Document), 50% (Concrete Dependency), and 46% (Concrete Cosine).

Participant	01	02	03	04	05	06	07	08	09	10	11
Abs-Doc	.519	.644	.553	.530	.448	.448	.456	.514	.490	.472	.572
Abs-Dep	.381	.570	.415	.600	.522	.544	.661	.551	.508	.461	.599
Abs-Cos	.568	.625	.547	.585	.433	.495	.539	.449	.423	.382	.667
Con-Doc	.404	.593	.472	.530	.416	.387	.437	.482	.456	.469	.527
Con-Dep	.380	.548	.466	.546	.566	.469	.445	.534	.553	.465	.486
Con-Cos	.492	.497	.487	.471	.391	.431	.435	.444	.466	.507	.407

Table 3.1: Accuracies of Models. This table includes a breakdown of the fraction of successes of each model of each word classification broken down by participant out of a total of K trials as described in the text.

3.2 Deviance From Chance

When comparing each word-model combination to the expected accuracy from chance (50%), one discovers that abstract words from all model types performed significantly better than chance (document, $p < .01$; dependency, $p = .03$; and similarity, $p < .01$) and that concrete words performed significantly worse than chance for two model types (document, $p < .01$; dependency, $p = .38$; and similarity, $p < .01$). A more sophisticated deviation-from-chance calculation could be performed in the future, but these initial investigations suggest that semantic space models might perform better with abstract words than with concrete words. Figure 3.1 demonstrates the performance of the models both at the participant level and overall.

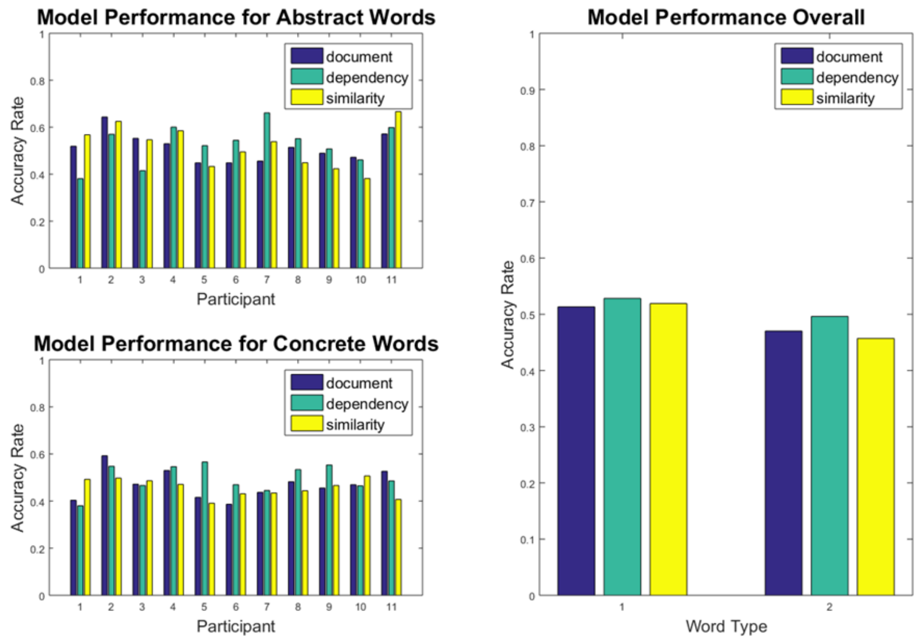

Figure 3.1: Performance of Models. The present figure demonstrates the results of the models. The image on the upper left panel shows performance of models with abstract words across participants; the image on the lower left panel shows performance of models with concrete words across participants; the image on the right panel shows the performance of all models averaged across participants. (1 = Abstract, 2 = Concrete.)

3.3 ANOVA Results

The ANOVA did not find any significant effect of word classification, model type, or the interaction thereof after correcting for Tukey's HSD. More information can be found in Table 3.2.

	SS	df	MS	F	p	HSDT
Word-Type	0.02	1	0.2	4.44	.0414	.04
Model-Type	<0.00	1	<0.00	<0.00	1	.04
Interaction	<0.00	1	<0.00	<0.00	1	.08
Error	0.18	40	<0.00			
Total	0.20	43				

Table 3.2: Results of ANOVA. Word-type refers to abstract versus concrete words, and model-type refers to document versus dependency space models. SS = Sum of Squares, df = degrees of freedom, MS = Mean Sum, F = F statistic, p = p-value, HSDT = Tukey's Honestly Significant Difference threshold.

3.4 ArtRepair

The general linear models included regressors for motion, but a concern arises that motion artefacts could still be a major contributor to noise, especially in light of the unexpected near-chance performances of the model. Thus, a repeat of the experiment with an attempt to reduce noise from motion artefacts, using the ArtRepair toolbox for spm12, was performed as a follow-up to the initial findings. The ArtRepair toolbox was used on participant 07 because participant 07 had the most movement and thus would be the most indicative of whether motion regressors in the general linear model were insufficient. Not only did the ArtRepair not improve the performance of the model, it decreased the average performance (from 50% to 42%); however, this difference was not found to be significant ($p = .173$). A visualization of the results is available in Figure 3.2, which shows a breakdown by word type and model type.

