Links between Short-Term Memory and Word Retrieval in Aphasia

Irene Minkina

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee

Diane L. Kendall, Chair

Nadine Martin

Kristie A. Spencer

Program Authorized to Offer Degree:

Speech & Hearing Sciences

University of Washington

Abstract

Links between Short-Term Memory and Word Retrieval in Aphasia

Irene Minkina

Chair of the Supervisory Committee
Diane Kendall, Ph.D., Professor & Chair
Department of Speech & Hearing Sciences

Effective assessment and treatment of aphasia requires specific and sensitive diagnosis of the linguistic level(s) wherein the linguistic impairment lies (e.g., semantic, lexical, phonological, etc.). Recent work in aphasiology argues that linguistic impairments at both the production and comprehension levels of language are supported, at least in part, by shared underlying processes. These processes aid in the access to and the temporary maintenance of linguistic elements needed to produce and comprehend language. One such process that has been a topic of recent research in aphasia is verbal short-term memory, or the temporary, limited-capacity storage mechanism through which information is briefly retained. A typical measure of verbal short-term memory is an immediate repetition span task: individuals listen to a list of digits or words and repeat them immediately after the list ends. People with aphasia demonstrate digit and word span lengths below that of neurologically healthy individuals. More specifically, resent research demonstrates that individuals with aphasia with greater language comprehension

impairments at the semantic level show different patterns of breakdown in serial recall tasks from that of individuals with aphasia with greater language comprehension impairments at the phonological level. Individuals with language comprehension impairments at the level of phonology tend to err on items at the beginning of a list and demonstrate higher repetition accuracy on high frequency and imageability words, while individuals with semantic language comprehension impairments tend to err on items at the end of a list and are not as susceptible to manipulations of frequency and imageability. The current study further explored the link between locus of linguistic impairment in aphasia and immediate serial recall performance by examining the relationship between type of word retrieval impairment (i.e., semantic or phonological) and two aspects of verbal short-term memory performance, 1) location of errors made on a serial recall task (word pair repetition) and 2) susceptibility to word frequency and imageability manipulations during word pair repetition. The extent to which a linguistically-specified short-term memory system, versus a domain-general one, supports the word retrieval process was also examined.

The results demonstrated that overall accuracy of word retrieval correlates positively with word span length in the absence of a correlation between word span length and a nonverbal (i.e., spatial) span length, providing support for the existence of a linguistically-specified short-term memory system. Additionally, while a relationship between type of word retrieval impairment and location of errors made on a word pair repetition task was not realized, a significant correlation between type of word retrieval impairment and susceptibility to word frequency and imageability manipulations during word pair repetition was found. As the number of phonological word retrieval errors, relative to semantic errors, increased, the bias towards correct repetition of high imageability over low imageability words increased. These finding are

interpreted in the context of an interactive activation model of word retrieval where activation spreads from lexical-semantic to phonological representations. While an error location bias on the word repetition task likely arises due to intricate timing of linguistic activation, a bias towards correct repetition of highly imageable and frequent words likely arises due to the overall activation strength of linguistic representations. Results suggest that word pair repetition, a classic short-term memory task, and picture naming, a classic language production task, may be supported by a shared temporary linguistic activation process dependent on the overall activation strength of linguistic representations. Theoretically, the findings demonstrate the need to rethink the classic distinction between short-term memory and language that views the processes as separable. Clinically, the findings suggest that the analysis of imageability biases in word pair repetition tasks may help determine word retrieval impairment type, an encouraging idea given the much more laborious and time-intensive nature of analyzing word retrieval errors.

Table of Contents

List of Tables and Figures

Figures

Tables

Acknowledgements

This dissertation was completed with support from an Institutional Training Grant from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health (T32000033) and a Speech and Hearing Sciences department teaching assistantship.

Completion of this doctoral dissertation was possible with the support and presence of many individuals from the University of Washington Speech and Hearing Sciences department and beyond. First and foremost, I am extremely and wholeheartedly grateful for the unyielding support of my mentor, Dr. Diane Kendall, who embodies the art of holistic mentorship. From day 1 of my graduate program, Diane made it clear to me that her mentorship role extends far beyond that of a research advisor, and that she is here to support me as a whole person. Diane has taught me, by example, to work hard and focus on the projects in front of me while always keeping in mind the bigger picture of why we do this work. She has constantly challenged and inspired me to think deeper, work through challenging theoretical concepts with an unyielding curiosity, and ask the difficult research questions. She has beautifully modeled the importance of balancing and integrating work and personal life, and of navigating life with fierce compassion, bravery, and grace. Diane has made it clear to me from the beginning that her mentorship will extend far beyond the completion of my Ph.D., and I am incredibly grateful for her support throughout my program and beyond.

I am very thankful for the encouragement and guidance of my dissertation committee members, Dr. Nadine Martin, Dr. Kristie Spencer, and Dr. Kathryn Yorkson, who have challenged, inspired, and pushed me to stretch my thinking and approach my research with clarity, integrity, and passion. I am grateful to Dr. Nadine Martin for her seminal work in the

area of short-term memory and aphasia, and for her guidance and encouragement of my research goals and desire to contribute to this work. I am delighted and honored to continue working with Dr. Nadine Martin as a postdoctoral fellow. I am thankful to Dr. Kristie Spencer for helping me approach my work systematically and narrow the focus of my dissertation, and for challenging me to expand my thinking. I am thankful to Dr. Yorkston for encouraging me to approach the theoretical research questions at the heart of this study through a clinical lens. I am grateful to all of my committee members for their time and patience in guiding me and helping me grow as a researcher and person in my doctoral program.

The collaborative and warm environment that Dr. Diane Kendall and the University of Washington Aphasia Research Laboratory have built has provided indispensable support throughout my doctoral program. To my lab "big sisters," Lauren Bislick, Liz Brookshire, Rebecca Hunting-Pompon, Megan Oelke, and JoAnn Silkes, with whom I have worked side by side throughout the duration of my doctoral program, I have no words to describe the joy it has been to work alongside you and the gratitude I have for your unyielding support, both professional and personal. I look forward to many more years of friendship and collaboration to come. Thank you, also, to the former and newer members of the laboratory, including Wesley Allen, who always reminds me to smile and live wholeheartedly, as well as Christina del Toro, Kerry Lam, Reva Robinson, and Janaki Torrence for their support and guidance. I am also thankful for my peers in the department of Speech & Hearing Sciences, Christina Defrancisci and Derek Isetti, for their friendship and support. Last but not least, I am thankful to Ernie Lefler, doctoral program advisor for the Speech & Hearing Sciences department, for providing guidance throughout the program, as well as many much needed stress-relieving laughs.

I am forever grateful to the individuals who participated in this study, as well as their families, for their time, efforts, and selfless willingness to contribute to aphasia research. They, and the larger Seattle aphasia community, have continually inspired me and modeled what it means to live with resilience, compassion, and grace. My research assistants, Heather Jackson, Olivia Kao, Elizabeth McKeen, Alyssa Shawler, and Josie Stump, who volunteered their time to help me process my dissertation data, also have my sincere appreciation and gratitude. I could not have completed my dissertation in this time frame without them.

Many people outside of the Speech & Hearing Sciences community have supported my completion of this work. I am grateful to have found a dance and movement community in Seattle that has provided me with incredible support, stress relief, and balance throughout my doctoral program. I am especially grateful to my very dear friends and mentors, Jennifer Lucero-Earle and Liziah Richards, for continually reminding me that I am capable of more than I think, inspiring me to work and live wholeheartedly, and dancing through life with me. Last, but certainly not least, I am grateful for the support of my parents, grandparents, and sisters. My mom, especially, has been a never-ending source of love and encouragement. My grandfather, Dr. Lev Minkin, an 86 year young theoretical linguistic living in Kharkov, Ukraine, has been one of my biggest supporters throughout my education. Despite the physical distance between us, his enthusiasm, love, and celebration of my work have been a constant source of support and encouragement throughout my doctoral program.

I am deeply grateful for the presence, support, and encouragement of the individuals mentioned above. Thank you for going on this journey with me. I am honored to know you.

Chapter 1: Introduction

Definitions and Clinical Implications

Aphasia, an acquired multimodality language processing disorder most commonly caused by a left cerebral hemisphere stroke, impacts over a million individuals in the United States ("Aphasia Fact Sheet," 2015). The disorder affects all linguistic domains (speaking, understanding, reading, and writing of language) (Rosenbek, LaPointe, & Wertz, 1989). Impairments to these domains arise from disrupted access to linguistic representations (e.g., semantic features, phonemes, etc.), which affects the ability to accurately and efficiently perform language functions (McNeil, Odell, & Tseng, 1991; McNeil & Pratt, 2001). Though several early definitions of aphasia emphasized that a loss of linguistic representations underlies the deficits seen in aphasia (Benson, 1979; Darley, 1982; Ryan, 1982, as cited by McNeil & Pratt, 2001), this all-or-nothing view of language breakdown has been challenged by repeated behavioral observations of individuals with aphasia that suggest otherwise. Two such observations are stimulability and variability of individuals with aphasia. Stimulability refers to the observation that language performance in people with aphasia can be manipulated through a variety of means, including loudness of stimuli, linguistic cues (e.g., giving an individual the first sound of a word in order to aid with picture naming), and size and color of the stimuli. These types of stimulus manipulations facilitate *access* to linguistic representations, indicating that they are not wholly lost but are more difficult to retrieve in individuals with aphasia. Variability refers to the inconsistency of language performance seen in people with aphasia over time: an individual with aphasia may be able to retrieve a word one day or moment but not do so on another occasion. Such a phenomenon would not be possible if linguistic representations were completely degraded (McNeil, Odell, & Tseng, 1991), and points instead to a breakdown in access to linguistic elements required for language processing. The scarcity of evidence in the extant literature for total dissociations between domains (e.g., completely spared language comprehension in the face of weak language production) provides additional evidence for a system in which underlying processes exists that support language more generally. Thus, the thorough study of language breakdown in aphasia must take into account not only the linguistic representations themselves, but also the processes that facilitate access to these elements.

The study of the ways in which typical speakers are able to access linguistic elements in order to easefully utter a word, phrase, or sentence, and the ways in which such a process breaks down, is necessarily a study of intricate temporal processes (N. Martin, 2000). The online production and comprehension of language requires speakers to access linguistic elements at a precise moment in time and to temporarily hold on to them until production or comprehension of a particular utterance is complete. If a problem of linguistic access underlies the impairments seen in aphasia, one major question of interest to clinical aphasiologists is, what processes fuel the ability to access the linguistic elements needed to produce and comprehend language? Researchers in pursuit of a better understanding of these underlying processes will help move the field of clinical aphasiology towards the development of more effective assessment and treatment tools for individuals with aphasia.

At the assessment level, a process-driven research program will help to elucidate *how* linguistic domains break down. While classic clinical tests of aphasia such as the Western Aphasia Battery (Kertesz, 1982) and the Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barressi, 2000) might give us a good idea of *what* domains are impaired (e.g., an individual might demonstrate fairly preserved word and sentence comprehension, with impairments in word production and repetition), they don't help us understand *how* processes supporting these language functions break down to cause linguistic impairments. Unlike these classic tests, The Comprehensive Aphasia Test (Swinburn, Porter, & Howard, 2004), a relatively new assessment of language in individuals with aphasia, is one of the few to systematically vary psycholinguistic variables (e.g., word imageability and frequency) of the test's stimuli. These carefully constructed stimuli give researchers and clinicians a window into *why* a certain linguistic domain is broken. For example, the Naming Objects subtest of the Comprehensive Aphasia Test (CAT) allows examiners to track how many high versus low imageability and frequency words were correctly retrieved. If individuals demonstrate a breakdown in the naming of low frequency and imageability words only, this may signal that though lexical-semantic activation is impaired, ease of access (i.e., activation of semantic and lexical features) for more salient words is preserved. This observation can guide the clinician to work on low imageability and frequency word production during treatment. Conversely, if word retrieval is impaired but a discrepancy between high and low imageability and frequency words is not observed, the individual may a) have a more global impairment in lexical-semantic processing, b) have an impairment at the phonological level, or c) a combination of the two. In this case, it may be more beneficial to work on high frequency and imageability words and/or phonologically simpler words. Such inferences cannot be drawn from assessments of aphasia that lack carefully constructed stimuli.

Though the CAT is an invaluable tool for gauging the nature of a linguistic impairment, the authors themselves agree that it is just one step in understanding what is broken in aphasia, and that more detailed tools are needed to further understand the processes behind an individual's language performance (Howard, Swinburn, & Porter, 2010). One such process which has received considerable attention in the recent literature is verbal short-term memory (VSTM), or the temporary activation of linguistic elements needed for language production and comprehension. The present study seeks to better understand the relationship between VSTM, a process believed to facilitate access to linguistic representations (N. Martin & Ayala, 2004; N. Martin & Gupta, 2004; N. Martin & Saffran, 1997), and word retrieval. In addition to contributing to a model of VSTM breakdown in aphasia, it aims to inform the development of specific, impairment-level assessments and treatment programs for anomia (i.e., word retrieval impairment), the most pervasive symptom of aphasia (Benson, 1988). Assessments and treatments designed for individuals with aphasia have recently begun targeting VSTM (for review, see Salis, Kelly, & Code, 2015), and further understanding of the role of VSTM in language processing, particularly the association between word retrieval and verbal short-term memory, can help to refine and expand upon these tools and treatment protocols for individuals with aphasia.

In the current study, VSTM is defined as the temporary activation of linguistic representations from long-term memory, an idea that stems from an interactive bidirectional activation model of word retrieval and VSTM (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; N. Martin & Saffran, 1997). The model argues that the temporary activation of a word's meaning, grammatical form, and sounds *is* the temporary storage (i.e., STM) process that supports the production and comprehension of words. Before the interactive activation model is described in detail, the assumptions and limitations of an influential classic model of working memory that eventually lead to a major shift in the definition of STM are discussed below. From here on out, the term STM will be used when referring more generally to a temporary storage mechanism, while the term VSTM will be used when referring to a linguistically-specified temporary storage system. The current study will argue that VSTM is a specific type of STM that supports language processing.

Classic Model of Working Memory

To contextualize the theoretical underpinnings of the present study, it is first necessary to describe the concept of working memory, which consists of limited-capacity temporary information storage (short-term memory) and attentional functions that enable currently stored information to be refreshed and manipulated in service of a given goal. In other words, the more a task depends on the updating and attentional control of information in short-term memory, the more work it requires (Cowan, 2008). Though it is reasonable to assume that all conscious tasks require some degree of attentional control, and that the complete separation of short-term memory and attention is impossible, many studies have demonstrated that studying short-term memory and working memory as separate constructs is theoretically valid (for review, see Unsworth & Engle, 2007). The use of the term short-term memory in the present study does not imply that attentional influences are absent. The focus of this work, however, is to investigate a process during which fleeting storage of information operates with minimal support from attentional processes.

Short-term memory has classically been defined as a fleeting, limited-capacity storage system used for the purpose of briefly retaining temporarily stored information (Baddeley & Hitch, 1974; Baddeley, 1986a, 1986b; Baddeley & Hitch, 1994). Earlier models of memory viewed the storage system as largely isolable from language (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974). Baddeley and Hitch (1974) introduced a model of working memory that remains influential in memory research today. Short-term memory (STM), the temporary storage component, constitutes a portion of this model, while working memory is a larger term that encompasses STM as well as attentional resources that more actively process information stored in STM. The storage mechanism (the STM buffer) is responsible for the brief maintenance of information, and without the support of attentional resources, this information quickly fades. The more active mechanisms are responsible for maintaining information in the STM buffer longer than its passive limits allow, as well as for manipulating this information in the service of a given goal (Baddeley, 1986a; Baddeley & Hitch, 1994; Caspari, Parkinson, LaPointe, & Katz, 1998).

Of the components in the working memory model, the phonological loop (Baddeley, 1986a), a speech-based mechanism responsible for the storage and maintenance of verbal information, is most relevant to the discussion of STM. The phonological loop consists of two subcomponents, a short-term store (the temporary storage component) and an articulatory mechanism that allows information in the short-term store to be refreshed, and therefore, maintained longer than the limits of the short-term store allow. The central executive, the third component, governs the short-term store and the articulatory mechanism, deciding what information needs to be maintained and/or manipulated in the service of a particular cognitive task. A key aspect of the working memory model is the code through which information is stored: information is held in the short-term store by means of a phonological code. Several pieces of evidence support the temporary storage of information through a phonological code. These include the phonological similarity effect, a phenomenon describing the finding that items that are phonologically similar are more difficult to recall in a sequence (Baddeley, 1966, 1986a) and the word length effect, the finding that shorter words are more likely to be correctly recalled in a sequence than longer words (Baddeley, 1986a; Baddeley, Thomson, & Buchanan, 1975; Watkins, 1972; Watkins & Watkins, 1973). The interference of phonologically similar items as well as the advantage of items containing fewer sounds observed during serial recall have frequently been used to argue for a phonologically-coded short-term store.

Though the phenomena discussed above provide compelling evidence for the involvement of a phonological code in the temporary storage of verbal information, several problems with the model have been acknowledged in the recent literature. These include a number of more recent experiments that failed to recreate the word length effect, which suggests that the effect is likely item-specific, as well as the finding that several recent studies that did show a word length effect may have done so due to other confounding linguistic factors (e.g., orthographic neighborhood size) rather than word length itself (for review, see Caplan, Waters, & Howard, 2012, pp. 281-282). Another issue is the model’s inability to precisely describe the interference mechanism that leads to the phonological similarity effect (Neath & Suprenant, 2003, pp. 76-78). The phonological similarity effect has been argued to occur during the retrieval stage of STM due to the stronger facilitation of retrieval through phonological retrieval cues when the list contains phonologically dissimilar items, and this finding has been used as evidence for a phonological store in STM (Baddeley, 1968, as cited by Caplan et al., 2012). The problem with this assumption is that many other non-phonological factors have been shown to affect immediate serial recall performance (e.g., familiarity, imageability, concreteness, etc.), so the presence of the phonological similarity effect during recall of span information does not necessitate a purely phonological short-term store (Caplan, et al., 2012, pp. 285-287). Another criticism is the model’s inability to account for the storage of information during discourse processing. To participate effectively in an ongoing conversation, one needs to store previously communicated information. A limited-capacity phonological store would not allow for this to occur, because the information would constantly need to be replaced with new information

(Crosson, 2000). To address these and other limitations of the original working memory model, Baddeley (2000) added a new component, the episodic buffer, defining it as a multimodal temporary storage system capable of integrating various types of information, including semantic information, to aid with recall. Though a model that accounts for the support of multiple types of information coding to aid with recall is more plausible than a model which solely relies on a phonologically-based code, the episodic buffer is a new, underspecified component of the working memory model, and the processes through which it operates have not been adequately described. Thus, though the addition of a buffer that is able to integrate multiple information types, including semantic information, to aid in recall is much more theoretically plausible, its vaguely described nature does not allow for the formation and testing of predictions concerning its role in VSTM. Since the inception of Baddeley's model of working memory, many others have addressed VSTM, and one model in particular has done so in the context of language breakdown in aphasia.

A Model of Word Retrieval and Verbal Short-Term Memory in Aphasia

The model at the forefront of the present study is Dell's two-stage interactive activation model of word retrieval (Dell, 1986; Dell, et al., 1997). Dell's model makes specific assumptions regarding the representation of lexical items, the nature of the connections between different levels of linguistic knowledge that support a given lexical item, and the steps through which a lexical item is retrieved. The model makes use of three levels of linguistic knowledge: word meaning is represented via a group of semantic units, or *nodes*, that hold conceptual information, the *lemma* (also called *lexical* or *word*) nodes constitute grammatical knowledge, and phonological nodes hold sound-level information. Nodes at neighboring levels of lexical representation are linked through bidirectional connections which are excitatory in nature, meaning that nodes at one level can activate, or excite, nodes at a neighboring level (Figure 1). These connections have two key properties: 1) connection weight, which describes the strength of activation between nodes and 2) decay rate, how quickly that activation dissipates over time.