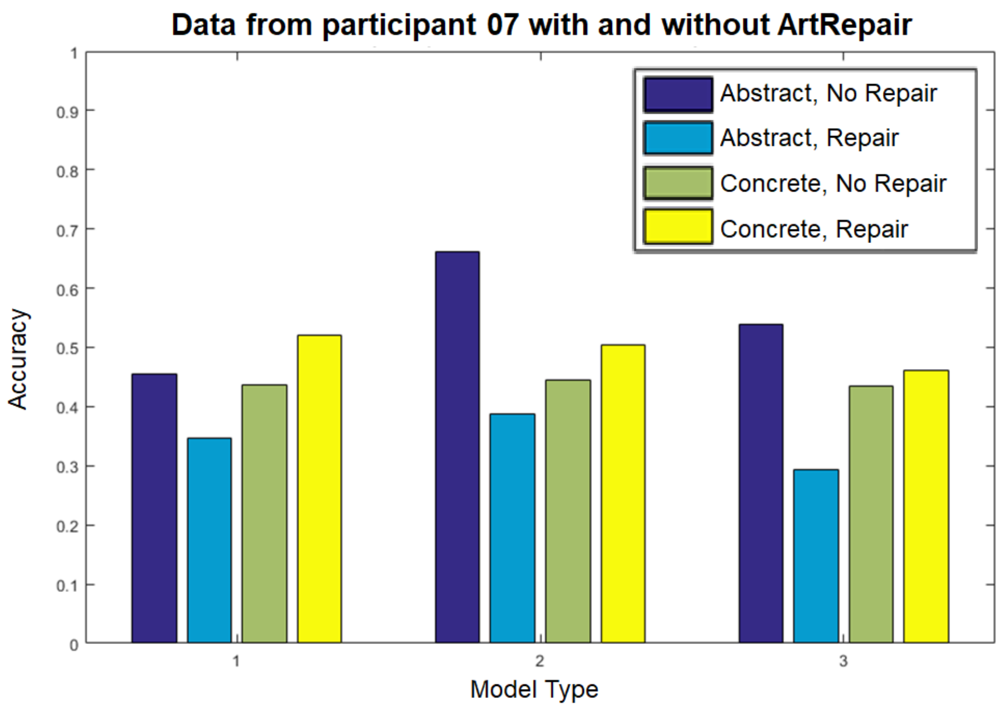

Figure 3.2: Effect of ArtRepair. The ArtRepair toolbox was used on participant 07's data to reduce motion artefacts; however, this technique did not significantly improve performance of the models. (1 = Document, 2 = Dependency, 3 = Cosine Similarity.)

Chapter 4

Discussion

The hypothesis of the present study is that there would be an interaction between word type and model type as motivated by the DRF hypothesis. The primary finding of the present study is that there was no significant effect of word type, model type, or any interaction thereof, which does not allow us to eliminate the possibility that these models might behave identically, regardless of word- or model-type. Furthermore, the most striking result was the inability to reproduce the general finding that these semantic space models perform at accuracy rates significantly better than chance. With the present data, the only significant improvements from chance were in models looking at abstract words; in fact, two models looking at concrete words performed worse than chance. However, these deviances, while significant, are small, and the effect of word-type was not significant after a multiple comparisons correction, so caution is urged when interpreting this finding. The remainder of the present chapter will discuss interpretations for the reliability of semantic space models and for the plausibility of the DRF hypothesis, considerations for the limits of the present study, predictions for the results of future studies, and applications of semantic modeling research to its greater scientific, technological, and clinical relevances.

4.1 Interpretations

Previous literature found that model accuracy rates were significantly better than chance when looking at fMRI data from multiple iterations of the same stimulus (i.e. 6 instances per word) as participants were involved in an intense semantic task (i.e. thinking deeply about a word's meaning) (Mitchell et al., 2008). This finding extended to a later adaptation of the model that used the document and dependency vectors provided in the Fyshe database (Schloss & Li, 2016). The present study used fewer iterations (i.e. 2 instances per word) and a much more superficial and less intense semantic task (i.e. concreteness judgment within 4 seconds). The present study, using these less-than-ideal circumstances, was unable to reproduce consistent better-than-chance findings for concrete words, which suggests that the reliability of these models cannot generalize to all lexical tasks in all quantities, and instead that these models might perform only in a restricted set of finely tuned experiments.

The finding that models performed better for abstract words than concrete words is cautiously intriguing. The separation of these classifications is slight, does not interact with the other model types tested, and might not be indicative of neural representations as much as inherent biases in the model, if not simply an artefact. Nevertheless, this finding provides hope that better methodology in data collection or data analysis could enhance the current model's concreteness-dependent results. At best, the present study is inconclusive in terms of the DRF hypothesis because, as long as the models do not perform consistently better than chance, no conclusions can be made that could not simply be accounted for by the random fluctuations of a noisy model.

Even if future models were to better account for noise and provided results in line with the present hypothesis, one would still have to be cautious in interpreting the results. A positive finding would support the DRF hypothesis, but it could also support other hypotheses as well. For instance, a positive result would not eliminate the possibility of a confound not considered in the present study, such as emotional valency or how polysemous a word is. One has to be cautious with the dichotomy of abstract and concrete words as it correlates with other measures and is arguably better perceived in terms of a multidimensional continuum instead of a binary distinction (Reilly & Kean, 2007; Binney et al., 2016).

Another question that the present study aimed to answer was whether these models could generalize to abstract words alone, which had never been attempted before. Due to the fact that most of the results were at or nearly close to chance levels, it is inconclusive whether or not these models, when in their ideal state, would perform equally as well for abstract words. Further study is needed to explore this possibility; however, the result that abstract words performed slightly better than chance in all model types is a promising result, which might indicate that the predictive power of semantic space models is not restricted to concrete words.

Finally, on the question of whether dependency models are a good enough proxy for similarity, the study's answer is likewise inconclusive because comparing across model types with close-to-chance results is as unreliable as comparing across word types. Further study is needed for this question as well.

4.2 Considerations

The present study used a different population of participants (i.e. older adults) than did previous studies. Due to the knowledge that there is large variation in functional and structural neuroanatomy even within a similar population, it is difficult for us to rule out the effect of individual variation or of age variation or of any other demographic that might not have been considered. Age is one such possible consideration because one cannot rule out the possibility that as one ages, semantic processing may change; however, in the present study, there is only a weak, nonsignificant negative correlation ($r = -.2704$, $p = .4213$) between participant age and the mean performance of their models. That said, the populations by the Mitchell and Sandberg labs may not be representative of each other or even the general population, perhaps due to age or some other demographic. It is important to consider how the population may have affected the results of these studies.

The present study also has a variety of other differences that might influence the findings. The study has a different task and different number of word repetitions. The task used in the present study (i.e., concreteness judgment) is not as semantically demanding as tasks in the previous literature, and the number of word repetitions is not as high as in previous studies. The present study also has a different set of stimuli, so it might be worth considering whether the words themselves are a contributing factor to the difference between the present results and the results in prior studies. Varying accuracies between the word lists or confounds like age of acquisition could also play a role in seeing differences between the word types. If one had to pick which of these possible considerations to revisit, the best would likely be semantic task, as the intensity of the semantic task seems to be the most profound difference between the present study and previous studies.

Another possible consideration is that the vectors chosen might not be perfect proxies for the types of representation (i.e. associativity and similarity) that the present study aimed to explore. For the purpose of the present study, the document vectors was assumed to be a good proxy for association as the co-occurrence of words in a text is nearly identical to the definition of associated words. The dependency vectors as a proxy for similarity is less ideal because although the dependency vectors involve a more syntagmatic nature of composition than document vectors do, they still ultimately depend on co-occurrences which is a hallmark of association. However, one would expect dependency vectors to be more similarly represented relative to the document vectors, which satisfies our purposes for making comparison, but caution should be taken when interpreting these findings so as not to make a direct connection between dependency vectors and similarity. In the future, a comparison between abstract and concrete words could be enhanced by choosing a more appropriate proxy for similarity.

The threshold for selection of voxels may have influenced the results; however, an attempt to increase the threshold did not result in drastically different findings. Still, we must consider the possibility that the voxel selection (especially given the fact that localization of function is important) could give out different results, explaining the discrepancy between the present and prior findings.