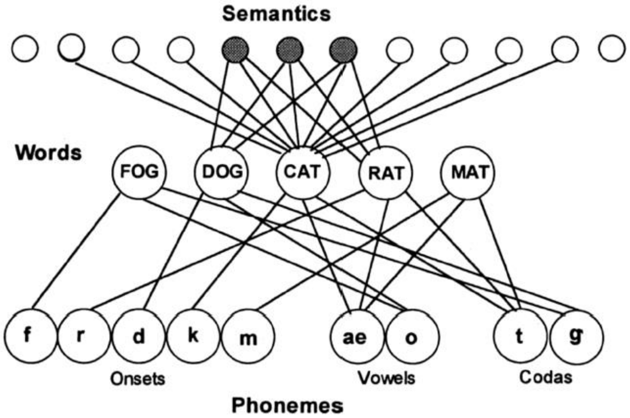


Figure 1. Dell’s interactive spreading activation model of word retrieval (Dell, Schwartz, et al., 1997)

In the interactive activation model, word retrieval occurs in 2 steps, which involve the selection of 1) a lemma, or lexical, node and 2) its phonemes. Word production begins with activation of semantic nodes which then spread activation to lexical nodes of the target word (e.g., cat), semantically related words (e.g., dog), and to their phonemes at the next level. As activation spreads to each level of linguistic representation, it begins to decay immediately but is replenished by feedback from the subsequent level of representation (phoneme to lexical nodes and lexical to semantic nodes). The first stage ends when the most highly activated lexical node is selected and sends activation to the word’s sounds (or phonemes), which are selected in the second stage. Thus, although the model includes two stages of retrieval (lexical and phonological), it is also interactive, in that lexical nodes in the first stage prime corresponding phoneme nodes which feed activation back to the lexical level. This feedback reinforces activation of the lexical nodes and activates other phonologically related lexical nodes before a word is selected. If activation between nodes is weak (i.e., reduced connection weight) and/or decays too quickly, the targeted linguistic nodes do not benefit from this feedback, and word retrieval will likely fail. A failure can result in the selection of a more active alternative, including a semantically related (e.g. cat→dog) or phonologically related (e.g. cat→mat) word, a mixed error with both a semantic and a phonological relationship to the word (e.g. cat→rat), a phonologically related nonword (cat→cag), an unrelated real word (e.g. cat→desk), or an unrelated nonword (e.g. cat→der).

Dell's model is inarguably influential in psycholinguistic research, and is used in a variety of ways to study word retrieval in both typical speakers and speakers with aphasia. Recent studies have used the model to inform studies of the interactivity among semantic, lexical, and phonological levels in word retrieval (Damian & Martin, 1999; Griffin & Bock, 1998; Jescheniak & Schriefers, 1998; Kittredge, Dell, Verkuilen, & Schwartz, 2008; Rapp & Goldrick, 2000; Schwartz, Dell, Martin, Gahl, & Sobel, 2006), competition among semantically (Belke, Meyer, & Damian, 2005; Schnur, Schwartz, Brecher, & Hodgson, 2006) and phonologically related nodes (Gordon, 2002; Navarette & Costa, 2005), the nature of semantic and phonological representations (Goldrick & Rapp, 2007; Vigliocco, Vinson, Lewis, & Garrett, 2004), and the representation and processing of words in bilingual speakers (Costa, Caramazza, & Sebastian-Galles, 2000; Costa, Santesteban, & Caño, 2005), among many other topics. The present study uses Dell's model to frame and set the stage for the study of the relationship between word retrieval and STM.

In the field of aphasiology, one influential model of STM is derived from the interactive activation model described above. Developed by Martin and colleagues (N. Martin & Saffran, 1997), the model assumes that STM is a property of cognitive processes that require a temporary storage system in order to function properly. Thus, VSTM encompasses the temporary activation of linguistic knowledge from long-term memory. Specific to word retrieval, STM constitutes the temporary activation of semantic, lexical, and phonological nodes needed to retrieve the lemma and phonemes of a given word. This definition can also be applied to comprehension, but the path of activation is reversed, because comprehension begins with the activation of sound representations. The model highlights the time-sensitive and fleeting nature of linguistic tasks, claiming that tasks that are classically labeled as STM tasks (e.g., immediate serial recall) and those labeled as language tasks (e.g., picture naming, sentence comprehension, etc.) are both dependent on and supported by the temporary activation of linguistic elements. Thus, tasks categorized as STM tasks and tasks labeled as language tasks are not wholly separable, but, rather, rely on a shared temporary storage buffer that briefly maintains activation of linguistic representations that are needed to produce and comprehend language (N. Martin & Gupta, 2004; N. Martin & Saffran, 1997; N. Martin, Saffran, & Dell, 1996).

The VSTM model developed by Martin and colleagues can be thought of as a linguistically-specified version of Cowan’s embedded processes model (Cowan, 1988, 1999) which, unlike Baddeley’s classic working memory model, emphasizes the temporary storage of information through a variety of codes (including nonverbal sounds, tactile, linguistic, etc.). Like Martin’s model, Cowan’s model views the mechanism of temporary storage as heightened activation of knowledge in long-term memory. Information can be brought to a heightened state of activation by external stimulation (for example, in the context of word retrieval, producing the word “cat” will activate semantic, lexical, and phonological features that underlie that word) or internally through thoughts that cause activation from other related representations to spread to the word “cat,” thereby activating linguistic knowledge representing that word. Decay functions in the same way as in the interactive activation model: activation begins to decay as soon as knowledge is activated, and attentional processes are needed in order to maintain and use this information beyond the limits of the short-term store. The critical similarity between Martin’s conceptualization of VSTM and Cowan’s more general STM model is the assumption that short-term memory is a property of the cognitive process that requires the storage of a given type of information, and therefore the process of short-term storage cannot be completely separated out from the cognitive process that the short-term storage works to support. Word retrieval is one cognitive process that requires temporary storage in order to function, and must therefore be supported by a STM system that is at least partly language-specific.

The parsimonious attribution of language processing and VSTM to the same underlying temporary activation process is supported by studies seeking to describe the neural substrates that underlie language and VSTM. One recent framework put forth by Cahana-Amitay and Albert (2015) describes the concept of *neural multifunctionality*, the idea that networks involved in the breakdown and recovery of language function in aphasia also support other functions that are commonly thought of as being separable from language functions. They cite evidence from a neuroimaging study that linked impairments of both language and VSTM (as measured by an immediate serial recall digit span task, described on p. 26) with damage to left inferior frontal and posterior temporal cortical networks (Koenigs, et al., 2011, as cited by Cahana-Amitay and Albert, 2015), as well as a voxel-based lesion-symptom mapping study that linked both digit span and auditory sentence comprehension performance with the preservation (or lack thereof) of posterior regions of the superior temporal gyrus and sulcus (Leff, et al., 2009, as cited by Cahana-Amitay and Albert, 2015). The framework does not propose completely inseparable short-term memory and language systems, but does argue for heavy overlap and interactivity between neural networks that support these functions.

If, as neural evidence suggests, word retrieval and verbal span break down and recover together in individuals with aphasia, how would Dell's interactive activation model account for the nature of this breakdown? In other words, how does the temporary linguistic activation process malfunction in aphasia? The two activation parameters that are vulnerable to damage in the model are connection weights, the strength of activation between linguistic nodes, and decay rate, the speed at which temporary activation of linguistic nodes dissipates and eventually returns to baseline. Weak connection weights, overly rapid decay, or a combination of the two breakdowns compromise the temporary linguistic activation (i.e., VSTM) process, thus contributing to impairments in both word retrieval and verbal span tasks in individuals with aphasia (Dell, et al., 1997; N. Martin & Saffran, 1997). The goal of the current study is not to distinguish between these different types of VSTM breakdown in aphasia, but rather to test the hypothesis that both word retrieval and verbal span tasks are subserved by a fleeting, linguistically-specified temporary activation process. Before the aims of the study are presented, extant work concerning the nature of VSTM in typical speakers and in speakers with aphasia is reviewed. Collectively, these studies suggest that a shared linguistically-specified temporary activation process might support both tasks typically used to test language (e.g., picture naming, word-to-picture matching, etc.) and tasks typically used to test VSTM (e.g., verbal span), an idea that sets the stage for the current study's goal of investigating the existence and nature of such a process.

Short-Term Memory: Methods of Study and Pertinent Research

STM in typical speakers: The serial recall task. The most common way to study short-term memory in typical speakers is with an immediate serial recall task. In an auditory immediate serial recall task, an individual hears a list of items (usually digits or words) and repeats them back in the order in which he/she has heard them (see Madigan, 1980 for a review). Stimuli are generally presented at a rate of 1 item per second and the speaker immediately recalls the list (this timing is critical when attempting to isolate the temporary storage component of working memory from attentional mechanisms that refresh and manipulate currently stored information). To determine a participant's span length, a typical measure of short-term memory capacity, the experimenter increases the list size until an individual consistently breaks down on a given list length. The capacity limit for the number of items a typical speaker can recall has been said to hover around seven, give or take two items (Miller, 1956). Proposed decades ago, this capacity limit has stood the test of time: Though several more recent studies have argued for a seemingly shorter limit of four chunks of information (e.g., Cowan, 2000), Mathy & Feldman (2012) have recently shown that a capacity limit of four information chunks is equivalent to about seven separate (un-chunked) pieces of information. Besides the observation of this limited capacity for immediately recalling a list of items, one observation that also remains very influential in the memory literature is the tendency for individuals to remember the first few items (*primacy effect*) and last few items (*recency effect*) better than the items presented in the middle of the list to be recalled. In the context of the Baddeley model of working memory (Baddeley & Hitch, 1974), the recency effect is explained by enhanced availability of the last few items in the phonological store, while the primacy effect is explained by deeper processing of the first few items due to the ability to subvocally rehearse these items longer than those presented later, leading to the transfer of these items into long-term memory.

The account of primacy and recency effects discussed above, which describes these phenomena in the context of phonological storage and rehearsal mechanisms, is different from an account which is described in detail later, and which is pertinent to the proposed work. Though they do not negate that rehearsal plays a role, Martin & Saffran (1997) propose that the primacy effect is also due to the strong semantic-level activation incurred by items presented early on in the string to be repeated, and that the recency effect is due to stronger phonological-level activation incurred by items presented later on in the list. They assume that storage of list items in short-term memory is not governed solely by phonological processes, but that semantic and lexical activation also support the storage of items in short-term memory. This view is supported by various findings related to serial recall accuracy, including the better recall of words over nonwords (Hulme, Maughan, & Brown, 1991), digits over words (Brener, 1940), high imageability words over low imageability words (Bourassa & Besner, 1994), high frequency words over low frequency words (Allen & Hulme, 2006; Hulme, et al., 1997; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002), concrete words over abstract words (Allen & Hulme, 2006; Brener, 1940; Walker & Hulme, 1999), and content words over function words (Bourassa & Besner, 1994; Brener, 1940). Both span length and serial position effects are critical for the discussion of STM breakdown in aphasia. Investigations of STM and its relationship to language breakdown in aphasia that have used span length as a measure are first discussed, followed by investigations that have looked at the nature (e.g., location) of errors in span tasks to more precisely delineate the mechanisms supporting both STM and language processing.

STM and language breakdown in aphasia.

Potential problem with repetition span tasks. The majority of studies investigating the relationship between STM and language processing in aphasia do so through the use of word and digit span paradigms like the ones described above. Typically, measures of repetition span length (digit and word) are used as a standard measure of short-term memory ability, and the relationship between these measures and more traditional language processing tasks is examined in order to investigate associations between short-term memory and language processes. Attributing the results of such measures to STM in individuals with aphasia, though, is not straight-forward, as repetition involves three processes: word recognition, temporary storage of the target word, and production of the target word (Dell, Martin, & Schwartz, 2007). Repetition spans, then, can make it difficult to determine whether the breakdown in performance truly occurred due to a breakdown in temporary storage, because it is possible that storage is preserved while recognition or production of the stored word are impaired. To partly circumvent this problem, pointing spans have been used with individuals with aphasia in addition to repetition spans, thus bypassing the production component of the repetition span task (N. Martin & Ayala, 2004; N. Martin & Gupta, 2004; R. C. Martin, Lesch, & Bartha, 1999). In a pointing span, subjects are presented with a list of items (digits or words) spoken at an approximated rate of 1 item per second, and are then asked to point to the items in the order in which they heard them on a visual array. Martin and Ayala (2004) showed that individuals did not differ in their performance on the two versions of the span task (repetition versus pointing), suggesting that both can be used to tap into STM ability in individuals with aphasia.

The use of span length in studies of STM breakdown in aphasia. Multiple studies have demonstrated that short-term memory span (digit and/or word span) is often impaired in individuals with aphasia as compared to the expected span length of neurologically healthy individuals (Albert, 1976; De Renzi & Nichelli, 1975; Gvion & Friedmann, 2012; N. Martin & Ayala, 2004; N. Martin & Saffran, 1997; N. Martin, et al., 1996; Potagas, Kasselimis, & Evdokimidis, 2011). Additionally, individuals with aphasia demonstrate lower span length than other brain-damaged individuals (with a locus of damage in one or both hemispheres) without aphasia (Lang & Quitz, 2012), individuals with right hemisphere brain damage, without aphasia (Laures-Gore, Marshall, & Verner, 2011), and individuals with left hemisphere brain damage, without aphasia (Kasselimis, et al., 2013). Together, these findings suggests that VSTM impairments found in individuals with aphasia are not simply the result of generalized slow processing following brain damage, but that VSTM and language impairments are likely subserved by overlapping neural networks, as Cahana-Amitay & Albert (2015) describe in their framework of Neural Multifunctionality.

Consistent with this claim, correlational studies of individuals with aphasia have demonstrated relationships between these traditional measures of short-term memory and language processing (N. Martin & Ayala, 2004; N. Martin & Gupta, 2004; Potagas, et al., 2011), suggesting that language and short-term memory performance are more tightly linked than early models of language processing (e.g., Baddeley & Hitch, 1974) would suggest. One such study, conducted by N. Martin and Ayala (2004), was unique in that it tested relationships between language comprehension and short-term memory *and* between word retrieval and short-term memory. N. Martin and Ayala (2004) tested a group of 46 individuals with aphasia on a battery of receptive phonological and semantic language tasks. Phonological tasks included phoneme discrimination with various intervals imposed between the two phonemes in the pair as well as auditory rhyme judgments. Semantic tasks included two word-to-picture matching measures and synonymy judgments (requiring the participant to determine the two words in a set of three that are most semantically similar). Based on their performance, participants were given a P score (quantifier of their phonological processing ability) and an S score (quantifier of their semantic processing ability). These measures were correlated with digit and word span length (both repetition and pointing). P scores correlated positively with all span measures and S scores correlated positively with the two pointing span measures. However, when individuals whose S scores were less than or equal to their P scores were studied as a subgroup, these individuals' S and P scores were shown to correlate positively with all four measures of span. The authors attributed this discrepancy in findings between the full group of participants and the subgroup to the different activation requirements of repetition versus pointing span tasks. Repetition span tasks depend more strongly on phonological information (semantics do not necessarily have to be accessed in order to repeat correctly) while pointing span tasks rely more equally on phonological and semantic activation (because of the word-to-picture matching step required in a pointing task), an observation that may explain why the correlation between semantic processing and repetition span was not borne out in the full group of participants. Individuals with particularly weak semantic processing (relative to phonological processing) may be strongly affected by the lack of extra support that intact semantic activation typically provides when performing repetition tasks, and thus, their relatively weak S scores correlate positively with repetition span tasks. Together, these results suggest that the same phonological and semantic networks that are activated during receptive language tasks may also be activated during short-term memory tasks, suggesting that processes governing receptive language and short-term memory tasks may be at least partly shared.

The correlational results summarized above concern language comprehension tasks, which have been more frequently used to study associations between language and aphasia than language production tasks. Using production tasks to classify language production impairment type is more challenging because production ability is arguably more difficult to quantify in relation to locus of impairment. N. Martin and Ayala (2004), however, have studied the relationship between VSTM and language production impairment type by correlating numbers of phonologically related nonword errors in picture naming and numbers of semantic errors in picture naming with verbal span tasks, and showed some notable associations. For individuals who were relatively more impaired in semantic input (i.e., comprehension) processing, repetition span tasks correlated negatively with # of phonological errors and all span tasks correlated negatively with # of semantic errors. For individuals who were relatively more impaired in phonological input processing, # of phonological errors correlated negatively with both repetition span measures and with the word pointing span task, and no significant correlations were found with # of semantic errors. The lack of a correlation between semantic errors and span tasks in the group of individuals who were more impaired on phonological input processing tasks relative to semantic input processing tasks might suggest that even if semantic processing appears strong (as defined by low numbers of semantic errors), the available semantic support may not be enough to overcome weak phonological-level activation during span tasks. Overall, these results, like the ones for the receptive tasks described above, suggest that language processing tasks (both comprehension and production) and short-term memory span tasks may rely on a common temporary storage process. Results from a related study, which found that two measures of picture naming accuracy (percentage of correct responses and percentage of omission, or no response, errors) correlated significantly with both digit and word pointing span, provide further support for a relationship between STM and word retrieval (N. Martin & Gupta, 2004).

Collectively, these studies suggest that tasks classically labeled as language tasks (e.g., picture naming, word-to-picture matching, phoneme discrimination, etc.) and tasks classically labeled as STM tasks (i.e., repetition and pointing spans) may rely on shared processes that subserve the temporary storage of both lexical-semantic and phonological information, and that language impairments in aphasia may be partly due to breakdowns of these shared processes. The involvement of lexical-semantic processes in STM is further supported by studies that demonstrate better span performance (i.e., greater span length) for words over nonwords (N. Martin, Dell, Saffran, & Schwartz, 1994; N. Martin, et al., 1996), high frequency words over low frequency words (N. Martin & Gupta, 2004; N. Martin & Saffran, 1990), and high imageability words over low imageability words (N. Martin, et al., 1994; N. Martin & Gupta, 2004; N. Martin, et al., 1996; R. C. Martin, et al., 1999) in individuals with aphasia.

While studies that use span length to investigate STM certainly have merit, the conclusions that can be made about the nature of the STM-language relationship from a demonstration of significant positive correlations between immediate span length and scores on classic language tasks or significantly worse performance on immediate span length measures in individuals with aphasia than in individuals without aphasia are limited. In terms of correlational research between span length and accuracy on language measures, all that can be concluded from the significant positive correlations that have been shown is that the greater the span, the greater the language task accuracy. These results do not tell us about the more precise nature of the established STM-language relationship, the *how* behind the STM-language associations (note, though, the exception of Martin and Ayala's (2004) study, which looked at subgroups of individuals based on type of language impairment (lexical-semantic versus phonological) to make conclusions about the possible processes driving the correlations between verbal span and language taks).

Furthermore, the language used in the interpretation of the relationship between span length and language processing in previous work is often imprecise and/or paradoxically suggestive of a separation between language and short-term memory despite the established correlational results. Laures-Gore et al. (2011), for example, found that individuals with aphasia performed worse on digit span tasks than right brain damaged individuals without aphasia, and hypothesized that such a result can be explained by a deficit in working memory, using vague terms such as “decreased attentional capacity,” “inefficient resource allocation,” or a “deficient phonological loop.” Though these constructs have all been implicated as possible aspects of aphasia, the absence of clear definitions of these terms and the lack of a discussion of how they might support language tasks in addition to digit span performance limits the usefulness of the results. Additionally, the authors’ use of the Baddeley model (Baddeley & Hitch, 1974) to frame these results does little to explain the findings, as this model does not make a direct link between immediate serial recall tasks and language processing beyond suggesting that both rely on sound-level information. Gvion & Friedmann (2012) similarly explain their findings that individuals with aphasia perform worse on immediate word span tasks than typical speakers through an impairment to the phonological loop, again relying on a model that does not directly link language and STM performance. Kasselimis et al. (2013), who demonstrate findings similar to Gvion & Friedmann (2012) and Laures-Gore et al. (2011) when comparing digit span performance of individuals with aphasia to left hemisphere brain damaged adults without aphasia, also use shallow terminology to interpret their results, and concludes that STM deficits are “dependent on the presense of aphasia” (p. 1776) without discussing how impaired short-term memory may impact performance on language tasks. The constraint of using span length as the sole STM measure, along with the use of vague or undefined terminology to interpret results, limit the theoretical depth of this work.

The literature review now turns to studies that look at other aspects of verbal span to attempt to understand the mechanisms driving the link between verbal span and language breakdown. These studies look closely at the nature of errors made on verbal span tasks and attempt to relate them to performance on classic language tasks (e.g., picture naming, word-to-picture matching), allowing for more precise conclusions to be made about the nature of the relationship between language processing and STM.

Beyond span length: Studies using qualitative aspects of span to assess STM in aphasia. Though studies of short-term memory in aphasia that attempt to relate immediate serial recall performance to more traditional language task performance are a valuable starting point for the exploration of processes that may underlie both types of tasks, other approaches have more precisely explored and compared the nature of their breakdown. One way this has been done is through a longitudinal study of patient NC (N. Martin, et al., 1994; N. Martin, et al., 1996), who exhibited both severe language deficits and VSTM deficits. NC’s repetition and pointing spans were limited to a single item, and he often demonstrated semantic errors during repetition tasks and formal errors during word retrieval (both patterns are atypical in the neurologically healthy system). NC’s word retrieval and repetition patterns were investigated in the context of Dell’s interactive activation model (Dell, 1986; Dell & O'Seaghdha, 1992; Dell, et al., 1997), which posits that word retrieval begins with the activation of semantic features, which in turn send activation to the target word’s lemma (which holds a word’s grammatical form) and semantically related lemmas. Activation from the selected lemma is sent to the word’s phonemes, which are selected in the second step. Activation is bidirectional, meaning that later levels of processing send feedback to earlier levels, and it is temporary, meaning that it decays rapidly unless its activation is replenished via communication with other linguistic nodes. Both the patient's initial repetition and initial naming patterns were explained by overly rapid decay of linguistic nodes. In the case of naming, overly rapid decay prevents the target lemma from holding on to the activation it receives from semantic nodes. Feedback activation from phonological nodes activates phonologically related lemmas. Because these competing lemmas are activated later than the target lemma, they are subject to less decay at the time of word retrieval, and are most likely to be selected, making formal errors very likely. In the case of repetition, the order of activation is reversed: phonological nodes are activated first, and semantic nodes are activated later. Due to the order of activation that occurs right before an auditorily presented word is repeated, phonologic nodes are subject to the most decay, while semantic nodes are more active at the time of recall. Because target semantic nodes send feedback to semantically-related lemmas, which are activated last in the input process that governs repetition, these lemmas are very likely to be selected in a system suffering from overly rapid decay of linguistic nodes.

The parsimonious attribution of repetition and naming errors to a breakdown of a single process (the process which temporarily holds linguistic representations active) suggests that both short-term memory (as tested by repetition) and language processing (as tested by word retrieval) may share a common temporary storage system. Analysis of NC's recovery provides further evidence for this claim. As the system recovered, NC demonstrated mostly semantic errors in naming and mostly formal and nonword errors in repetition, a pattern that is more consistent with that of typical speakers. The authors claimed that this occurred due to recovery from overly rapid decay: in the case of milder impairment where the decay process is functioning more effectively, nodes activated earlier (phonological nodes in repetition and semantic nodes in naming) are not as likely to be beat out by nodes activated later on, so errors that occur are more likely to be related to the earlier activated nodes, yielding semantic errors in naming and formal and nonword errors in repetition. This theory was tested with a computational simulation of Dell's interactive activation model, which was able to predict both the changes in NC's repetition and naming through changes to a single parameter, providing further evidence for the mutual support of STM and word retrieval by a shared, activation-based temporary storage system. A related study added to this idea with the finding that the interactive activation model successfully predicted the repetition of 65 individuals with aphasia from their naming performance (Dell, et al., 1997).

The studies discussed above demonstrate the importance of looking beyond span length when investigating STM and its relationship to language breakdown in aphasia. They have also shown that examining error types in repetition performance can yield valuable insights into the nature of the process that supports both STM, as measured by verbal span tasks, and language processing. Several studies have also looked at errors in repetition span from the perspective of error location (where in the repetition string the error occurred). This work is related to recency and primacy effects that are seen in typical speakers when they perform serial recall tasks (see p. 26 for details). Several case studies of individuals with aphasia have shown that the loss of recency and primacy effects in repetition may be associated with type of language processing deficit (i.e., semantic versus phonologic) (N. Martin & Saffran, 1990; Saffran, 1990), and argue that the associations between recency/primacy effects and language processing provide support for the existence of a common underlying process that supports both language and short-term memory.

This argument has been most convincingly shown through a study which linked language comprehension to repetition performance (N. Martin & Saffran, 1997). The authors assessed lexical-semantic and phonological abilities in several receptive language tasks in 15 people with aphasia (PWA) and used an immediate repetition paradigm (with word pairs and single words) to assess VSTM performance. In the context of a spreading activation model (Dell, 1986; Dell & O'Seaghdha, 1992; Dell, et al., 1997), they predicted that individuals who were more impaired in semantic input processing (relative to phonological, as determined by a scored derived from a battery of semantic and phonological input processing tasks) would demonstrate different repetition profiles than individuals who were more impaired in phonological processing. Individuals with impaired phonology and relatively preserved lexical-semantic processing should be more accurate in retrieving earlier segments, whose semantic activation has surpassed phonological activation (primacy effect). Individuals with impaired semantics and relatively preserved phonology should show the opposite effect: they should be more accurate in retrieving later segments, which have received strong phonological activation but have not built up strong semantic activation (recency effect). Additionally, individuals with impaired phonology and relatively preserved lexical-semantics should be more susceptible to manipulations of word imageability and frequency because more highly imageable and frequent words are more easily activated at the semantic and lexical levels of processing, respectively. These predictions were borne out: individuals with relatively strong semantic input processing demonstrated a bias towards remembering information at the beginning of a string and their repetition accuracy improved when highly imageable and frequent words were presented, while individuals with relatively strong phonological input processing demonstrated a bias towards remembering information at the end of a string and their repetition was less affected by manipulations of imageability and frequency. These results lent support to the idea of a common activation maintenance process that at least partly supports both STM and language through the temporary engagement of linguistic representations (See Figure 2a for a visual illustration of these results).

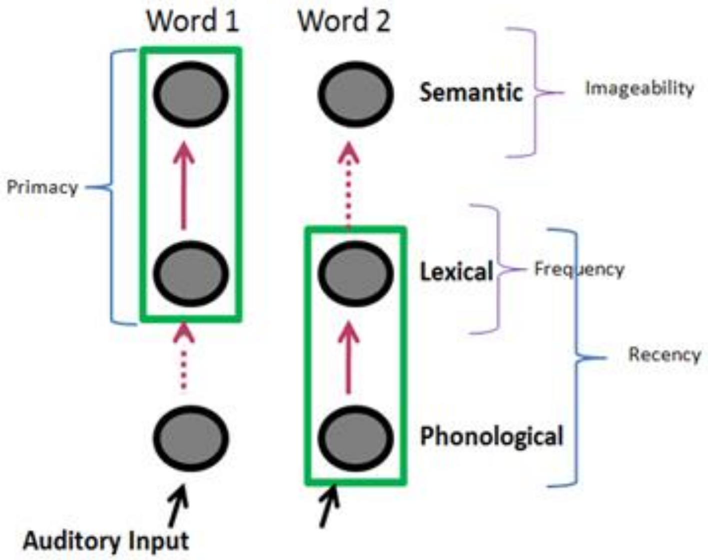


Figure 2a. Interactive activation model of verbal short-term memory. The linguistic activation pattern of an auditorily presented word pair immediately prior to repetition is shown above. The interactive activation model of language assumes that as a word pair is heard, activation travels from the phonological level up to the lexical and semantic levels. Because word 1 is spoken first, its activation has more time to spread up to the semantic level by the time the word pair is said, while its phonological activation will begin to decay by the time word 2 is spoken. Conversely, because word 2 is spoken last and there is not as much time between its presentation and recall, its activation will be strongest at the phonological level and will not have much time to spread to the semantic level. Thus, word 1 will be more strongly supported by semantic activation at time of recall (*primacy*), while word 2 will be more strongly supported by phonological activation at time of recall (*recency*). Additionally, imageability and frequency manipulations are thought to affect the first step of word retrieval (called *lemma* or *lexical* access), and thus highly imageable and frequent words are more easily activated at the semantic and lexical levels, respectively (figure and theory adapted from Martin & Saffran, 1997).

More recently, the study discussed above was replicated with a larger number (n=46) participants (N. Martin, Ayala, & Saffran, 2002). Primacy was related to semantic ability (participants with relatively strong semantic comprehension abilities were more likely to demonstrate a bias towards remembering information at the beginning of a string) and recency was related to phonological comprehension ability (participants with high phonological abilities were more likely to demonstrate a bias towards remembering information at the end of a string), though the recency correlation did not reach significance. These results provide further evidence of links between language comprehension and short-term memory; however, they also signal the need for much more exploration of this topic.

It is important to note that though the authors suggested that common underlying temporary activation mechanisms might support short-term memory, language comprehension, *and* language production, participants in the above two studies were characterized only by language comprehension tasks. Extant work addressing the relationship between language production and short-term memory is scarce, though two case studies have looked beyond span length to begun to elucidate their connection. In parallel to Martin and Saffran (1997), Wilshire, Keall, and O'Donnell (2010) characterized two participants with aphasia, who showed both impaired naming and impaired repetition spans, by type of word retrieval deficit. Patient NP demonstrated mostly semantic and other whole word substitutions on naming tasks. His word comprehension performance appeared to be relatively strong. The authors concluded that his main impairment involved the first step of word production (lemma selection). Patient TV, on the other hand, demonstrated mostly phonological nonword errors in word retrieval and relatively well preserved word comprehension, and authors concluded that his main impairment involved the second step of word retrieval (phonological selection). Patients performed serial recall tasks, and two important observations were noted: Patient NP demonstrated no evidence of a primacy effect and no susceptibility to imageability manipulations on span performance. Patient TV demonstrated no evidence of a recency effect and showed a susceptibility to imageability manipulations (repetition was more accurate with highly imageable words). These results are consistent with previous work on serial position effects in aphasia: patient NP, whose naming was more semantically impaired, demonstrated a loss of primacy and a lack of a response to a semantic manipulation (imageability), while patient TV, whose naming was more phonologically impaired, demonstrated a loss of recency and a susceptibility to imageability manipulations due to his more preserved lexical-semantic network. This work suggests that language production may have a similar relationship to STM (as measured by repetition span performance) as was earlier suggested for language comprehension.

Verhaegen, Piertot, & Poncelet (2013) further described the STM-word retrieval relationship by qualitatively comparing the performance of 2 individuals with aphasia on picture naming and VSTM tasks. The two participants differed in the types of naming errors they produced. Patient BN demonstrated frequent phonological paraphasias during picture naming and picture description tasks, and did not benefit from phonemic cues. Patient TM demonstrated frequent semantic paraphasias and circumlocutions, and demonstrated increased accuracy when phonemic cues were provided. Based on these observations the authors concluded that patient BN's locus of breakdown was at the phonological level, while patient TM's locus of breakdown was at the lexical-semantic level. The patients also demonstrated divergent VSTM profiles. Both patients completed two VSTM tasks, one in which they heard a list of items and then had to decide whether a probe word rhymed with any word in the list (rhyme probe task), and one in which they heard a list of items and had to decide whether a probe word belonged in the same semantic category as any word in the list (category probe task). BN was impaired on the rhyme probe task but performed within normal limits on the category probe task. Conversely, TN was impaired on the category probe tasks but performed within normal limits on the rhyme probe task. Thus, the participants' VSTM impairments were consistent with their word production error profiles: BN demonstrated breakdowns at the phonological level on both VSTM and word retrieval tasks, while TM demonstrated breakdowns at the lexical-semantic level on both VSTM and word retrieval tasks. These results suggest that VSTM and word retrieval may be supported by a shared underlying process. Though small in scale, these findings, along with those of Wilshire et al. (2010), warrant further research in order to better understand the possible interdependence of language production and short-term memory on a shared temporary storage system, and to better understand the nature of this system if it exists.

The specificity of short-term memory in language processing. The studies discuss above deal with a specific type of STM, the temporary activation of *verbal* representations. Other studies have looked at performance of individuals with aphasia on *nonverbal* short-term memory measures to determine whether a modality dependent or independent temporary storage mechanism supports language processing. The theories that frame the present work (i.e., Martin and colleagues' interactive activation model of verbal short-term memory, Cowan's Embedded Processes model, and the *Neural Multifunctionality* framework (Martin & Saffran, 1997; Cowan, 1988, 1999; Cahana-Amitay & Albert, 2015, respectively), are consistent with a language system that is supported by a modality-specific (i.e., verbal) STM, because all three theories argue that short-term memory is a property of the cognitive system whose function it supports. In extant literature, conflicting views arise on this matter. The most typical nonverbal span task studied in individuals with aphasia is the Corsi Block Span Task (De Renzi & Nichelli, 1975). This span task requires individuals to point to the same sequences of blocks to which the experimenter points, and sequence length is gradually increased to determine an individual's nonverbal span. While individuals with aphasia demonstrate reduced spatial spans in comparison to controls (N. Martin & Ayala, 2004) and left hemisphere brain damaged adults without aphasia (Kasselimis, et al., 2013), individuals generally demonstrate much stronger performance on spatial span tasks relative to verbal span tasks (see Martin & Ayala, 2004, p. 472, for a discussion of this issue). Additionally, though a previous study used the finding that the Corsi Block Span Task correlates with classic tests of language impairment in aphasia to support a modality-independent view of STM, the concurrent finding that digit span length, a measure of *verbal* short-term memory, also correlates with language impairments but *not* with the Corsi Block Span Task, proves to be problematic for this argument (Potagas, et al., 2011). Additionally, though Martin & Ayala (2004) found that repetition digit span and word span tasks were found to correlate positively with the Corsi Block Span Task, these correlations were not borne out for pointing versions of the digit and word span tasks, arguably purer measures of VSTM. Together, these results signal the need for additional research to determine the extent to which language processing is supported by a linguistically-specified short-term memory system.

Specific Aims

The goal of this study is to expand on the aforementioned work by investigating whether VSTM is related to language production (specifically, word retrieval impairment type) in individuals with aphasia. As described in the review of extant work above, traditional measures of verbal span length (both repetition and pointing varieties) correlate with both language comprehension and word retrieval tasks. Additionally, word retrieval and VSTM, as measured by repetition span performance, seem to recover together, and both word retrieval and repetition have been computationally modeled successfully using the same parameters. Perhaps the most compelling link between VSTM and language, though, has been shown through correlations between serial position (primacy and recency) effects in repetition and language comprehension impairment type (i.e., lexical-semantic versus phonological). Additionally, links between susceptibility to linguistic manipulations (i.e., frequency and imageability) in repetition and language comprehension impairment type have demonstrated strong ties between VSTM and language processing. The present study sets out to expand this work by investigating the relationship between VSTM and word retrieval impairment type. To that end, the aims are as follows:

Specific Aim 1: Investigate whether word retrieval is supported by a linguistically-specified STM system by determining whether 1) word retrieval impairment severity is associated with word pointing span length and 2) whether word pointing span length is associated with nonverbal (i.e., spatial) span length. *Predictions*: Word pointing span length will be positively and significantly correlated with word retrieval severity in individuals with aphasia and will not be significantly correlated with a nonverbal spatial STM task. This prediction is rooted in the idea that STM is a property of the cognitive system it supports, so a STM task that relies on temporary storage of verbal information (i.e., VSTM) should be associated with a language production task but not a task which does not require linguistic storage. This first aim will set the stage for more intricate analyses (specific aims 2 and 3), which examine the nature of VSTM as it relates to word retrieval impairment type.

Specific Aim 2: Determine whether type of word retrieval impairment (lexical-semantic versus phonological) is associated with VSTM performance as measured by primacy/recency effects observed in a word pair repetition task. *Predictions*: Individuals who produce relatively more semantic than phonological nonword naming errors will demonstrate a recency effect, while individuals who produce relatively more phonological nonword naming errors than semantic errors will demonstrate a primacy effect. This is predicted because short-term memory of the first word (primacy) has been associated with semantic activation and short-term memory of the second word (recency) has been associated with phonological activation. See Figure 2b for a visual illustration of these predictions.

Specific Aim 3: Determine whether two psycholinguistic manipulations known to affect language processing will differentially affect VSTM performance in individuals with different types of word retrieval impairments (lexical-semantic versus phonological). This aim will explore the relationship between word retrieval impairment and word repetition performance in conditions that manipulate word frequency and imageability. *Predictions:* Individuals who produce more phonological than semantic naming errors will demonstrate more accurate repetition on both high imageability and high frequency words than on low imageability and low frequency words, while individuals who produce more semantic than phonological naming errors will not be as susceptible to these manipulations. This is predicted because imageability and frequency manipulations have been shown to affect activation at the first stage of word retrieval, which largely involves semantic and lexical processing. Therefore, individuals with relatively stronger semantics (i.e., relatively more phonological errors than semantic errors) should be more susceptible to these manipulations, while individuals with relatively stronger phonological processing should not be as susceptible. See Figure 2b for a visual illustration of these predictions.

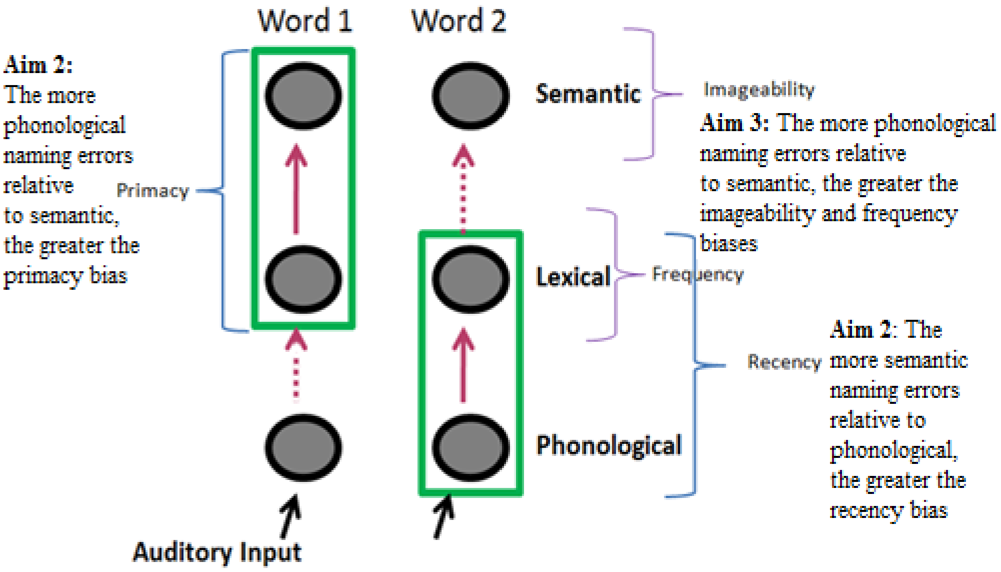


Figure 2b. Interactive activation model of verbal short-term memory with specific aim 2 and 3 predictions.

Primacy bias is the tendency to correctly repeat more first words than the second words. *Recency bias* is the tendency to correctly repeat more second words than first words. *Imageability bias* is the tendency to correctly repeat more high imageability words than low imageability words. *Frequency bias* is the tendency to correctly repeat more high frequency words than low frequency words.

Chapter 2: Method

This project was approved by the University of Washington Institutional Review Board (IRB #48385), and informed consent was obtained from each participant.

Participants: Number and Inclusion/Exclusion

Of the 26 individuals with aphasia who were recruited for this study, 24 individuals with aphasia following left hemisphere damage due to a cerebrovascular accident, confirmed through diagnostic imaging report, met all study criteria (see Figure 3 for illustration of protocol steps, inclusionary criteria, and reasons for exclusion). The target number of participants (n = 24) for this study was derived from a power analysis based on Pearson correlation coefficient values taken from N. Martin and Saffran (1997). The power analysis was conducted via an online statistical calculator (A. Chang, 2015). Because the aforementioned study used one-tailed tests of significance while the current study used more conservative two-tailed tests, the alpha level was set at .025 for the power analysis. Power was set at 80%. The correlation coefficient was set at .54 based on the correlation between primacy (tendency to preserve the first item in a repetition string) and receptive language impairment type found by N. Martin & Saffran (1997), p. 666. The decision to use the correlation coefficient value found between primacy and receptive language impairment type was based on conservativeness of the sample size estimation, as the correlation coefficients found by Martin & Saffran (1997) for imageability and receptive language impairment type were generally larger, meaning that a smaller sample size would have been derived.

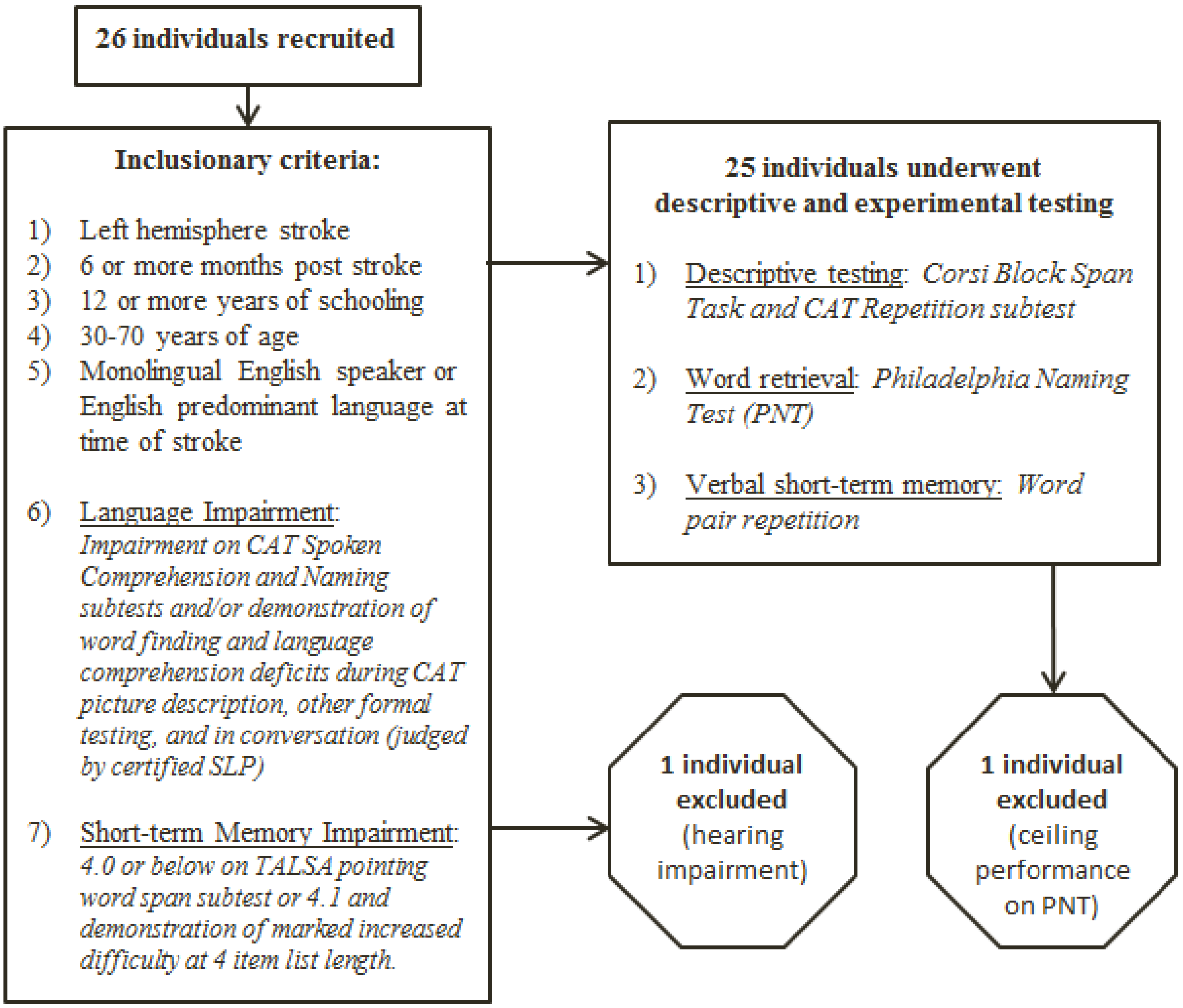


Figure 3. Flow chart illustrating protocol, inclusionary criteria, and excluded participants.