The method of comparing predicted vectors is slightly different than Mitchell and Schloss's method; however, this is unlikely to have caused a major difference between the results. An attempt to use the previous method on a subset of the present data seemed to show that there was not a great difference in the accuracy values that resulted between these two methods of comparison. Nevertheless, determining the most appropriate way to count "successes" and "failures" is a worthy consideration for these models. In the present study, a success was counted if the sum of cosines of angles between the predicted and actual vectors was higher with the correct pairs than with the incorrect pairs. In other research, the counting of successes might have been assessed differently, such as counting whether a predicted vector was closer to a correct actual vector or an incorrect actual vector. It is unlikely that these different methods would give drastically different results, but it is worth noting that they would not necessarily give identical results either.

Another possible consideration is that motion artefacts might have distorted the data and made the models perform less reliably. A follow-up analysis, as shown in the results, on an individual participant seems to suggest that motion is not a primary factor for the near-chance results; however, attempts to improve data collection and analysis that will provide cleaner data provide another possible way to improve the reliability and performance of these models.

A final consideration is that Semantic Space Models may not be able to capture the association-based and similarity-based types of representation in the brain even if they exist (or at least the model accuracy rate is not discriminative of it); it may be that some other factor is being captured by the model which is overpowering an effect of associativity and similarity representation. It is however too soon to be able to tell if the discrepancies in the results are due to participant demographics, task, stimuli, the modeling technique, or any other factor. As such, future work is warranted.

4.3 Future Work

Future studies could take many different directions and could provide some interesting results and perhaps illuminate more about the results found in both the present study and previous literature.

For instance, different masks could be attempted; perhaps looking specifically at voxels that are significant when the specific stimulus or the stimulus's classification is subtracted from the control might have a different outcome. If the model still performs at chance levels with different masks, then that suggests that the selection of voxels may not explain the difference in results. Future studies may apply a mask with a contrast of abstract words only to the control or with a contrast of concrete words only to the control.

Different models could be attempted as well. Future studies could look into results from other vectors (e.g., those based on Florida association norms) or a concatenation of vectors from different model types. It is possible that a poor selection of semantic vectors could be at play with the present work. If better performance exists with different semantic vectors and the same data set, that would suggest that the near-chance findings found here have more to do with the selection of semantic vectors than with the data set itself. If the finding is found with other model types, then that might suggest that something different about the data collection of the different studies is more explanatory for the different findings.

Another possible avenue would be to analyze different tasks, such as lexical decision, similarity judgment, or association judgment tasks. Similarity judgment tasks are where participants are given a word pair and asked if the the words in the pair are similar; association judgment tasks, if the word in the pair are associated. One would expect that lexical decision would be similar to chance if chance results were related to level of semantic processing as lexical decision is not as semantically intensive; whereas similarity and association judgment might have an interaction with the difference in model type. Such a study would provide an additional way to test the DRF hypothesis and semantic space model's ability to detect it. In addition, one could collect data from different participants, perhaps even in a new experiment designed with these types of models in mind might be an ideal way forward to investigate these models. A new experiment would also allow for a better selection of words, which could reduce confounds.

Future studies may even wish to look at different brain imaging or recording techniques, which would be akin to using a different "brain space" for the activation data. One such example would be electroencephalography (EEG), which records electrophysiological data from participants from electrodes place on their scalp. It is worth noting that EEG data does not have the same quality of spatial resolution that fMRI has, but it has better quality temporal resolution, which might allow it to uncover something that cannot be detected through fMRI. If one were to expect that associative and similarity processing took different amounts of time, then models using EEG recording vectors instead of fMRI activation vectors would be able to discriminate between the two different representational frameworks and give insight into how these frameworks relate to concreteness. The fact that participants' reaction times varied in tasks with abstract versus concrete words suggests that there may also be something temporal at play as well as spatial. It might be intriguing to see whether an adaptation of this model to EEG can extend the predictive power of semantic space models beyond fMRI data.

Future investigations should seek to answer whether there is any difference between abstract and concrete words and whether certain model types perform better than others. After this is established, studies could also see whether the model can perform a concreteness judgment task by determining whether a word is abstract or concrete. The model can then be studied as it relates to focus on specific brain regions. The model could then be lesioned, and the future study can investigate whether the outcomes simulate what is seen in aphasia by comparing the lesioned models with data from PWA.

4.4 Applications

The present study has scientific, technological, and clinical relevances.

The scientific relevance is that knowing more about semantic processing will help researchers to understand more about how the brain works. With the continued debates in the field over the concreteness effect and with a lack of knowledge more generally about how semantic processing in the brain operates, these types of studies will be very critical to a better understanding of these effects and processes, which will expand our neuroscientific knowledge overall. The present study assists in understanding the controversies in the field, and future studies have the promise to do even more so.

The technological relevance is that semantic space models may be used to develop technologies like content-based recommender systems and natural language processing systems. For instance, content-based recommender systems use user ratings of articles or books to determine which other articles or books are similar to the ones rated highly to recommend new articles or books to that user (de Gemmis et al., 2015). Other technologies that benefit from better understanding of semantics are translation services and database categorizations. If better machine learning technologies can be created in ways that fit with neuroscientific evidence, then these technologies may be better developed to fit human needs.

Finally there is clinical relevance to studying abstract and concrete words, especially in considering aphasia. More knowledge about the neuroscientific bases of semantics can help inform therapies for people with communication impairments like aphasia. For instance, since previous studies suggest that more complicated linguistic tasks require more complicated processing than simpler linguistic tasks, research into therapies for PWA suggest that training these more complicated tasks often helps PWA respond to simpler tasks, but not necessarily vice versa. In a similar fashion, training PWA with abstract words often gives them benefits to concrete words that is not seen vice versa. (Kiran et al., 2009; Sandberg & Kiran, 2014). Although most therapies for PWA focus on concrete words, previous research suggests therapies with abstract words could have benefits that concrete words do not have. Training PWA with abstract words were shown to generalize to untrained concrete words, but vice versa (training concrete words to generalize onto abstract words) was not found (Kiran et al., 2009; Sandberg & Kiran, 2014).