Inclusionary criteria (general). All participants had left hemisphere damage due to at least one cerebrovascular accident (individuals with multiple left hemisphere cerebrovascular accidents were included). The neurologic injury was documented by a CT or MRI scan and interpreted by a radiologist or behavioral neurologist. All participants were at least 6 months post their most recent cerebrovascular accident. All participants had to have at least 12 years of schooling (high school diploma) and had to be between the ages of 30 and 70 to be included. All participants were required to be monolingual English speakers *or* to have used English as their predominant language at the time of their stroke. English language predominance was confirmed as follows: multilingual individuals were asked about the age of acquisition of English and which language they currently use most of the time. English language predominance was confirmed by a short interview with an individual who has known the participant for at least a year prior to their cerebrovascular accident. Though premorbid right handedness was a preferred characteristic, left handed individuals who demonstrated language profiles consistent with a diagnosis of aphasia were judged for inclusion in the manner described above.

Inclusionary criteria (language and short-term memory). To confirm presence of aphasia, participants had to demonstrate holistic language deficits, consistent with a process-driven definition of aphasia that assumes that both comprehension and production of language are disrupted in aphasia. All participants demonstrated presence of language production and comprehension difficulties as measured by the Comprehensive Aphasia Test (CAT; Swinburn, Porter, & Howard, 2004), confirmed through two steps. First, two modality subscores of the CAT, Comprehension of Spoken Language and Naming, were used. Individuals who scored at or below 56 out of 66 total points on the CAT Comprehension of Spoken Language subtest and at or below 69 on the CAT Naming subtest were classified as aphasic and automatically met the language criteria for the study. Second, individuals who fit the neurological criteria for inclusion and did not meet one or both of the standard CAT cutoff scores listed above but demonstrated marked word finding and language comprehension difficulties in conversation, during discourse testing (as measured by the CAT Picture Description), and during the full language testing battery (see details of descriptive testing, below) were also included in the study. Presence of aphasia in the latter group of individuals (i.e., those who did not fall below both cutoff scores on the CAT) was confirmed by clinical judgment of a certified speech language-pathologist at the

University of Washington Aphasia Laboratory who took into account both video and audio recordings of the formal language evaluation as well as discourse characteristics during informal interactions with participants. Of the 24 participants who completed the study, 2 demonstrated Naming scores slightly above the cutoff, 4 demonstrated Comprehension of Spoken Language scores above the cutoff, and 1 demonstrated both Naming and Comprehension of Spoken Language scores above the cutoff. All of these participants demonstrated marked anomia and slow, halting, and effortful speech in conversation and during the CAT picture description test, and slowed processing on formal and informal language comprehension tasks.

In addition to demonstrating aphasia according to methods described above, participants had to demonstrate impaired verbal short-term memory as measured by the word pointing span subtest of the Temple Assessment of Language and Short-Term Memory in Aphasia (TALSA; N. Martin, 2012). The pointing span task was selected over a repetition version of the task because it is arguably a somewhat purer measure of verbal short-term memory (i.e., temporary activation) impairment in that it does not require a verbal response. In this task, a closed set of nine words was used. Participants heard a list of words and were required to point to the pictures corresponding to the words in the order in which they heard them. Participants performed 10 trials at each list length until they achieved an accuracy level below 50% at a given list length. Span length was calculated by the following formula: list length at which at least 50% of lists were reproduced correctly + .50 of the proportion of lists recalled at the next list length (Shelton, Martin, & Yaffee, 1992). Pointing span was judged to be impaired using two steps. First, individuals whose spans were calculated at 4.0 or below automatically met the criteria for verbal short-term memory impairment and were included in the study. Next, participants who achieved a score of no more than 4.1 (i.e., no more than 2/10 items correct on the 5 item list length) *and* a marked increase in difficulty at the 4 item list length (e.g., observed through verbal report of increased effort/fatigue, significantly slowed response time, and/or a significant increase in the omission of list items in their responses) were also included in the study.

Exclusionary criteria. Individuals were excluded if they had a history of right hemisphere cerebrovascular accident, degenerative neurological illnesses (e.g., dementia or Parkinson's disease), and/or currently untreated depression or other psychiatric illness. Raven's Progressive Matrices (Raven, Raven, & Court, 1998) was used as a general measure of non-linguistic processing impairments, and individuals with a score below 23 were excluded. Individuals with a visual impairment (as defined by failure on line bisection screening and/or failure to read the second-to-bottom line of the Tumbling E eye chart at a distance of 20 feet) were excluded (D. F. Chang, 1995). Individuals with a hearing impairment (as defined by failure to pass an audiometric pure-tone, air conduction screening at HL 35 dB at 500, 1K, and 2K Hz for at least one ear) were also excluded. Of the 26 recruited individuals, one individual was excluded due to a hearing impairment.

Because the experimental tasks of the present study depended on verbal output (repetition and verbal picture naming), it was critical to determine whether each potential participant had an apraxia of speech in addition to their language impairment. Subjects were excluded if they demonstrated a severe apraxia of speech, as judged by 3 certified speech-language pathologists' assessments of videos of participants' performance on repetition, picture description, and other tasks from the Comprehensive Aphasia Test. The videos were evaluated by 3 certified speech-language pathologists using consensus judgments for the following behaviors: slow rate, prolonged segment durations and intersegment durations (including intrusive schwa), distortions, and prosodic abnormalities. To further minimize the influence of apraxia of speech, productions (i.e., repetition and word retrieval) of the 4 participants judged to have apraxia of speech who passed the aforementioned criteria and were included in the study followed additional strict, conservative coding guidelines to reduce the possibility that motoric errors were incorrectly classified as aphasic errors (for details, see p. 56).

Descriptive Testing

In order to further describe the nature and severity of the participants' language impairments, the Repetition subtest of the CAT was administered in addition to the previously mentioned subtests (i.e., Language Comprehension, Naming, and Picture Description).

Participant Characteristics Summary

The 24 participants who completed the study were on average 61.83 years of age (SD = 8.04), had 15.02 years of education (SD = 2.37), and were 64.09 months post cerebrovascular accident onset (SD = 32.11). All but three participants were premorbidly right handed, and one of the three left handed participants reported relative ambidexterity, though this was not confirmed through formal testing. Individual demographic characteristics and standardized test scores are reported in Tables 1 and 2, respectively.

Table 1. *Participant Demographic Characteristics*

ID	**Sex**	**Age**	**MPO**	**Lang 1**	**Hand**	**Edu**	**Etiology**	**AOS characteristics**
1	M	70	107	Eng	L	19	Remote left temporofrontal infarct w/ associated encephalomalacia	None
2	F	64	120	Eng	R	18	Left frontal, temporal, parietal infarct, prior to aneurysm clipping of left MCA (Hemorrhage, craniotomy x 2).	None
3	M	54	74	Eng	R	16	Small left cortical lateral frontal infarct	None
4	M	68	36	Eng	R	12	Left MCA infarct. Temporal lobe, insula, frontal, and parietal lobes affected.	None
5	F	71	125	Eng	R	14	Left MCA infarct involving basal ganglia and cortical grey matter	None
6	F	60	70	Eng	R	12	Large left infarct with involvement of temporal pole and frontal lobe (extending from orbitofrontal cortex to Broca's area and to superior frontal sulcus).	None
7	M	61	77	Eng	R	16	Left hemorrhage with involvement of basal ganglia, posterior frontal, temporal, and parietal lobes.	None
8	F	45	10	Eng	R	16	Left MCA infarct involving insula, operculum, modest extent of superior temporal gyrus, and extensive frontal cortex	None
9	F	54	54	Eng	R	14	Left MCA aneurysmal subarachnoid hemorrhage	Moderate. Slowed rate, segmented speech, increased difficulty on consonant clusters; presence of schwas and distorted substitutions
10	M	66	60	Eng	R	16	Left MCA involving anterior left parietal lobe extending to the level of the sylvan fissure with question of component of extension into the left temporal lobe. Mild effacement upon left lateral ventricle.	None
11	F	59	51	Eng	R	20	Decompressive craniotomy after large left CVA. Left internal carotid occlusion (likely a very large ischemic infarct s/p decompressive craniotomy).	None
12	M	72	64	Eng	R	23	Left hemorrhagic CVA with involvement of basal ganglia, adjacent insular cortex, left corona radiata, with ischemic damage to frontal lobe	Mild. Slightly abnormal rate; mild disruption to prosody and articulation
13	F	51	69	Gujarati	R	12	Left basal ganglia hemorrhagic infarct, with ischemic damage to overlying perisylvian cortex	None

14	M	61	116	Eng	R	19	Left MCA infarct, large territory hypodensity involving the left cerebral hemisphere in the frontal, temporal, and parietal lobes.	Moderate. Distorted substitutions, errors increase with increasing word length, abnormal prosody
15	F	71	41	Eng	R	14	Surgery to "tie off" aneurysm led to bleeding that required a 2nd surgery. Distal left M1 embolic event. Infarct involves part of putamen, insula, operculum (though minimal frontal), and extensive temporal and parietal lobe.	None
16	F	70	85	Eng	R	12	Hx of two strokes (or one stroke and 1 TIA). Left MCA territory infarcts including the temporoparietal junction, coronal radiata, and subinsular region (ischemic).	None
17	M	62	42	Eng	R	13	Distal L M1 event with extensive infarction of temporal lobe and insula and modest infarction of frontal and parietal operculum.	None
18	M	57	128	Eng	L	18	Large left MCA infarct with involvement in head of the caudate, most of left lenticular nucleus, the insular cortex, and posterior half of the left frontal lobe extending into the anterior portion of the left parietal lobe.	None
19	F	64	42	Eng	R	17	Large left basal ganglia hemorrhage with intraventricular extension.	None
20	M	60	32	Eng	R	16	Large left basal ganglia hemorrhage	None
21	M	70	59	Eng	R	16	Suggestion of subtle loss of gray-white matter differentiation in the left basal ganglia. Occlusive thrombosis within proximal aspect of left middle CVA.	None
22	F	72	34	Eng	R	16	Left M2 region infarct with loss of left insular ribbon, loss of gray white differentiation in the left frontal operculum and left parietotemporal lobes (ischemic).	Mild. Distorted substitutions, segmented syllables, increase in errors with increasing word length, slowed speech rate
23	F	56	17	Eng	R	12	Left MCA ischemic event. Infarct extending deep from lateral left frontal pole.	None
24	M	46	86	Eng	L	13	Massive left MCA infarct with subsequent surgical decompression. Left hemisphere is gone, expect for ACA and PCA territories.	None
AVE		61.83	66.63			15.56		
SD		8.04	33.77			2.95		

MPO = months post onset; Lang1 = native language; Hand = handedness; Edu = education; AOS = apraxia of speech

Table 2. *Inclusionary and Descriptive Short-Term Memory and Language Test Scores*

				CAT Naming				***CAT Spoken Comprehension***				***CAT Repetition***						***CAT PD***
ID	**Word Span**	**Corsi Block**	**RPM**	**Fluency**	**Obj.**	**Act.**	**Total**	**Words**	**Sent.**	**Par.**	**Total**	**Words**	**CW**	**NW**	**DS**	**Sent.**	**Total**	**Total**
01	2.1	4	35	8	16	1	25	28	18	4	50	25	2	2	6	0	35	6
02	2	5	32	18	34	0	52	26	17	4	47	32	4	8	6	6	56	13.5
03	2.15	5	33	19	44	4	67	27	28	4	59	31	6	4	12	8	61	49
04	3.2	4	29	23	44	5	72	26	26	3	55	32	6	6	10	10	64	51
05	2.2	4	33	19	37	3	59	29	23	3	55	24	4	0	8	6	42	26
06	1.15	5	29	13	21	0	34	21	9	0	30	26	3	5	6	6	46	19.5
07	3.1	3	32	20	43	7	70	23	26	3	52	31	6	10	10	8	65	44
08	4	5	36	11	36	10	57	30	27	4	61	32	6	10	8	12	68	27
09	2.1	4	29	13	35	6	54	26	11	4	41	30	6	9	8	6	59	10
10	1.1	4	30	0	0	1	1	17	20	1	38	19	0	5	2	0	26	16
11	2.1	4	35	8	40	2	50	23	13	13	49	28	6	8	8	6	56	17.5
12	3.05	5	33	16	46	10	72	29	30	4	63	28	6	9	10	12	65	23
13	4.05	4	33	20	39	2	61	26	25	2	53	32	6	10	10	12	70	32.5
14	2.1	5	33	14	34	2	50	29	20	4	53	20	0	0	8	6	34	7.5
15	3.05	5	34	20	40	9	69	27	25	4	56	30	4	6	8	6	54	53
16	1	3	23	3	3	0	6	22	16	3	41	27	1	2	4	6	40	11
17	2.05	4	35	11	32	0	43	24	25	4	53	28	6	2	4	6	46	15.5
18	2	3	25	4	27	1	32	22	19	3	44	32	4	4	6	6	52	6
19	4	3	30	7	36	10	53	29	28	2	59	32	6	10	12	12	72	64.5
20	2	5	32	8	5	1	14	26	27	3	56	14	0	0	4	0	18	3.5
21	3.2	5	32	7	33	0	40	26	27	3	56	32	6	10	10	12	70	35
22	4	3	34	13	45	9	67	28	30	4	62	32	6	8	10	12	68	30.5
23	2.2	4	27	4	42	8	54	29	24	2	55	32	6	10	6	8	62	26
24	1.05	3	31	2	12	0	14	22	16	0	38	30	4	8	0	6	48	2
AVE	2.46	4.13	31.46	11.71	31	3.79	46.50	25.63	22.08	3.38	51.08	28.29	4.33	6.08	7.33	7.17	53.21	24.56
SD	0.96	0.80	3.23	6.66	13.98	3.80	21.46	3.25	6.04	2.39	8.54	4.86	2.20	3.59	3.05	3.73	14.82	17.42

RPM = Raven's Progressive Matrices; Obj. = Object; Act. = Action; Sent. = Sentences; Par. = Paragraphs; CW = Complex Words; NW = Nonwords; DS = Digit Strings Sent. = Sentences; PD = Picture Description ● CAT cut-off scores: Naming ≤ 69; Spoken Comprehension ≤ 56; Repetition ≤ 67; Picture Description ≤ 3

Classification of Word Retrieval Impairment

Stimuli. In order to determine overall word retrieval accuracy and quantify the relative contributions of semantic and phonological breakdown to each participant's word retrieval impairment, the Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) was administered to all participants. The PNT is a computer-based confrontation naming test comprised of 175 high-, medium-, and low-frequency nouns ranging from 1 to 4 syllables, and provides ample error opportunities and strict, thorough error coding guidelines based on Dell's spreading activation model of word retrieval (Dell, et al., 1997). The PNT has been successfully used with individuals with aphasia to elicit the same error types as those of interest in this proposal (N. Martin & Ayala, 2004).

Administration. The PNT was presented on a computer via Microsoft PowerPoint 2010. The participants saw black and white line drawings representing each word. Participants saw the following instructions on the screen prior to beginning the task: *I am going to ask you to name some pictures. When you hear a beep, a picture will appear on the computer screen.* ***Your job is to name the picture as soon as you see it.*** *Please use only one word. We'll practice several pictures before we begin*. The participants were asked to perform at least 2 practice trials prior to the administration of the test to assure understanding of task requirements. Once the test commenced, a timeframe of 30 seconds was permitted to name each picture. Verbal productions of participants' responses were audiorecorded for subsequent analysis with an Olympus Digital Voice Recorder, model VN-722PC, for subsequent analysis.

Scoring. For each trial, the first complete attempt, as defined by PNT scoring instructions, was coded as correct or incorect by the lead researcher or a trained reseach assistant competent in broad phonetic transcription. Incorrect responses were coded for error type according to the scoring guidelines of the PNT (see Appendix I for guidelines for determining first complete attempt and error codes). The main errors of interest were semantic errors and phonologically-related nonword errors, though formal and mixed errors were also coded and tracked. All other errors were coded in the "other" category. Distorted phonemes were considered to be correct. To minimize the possibility of apraxic errors being mistakenly coded as phonologic nonword errors in the 4 participants identified as apraxic (see Table 1, participant IDs 9, 12, 14, and 22), the following distorted substitutions were scorred as correct: substitutions of voiced consonants for unvoiced, and vice versa, if place and manner were preserved, and substitutions of stop consonants for nasal consonants, and vice versa, if place and voicing were preserved. This decision was made based on literature that argues that these errors are more likely to be attributed to AOS than phonologic paraphasias (Itoh, Sasanuma, & Ushijima, 1979; Mauszycki, Dromey, & Wambaugh, 2007; Odell, McNeil, Rosenbek, & Hunter, 1990), as well as the observation that 2 of the individuals with AOS (see Table 1, participant IDs 9 and 14) consistently distorted these types of substitutions, thereby providing evidence that these utterance were motor-based. Additionally, schwa insertions were scorred as correct, due to evidence suggesting that this error type is motor-based (e.g., Ballard, Granier, & Robin, 2000).

In order to classify each participant's naming impairment, the total numbers of semantic errors and phonologically-related nonword errors (called S and P output scores, respectively), as defined by the PNT error coding guidelines, were calculated. A P-S output index was calculated as follows: total # of phonologically related nonword errors – total # of semantic errors. To establish intra- and inter-rater reliability for error type coding, 15% of each participant's responses were randomly selected and coded by the same research assistant and another research assistant competent in broad phonetic transcription, respectively. Cohen's Kappa (Landis & Koch, 1977) was calculated on the coded responses to quantify reliability.

Of the 25 individuals who met inclusionary criteria for the study, one additional individual was excluded due to ceiling performance on the PNT.

Classification of Verbal Short-Term Memory Impairment

For research question 1, the word pointing span subtest of the TALSA (N. Martin, 2012) was administered to the participants and span length was calculated using the methods described on p. 49. For research questions 2-3, the more involved method for classifying VSTM impairments is described below.

Stimuli. In order to quantify the presence and nature of VSTM impairment, two-hundred forty 2- and 3-syllable nouns varied for frequency and imageability were selected. From this corpus of words, four lists of 60 individual words were created: high frequency-high imageability words, low frequency-high imageability words, high frequency-low imageability words, and low frequency-low imageability words (Figure 4). Words were deemed high frequency if they occurred more than 40 times per million. Words were deemed low frequency if the occurred less than 25 times per million (based on ratings from SUBTLEX-US database, Brysbaert & New, 2009). Words were deemed high imageability if the imageability rating was greater than 497, and low imageability if the imageability rating was less than 497 (based on ratings from MRC psycholinguistic database, Coltheart, 1981). Stimulus numbers and construction details were based on N. Martin & Saffran (1997) and N. Martin et al. (2002). However, different frequency and imageability databases were used to select the words used for the verbal short-term memory task in this study. SUBTLEX-US was used to determine frequency measures because it is a database of verbal rather than written frequencies, so it more closely matches the experimental task than the Kucera & Francis (1967) database used by N. Martin and Saffran (1997). The MRC Psycholinguistic Database was used to determine imageability, as it is more recent and includes a larger set of items than the Paivio, Yuille, and Madigan (1968) database used by Martin and Saffran (1997).

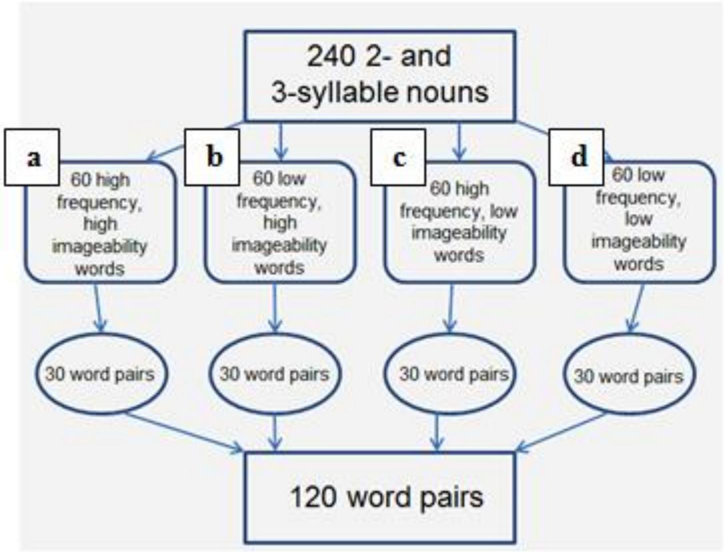


Figure 4. Word pair repetition stimuli construction.

Since frequency and imageability manipulations are critical to research question 3, t-tests were conducted to assure that the subsets of high imageability and low imageability words (Figure 4, lists a+b and c+d, respectively) did not differ on frequency, $t(238) = .031$, $p = .976$, alpha = .05, and that subsets of high frequency and low frequency words (Figure 4, lists a+c and b+d, respectively) did not differ on imageability, $t(238) = .469$, $p = .640$. Additionally, phonotactic probability ratings for the selected words were obtained from the Irvine Phonotactic Online Dictionary, version 1.4 (Vaden, Halpin, & Hickok, 2009). The high and low imageability words did not differ on phonotactic probability $t(237) = -1.89$, $p = .060$, alpha = .05, the high and low frequency words did not differ on phonotactic probability $t(237) = .902$, $p = .368$, alpha =

.05, and each word pair was matched for phonotactic probability as closely as possible.

From the four lists of 60 total words, 30 word pairs were created, for a total of 120 word pairs. Words in each pair were matched for syllable length. Words in each pair did not have a clear semantic relationship. Productions of each of these word pairs were audio recorded by a native male speaker of English with a Marantz Professional Solid State Recorder, model PMD671. Each word in each word pair was recorded separately before splicing the recordings together to minimize differences in intonation between the first and second word. To ensure intelligibility of the recordings, 3 native speakers of English were asked to listen to the recordings and transcribe them. Recordings that were not correctly transcribed by all 3 raters were re-recorded and re-transcribed until 100% accuracy among the 3 raters was reached.

Administration. Stimuli were administered in two sets (set A and set B) of 120 word pairs, with a short break in between the two sets. The two sets contained the same pairs of items, but pairs were presented in reverse order. The order of the two sets was randomized for each participant, and the order of the word pairs was randomized within each set. Half of the participants received set A first, and half of the participants received set B first. Words in each pair were presented at a rate of 1 word per second. Participants were instructed to repeat each word pair immediately after they heard it in the order in which they heard it. Participant's responses were audio recorded with an Olympus Digital Voice Recorder, model VN-722PC for subsequent analysis.

Scoring. The audio recordings were analyzed by the lead researcher or a trained reseach assistant competent in broad phonetic transcription. The first and second word in each set were scored for whole word accuracy (correct/incorrect). Only the first repetition attempt was scored. Individuals were required to initiate repetition within 5 seconds of hearing each word pair in order to get credit for a given trial. For a word to be considered correct, all phonemes had to be correctly produced in the correct positions, and the word had to be produced in the correct serial position. In the event that participants only produced the second word in the two-word string, the word was assumed to occur in the correct serial position, as there was ample behavioral evidence that individuals who did this frequently were aware that they were omitting the first word. Distorted phonemes were considered to be correct, and the modified scoring rules for individuals with AOS described in the previous section (p. 56) were also used in word pair repetition scoring.

A primacy bias score was calculated for each participant as follows: # first words correct/(# first + second words correct). An imageability bias score was calculated as follows: # of high imageability words correct/total words correct. To calculate imageability bias and frequency bias scores for each participant, words were collapsed across serial position. A frequency bias score was calculated as follows: # of high frequency words correct/total words correct. Consistent with N. Martin & Saffran (1997), a recency bias score was not calculated, as the correlation between recency bias score and the output language bias score (P-S) would be the reciprocal of the correlation with primacy bias score.

Classification of Nonverbal Short-Term Memory Impairment

In order to examine the relationship (or lack thereof) between verbal and nonverbal STM, the Corsi Block Span Task (De Renzi & Nichelli, 1975), a spatial span task, was administered. During the task, subjects were seated across from the experimenter, who pointed to a sequence of nine wooden blocks laid out in an asymmetric configuration mirroring the setup of DeRenzi & Nichelli (1975). The asymmetric set-up helped to prevent linguistic coding (e.g., subvocalizing “top right corner” as the experimenter pointed to a block). Sequences began with one and advanced to greater span lengths until the participant missed two consecutive trials on a given list length. Each list length had six total trials. Nonverbal span length score was defined as one less than the length at which two consecutive trials were missed.

Data Analysis

Table 3 outlines the research questions, tasks, outcome measures, statistical analyses, and predictions.To answer research question 1, word pointing span length was correlated with PNT accuracy and Corsi Block (nonverbal) span length. To answer research question 2, P-S scores were correlated with primacy to determine the relationship between type of word retrieval impairment and verbal short-term memory. To answer research question 3, P-S scores were correlated with imageability and frequency bias scores. Effect sizes for these correlations were determined as follows: r = .1: small, r=.3: medium, r=.5: large (Cohen, 1988). Because these main analyses were based on a priori hypotheses and motivated by extant theoretical and empirical work, the alpha value for each of these three analyses was set at .05 (2 tailed test).

Table 3. *Summary of Research Questions, Methods, and Predictions*

Research Question	Task	Outcome Measure	Statistical Analysis	Predictions
1. Is word retrieval accuracy associated with a VSTM task in the absence of a correlation between linguistic STM and a nonlinguistic (i.e., spatial) STM task?	*Word retrieval task*: Naming of 175 1-4 syllable nouns (PNT) *Verbal short-term memory task*: Word pointing span (TALSA) *Spatial short-term memory task*: Corsi Block Span Task	*Word Retrieval*: # of correctly named pictures *Verbal short-term memory*: Span length on TALSA word pointing span, calculates as follows: list length at which 50% of lists correctly repeated + .5 of the proportion of lists recalled at the next list length *Spatial short-term memory*: Span length on Corsi Block Span Task, defined as one less than the length at which 2 consecutive trials were missed	Word retrieval accuracy will be correlated (Pearson) with VSTM span length VSTM span length will be correlated with Corsi Block span length	Word retrieval accuracy scores will positively correlate with VSTM span length. VSTM span will not significantly correlate with Corsi Block span length.
2. Is type of word retrieval impairment associated with primacy/recency effects observed in a VSTM task?	*Word retrieval task*: Naming of 175 1-4 syllable nouns (PNT) *Verbal short-term memory task:* Repetition of two- and three-syllable word pairs	*Word retrieval:* # of semantic errors (S score), # of phonological nonword errors (P score) and relative score reflecting these error types (P-S) *Verbal short-term memory*: Primacy: # first words correct/# total words correct	The 3 word retrieval measures will be correlated (Pearson) with primacy.	P-S scores and P scores will positively correlate with primacy. S scores will negatively correlate with primacy.
3. Is type of word retrieval impairment associated with imageability and frequency effects observed in a VSTM task?	*Word retrieval task*: Naming of 175 1-4 syllable nouns (PNT) *Verbal short-term memory task:* Repetition of two- and three-syllable word pairs (varied for frequency and imageability)	*Word retrieval:* # of semantic errors (S score), # of phonological nonword errors (P score) and relative score reflecting these error types (P-S) *Verbal short-term memory:* Imageability effect score: # high imageability words correct/# total words correct Frequency effect score: # high frequency words correct/# total words correct	The 3 word retrieval measures will be correlated (Pearson) with imageability effect scores and frequency effect scores.	P-S scores and P scores will positively correlate with imageability and frequency. S scores will negatively correlate with imageability and frequency.

Chapter 3: Results

Aim 1: Word Retrieval Accuracy, Word Pointing Span Length, and Spatial Span Length

Prior to investigating the primary research questions regarding the relationship between intricate measures of VSTM (i.e., 1) primacy/recency effects and 2) susceptibility to psycholinguistic manipulations in word pair repetition) and word retrieval, several analyses were conducted to contribute to questions asked in previous literature about the relationship between verbal short-term memory span length, nonverbal short-term memory span length, and language performance, and to set the stage for specific aims 2 and 3. Specific aim 1 explored the extent to which word retrieval is linked to a linguistically-specified temporary storage (i.e., VSTM) system, while the subsequent questions more specifically explored the nature of this link. First, the relationship between word pointing span length and word retrieval (as measured by overall accuracy on the PNT) was investigated. Pearson's *r* was computed to assess this link, which has been found in a previous study of individuals with aphasia (Martin & Gupta, 2004), and a significant, positive relationship was predicted. The result was consistent with previous literature (Martin & Gupta, 2004), $r(22) = .732$, $p = .000$; large correlation (Figure 5).

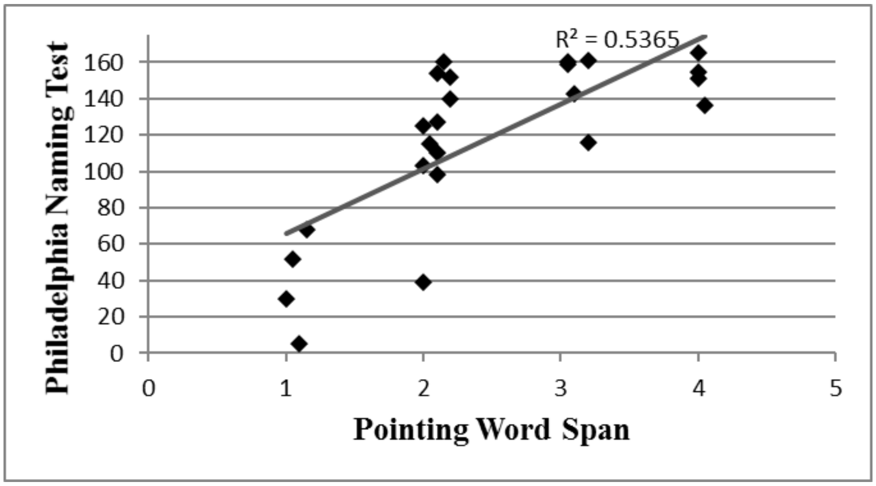


Figure 5. Positive correlation between word retrieval and pointing word span.

Regression analysis. To more precisely describe the significant positive relationship between VSTM (as measured by the word pointing span task) and word retrieval (as measured by overall accuracy on the PNT), a simple linear regression analysis was conducted. Word pointing span scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., word pointing span) and the outcome (i.e., word retrieval). Full regression results are outlined in Appendix IIa. The regression analysis yielded the following explained variance value: $R^2 = 0.54$, $F(1, 22) = 25.46$, $p < .001$, $R^2_{\text{adjusted}} = .52$. The model estimate showed that mean PNT word retrieval accuracy was 117.67 points ($SE = 6.63$) for individuals with average word pointing span performance. Word pointing span had a positive effect on PNT accuracy ($b = 34.17$, $SE = 6.77$), $t(22) = 5.05$, $p < .001$. Specifically, there is an estimated mean increase of 34.17 points on the PNT for individuals who performed one standard deviation higher than average on the word pointing span.

Next, an analysis was conducted to determine whether a relationship existed between verbal and nonverbal short-term memory span. The interactive activation model that drives this work (Dell, 1986; Dell, et al., 1997; N. Martin & Saffran, 1997) argues that the nature of short-term memory depends on the processes it supports, and therefore argues against a general, amodal short-term storage system that supports the activation of both linguistic and nonlinguistic information. Thus, a task such as word pointing span, which is heavily reliant on the temporary activation of linguistic elements, should not be related to a task such as the Corsi Block Span Task, which relies on memory for spatial patterns and minimizes linguistic influence. Pearson's *r* was computed to assess the relationship between the word pointing span task and the Corsi Block Span Task. As expected, there was no significant correlation between the two different types of STM tasks, *r*(22) = .010, *p* = .962 (Figure 6). Though not probable, a significant correlation between the Corsi Block Span Task and PNT word retrieval accuracy could still be found, so an additional post hoc analysis was added to calculate this correlation. As expected, there was no significant correlation between the Corsi Block Span Task and PNT word retrieval accuracy, *r*(22) = .13, *p* = 0.53 (Appendix III)

Additionally, though these tests were not compared directly via a *t* test due to their differing standard administration and scoring rules (i.e., the word pointing span task tests up to a span of 7 and allows for more precise scoring than the Corsi Block Span Task, which tests up to a span of 9 and allows whole number scores only), the data demonstrate differential impairments between the two types of spans. Word pointing span length ranged from 1-4.05, with a mean of 2.45, SD = 0.96 (typical performance mean is 5.53, SD = 0.85), while Corsi Block Span ranged from 3-5, with a mean of 4.13, SD = 0.80 (typical mean performance is 5.92, though this number should be interpreted cautiously, as individuals who were tested in order to determine test norms only received 2 trials at each list length; DeRenzi & Nichelli, 1975). The notably higher scores on the spatial span task are consistent with the metacognitive observations made by participants during its administration, including "this is easy for me," and "I'm better at seeing thing," referring to the Corsi Block task when comparing the relative ease of the two STM span tasks. Even the participants who topped out at 3 items on the Corsi Block Span Task, the lowest observed score, demonstrated little to no fatigue or frustration on this task; however, all participants demonstrated marked difficulty on the word pointing span task, albeit at varying span lengths. Typically, participants were able to perform their second to last word span length with relative ease, and demonstrated a dramatic drop in performance and an increase in effort during the last word span length that they reached.

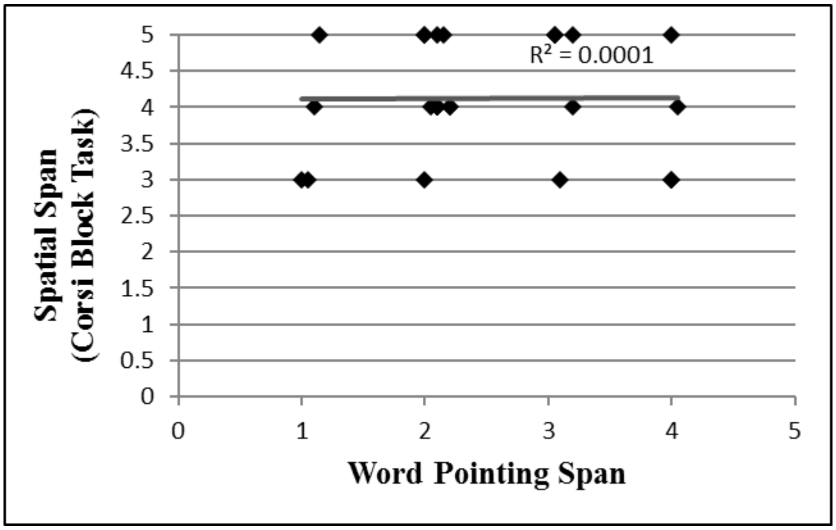


Figure 6. No significant correlation between word pointing span and spatial span.

These analyses demonstrated that verbal and spatial STM abilities, as measured by the TALSA word pointing span and Corsi Block Span Task, are not related to each other, providing precedence for the context-dependent study of STM through the consideration of processes that the temporary storage system supports (e.g., word retrieval). Additionally, the large positive correlation between word pointing span performance and PNT performance established a strong, specific link between VSTM and word retrieval, setting the stage for a more intricate exploration of the nature of this relationship.

Aim 2: Primacy/Recency Effects in Repetition as they Relate to Word Retrieval Breakdown

The second specific aim was the exploration of a possible relationship between location of errors in a word pair repetition task and types of errors made on a word retrieval (PNT) task. A primacy bias score (# first words correct/(# first + second words correct)) was used to determine positional repetition bias in the verbal STM (i.e., word pair repetition) task. Three scores were used to quantify word retrieval impairment type on the PNT: S score (total # of semantically related naming errors), P score (total number of phonologically related nonword errors), and P-S score (a measure of relative impairment, where a positive score indicates greater phonological impairment and a negative score indicates greater semantic impairment). Intra-rater and inter-rater reliability calculations (Cohen's kappa, Landis & Koch, 1977) on the coded word retrieval and word pair repetition responses demonstrated that these measures were highly reliable. A kappa coefficient of .840 (almost perfect agreement) was found for intra-rater naming error coding reliability and of .958 (almost perfect agreement) for intra-rater word pair repetition coding reliability. A kappa coefficient of .795 (substatial agreement) was found for inter-rater naming error coding reliability and of .919 (almost perfect agreement) was found for inter-rater word pair repetition coding reliability.

Word retrieval and primacy score distribution. Raw accuracy on the Philadelphia Naming Test, which consists of 175 total items, ranged from 5 to 165. Numbers of phonologically related nonword errors (P scores) ranged from 0 to 46, with all but 2 participants producing at least 1 error of this type. Numbers of semantically related errors ranged from 1 to 27. P-S scores ranged from -17 (highest semantic error bias) to 39 (highest phonologically related nonword error bias), with 15 participants showing a bias towards semantic errors, 1 participant showing no bias, and 8 participants showing a bias towards phonologically related nonword errors (Figure 7a illustrates this distribution). On the word pair repetition task, which consisted of 240 word pairs, word 1 total accuracy ranged from 0 to 230 words correct, and word 2 total accuracy ranged from 0 to 231 total words correct. Primacy bias scores ranged from 0% (highest biast towards correct repetition of word 2) to 96.50% (highest bias towards correct repetition of word 1), with 11 individuals falling below 50% (i.e., demonstrating a recency bias) and 13 individuals falling above 50% (i.e., demonstrating a primacy bias). Though all participants demonstrated a bias (nobody received a primacy bias score of 50%), half of the participants howevered within 3% of the 50% score, while the other half showed more extreme biases towards correct first words (5 participants) or second words (7 participants). Figure 7b illustrates this distribution. Individual word retrieval type scores and primacy bias scores are listed in Table 4.