The speculative explanation for one-way generalization is that abstract word processing is a complicated task that involves accessing the concrete words, perhaps in an associative manner. Meanwhile, concrete word processing is not believed to involve accessing abstract words, perhaps because the information for concrete words is coded in sensory experiences, or with a self-contained context, or through paradigmatic representation, which allows the brain to bypass the need to access abstract words during concrete word processing. Understanding what is happening in this situation is important in research into therapies for PWA, as language recovery requires using the neuroplasticity of the underlying brain networks to recreate these semantic representations. If we can better understand how these representations work in neurotypical brains, then perhaps new therapies could be created that try to capitalize on that knowledge when developing new therapies. Though the authors suggest this idea as a possible reason, it is still unknown entirely why generalization of one word type to another only went in one direction. Further analysis, such as in the present study, will help us to understand therapy-related phenomena better. In addition, models that can predict and simulate aphasia allow for experiments that would be unethical or unfeasible to perform in humans. These testable predictions can then be confirmed or rejected by empirical results, and then old therapies can be scrutinized and new therapies may develop.

4.5 Summary

Psycholinguists have long observed differences between abstract and concrete words, especially the concreteness effect where people's performances in various tasks tend to be faster and more accurate with concrete words than with abstract words. Several theories, such as DCT and CAT, have been offered to explain the concreteness effect, but these explanations still remain controversial in the field. Furthermore, the neural bases of abstract and concrete words have also been explored, including the double dissociations seen where PWA often have more deficits with abstract words while some PWSD show more deficits with concrete words. Meta-analyses of neuroimaging studies on NTA suggest that abstract and concrete word processing may be localized in different brain regions. The combination of the concreteness effect, double dissociations, and neuroimaging evidence paved the way for the DRF hypothesis, which suggests that abstract and concrete word processing not only occurs in distinct locations but also that it occurs through distinct representational frameworks: namely, associativity and similarity, respectively.

Over the past few decades, there has been growing interest in using computational models to explore phenomena in multitude of disciplines, including linguistics and neuroscience. Due to their fast processing time and spacious memories, computational models will be helpful in understanding the most complicated biological organ known to exist, especially if models and experiments exist in a symbiotic relationship, where one can motivate and explain the findings of the other. A specific type of computational model that is applicable to the topic of interest is semantic space models, which can be used to predict imaging data from corpus data. The performance of these models in different schemes may provide insight about brain function, and the present study operated under the hypothesis that the models would capture an interaction between model types and concreteness in a manner similar to the DRF hypothesis.

The present study looked at the performance (i.e. predictive accuracy) of neuroimaging-predictive semantic space models in three different modeling types (i.e. document vectors, dependency vectors, and cosine similarity vectors) and two different word classifications (i.e. abstract words and concrete words). The present study found that models with abstract words performed significantly better than chance while two models with concrete words performed significantly worse than chance. There were no effects of word classification, model type, or any interaction, which does not seem to align with the DRF hypothesis nor previous findings about the reliabilities of these models. The results of the present study are primarily inconclusive, but they do provoke considerations for how to improve the methodology and implications on the importance of crafting experiments that can be compatible with semantic space models.

The present study has scientific, technological, and clinical relevance. Firstly, uncovering the bases of semantic processing is an eventual goal understanding language in neuroscience. The combination of neuroimaging evidence and computational models bring us a few short steps closer to that eventual goal. Secondly, semantic space models have had a known application for recommender systems, and it is likely that semantic space models may be used for other technologies related to natural language processing. Finally, knowledge about semantic processing would be useful to speech language pathologists in developing and modifying therapies for people with communication disorders like aphasia. Future models may be able to simulate lesions and therapies to allow for preliminary *in silico* investigations into aphasia. Overall, the present thesis provides scientifically, technologically, and clinically relevant results from an empirically-derived model that illuminates how to design future experiments and improve computational models in a way that may be able to better explain the semantic processing of concreteness and various other related neurolinguistic phenomena.

Appendix A

Lexical Inventory

The present appendix provides lists of words that were included in the model (Tables A.1, A.2, A.3, and A.4) as well as words that were included in the study but were excluded from the model (Table A.5). These words exist across two categories (i.e., words from courthouse and hospital settings) and across word classifications. The study looked at 60 abstract and 60 concrete words; in the end, the model included 49 abstract and 49 concrete words.

acquittal	judgment	proof
adjourn	judicial	rights
equality	justice	sustained
guilt	legal	tried
illegal	order	truth
injustice	perjury	law
innocence	plead	sue
interrogate		

Table A.1: Abstract courthouse words included. The present table shows a list of all the abstract words included in the model in the courthouse category.

anxiety	hygiene	science
boredom	illness	sick
care	medical	sterile
clinical	mortality	technology
compassion	pathology	therapy
condition	quiet	trauma
emergency	recovery	treatment
healing	relapse	weak
health	research	welfare

Table A.2: Abstract hospital words included. The present table shows a list of all the abstract words included in the model in the hospital category.

attorney	flag	policeman
bench	ticket	prison
bible	handcuffs	prisoner
cases	jail	prosecutor
chambers	judge	records
clerk	jury	robes
criminals	lawyer	subpoenas
defendant	plaintiff	witnesses
detective		

Table A.3: Concrete courthouse words included. The present table shows a list of all the concrete words included in the model in the courthouse category.

ambulance	gown	surgeon
bandage	mask	syringe
blood	medication	technician
chart	needle	thermometer
doctor	nurse	tweezers
drugs	patient	visitor
gauze	physician	volunteer
glove	scrubs	wheelchair

Table A.4: Concrete hospital words included. The present table shows a list of all the concrete words included in the model in the hospital category.

Word	**Classification**	**Category**	**Reason**
appeal	abstract	courthouse	B
claim	abstract	courthouse	B
confession	abstract	courthouse	B
dismiss	abstract	courthouse	B
divorce	abstract	courthouse	B
motion	abstract	courthouse	B
oath	abstract	courthouse	B
pardon	abstract	courthouse	B
admission	abstract	hospital	B
diagnosis	abstract	hospital	B
nauseous	abstract	hospital	N
bailiff	concrete	courthouse	N
fugitives	concrete	courthouse	N
gavel	concrete	courthouse	N
stenographer	concrete	courthouse	N
bedpan	concrete	hospital	N
crutches	concrete	hospital	N
gurney	concrete	hospital	N
paramedic	concrete	hospital	N
scalpel	concrete	hospital	N
stethoscope	concrete	hospital	N

Table A.5: Excluded words. The present table shows all the words that were in the fMRI study but were excluded from the model either because they were not in the Fyshe database (N) or to balance abstract and concrete words studied (B).

In addition to excluding 11 words per word type from all participants, the model also excluded any data from words where the participant responded to the concreteness judgment task incorrectly in one or both of the scans. This exclusion was performed on a per-participant basis and is summarized in Table A.6. Each participant had a certain number of inaccurate words excluded per word type (ranging from 0 to 22 words) which resulted in a fewer number of possible trials in the leave-out-two paradigm (ranging from 351 to 1176 trials). The full details are provided in more depth in the table.