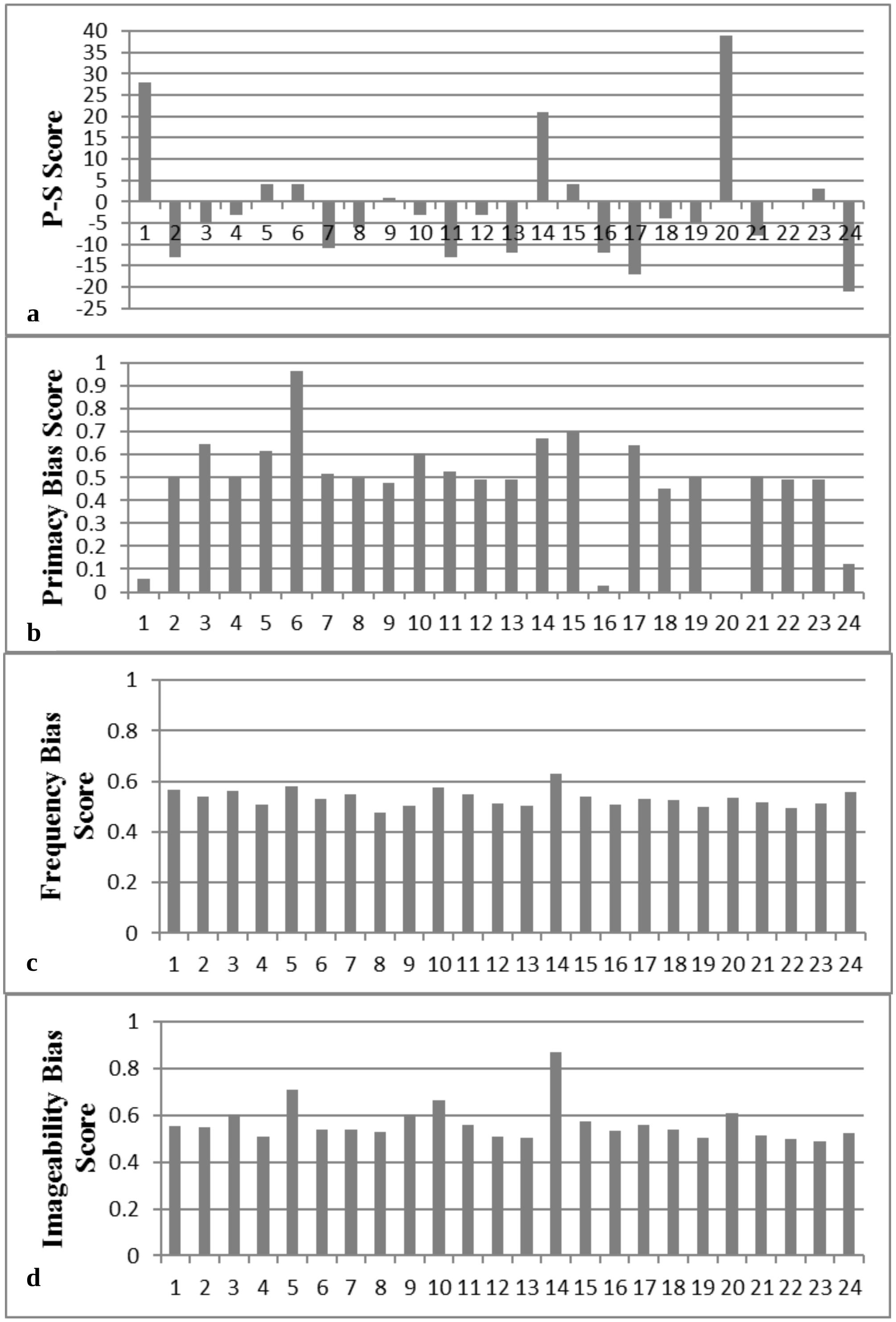


Figure 7. Distributions of word retrieval and short-term memory impairment scores, with all scores organized by participant ID number. *a)* P-S scores; *b*) primacy scores; *c*) frequency scores; *d)* imageability scores

a) P-S = # phonologic naming errors-# semantic naming errors; *b) Primacy score* = # first words correctly repeated /total words correct; *c) Frequency score* = # high frequency words correctly repeated/total words correct; *d) Imageability score* = # high imageability words correctly repeated/total words correct

Table 4. *Individual data: Primacy bias and word retrieval type scores*

	Word Pair Repetition			***Picture Naming (PNT)***			
ID	Word 1 Raw #	Word 2 Raw #	**Primacy Bias**	Raw #	**P score**	**S score**	**P-S Score**
01	9	141	0.06	98	32	4	28
02	98	97	0.50	125	2	15	-13
03	171	93	0.65	160	4	9	-5
04	230	229	0.50	161	1	4	-3
05	100	62	0.62	152	7	3	4
06	138	5	0.97	68	18	14	4
07	198	185	0.52	143	5	16	-11
08	190	194	0.49	165	0	6	-6
09	156	171	0.48	154	6	5	1
10	20	13	0.61	5	2	5	-3
11	179	162	0.52	127	4	17	-13
12	202	209	0.49	160	3	6	-3
13	225	231	0.49	136	1	13	-12
14	51	25	0.67	110	27	6	21
15	195	83	0.70	159	5	1	4
16	4	142	0.03	30	4	16	-12
17	112	63	0.64	115	3	20	-17
18	166	203	0.45	103	6	10	-4
19	227	225	0.50	151	0	5	-5
20	0	28	0	39	46	7	39
21	219	215	0.50	116	3	11	-8
22	218	225	0.49	155	6	6	0
23	211	216	0.49	140	10	7	3
24	16	117	0.12	52	6	27	-21
AVE	138.96	138.92	0.481	117.67	8.38	9.71	-1.33
SD	81.69	77.21	0.22	46.65	11.30	6.32	13.88

Correlational analysis. The main analysis concerned the relationship between P-S score and primacy bias score. Prior to the main analysis, P and S scores were correlated with primacy bias score, as was conducted in Martin & Saffran (1997) with receptive language scores. Though these scores are less informative in that they represent raw numbers of errors only and do not reflect the relative dominance of a given error type, they provide information about frequency of the error, which the relative score does not. Thus, all 3 word retrieval scores (P, S, and most importantly, P-S) were of interest, and were correlated with primacy score.

The results of the correlational analyses are outlined in Table 5. No significant correlation was found between P score and primacy, *r*(22) = -.367, *p* = .078, though a trend was found in the opposite of the predicted direction: as the number of phonologically related nonword errors (i.e., P score) increased, primacy bias (i.e., tendency to correctly repeat the first word in a word pair stimulis) decreased. No significant correlation was found between S score and primacy, *r*(22) = -.170, *p* = .426, though the negative correlation was consistent with the predicted direction. Finally, the main analysis showed no significant correlation between P-S score and primacy, *r*(22) = -.221, *p* = .299 (Figure 8), and the negative correlation was inconsistent with the predicted direction.

Table 5. *Pearson correlation table: Primacy bias and word retrieval impairment type*

	M	*SD*	*n*	P score	S score	P-S score	Primacy
P Score	8.38	11.30	24	--	-.18	.89 **	-.37
S Score	9.71	6.32	24	-.18	--	-.60 **	-.17
P-S Score	-1.33	13.88	24	.89 **	-.60 **	--	-.22
Primacy Bias	0.48	0.22	24	-.37	-.17	-.22	--

** *p* < .01

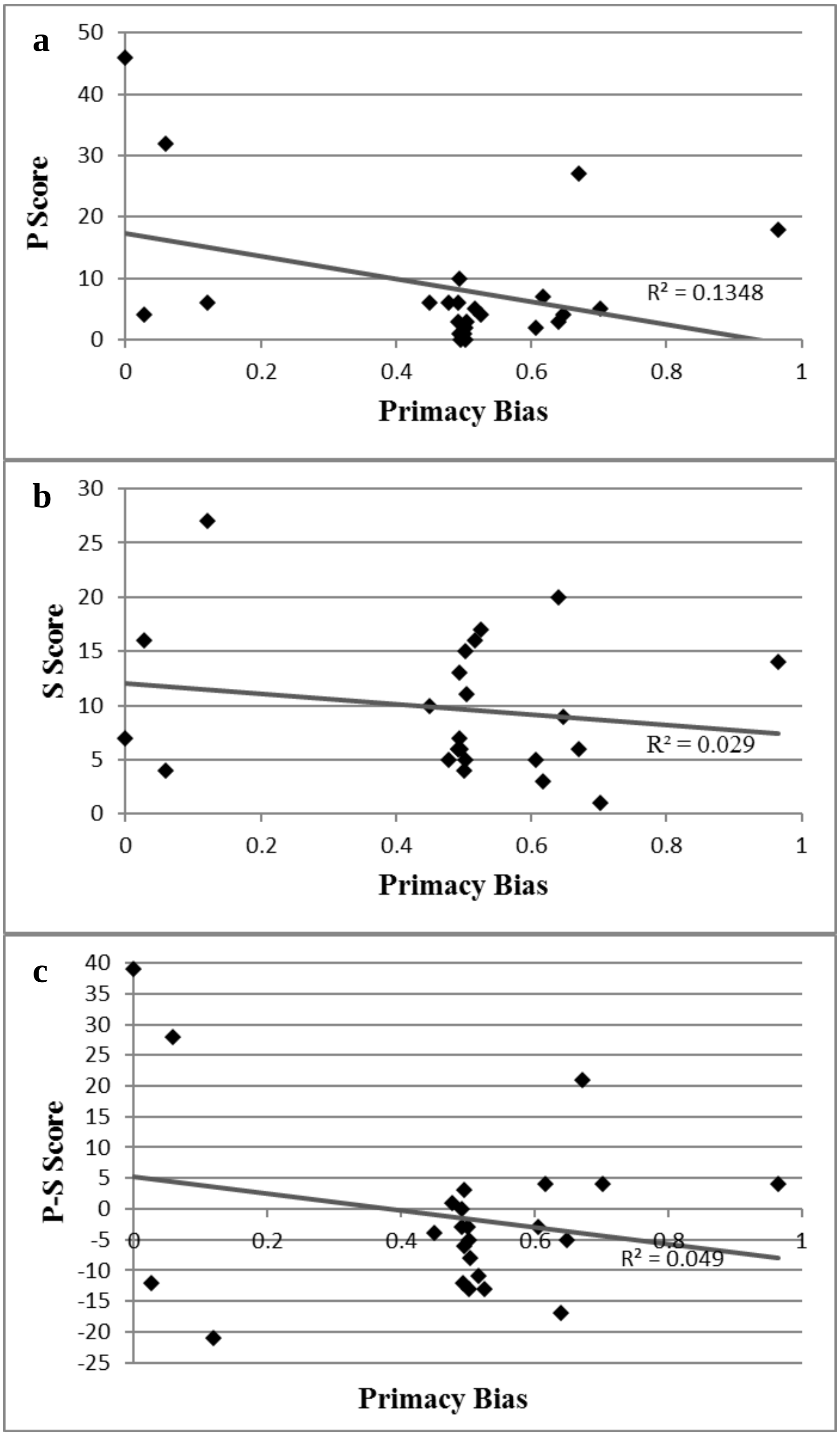


Figure 8. Correlation graphs for P scores and primacy (a), S scores and primacy (b), and P-S scores and Primacy (c). No significant correlations were found.

Aim 3: Imageability and Frequency Effects in Repetition and Word Retrieval Breakdown

The purpose of specific aim 3 was an exploration of a possible relationship between susceptibility to word frequency and imageability manipulations in a word pair repetition task and types of errors made on a word retrieval (PNT) task. A frequency bias score (# high frequency words correct/(# high frequency words correct + # low frequency words correct)) was used to determine bias towards correct repetition of high frequency words on the verbal STM (i.e., word pair repetition) task. An imageability bias score (# high imageability words correct /(# high imageability words correct + # low imageability words correct)) was used to determine bias towards correct repetition of high imageability words in the VSTM (i.e., word pair repetition) task. Words were collapsed across serial position in the calculations of frequency and imageability bias scores. The main analysis concerned the relationship between the P-S score (relative word retrieval impairment) and the psycholinguistic manipulation bias scores (i.e., frequency and imageability). Prior to the main analysis, P and S scores were correlated with frequency and imageability bias scores, as was conducted in Martin & Saffran (1997) with receptive language scores. Thus, all 3 word retrieval scores (P, S, and most importantly, P-S) were of interest, and were correlated with frequency and imageability bias scores.

Frequency bias and imageability bias score distribution. Accuracy on high frequency words (240 total words) ranged from 15 words correct to 233 words correct. Accuracy on low frequency words ranged from 13 words correct to 227 words correct. Frequency bias scores ranged from 47.66% (slight bias towards correct responses to low frequency items) to 63.16% (bias towards correct responses to high frequency items), with 21 out of 24 participants demonstrating the expected bias for correct responses on high frequency items, 2 participants demonstrating a slight bias towards correct responses on low frequency items, and 1 participant demonstrating no bias. Figure 7c illustrates this distribution. Accuracy on high imageability words (240 total words) ranged from 17 words correct to 234 words correct. Accuracy on low imageability words ranged from 11 words correct to 226 words correct. Imageability bias scores ranged from 48.71% to 86.84%, with 22 out of 24 participants demonstrating the expected bias for correct responses on high imageability items and 2 participants demonstrating a slight bias towards correct responses on low imageability items. Figure 7d illustrates this distribution. Individual frequency and imageability bias scores, along with P-S word retrieval error type scores, are listed in Table 6.

Correlational analysis: Frequency bias and word retrieval impairment type. A significant positive correlation was found between P score and frequency bias, $r(22) = .406$, $p = .049$ (medium correlation): as the number of phonologically related nonword naming errors increased, tendency to correctly repeat more high frequency words than low frequency words increased. This result is consistent with the predicted results and the positive correlation found in Martin & Saffran (1997) between frequency bias and receptive word retrieval impairment type. No significant correlation was found between S score and frequency bias, $r(22) = .020$, $p = .927$, and the positive correlation is inconsistent with the predicted direction. No significant correlation was found between P-S score and frequency bias, $r(22) = .322$, $p = .126$, though the positive correlation is consistent with the predicted direction. For full correlation results, see Table 7. For correlation graphs, see Figure 9.

Table 6. *Individual data: Frequency Bias, Imageability Bias, and Word Retrieval Type Scores*

ID	High Frequency Raw Accuracy	Low Frequency Raw Accuracy	**Frequency Bias**	High Imageability Raw accuracy	Low Imageability Raw Accuracy	**Imageability Bias**	**P-S Score**
1	85	65	0.57	83	67	0.55	28
2	105	90	0.54	107	88	0.55	-13
3	149	115	0.56	159	105	0.60	-5
4	233	226	0.51	234	225	0.51	-3
5	94	68	0.58	115	47	0.71	4
6	76	67	0.53	77	66	0.54	4
7	211	172	0.55	207	176	0.54	-11
8	183	201	0.48	203	181	0.53	-6
9	165	162	0.50	197	130	0.60	1
10	19	14	0.58	22	11	0.67	-3
11	187	154	0.55	191	150	0.56	-13
12	211	200	0.51	210	201	0.51	-3
13	229	227	0.50	230	226	0.50	-12
14	48	28	0.63	66	10	0.87	21
15	150	128	0.54	160	118	0.58	4
16	74	72	0.51	78	68	0.53	-12
17	93	82	0.53	98	77	0.56	-17
18	194	175	0.53	199	170	0.54	-4
19	226	226	0.50	227	225	0.50	-5
20	15	13	0.54	17	11	0.61	39
21	225	209	0.52	223	211	0.51	-8
22	220	223	0.50	220	223	0.50	0
23	218	209	0.51	208	219	0.49	3
24	74	59	0.56	70	63	0.53	-21
AVE	145.17	132.71	0.53	150.04	127.83	0.57	-1.33
SD	72.53	74.09	0.03	71.45	76.16	0.08	13.88

Table 7. *Pearson correlations: Frequency, Imageability, and Word Retrieval Impairment Type*

	M	*SD*	*n*	P score	S score	P-S score	Frequency	Imageability
1. P Score	8.38	11.30	24	--	-.18	.89 **	.41 *	.39
2. S Score	9.71	6.32	24	-.18	--	-.60 **	.02	-.25
3. P-S Score	-1.33	13.88	24	.89 **	-.60 **	--	.32	.43 *
4. Frequency Bias	0.53	0.03	24	.41 *	.02	.32	--	.82 **
5. Imageability Bias	0.57	0.08	24	.39	-.25	.43 *	.82 **	--

* *p* < .05, ** *p* < .01

Regression analysis: Frequency bias and P score. To more precisely describe the significant positive relationship between frequency bias and P score, a simple linear regression analysis was conducted. Frequency bias scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., frequency bias) and the outcome (P score). The regression analysis yielded the following explained variance value: $R^2 = .17$, $F(1, 22) = 4.34$, $p < .05$, $R^2_{\text{adjusted}} = .13$. The model estimate showed that mean P score was 8.38 ($SE = 2.16$) for individuals with average frequency bias. Frequency bias had a positive effect on P score ($b = 4.59$, $SE = 2.20$), $t(22) = 2.08$, $p < .05$. Specifically, there is an estimated mean increase of 4.59 phonologically related nonword naming errors on the PNT for individuals who demonstrated a one standard deviation-higher than average frequency bias score.

Correlational analysis: Imageability bias and word retrieval impairment type. No significant correlation was found between P score and imageability bias, $r(22) = .385$, $p = .063$, though a trend was found in the predicted direction: as the number of phonologically related nonword naming errors increased, tendency to correctly repeat more high frequency words than low frequency words increased. No significant correlation was found between S score and imageability bias, $r(22) = -.253$, $p = .232$, though the negative correlation is consistent with the predicted direction. A significant positive correlation was found between P-S score and imageability bias, $r(22) = .429$, $p = .036$ (medium correlation): as the bias to produce phonologically related nonword errors over semantically related errors increased, imageability bias increased. This result was consistent with the predicted results and the positive correlation found in Martin & Saffran (1997) between imageability bias and receptive word retrieval impairment type. For correlation graphs, see Figure 10.

Regression analysis: Imageability bias and P-S score. To more precisely describe the significant positive relationship between imageability bias and P-S score, a simple linear regression was conducted. Imageability bias scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., imageability bias) and outcome (P-S score). The regression analysis yielded the following explained variance value: $R^2 = .18$, $F(1, 22) = 4.96$, $p < .05$, $R^2_{adjusted} = .15$. The model estimate showed that the mean P-S score was -1.33 (*SE* = 2.62) for individuals with average imageability bias. Imageability bias had a positive effect on P-S score ($b = 5.95$, $SE = 2.67$), $t(22) = 2.23$, $p < .05$. Specifically, there is an estimated mean increase of 5.95 points on the PNT P-S score for individuals who demonstrated a one standard deviation-higher than average imageability bias score.

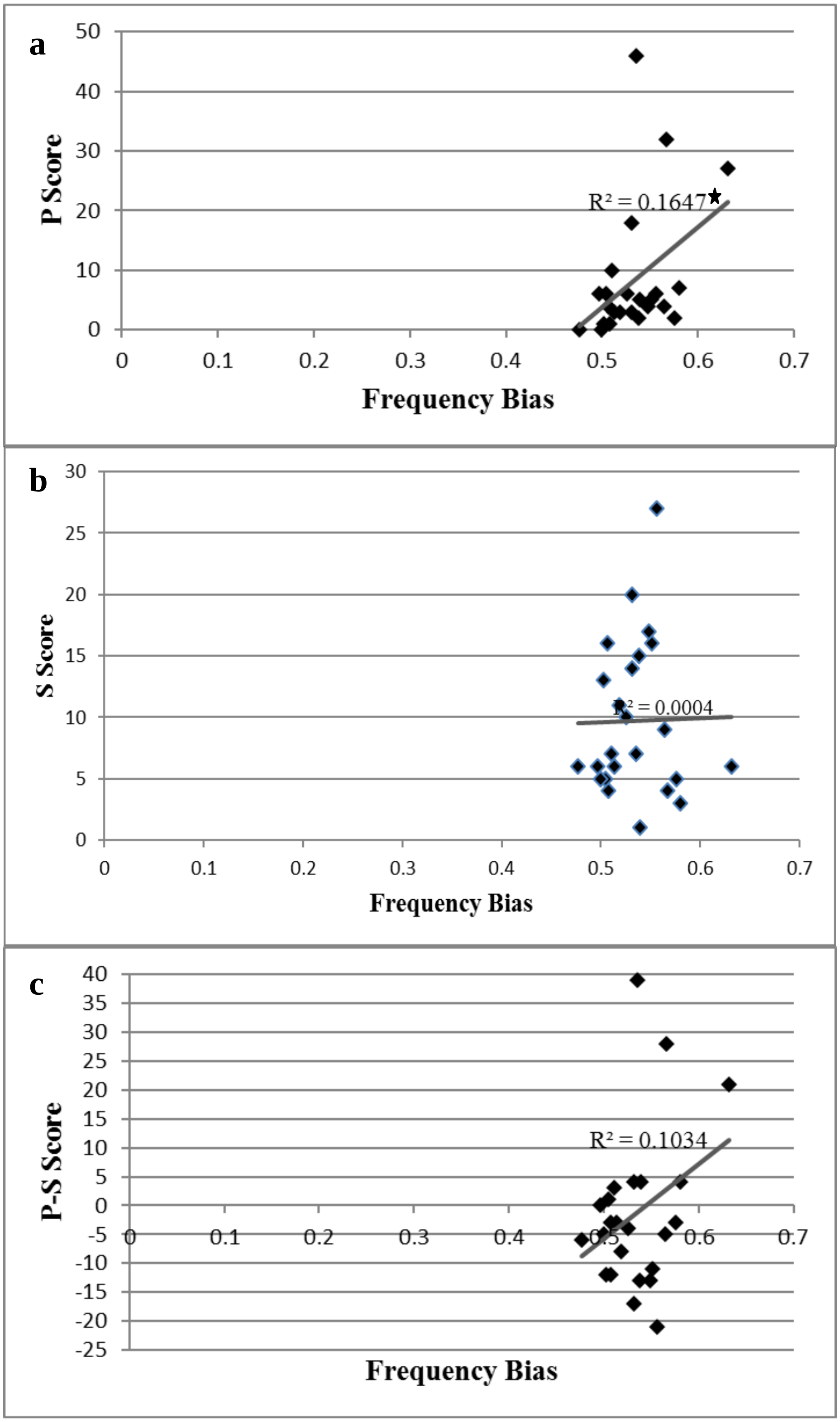


Figure 9. Correlation graphs for P score and frequency bias (a), S score and frequency bias (b), and P-S score and frequency bias (c). Significant R^2 values are starred.

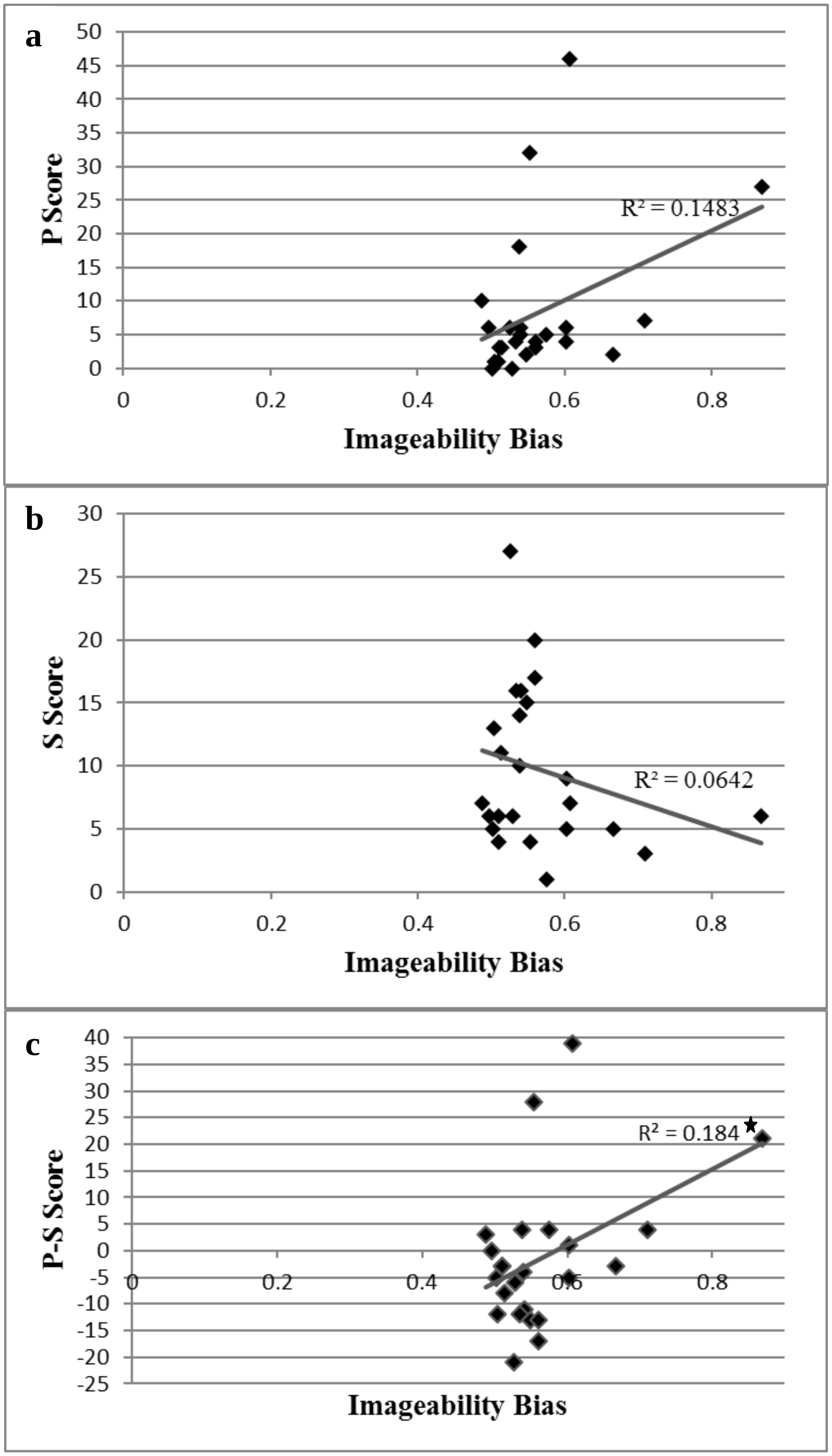


Figure 10. Correlation graphs for P score and imageability bias (a), S score and imageability bias (b), and P-S score and imageability bias (c). Significant R^2 values are starred.

Post Hoc Analyses

Several post hoc analyses that were driven by behavioral observations of participants' performance on the word retrieval and VSTM tasks were conducted. The first of these analyses concerned a change to the classification rules of P errors, this time including formal errors (i.e., phonologically related real words) in the calculation of the P score. The choice to encompass formal errors in the P score calculation was based both on behavioral observations of the participants during the PNT as well as previous literature suggesting that formal errors are likely to be devoid of true lexical influence despite their status as real words, and thus often stem solely from phonological-level breakdown (Nickels & Howard, 1995b). Behaviorally, participants frequently made formal errors that differed greatly from the target in terms of their function (e.g., different part of speech). Some examples of our participants' errors classified as formal errors that were suspected to arise from phonological influence only were as follows, with the target on the left and the error on the right: frog→for, crutches→crush, eye→hi, cross→crass. Additional evidence of the frequent incorrect attribution of errors that are solely phonologic in origin to both lexical and phonological influences was demonstrated by Nickels & Howard (1995b), an argument supported by two main observations of aphasic errors. The first is that formal errors occur more frequently on short than long words, which is consistent for what would be expected if they occurred by chance, given that shorter words have more phonological neighbors than long words. The second finding concerned the creation of a pseudocorpus based on actual phonological errors in individuals with aphasia. Based on the observed errors, Nickels & Howard (1995b) created their own pseudocorpus by randomly replacing wrongly substituted sounds for others that were phonotactically legal in that location, and found that the pseudocorpus and actual corpus of errors did not differ significantly in the numbers of phonologically related nonwords and phonologically related real words (i.e., formal errors). They concluded that formal errors are most likely real words by chance. Though it is not possible to definitively determine whether a formal error stems from lower level phonological breakdown or a breakdown that starts higher in the production process and has lexical influences, these observations provide support for rethinking the P score calculation to include formal errors. Thus, the main analyses described above were rerun using an adjusted P score (PA Score), which represented the combined total number of phonologically related nonword errors and formal errors (Post hoc analysis *a*).

While post hoc analysis *a* focused on rethinking error typing on the word retrieval task, post hoc *b* focused on rethinking primacy/recency predictions on the VSTM (i.e., word pair repetition) task. While both phonological and semantic representations are likely active during repetition tasks, only phonological representations are required to correctly produce an auditorily presented target. Thus, it is possible that, consciously or not, some participants were relying on semantic support during repetition more than others. Due to the possibility that even individuals who are stronger in semantic processing may not be optimally using semantic activation to boost repetition performance, a subgroup analysis of specific aim 2 (relationship between primacy effects in repetition and word retrieval impairment type) was performed. During the word pair repetition task, some individuals produced semantically related errors, while other individuals did not produce any semantically related errors (error coding followed PNT rules; see Table 8 for a list of these errors). Thus, post hoc analysis b looked at subgroups to determine whether the relationship (or lack thereof) between primacy and word retrieval error type (i.e., P-S scores) was modulated by the types of errors individuals demonstrated on repetition tasks. Individuals who demonstrated semantically related (i.e., semantic and mixed) errors in repetition tasks may have had more access to semantics during repetition than those who did not. If types of errors made on the repetition task are a window into reliance on semantic activation in repetition, then the group that made semantically related errors on the word pair repetition task might demonstrate the predicted positive relationship between primacy and P-S scores, but the latter group (i.e., no semantically related errors in repetition) may actually demonstrate the opposite relationship (negative correlation between primacy and P-S scores). This prediction is due to the timing of phonological spread of activation immediately before an utterance is repeated: in the case of sole reliance on phonological representations, it should always be easier to repeat the second word, whose phonemes are more active at the time of recall due to recency of the stimulus. Thus, specific aim 1 analyses were rerun using a subgroup analysis for post hoc *b*.

Table 8. *Numbers and types of errors made by the subgroup of participants who produced semantically related errors on word pair repetition (error appears to right of target)*

ID	# Semantic	Semantic Error Transcription	# Mixed	Mixed Error Transcription	Total #
02	1	circus→clown	0	NA	1
03	2	cigarette→fire visitor→friend	3	woman → wife officer → soldier mother → father	5
05	4	wedding→marriage sunburn→sunblock pigeon→bird wisdom→knowledge	2	pigeon → parrot alcohol → liqueur	6
07	2	movie → theater hockey → soccer	0	NA	2
12	0	portion → section	2	pudding → pie	2
14	0	NA	1	hospital → doctor	1
15	5	detective →policeman pigeon → rooster soda → drink hero → winner congress → senator	1	memory → remember	6
17	3	breakfast → dinner summer → winter engine → repair	1	question →answer	4
20	1	pigeon → bird	0	NA	1
23	2	cellar → garage (x2)	0	NA	2
24	2	engine → firetruck tennis → bowling	0	NA	2

Post hoc a.: **Re-analysis of aims 2 and 3 with adjusted P scores.** All previous analyses were performed with PA scores. Thus, word retrieval phonological impairment measures (i.e, adjusted P and adjusted bias score [PA-S]) were correlated with the VSTM impairment measures (primacy bias, frequency bias, and imageability bias). Full results of the correlational analyses are shown in Table 9. The correlation between primacy bias and PA remained insignificant and opposite of the predicted direction. Consistent with results of the main analysis (i.e., using unadjusted P score), frequency correlated positively with PA score. Additionally, imageability correlated positively with PA score (*r* = .45, medium correlation), while the main analysis demonstrated a trend approaching significance in the same direction. The correlation between primacy bias and PA-S remained insignificant and opposite of the predicted direction. Consistent with the main analysis and predictions, the correlation between imageability bias and PA-S score is positive and significant. Unlike in the main analysis, the correlation between frequency bias and PA-S score is also positive and significant (*r* = .47, moderate correlation), which is consistent with the predicted results. To more precisely describe the significant correlations, simple linear regression analyses were conducted.

Table 9. Pearson Correlation Analysis Table: Post Hoc A (Re-analysis with adjusted P score)

	M	*SD*	*n*	PA Score		PA-S Score		Primacy	Frequency		Imageability	
1. PA Score	14.13	16.78	24	--		.94	***	-.22	.48	*	.45	*
2. PA-S Score	4.42	18.86	24	.94	***	--		-.14	.41	*	.49	*
3. Primacy Bias	0.48	0.22	24	-.22		-.14		--	.13		.21	
4. Frequency Bias	0.53	0.03	24	.48	*	.41	*	.13	--		.82	**
5. Imageability Bias	0.57	0.08	24	.45	*	.49	*	.21	.82	**	--	

PA Score = Adjusted P Score (includes both phonologically related nonwords and formal errors)
* *p* < .05,**p<.01, *** *p* < .001.

Regression Analysis: Frequency Bias and Adjusted P. Frequency bias scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., frequency bias) and outcome (adjusted P, or PA, score). The regression analysis yielded the following explained variance value: $R^2 = .22$, $F(1, 22) = 6.35$, $p < .05$, $R^2_{adjusted\,=}$ .19. The model estimate showed that the mean PA score was 14.13 ($SE = 3.09$) for individuals with average frequency bias. Frequency bias had a positive effect on PA score ($b = 7.94$, $SE = 3.15$), $t(22) = 2.52$, $p < .05$. Specifically, there is an estimated mean increase of 7.94 phonologically related errors (including both nonword and formal errors) for individuals who demonstrate a one standard deviation-higher than average frequency bias score.

Regression Analysis: Imageability Bias and Adjusted P. Imageability bias scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., imageability bias) and outcome (adjusted P, or PA, score). The regression analysis yielded the following explained variance value: $R^2 = .20$, $F(1, 22) = 6.35$, $p < .05$, $R^2_{adjusted} = .17$. The model estimate showed that the mean PA score was 14.13 ($SE = 3.13$) for individuals with average imageability bias. Imageability bias had a positive effect on PA score ($b = 7.57$, $SE = 3.19$), $t(22) = 2.37$, $p < .05$. Specifically, there is an estimated mean increase of 7.57 phonologically related errors (including both nonword and formal errors) for individuals who demonstrate a one standard deviation-higher than average imageability bias score.

Regression Analysis: Imageability Bias and Adjusted P-S Score. Imageability bias scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., imageability bias) and outcome (adjusted P-S, or PA-S, score). The regression analysis yielded the following explained variance value: $R^2 = .24$, $F(1, 22) = 6.81$, $p < .05$, $R^2_{adjusted} = .20$. The model estimate showed that the mean PA-S score was 4.42 ($SE = 3.44$) for individuals with average imageability bias. Imageability bias had a positive effect on PA-S score (b = 9.17, SE = 3.51), $t(22) = 2.61$, $p < .05$. Specifically, there is an estimated mean increase of 9.17 points on the PNT PA-S score for individuals who demonstrate a one standard deviation-higher than average imageability bias score.

Regression Analysis: Frequency and Adjusted P-S Score. Frequency bias scores were transformed into z-scores for ease of interpretation of the relationship between the predictor (i.e., frequency bias) and outcome (adjusted P-S, or PA-S, score). The regression analysis yielded the following explained variance value: $R^2 = .17$, $F(1, 22) = 4.56$, $p < .05$, $R^2_{adjusted} = .13$. The model estimate showed that the mean PA-S score was 4.42 (SE = 3.58) for individuals with average frequency bias. Frequency bias had a positive effect on PA-S score (b = 7.82, SE = 3.66), $t(22) = 2.14$, $p < .05$.Specifically, there is a mean estimated increase of 7.82 points on the PNT PA-S score for individuals who demonstrate a one standard deviation-higher than average frequency bias score.

Post hoc b.: **Re-analysis of aim 1 with repetition subgroups.** Participants were divided into two subgroups: individuals who made at least one semantically related error (i.e., either a semantic or mixed error, based on PNT coding guidelines) on the word pair repetition task (SRE group, n = 11) and individuals who made no semantic or mixed errors on the word pair repetition task (NSRE group, n = 13). For each subgroup, Pearson correlations were performed between primacy bias and P, P-S, PA, and PA-S scores. No significant correlations for the measures of interest emerged (Tables 10 and 11).

Table 10. Pearson Correlation Analysis Table: Post Hoc B (SRE group)

	M	*SD*	*n*	P	S	P-S	PA	PA-S	Primacy
1. P Score	10.73	13.61	24	--	-.27	.90 **	.99 **	.92 **	-.54
2. S Score	10.64	7.92	24	-.27	--	-.66 *	-.29	-.59	-.38
3. P-S Score	0.09	17.48	24	.90 **	-.66 *	--	.90 **	.99 **	-.25
4. PA Score	17.09	19.41	24	.99 **	-.29	.90 **	--		-.45
5. PA-S Score	6.45	23.023	24	.92 **	-.59	.99 **	.95 **	--	-.25
6. Primacy Bias	0.49	0.23	24	-.54	-.38	-.25	-.45	-.25	--

* p < .05, ** p < .01

Table 11. Pearson Correlation Analysis Table: Post Hoc B (NSRE group)

	M	*SD*	*n*	P	S	P-S	PA	PA-S	Primacy
1. P Score	6.38	9.00	24	--	-.08	.89 **	.96 **	.94 **	-.21
2. S Score	8.92	4.79	24	-.08	--	-.52	.01	-.31	.08
3. P-S Score	-2.54	10.54	24	.89 **	-.52	--	.82 **	.94 **	-.21
4. PA Score	11.62	14.53	24	.96 **	.01	.82 **	--	.95 **	-.01
5. PA-S Score	2.69	15.27	24	.94 **	-.31	.94 **	.95 **	--	-.03
6. Primacy Bias	0.47	0.23	24	-.21	.08	-.21	-.01	-.03	--

** p < .01

Chapter 4: Discussion

Introduction to the Discussion

The purpose of this study was to determine whether a linguistically-specified STM system subserves word retrieval and to examine the nature of this mechanism if it exists. Essentially, the findings were that 1) word pointing span length *is* positively correlated with picture naming accuracy and *not* correlated with nonverbal span length, 2) primacy bias during word pair repetition is *not* correlated with word retrieval impairment type (P-S score) and 3) imageability bias during word pair repetition *is* positively correlated with word retrieval impairment type (P-S score), while frequency bias is *not* significantly correlated with word retrieval impairment type. The results are interpreted below in the context of an interactive linguistic activation process that partly, though not completely, subserves both word pair repetition (i.e., the current study's VSTM task) and picture naming performance. Thus, the temporary activation of semantic, lexical, and phonological representations is argued to be the VSTM system that supports word retrieval, and the discussion will propose that it is the overall activation strength of linguistic representations rather than the timing of spreading activation from one linguistic level to another that is most critical to successful word retrieval, at least in the context of the confrontation picture naming task used in this study. The discussion argues that the classic distinctinction and assumption of the separability of STM and language is inconsistent with the current study's findings and extant work, detrimental to the advancement of both theoretical and clinical knowledge, and in need of reconsideration. Clinical implications of a VSTM system supported by interactive linguistic activation are discussed, with a special focus on the impact on diagnosis of word retrieval impairments in aphasia. Finally, recommendations for future research concerning the role of VSTM in both language production and comprehension are discussed.

Specific Aim 1: Specificity of STM and its Breakdown in Aphasia

The large positive correlation between word pointing span length, which measured the construct of VSTM, and number of correct items on the PNT, which measured the construct of word retrieval, demonstrated a strong surface relationship, thus providing a justification for asking further questions about the nature of this link. The fact that word pointing span explained 53.65% of the variance in word retrieval scores is an undoubtedly poignant sign that the two seemingly separable constructs are strongly associated. It is important to note that because word pointing span does not require an overt response, it more successfully isolates the temporary storage component (i.e., STM) than the more common word repetition span task. Thus, the significant correlation between word pointing span length and word retrieval is not influenced by the tasks both requiring a verbal output. Alongside this result, the finding that word pointing span and the Corsi Block Span Task, a measure of nonverbal STM, did *not* significantly correlate with each other suggests that the link between word pointing span and word retrieval is likely attributable to a linguistically-specified temporary storage mechanism rather than a domain-general one.

The few past studies that have dealt with the relationship between STM (both verbal and nonverbal) and word retrieval are equivocal on the issue of a linguistically-specified STM. Consistent with the current study's findings, Martin & Gupta (2004) found a significant positive correlation between percent correct on the Philadelphia Naming Test and word pointing span ($r(48) = .52$, $p < .01$). In terms of the specificity of this link, previous studies present conflicting results. Martin & Ayala (2004) found no correlation between word pointing span and the Corsi Block Span Task (p. 472), no correlation between the Corsi Block Span Task and 4 out of 6 language tasks (i.e., phoneme discrimination with a filled interval between the phonemes, synonymy judgments, and two word-to-picture matching tasks), and significant positive correlations between the Corsi Block Span Task and 2 measure of phonological processing (i.e., phoneme discrimination with an unfilled interval and auditory rhyme judgments) in a group of 46 individuals with aphasia. All correlations between word pointing span and the 6 language tasks were positive and significant. Additionally, correlations between word pointing span and both numbers of semantic errors and phonological nonword errors made on the PNT were negative and significant, in line with the current study's results. Thus, most of their results support a linguistically-specified view of STM as it relates to language processing.

Though a more recent study by Potagas et al. (2011) claimed that they found results pointing to a modality-independent short-term memory, their resuts contradicted this argument. The authors found a significant correlation between Aphasia Score, a score representing combined totals from the auditory sentence comprehension and oral expression subtests of the Boston Diagnostic Aphasia Examination (Goodglass, et al., 2000), and a repetition version of the digit span. Additionally, they found a significant positive correlation between Aphasia Score and the Corsi Block Span Task. Though the authors used these results to argue that language processing is supported by a domain general short-term memory system, the falacy in this argument is that they also found that the Corsi Block Span Task does not correlate with the word pointing span task, suggesting that they measure different abilities. Though their results may suggest that the nonverbal span task and language tasks do share some processing mechanisms, they are not likely to be the same mechanisms that are shared between the language and verbal span tasks. The lack of a correlation between the verbal and nonverbal span tasks suggests that different storage systems support STM for linguistic and nonlinguistic information, and is consistent with the results of specific aim 1.

Though specific aim 1 demonstrated only a surface association between VSTM and word retrieval, what separates the results from some of the previous work is the analysis of this association in context of insignificant relationships between a verbal and nonverbal span task and between word retrieval accuracy and a nonverbal span task. These results together suggest that there is in fact something specific about the temporary storage mechanism that supports word retrieval, and justifies an investigation of its intricacies. Specific aims 2 and 3, which looked at more detailed aspects of verbal span performance (i.e., positional bias and susceptibility to word frequency and imageability manipulations), offer a much more in depth look at the mechanisms driving the VSTM-word retrieval relationship.

Specific Aim 2: Primacy Bias and Word Retrieval Impairment Type

The lack of a positive correlation between primacy bias and word retrieval impairment type (i.e., P-S score) brings up several interesting theoretical and empirical issues. Two possible reasons why the relationship may not have been realized are that 1) the timing of linguistic spreading activation that supports word retrieval differs from that which supports word pair repetition and 2) the time course over which positional biases are observed may partly depend on an individual’s span length capacity, and thus may not be observed in a simple word pair repetition task in individuals with relatively mild span impairments. The arguments for both of these possibilities are developed below after the finding is framed in the context of previous work.