Participant	**Number Inaccurate**	**Possible Trials**	**Inaccurate Words**
01	7 abstract 0 concrete	861 abstract 1176 concrete	illness justice order relapse research rights treatment
02	3 abstract 21 concrete	1035 abstract 378 concrete	perjury proof sustained attorney cases clerk criminals defendant detective doctor drugs judge lawyer medication nurse patient

Participant	Number Inaccurate	Possible Trials	Inaccurate Words
			physician plaintiff policeman prisoner prosecutor scrubs surgeon technician
03	0 abstract 2 concrete	1176 abstract 1081 concrete	cases patient
04	0 abstract 1 concrete	1176 abstract 1128 concrete	cases
05	3 abstract 1 concrete	1035 abstract 1128 concrete	proof recovery treatment cases
06	20 abstract 0 concrete	406 abstract 1176 concrete	acquittal adjourn clinical emergency hygiene illness interrogate judicial law order pathology perjury plead recovery sick sue sustained therapy treatment truth
07	22 abstract 10 concrete	351 abstract 741 concrete	adjourn condition illness interrogate judgment

Participant	Number Inaccurate	Possible Trials	Inaccurate Words
			judicial law mortality order pathology perjury plead recovery relapse research science sterile sue sustained technology trauma tried chart clerk criminals defendant detective drugs medication patient policeman prisoner
08	14 abstract 0 concrete	595 abstract 1176 concrete	acquittal adjourn illegal illness interrogate judgment judicial law medical mortality pathology perjury sue truth

Participant	Number Inaccurate	Possible Trials	Inaccurate Words
09	12 abstract 1 concrete	666 abstract 1128 concrete	acquittal adjourn interrogate judgment judicial law order pathology perjury plead sue sustained criminals
10	0 abstract 0 concrete	1176 abstract 1176 concrete	
11	9 abstract 1 concrete	780 abstract 1128 concrete	emergency health illness interrogate proof sick technology trauma truth cases

Table A.6: Inaccurate words. The present table indicates which words were answered incorrectly in one or both scans, broken down by participant. Counts are given for numbers of inaccurate abstract words and concrete words ("Number Inaccurate") as well as for the numbers of total trials in the leave-out-two paradigm ("Possible Trials"). The total possible trials are equal to $(49 - x)(49 - x - 1)/2$ where x is the number of inaccurate words of the 49 total words per word type.

Appendix B

MATLAB scripts for General Linear Model

The present appendix provides two of the scripts that were used in the present study: `auto.m`, which uses parallel processing to run spm jobs for all participants, and `job01.m`, which is the job for participant 01 to perform the general linear model (GLM) and several contrasts. To create, `job02.m` and beyond, simply change the string in the sixth line and save as an additional file. These scripts also require other files such as `getPath.m` which returns a string of the path of the file and `getOnsets.m` which returns the onsets of a word or a group of words or nonwords (i.e., abstract, concrete, or control). Additionally, the scripts require `.mat` files which contain an array of participant numbers and an array of words. For sake of space, these additional pieces are not included in the appendix; however, with the appropriate MATLAB knowledge, they can be recreated for custom use. The inclusion of `auto.m` and `job01.m` should suffice in understanding how the MATLAB scripts for the GLM and contrasts were made for the present study.

```matlab
% auto.m by Dominick DiMercurio II, May 2017
% This script calls jobs for all participants in parallel
% Does jobs of GLM & contrasts for every word.
% Requires getPath.m, participants.mat, and words.mat

addpath(getPath('spm')); % adds SPM to path
spm('defaults', 'FMRI'); % starts SPM
spm_jobman('initcfg')

path = getPath('job'); % gets the path of job.m

% Saved participants and word files are loaded.
load participants.mat;
load words.mat;

%Requests one processor per participant and starts loop
parpool(length(participants))
parfor i=1:length(participants)
    for j=1:length(words)

        % Saves the word is currently working on.
        % This is later accessed by job##.m
        p = char(participants(i));
        word = char(words(j));
        save(['word' p '.mat', 'p', 'word']);

        % Calls the job##.m with spm_jobman
        spm_jobman('run', job);
    end
end

% exits parallel processing
delete(gcp('nocreate'));

% Displays a completion message
disp('This was a triumph!')
```

```matlab
% job01.m by Dominick DiMercurio II, May 2017
```

```
% This script is for participant 01, modified from spm
% Performs GLM & contrasts for a single word
% Requires getPath.m, getOnsets.m, w01.mat

p = '01';
participant = p;
path = getPath('data'); % gets path of data
load ['word' p '.mat']; % loads the current word

% uses the SSM directory for later storage
pathfile = strcat(path, participant, '\SSM');

% the matlabbatch as created by spm
matlabbatch{1}.cfg_basicio.file_dir.dir_ops.cfg_mkdir.parent = {pathfile};
matlabbatch{1}.cfg_basicio.file_dir.dir_ops.cfg_mkdir.name = strcat('SSM_GLM_',word); % modified to be specific to the word
matlabbatch{2}.spm.stats.fmri_spec.dir(1) = cfg_dep(strcat('Make Directory: Make Directory ''SSM_GLM_', word, ''''), substruct('.','val', '{}',{1}, '.','val', '{}',{1}, '.','val', '{}',{1}, '.','val', '{}',{1}), substruct('.','dir'));
matlabbatch{2}.spm.stats.fmri_spec.timing.units = 'secs';
matlabbatch{2}.spm.stats.fmri_spec.timing.RT = 2.5;
matlabbatch{2}.spm.stats.fmri_spec.timing.fmri_t = 16;
matlabbatch{2}.spm.stats.fmri_spec.timing.fmri_t0 = 8;
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).scans = {
    [pathfile '\scan1\wrajudgement1s003a001.nii,1']
    %... files not shown here for ease of reading
    [pathfile '\scan1\wrajudgement4s006a001.nii,127']
                                                                        };
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(1).name = word;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(1).onset = getOnsets(participant, 'scan1', word, 'word');
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(1).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(1).tmod = 0;
```

```
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(1).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(1).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(2).name = 'abstract';
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(2).onset = getOnsets(participant, 'scan1', word, 'abstract');
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(2).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(2).tmod = 0;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(2).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(2).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(3).name = 'concrete';
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(3).onset = getOnsets(participant, 'scan1', word, 'concrete');
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(3).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(3).tmod = 0;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(3).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(3).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(4).name = 'control';
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(4).onset = getOnsets(participant, 'scan1', word, 'control');
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(4).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(4).tmod = 0;
```

```
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(4).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(1).cond(4).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(1).multi = {''};
matlabbatch{2}.spm.stats.fmri_spec.sess(1).regress = struct('name', {}, 'val', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(1).multi_reg = {[pathfile '\scan1\rp.txt']};
matlabbatch{2}.spm.stats.fmri_spec.sess(1).hpf = 128;
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).scans = {
    [pathfile '\scan2\wrajudgement1s003a001.nii,1']
    %... files not shown here for ease of reading
    [pathfile '\scan2\wrajudgement4s006a001.nii,127']
                                                            };
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(1).name = word;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(1).onset = getOnsets(participant, 'scan2', word, 'word');
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(1).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(1).tmod = 0;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(1).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(1).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(2).name = 'abstract';
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(2).onset = getOnsets(participant, 'scan2', word, 'abstract');
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(2).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(2).tmod = 0;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(2).pmod = struct('name', {}, 'param', {}, 'poly', {});
```

```
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(2).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(3).name = 'concrete';
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(3).onset = getOnsets(participant, 'scan2', word, 'concrete');
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(3).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(3).tmod = 0;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(3).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(3).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(4).name = 'control';
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(4).onset = getOnsets(participant, 'scan2', word, 'control');
%%
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(4).duration = 4;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(4).tmod = 0;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(4).pmod = struct('name', {}, 'param', {}, 'poly', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(2).cond(4).orth = 1;
matlabbatch{2}.spm.stats.fmri_spec.sess(2).multi = {''};
matlabbatch{2}.spm.stats.fmri_spec.sess(2).regress = struct('name', {}, 'val', {});
matlabbatch{2}.spm.stats.fmri_spec.sess(2).multi_reg = {[pathfile '\scan2\rp.txt']};
matlabbatch{2}.spm.stats.fmri_spec.sess(2).hpf = 128;
matlabbatch{2}.spm.stats.fmri_spec.fact = struct('name', {}, 'levels', {});
matlabbatch{2}.spm.stats.fmri_spec.bases.hrf.derivs = [0 0];
matlabbatch{2}.spm.stats.fmri_spec.volt = 1;
matlabbatch{2}.spm.stats.fmri_spec.global = 'None';
```

```
matlabbatch{2}.spm.stats.fmri_spec.mthresh = 0.8;
matlabbatch{2}.spm.stats.fmri_spec.mask = {''};
matlabbatch{2}.spm.stats.fmri_spec.cvi = 'AR(1)';
matlabbatch{3}.spm.stats.fmri_est.spmmat(1) = cfg_dep('fMRI model specification: SPM.mat File', substruct('.','val', '{}',{2}, '.','val', '{}',{1}, '.','val', '{}',{1}), substruct('.','spmmat'));
matlabbatch{3}.spm.stats.fmri_est.write_residuals = 0;
matlabbatch{3}.spm.stats.fmri_est.method.Classical = 1;

% contrasts made for wd > ctrl, abs > ctrl, con > ctrl, abs + con > ctr, wd + abs + con > ctrl (note that abs and con exclude the wd itself even if it is part of that category)
matlabbatch{4}.spm.stats.con.spmmat = {strcat(path, participant, '\SSM\SSM_GLM_', word, '\SPM.mat')};
matlabbatch{4}.spm.stats.con.consess{1}.tcon.name = word;
matlabbatch{4}.spm.stats.con.consess{1}.tcon.weights = 1;
matlabbatch{4}.spm.stats.con.consess{1}.tcon.sessrep = 'repl';
matlabbatch{4}.spm.stats.con.consess{2}.tcon.name = 'abstract>control';
matlabbatch{4}.spm.stats.con.consess{2}.tcon.weights = [0 1 0 -1];
matlabbatch{4}.spm.stats.con.consess{2}.tcon.sessrep = 'repl';
matlabbatch{4}.spm.stats.con.consess{3}.tcon.name = 'concrete>control';
matlabbatch{4}.spm.stats.con.consess{3}.tcon.weights = [0 0 1 -1];
matlabbatch{4}.spm.stats.con.consess{3}.tcon.sessrep = 'repl';
matlabbatch{4}.spm.stats.con.consess{4}.tcon.name = 'abstract+concrete>control';
matlabbatch{4}.spm.stats.con.consess{4}.tcon.weights = [0 1 1 -2];
matlabbatch{4}.spm.stats.con.consess{5}.tcon.sessrep = 'repl';
matlabbatch{4}.spm.stats.con.consess{5}.tcon.name = [word 'abstract+concrete>control'];
matlabbatch{4}.spm.stats.con.consess{5}.tcon.weights = [1 1 1 -3];
```

```
matlabbatch{4}.spm.stats.con.consess{5}.tcon.sessrep = 'repl';
matlabbatch{4}.spm.stats.con.delete = 1;
```

Appendix C

MATLAB scripts for Semantic Space Models

The present appendix contains some of the functions used for the semantic space model: `model.m`, where the model is run, `myregress.m` where the regressions are performed, and `predict.m` where the model makes its predictions. As with the files in Appendix B, the files in the present appendix also require other files to run based on the custom needs of the study. For the sake of space, those files are not included, but they should be recreatable. The files `model.m`, `myregress.m`, and `predict.m` should be sufficient in understanding how the present study performed the semantic space models.

```
% model.m by Dominick DiMercurio, May 2017
function model()
%Lets user know the function has started
clc, disp('The mission has begun.')

```

```
% Pathfiles for Model folder and for psumn folders
pathModel = getPath('model');
pathPSUMN = getPath('data');
addpath(pathModel);

% Saved participant and word files are loaded.
load participants.mat
load words.mat

% Begins parallel processing with each participant
parpool(length(participants));
parfor z=1:length(participants)
% use function to avoid transparency problems
transparency(z,pathModel,pathPSUMN,participants,words.words);
end
delete(gcp('nocreate')) %turn off parallel processing
disp('The mission has completed.') %notify user of completion
end

function transparency(z,pathModel,pathPSUMN,participants,words)

    p = char(participants(z));
    addpath([pathPSUMN p]);

    thresh = 2.33; % uncorrected p < 0.01
    suprathresh = []; % indices of all voxels suprathresh
    maskpath = [pathPSUMN p '/SSM/SSM_GLM_ambulance/spmT_0005.nii'];
    maskdata = readnifti(maskpath);
    masksize = size(maskdata);

    for i = 1:masksize(1)
        for j = 1:masksize(2)
            for k = 1:masksize(3)
                if maskdata(i,j,k)>thresh
                    suprathresh = [suprathresh; i j k]; %#ok<AGROW>
                end
            end
```

```
        end
    end
    save([pathModel '/hi-suprathresh' p '.mat'], 'suprathresh');

    % Only get indices from accurate words
    wordindices = 1:length(words);
    wordindices = wordindices(~ismember(wordindices, getInaccurate(p)));

    data = zeros(length(words),length(suprathresh(:,1)));

    for j=wordindices
        w = char(words(j));

        % Retrieve the contrast data from the nifti file
        conpath = [pathPSUMN p '/SSM/SSM_GLM_' w '/con_0001.nii'];
        condata = readnifti(conpath);

        % Get only the data that is in suprathresh voxels
        for k=1:length(suprathresh(:,1))
            data(j,k) = condata(suprathresh(k,1), suprathresh(k,2),suprathresh(k,3));
        end

        disp(['Collecting contrast data for ' upper(w) ' from P' p]);
    end

    % Saves data to dataP##.mat
    save([pathModel '/data-psumn' p '.mat'],'data');
    disp(['Saving all the contrast data to data-psumn' p '.mat']);

    aDocTrials = [];
    aDepTrials = [];
    aSimTrials = [];
    cDocTrials = [];
    cDepTrials = [];
    cSimTrials = [];

```

```
    absindices = wordindices(wordindices<50);
    conindices = wordindices(wordindices>49);

    for t = absindices
        for u = absindices(absindices>t)
            % indices, removing inaccurate and two for the trial
            leaveTwo = absindices(~ismember(absindices,[t u]));

            % perform regression!
            [doc, dep, sim] = myregress(data,leaveTwo);


            % make predictions
            [tDocPred, tDepPred, tSimPred] = predict(doc, dep, sim, t);
            [uDocPred, uDepPred, uSimPred] = predict(doc, dep, sim, u);

            % get actual results
            tActual = data(t,:);
            uActual = data(u,:);

            % test the trials
            docTrial = cossim(tDocPred,tActual)+cossim(uDocPred,uActual)>cossim(tDocPred,uActual)+cossim(uDocPred,tActual);
            depTrial = cossim(tDepPred,tActual)+cossim(uDepPred,uActual)>cossim(tDepPred,uActual)+cossim(uDepPred,tActual);
            simTrial = cossim(tSimPred,tActual)+cossim(uSimPred,uActual)>cossim(tSimPred,uActual)+cossim(uSimPred,tActual);

            %docTrial = [cossim(tDocPred,tActual)>cossim(tDocPred,uActual) cossim(uDocPred,uActual)>cossim(uDocPred,tActual)];

            % update the trials
            aDocTrials = [aDocTrials docTrial];
            aDepTrials = [aDepTrials depTrial];
```

```
            aSimTrials = [aSimTrials simTrial];

        end
    end

    for t = conindices
        for u = conindices(conindices>t)
            % indices, removing inaccurate and two for the
               trial
            leaveTwo = conindices(~ismember(conindices,[t u
               ]));

            % perform regression!
            [doc, dep, sim] = myregress(data,leaveTwo);

            % make predictions
            [tDocPred, tDepPred, tSimPred] = predict(doc,
               dep, sim, t);
            [uDocPred, uDepPred, uSimPred] = predict(doc,
               dep, sim, u);

            % get actual results
            tActual = data(t,:);
            uActual = data(u,:);

            % test the trials
            docTrial = cossim(tDocPred,tActual)+cossim(
               uDocPred,uActual)>cossim(tDocPred,uActual)+
               cossim(uDocPred,tActual);
            depTrial = cossim(tDepPred,tActual)+cossim(
               uDepPred,uActual)>cossim(tDepPred,uActual)+
               cossim(uDepPred,tActual);
            simTrial = cossim(tSimPred,tActual)+cossim(
               uSimPred,uActual)>cossim(tSimPred,uActual)+
               cossim(uSimPred,tActual);

            % update the trials
            cDocTrials = [cDocTrials docTrial];
            cDepTrials = [cDepTrials depTrial];
            cSimTrials = [cSimTrials simTrial];
        end
    end
```

```
    % Saves runs to trials-psumn##.mat
    save([pathModel '/trials-psumn' p '.mat'],'aDocTrials',
       'aDepTrials','aSimTrials','cDocTrials','cDepTrials',
       'cSimTrials');
    disp(['Saving all the trial runs to trials-psumn' p '.
       mat']);

end
```

```
% myregress.m by Dominick DiMercurio II, May 2017
% Takes in activation as a words-by-voxels matrix
% Takes in indices as a vector for which words are used
% Gives out learned scale parameters in voxels-by-dimension
function [doc, dep, sim] = myregress(activation, indices)
% Loads semantic vectors retrieved from Fyshe
load doc300.mat
load dep300.mat
load cos300.mat % developed using cosine similarity

% uses 25 predictors with partial least squares regression
   on activation vectors for each model
[XL,YL,XS,YS,doc]=plsregress(activation(indices,:), doc300(
   indices,:), 25);
[XL,YL,XS,YS,dep]=plsregress(activation(indices,:), dep300(
   indices,:), 25);
[XL,YL,XS,YS,sim]=plsregress(activation(indices,:), cos300(
   indices,:), 25);

% removes first beta as it is the intercept
doc = doc(2:end,:);
dep = dep(2:end,:);
sim = sim(2:end,:);

end
```

```
% predict.m by Dominick DiMercurio II, May 2017
% xdoc, xdep, xsim is voxels-by-features (form myregress.m)
% index indicates word to predict
function [pdoc, pdep, psim] = predict(xdoc, xdep, xsim,
   index)
% Loads semantic vectors retrieved from Fyshe
```

```
load doc300.mat
load dep300.mat
load cos300.mat % developed using cosine similarity
% Predicts the word using learned scaled parameters as
    multiplied by semantic features of word (at index)
pdoc = xdoc*doc300(index,:)';
pdep = xdep*dep300(index,:)';
psim = xsim*cos300(index,:)';
% Transposes the predicted vectors
pdoc = pdoc';
pdep = pdep';
psim = psim';
end
```

References

Allport, D. A. (1985).
In S. K. Newman & R. Epstein (Eds.), *Current perspectives on dysphasia* (chap. Distributed memory, modular subsystems, and dysphasia). New York, NY: Churchhill Livingstone.

Arbib, M. (2015).
In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (chap. Neurolinguistics). Oxford, UK: Oxford University Press.

Bandettini, P. A. (2012). Functional MRI: A confluence of fortunate circumstances. *NeuroImage*, *61*(2), A3–A11.

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: Erp and behavioral effects in a lexical decision task. *Brain and language*, *125*(1), 47–53.

Begg, I. (1972). Recall of meaningful phrases. *Journal of Verbal Learning & Verbal Behavior*, *11*(4), 431–439.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–96.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & A, M. D. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917.

Binney, R. J., Zuckerman, B., & Reilly, J. (2016). A neuropsychological perspective on abstract word representation: From theory to treatment of acquired language disorders. *Current Neurology and Neuroscience Reports*, *16*, 79.

Boulenger, V., DÃľcoppet, N., Roy, A. C., Paulignan, Y., & A, N. T. (2007). Differential effects of age-of-acquisition for concrete nouns and action verbs: evidence for partly distinct representations? *Cognition*, *103*(1), 131–146.

Brown, G. D. A. (2007). A frequency count of 190,000 words in the london-lund corpus of english conversation. behavioural research methods instrumentation and computers. *Behavioural Research Methods Instrumentation and Computers*, *16*(6), 502–532.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia*, *39*, 1224–1239.

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

Coltheart, M. (2000). Deep dyslexia is right-hemisphere reading. *Brain and language*, *71*(2), 299–309.

Crutch, S. J., Connell, S., & Warrington, E. K. (2009). The different representational frameworks underpinning abstract and concrete knowledge: Evidence from odd-one-out judgements. *Quarterly Journal of Experimental Psychology*, *62*(7), 1377–1390.

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

de Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2015). In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (chap. Semantics-Aware Content-Based Recommender Systems). New York, NY: Springer US.

Fogarty, M. (2014). *Nouns: Concrete, abstract, collective, and compound.* Retrieved 2014-10-20, from `http://www.quickanddirtytips.com/education/grammar/nouns-concrete-abstract-collective-and-compound`

Fyshe, A., Talukdar, P., B, M., & Mitchell, T. (2013). *Documents and dependencies: an exploration of vector space models for semantic composition.* Sofia, Bulgaria.

Hoffman, P., Jones, R. W., & Lambon Ralph, M. A. (2010). Ventrolateral prefrontal cortex plays an executive regulation role in comprehension of abstract words: Convergent neuropsychological and repetitive TMs evidence. *Journal of Neuroscience*, *30*, 15450–15456.

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

Jeffries, E., Bateman, D., & Lambon Ralph, M. (2005). The role of the temporal lobe semantic system in number knowledge: Evidence from late-stage semantic dementia. *Neuropsychologia*, *43*, 887–905.

Jeffries, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology*, *23*(4).

Jeffries, E., Sage, K., & Lambon Ralph, M. A. (2007). Do deep dyslexia, dysphasia and dysgraphia share a common phonological impairment? *Neuropsychologia*, *45*(7), 1553–1570.

Kemmerer, D. (2015). *Cognitive neuroscience of language.* New York, NY and East Sussex, UK: Psychology Press.

Kiran, S., Sandberg, C., & Abbot, K. (2009). Treatment for lexical retrieval using abstract and concrete words in persons with aphasia: Effect of complexity. *Aphasiology*, *23*, 835–853.

Kiran, S., & Tuchtenhagen, J. (2005). Imageability effects in normal spanishâĂŞenglish bilingual adults and in aphasia: Evidence from naming to definition and semantic priming tasks. *Aphasiology*, *19*(3), 315–327.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34.

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *12*(1), 97–107.

Kucera, H., N, F. W., B, C. J., & F, T. W. (1967). *Computational analysis of present-day american english.* Providence, RI: Brown University Press.

Lambon Ralph, M. A., Pobric, G., & Jeffries, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex*, *19*, 832–838.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lowe, W. (2001). Towards a theory of semantic space. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *23*, 832–838.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203âĂŞ-208.

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: evidence from a single case study. *Brain Language*, *43*(2), 240–274.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain actitivity associated with the meanings of nouns. *Science*, *320*, 1191.

Muller, N. G., & Kleinschmidt, A. (2004). The attentional 'spotlight's' penumbra: center-surround modulation in striate cortex. *Nueroreport*, *15*, 977–980.

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience*, *6*(275).

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

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: a critical review of kousta et al. (2011). *Journal of Experimental Psychology: General*, *142*, 282–287.

Paivio, A., & Rowe, E. J. (1970). Noun imagery, frequency, and meaningfulness in verbal discrimination. *Journal of Experimental Psychology*, *85*, 264–269.

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

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*(5), 377–500.

Pobric, G., Jeffries, E., & Lambon Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences*, *104*, 20137–20141.

Pobric, G., Jeffries, E., & Lambon Ralph, M. A. (2009). The role of the anterior temporal lobes in the comprehension of concrete and abstract words: rTMS evidence. *Cortex*, *45*, 1104–1110.

Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science*, *31*, 157–168.

Romani, C., McAlpiine, S., & Martin, R. C. (2007). Concreteness effects in different tasks: Implications for models of short-term memory. *Quarterly Journal of Experimental Psychology*, *31*, 16949–16957.

Saffran, E. M., & Sholl, A. (1999). In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (p. 245). Oxford, UK: Oxford University Press.

Sahlgren, M. (2005). *An introduction to random indexing.* Copenhagen, DK.

Sahlgren, M. (2006). *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces.* Unpublished doctoral dissertation, Stockholm University.

Sandberg, C. (in preparation). Changes in task-based and resting-state functional connectivity in the default mode and semantic networks in persons with aphasia after word-finding therapy. *Unpublished manuscript.*

Sandberg, C., & 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.

Schloss, B., & Li, P. (2016). Disentangling narrow and coarse semantic networks in the brain: The role of computational models of word meaning. *Behavior Research Methods*, 1–15.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Pittsburgh, PA: Psychology Software Tools Inc.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1998). Context availability and lexical decisions for abstract and concrete words. *Journl of Memory and Language*, *27*, 499–520.

Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P. F., Adriany, G., Hu, X., & Ugurbil, K. (2002). Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, *36*, 1195–1210.

Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology*, *2*, 167–182.

Ter Doest, L., & Semin, G. R. (2005). Retrieval contexts and the concreteness effect: Dissociations in memory for concrete and abstract words. *European Journal of Cognitive Psychology*, *17*, 859–881.

Thompson-Schill, S. L., Kan, I. P., & Oliver, R. T. (2006). In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (chap. Functioning neuroimaging of semantic memory). Cambridge, MA: MIT Press.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York, NY: Teachers College, Columbia University.

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*, 1409–1422.

van Dijk, K., Hedden, T., Venkataraman, A., Evans, K., Lazar, S., & Buckner, R. (2013). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of Neurophysiology*, *103*(1), 297–321.

Vigliocco, G., Kousta, S., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural represetnation of abstract words: The role of emotion. *Cerebral Cortex*, *24*, 1767–1777.

Vigliocco, G., Kousta, S., Vinson, D., Andrews, M., & Del Campo, E. (2013). The representation of abstract words: What matters? reply to paivio's (2013) comment on kousta et al. (2011). *Journal of Experimental Psychology: General*, *142*, 288–291.

Vinson, D., Ponari, M., & Vigliocco, G. (2015). How does emotional context affect lexical processing? *Cognition and Emotion*.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256–1271.

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Human Brain Mapping*, *31*(10), 1459–1468.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*, 143–154.