The prediction of a positive correlation was motivated primarily by the finding of Martin & Saffran (1997), who showed a large positive correlation between primacy bias and language comprehension impairment type (i.e., S-P score, where a positive score denotes *stronger* semantic processing and a negative score denotes *stronger* phonological processing), $r(13) = .54$, $p < .05$., as well as by the findings of Wilshire et al. (2010), who reported on two individuals with aphasia, one who produced mostly semantic paraphasias and demonstrated a recency bias, and one who produced mostly phonological paraphasias and demonstrated a primacy bias. Martin & Saffran (1997) interpreted their results as evidence for a linguistically-specified storage system that utilizes the same interactive activation mechanism as is commonly attributed to word retrieval (Dell, et al., 1997). Thus, they argued that the surface link that is frequently found between span length and language processing tasks is at least partly driven by an intricate temporary activation process that allows phonological, lexical, and semantic information to be temporarily activated in an intricately timed manner. As suggested in the introduction to the discussion, this result strongly suggests that tasks commonly categorized as STM tasks (i.e., verbal span tasks) and those classified as language tasks are not as separable as has been assumed in past work.

Though the authors used only language comprehension tasks to classify participants (phonological tasks included phoneme discrimination in minimal pairs with and without a delay between the pairs, as well as a rhyme judgment task, and semantic measures included word-to-picture matching and synonymy judgment tasks), their theoretical conclusions seem to parsimoniously suggests that similar findings might be expected for language production tasks. An association between word retrieval impairment type and primacy bias, a measure that is heavily dependent on an intricately timed spread of activation between levels of linguistic processing, was not realized in the current study. One major issue that deserves consideration when interpretating this discrepancy are the differences in timing of temporary activation in comprehension versus production tasks, both those created in the laboratory and those that represent what speakers do in natural conversation. If the assumption that VSTM tasks such as the word pair repetition task used in the current study depend on a temporary linguistic activation process is correct, it is important to think about how this process might play out in the word pair repetition task in relation to both language comprehension and production tasks. In language comprehension tasks like those used by Martin & Saffran (1997), a fleeting auditory input must be processed quickly in order to correctly respond (e.g, making a rhyme judgment on two words immediately after hearing them). This is similar to the first step of a repetition span task, where the targets are heard and must be quickly repeated (note that both Martin & Saffran (1997) and the current study required individuals to repeat the targets immediately after hearing them). Both the VSTM and language comprehension tasks, then, likely depend on a very rapid and intricately timed spread of activation between phonological, lexical, and semantic nodes in order to successfully perform the tasks.

How do word retrieval tasks compare to comprehension and verbal span tasks? First, it is important to recognize that both the word pair repetition and word retrieval tasks used in this study have a production component. Word retrieval is arguably solely a production task (though see Nickels & Howard (1995b) for discussion of the possible use of a comprehension-based monitor for error detection in word production), while the last step of word pair repetition depends on verbal production. The time course over which these processes function, though, is likely very different. In the word pair repetition task, where individuals were required to respond within 5 seconds of hearing the word pairs in order to get credit for a repetition trial, this flow from phonological to semantic nodes during the comprehension portion of the task and from semantic to phonological nodes during the production process of a task must occur very quickly for a successful response to occur. In the word retrieval (i.e., PNT) task, where participants saw a picture on a screen and had up to 30 seconds to respond to the picture, timing likely functions on a much more flexible level. Though only first complete responses were scored, it is impossible to know what occurred internally when individuals took many seconds to initiate a response. Error detection/correction mechanisms that slowed down or re-initiated the activation stream could have been in play, and/or individuals could have taken much longer to activate the semantic, lexical, and phonological nodes needed to produce a response, so the overall spread of activation was slowed down (i.e., weakened connection weights). Primacy and recency effects depend on a very rapid spread of linguistic activation, especially when only two words are to be repeated, while word retrieval, especially in the task used in the current study, may follow a much more flexible time course. Thus, it is feasible to conclude that even though the same spread of activation process subserves both tasks, they likely operate on different time courses, a conclusion that is consistent with the lack of association found between primacy bias and word retrieval impairment type. Thus, even if an individual demonstrates a heavy bias towards phonological errors in word retrieval, the different time courses over which the two tasks likely function means that the expected primacy bias may not be realized in a word pair repetition task. If the word retrival task was constrained either by limiting the amount of time that the picture to be named appeared on the screen (as is done in subtest 7 of the TALSA, N. Martin, 2012) or by shortening the allotted response time, the activation time course of the word retrieval task may have better matched the activation time course of the word pair repetition task, possibly yielding different results.

A related issue to that of different time courses for word pair repetition and word retrieval is the time course over which primacy and recency effects may be expected to be observed in individuals with aphasia. Though Martin & Saffran (1997) successfully correlated primacy bias in word pair repetition with language comprehension impairment type, a later study by N. Martin et al. (2002) replicated the prior study with longer repetition strings and found that primacy bias during repetition of four word strings correlated with Composite S score (a score derived from a battery of semantic comprehension tasks, with a higher score denoting better semantic processing). Interestingly, this result was not found when only subjects with word span lengths between 2 and 3.5 were analyzed, suggesting that the relationship between primacy/recency biases and language impairments may be affected by the subjects' word span length capacity. In individuals with larger span lengths, positional biases may only be observed at larger word strings. In the current study, word pointing span lengths of the 24 participants ranged from 1 to 4.05 (AVE = 2.46; SD = .096), and these individual differences may be one explanatory factor as to why significant results were not realized. Also, given the claim that word retrieval, at least as measured by the PNT in the current study, likely functions over a slower time course than repetition span, it is possible that a longer repetition string may have yielded the expected associations between primacy bias and word retrieval impairment type. Thus, the possibilities that 1) linguistic spreading activation functions differently in word retrieval tasks than it does in repetition span tasks and 2) the word span length at which primacy/recency biases are observed in individuals with aphasia may partly depend on word span length capacity are both critical issues to consider when interpreting the null correlation between primacy bias and word retrieval impairment type.

Because the overarching goal of this work is to inform clinical practice, thinking about how the controlled word retrieval task relates to a more natural conversation is important. In comparison to language comprehension, discourse production in a conversation is likely still a more flexibly timed task than discourse comprehension because there are often multiple ways and a variety of words than can be used to communicate a message, and the speaker has a lot of control over timing and word choice on the production end and little control of timing or word choice on the comprehension end. On the comprehension side, words must be processed as they are spoken, and impaired comprehension will lead to parts of the message being missed or poorly understood. On the production side, if word and/or sentence retrieval are impaired, the speaker may be able to strategically select words and sentences that are easier to access to get the message across, which may be achieved through the slowing down of spreading activation in order to be able to internally correct for potential errors and instead select words on which errors are unlikely (those words whose semantic, lexical, and phonological features are easier to activate). Nickels & Howard (1995b) term these internally detected errors "covert errors" because they are not ever articulated and therefore very elusive to study. Thus, in natural speaking environments, word production may also be subserved by a more flexible and possibly slower temporary linguistic spreading activation system than word comprehension.

Additional factors that could have impacted our results and motivated several posthoc analyses are 1) the possible *over*coding of semantic errors and *under*coding of phonological errors and 2) the possibility that some individuals rely more on semantics than others in repetition, regardless of the level of phonological processing impairment. The error coding issues, enveloped in the more general limitation of using error type coding as a window into word retrieval impairment type, are discussed first, and are important because if P-S scores were skewed, they may have led to inaccurate correlational results. Because the current study aimed to precisely pinpoint the degree to which semantic and phonological processing contributes to word retrieval impairment, only semantic and nonword phonological errors were used because they are the most unequivocally atrributable to breakdowns in these respective processes. To say that such errors conclusively correspond to the degree of semantic and phonological impairment, however, would be false. Several theoretical issues with error coding could have led to overestimation and underestimation of semantic and phonological word retrieval impairments, respectively.

Semantic errors are almost always attributed to a breakdown in semantic activation in word retrieval, and it is fair to say that there is a general consensus that most individuals who produce these errors do so because of a breakdown at the first step of word retrieval, which involves activation of semantic nodes and lexical items that correspond to those nodes (Nickels, 1997). There is, however, an alternative possible explanation for the production of semantic errors which was proposed by Caramazza & Hillis (1990, as cited by Nickels & Howard, 1994 and Nickels, 1997), who reported on two individuals who demonstrated semantic errors in production but no semantic errors in written language and no evidence of a semantic deficit in spoken or written comprehension. They argued that the most likely explanation for the production of semantic errors in spoken naming was that these errors were actually driven by phonological impairment. Their logic was that cascading activation that traveled from the lexical level to the phonological level before lemma selection was complete allowed these individuals avoid lemmas whose phonemes were difficult to retrieve. Feedback from the difficult-to-activate phonological nodes led to the selection of the most highly active semantically related lemma whose phonemes could be retrieved with more ease.

In the context of the interactive activation model of word retrieval (Dell et al., 1997), it is not wholly clear how selecting the phonemes of one word would be so much more difficult than another that an individual would repeatedly make semantically related real word errors to avoid the phonemes of the difficult-to-retrieve words. Another potential issue with the conclusions made by Caramazza & Hillis (1990) concerns task equivalence. It is possible that the reported individuals did have comprehension deficits at the semantic level, but they were not picked up as readily as production deficits because the production tasks were more difficult, and/or because their comprehension impairments were more mild (see McNeil et al., 1991 for a discussion of the issue of task inequivalence when comparing different linguistic domains). This notion is consistent with the Comprehension Aphasia Test scores of our participants, four of whom demonstrated spoken comprehension scores slightly above the cut-off along with naming scores below the cut-off, but who all demonstrated comprehension deficits in conversation and haulting, slow, and effortful responses when completing the comprehension tasks. Though it is possible that some of our participants did make semantic errors that were influenced by phonological-level impairment, the argument made by Caramazza & Hillis (1990) lacks theoretical depth and seems to apply to individuals with seemingly domain-specific deficits (incongruous to that of the participants in the current study), and thus does not provide strong evidence that overcoding of semantic errors was an issue in the current study.

Along with the possible, though not probable, overcoding of semantic errors, a more important issue to consider is the possible *under*coding of phonological errors. While only nonword phonological errors were incorporated into the P score in order to isolate true sound-based impairments, Nickels & Howard (1995b) made a convincing argument that formal errors (i.e., phonologically related real words), which are considered to have a lexical component and thus to indicate both lexical- and phonological-level impairments, are often real words only by chance. Thus, it is possible that some formal errors resulted due to a solely phonological-level impairment, and phonological errors were underestimated in our sample as a result. Nickels & Howard (1995b) make their case on the basis of two points. First, the authors' observation that individuals with aphasia made formal errors more frequently on short words than on long words is a pattern that would be expected by chance, because short words have a greater number of phonologically-related neighbors than long words. Second, when Nickels & Howard (1995b) took a corpus of phonologically related (both formal and nonword) aphasic errors and randomly replaced the incorrect sounds with other phonotactically legal alternatives, they found that their pseudocorpus did not significantly differ in the numbers of phonologically related nonword and real word (i.e., formal) errors, thereby providing additional evidence that formal errors often stem from a breakdown that is isolated to the phonological level of processing. Based on the findings of Nickels & Howard (1995b), as well as the observation that errors in our study that we classified as formal on the basis of PNT rules often differed from the target in part of speech, and were thus not likely to have lexical-level involvement, P scores were recalculated to include formal errors (adjusted P score, or AP), and both AP and AP-S scores were correlated with primacy bias. Like in the main analysis, no significant results were found; thus, it is reasonable to conclude that limitations of error type coding did not affect the results.

Last but not least, the issue of the extent to which the participants in our study used lexical-semantics to supplement phonological activation when completing the word pair repetition task needs to be addressed. Because primacy/recency effect predictions depend on the dynamic interactive activation between phonological, lexical, and semantic levels of processing, and therefore, on the assumption that all three levels are involved in the VSTM process aiding word pair repetition, it is important to address the possibility that some individuals were using semantic information more than others to support their repetition. It could be that, regardless of the level of semantic or phonological processing impairment, some individuals are able to recruit the support of lexical-semantic networks more than others in a task that technically could be accomplished solely through phonological-level temporary activation. Though studies modeling repetition as well as those that study it behaviorally have demonstrated strong influence of lexical semantic processing, (see pp. 28-34), it is important to consider individual differences in the dependence on lexical-semantics in the word pair repetition task.

For individuals with very little dependence on semantic processing in the word pair repetition task, the predicted positive correlation between primacy bias and P-S score would *not* be expected, because, if they are relying mostly on phonological processing to complete the task, the second word should always be easier to repeat because its phonemes are activated closest to the time of recall. Thus, a correlation in the *opposite* direction should be expected: the more phonological errors are made relative to semantic errors, the less the primacy bias. In other words, individuals with weaker phonological processing should be more biased towards repeating the second word correctly over the first because its more recently activated phonemes should make the last word easier to repeat. This prediction is consistent with the negative (though insignificant) correlation between primacy bias and P-S score and the trend towards a negative correlation between primacy bias and P score. In addition to these findings, the observation that some of our participants made semantic errors on the repetition task, while others did not, led to the subgroup analysis in which individuals who made any number of either semantic or mixed (i.e., both semantically and phonologically related) errors on the word pair repetition task were analyzed separately from individuals who did not make any semantically related errors on the repetition task. The logic behind this post hoc analysis was that semantically related errors may be a window into the dependence on semantic information during repetition tasks, with those individuals who made semantic errors on the word pair task possibly demonstrating a greater reliance on lexical-semantic activation in the VSTM task than those who did not. When the subgroup analysis was performed, no significant correlations emerged. Thus, it does not appear that one of these subgroups was skewing the results.

Based on these findings as a whole, VSTM (as measured by word pair repetition) and word retrieval impairment type do not appear to be linked through the same intricate linguistic activation *timing* process. It is possible, however, that the same lingustic nodes are active during verbal span tasks and word retrieval tasks, but that different time courses subserve the two tasks. The discussion now turns to specific aim 3 to explore this possibility.

Specific Aim 3: Imageability & Frequency Biases and Word Retrieval Impairment

The core predictions for imageability and frequency bias correlations with word retrieval impairment type were the same as those of primacy bias and word retrieval impairment type (i.e., P-S scores): the associations should be positive because, like primacy bias, imageability and frequency biases have been linked with strong lexical-semantic processing (Martin & Saffran, 1997). The positive significant correlation between imageability bias and word retrieval impairment type found in the current study is consistent with this prediction, while the lack of a significant correlation between frequency bias and word retrieval impairment type is not. Before interpreting specific results, it is important to describe why imageability and frequency biases might arise. In the interactive activation model of word retrieval, all nodes have a resting activation value, and not all nodes rest at the same activation levels (Dell et al., 1997). What might impact a node's resting activation level? In the case of lemma or lexical nodes, the number of encounters with the word (e.g., in conversation or in written form) affect the word's lemma's resting activation level. Thus, word frequency contributes to variations in the resting state of lemma nodes. In the case of the semantic nodes, the depth and abundance of semantic features connected to a word affects that word's resting activation level. In other words, the extend to which a world is picturable, or imageable, affects the strength of that word at rest because a more extensive semantic network supports its activation (Nickels & Howard, 1994, p. 315). The higher a node's resting activation level, the more efficiently and strongly it can activate nodes at neighboring levels when its production is initiated (N. Martin & Saffran, 1992, as cited by Nickels & Howard, 1994). Additionally, nodes that are more active at rest can be kept active longer, because their higher resting levels of activation assure that they don't decay back to baseline activation as quickly. Thus, though the predictions of specific aims 2 and 3 are similar, the two aims likely get at two diffent aspects of the temporary linguistic activation process. While the primacy bias studied in specific aim 2 gets at an intricate spreading activation process that depends on activation transmission and *timing*, the frequency and imageability biases studied in specific aim 3 are likely accounted for by overall temporary activation *strength*. With this distinction in mind, results of specific aims 3a (frequency bias and word retrieval impairment type) and 3b (imageability bias and word retrieval impairment type) are interpreted, and results are then discussed together with specific aims 1 and 2 in order to explore more general questions and conclusions about the function of temporary storage in word retrieval.

The main analysis of aim 3a, which looked at the association between P-S score and frequency bias, showed no significant correlation (though the results yielded a positive correlation in the predicted direction). Besides the main correlational analysis, frequency bias was also correlated with raw P and S scores, and a significant medium positive correlation was found for frequency bias and P scores. S scores did not significantly correlate with frequency bias. To interpret these results, it is important to look closely at the results of Martin & Saffran (1997) and other related studies, and to consider the caveats of word frequency effects. Martin & Saffran's (1997) frequency bias results are discussed first. The findings were very modest, with a positive and significant finding between S-P comprehension score (note that in this case, a positive number indicates stronger processing at the semantic level) and frequency bias at the single word repetition level only. The S-P comprehension score and frequency bias at the word pair repetition level were correlated positively, though not significantly.

The current study's finding of a positive correlation between raw number of phonological errors and frequency bias is consistent with Martin & Saffran's (1997) prediction that individuals with weak phonological processing should rely more on lexical-semantics during tasks requiring temporary storage of linguistic information, and thus, highly frequent words, which are associated with strong lexical-level activation, should be easier to repeat. Thus, their prediction, as well as those of the present study, rest on the assumption that word frequency plays a role predominately at the lemma level. In Dell's model (Dell et al., 1997), lemma access marks the first step of the word retrieval process, so it makes sense that a psycholinguistic manipulation that is thought to affect the lemma level would be assumed to affect lexical-semantic processing. One study, however, demonstrated that individuals with aphasia made significantly fewer semantic *and* phonological errors on high frequency words than on low frequency words (Kittredge, et al., 2008), a finding that may be explained by a holistic influence of frequency on linguistic activation, including the activation of phonological nodes (Bastiaanse, Wieling, & Wolthuis, 2015). Additionally, Nickels & Howard (1995a) argued that the effect of word frequency on the production of semantic errors, while present in some individuals with aphasia, is modest when other variables (e.g., imageability) are controlled for, as was done in the current study. Together, these results might explain the modest associations between frequency and language processing impairment type found by Martin & Saffran (1997) and in the current study.

Can Dell's interactive activation model account for a more global effect of frequency, rather than just an effect at the lexical-semantic processing step? The lexical, or lemma, level certainly plays a major role in the first step of word retrieval, where the lemma itself is selected. In the case of a higher frequency word, the resting activation of its lemma should be boosted by its frequent selection. Thus, lemma selection of higher frequency words is likely accomplished more quickly and easily than that of low frequency words due to this boost. Might the same lemma's high frequency status influence the selection of a word's phonemes? Once the lemma node is selected, that lemma sends activation to the corresponding phonemes. A higher frequency lemma may support a stronger jolt of activation to the phonological level due to the summation of its higher resting activation level with the large additional activation jolt that supports its selection at the first step. This holistic explanation of the influence of word frequency is consistent with the modest and limited associations between frequency bias and lexical-semantic impairments found by Martin & Saffran (1997) and in the current study.

While manipulations of word frequency may be more globally influential than was initially assumed in the current study's predictions, the imageability manipulation works more specifically at the level of the semantic nodes, since the measure reflects how picturable a word is, a concept that is directly connected to how richly the word is represented (how many semantic features are associated with that word) (N. Martin & Saffran, 1992; Nickels & Howard, 1994). In the current study's preliminary analyses, a trend towards a positive correlation between imageability bias and raw number of phonological errors emerged. In other words, as numbers of phonological errors increased, individuals were more likely to demonstrate better repetition performance on high over low imageability words. No significant correlation emerged between imageability bias and raw number of semantic errors, though the negative correlation was consistent with the predicted direction. As for the main analysis of the association between imageability bias and P-S scores, a significant medium positive correlation was found ($r(22) = .429$, $p = .036$), a result consistent with the positive correlation between imageability and S-P comprehension score found by Martin & Saffran (1997).

The finding that word retrieval error type (i.e, P-S score) was related to imageability bias suggests that word pair repetition, a task classically categorized as a STM task, and word retrieval, a task classically categorized as a language production task, may share a common underlying process related to overall activation strength of linguistic representations. Perhaps temporary activation of the same representations subserves both types of tasks, so that individuals who make predominately phonological errors in naming and demonstrate a bias towards highly imageable items on repetition tasks do so in part due to diminished activation strength at the phonological level of processing. The parsimonius attribution of two commonly separated tasks (i.e., verbal span and picture naming) to the breakdown of a single activation process is further supported by the posthoc analysis, in which the P score was adjusted to include formal errors (see discussion of Specific Aim 2, pp. 98-99) for the reasoning behind this adjustment). With the adjusted P score (i.e., PA score), the significant correlations between frequency bias and number of phonological errors and between imageability bias and P-S score remained significant and of the same magnitude, while two additional significant correlations consistent with the predictions emerged. Medium correlations between imageability bias and PA score and between frequency bias and PA-S score were found, providing further support for a common underlying linguistic activation process that provides support to both word span and naming tasks. Of corse, these additional correlations should be interpreted with caution, as it is possible that the adjusted P score *over*estimated the production of purely phonological errors, as some of the formal errors that our participants produced most likely did have a lexical component. Still, taken together, the results demonstrate an intricate link between VSTM and language processing, demonstrating that the STM system that supports word retrieval is, at least in large part, language specific. Because word imageability and, to a lesser extent, word frequency manipulations have been linked to the overall activation strength of linguistic representations, the temporary activation process that underlies both repetition-based and naming tasks likely involves the activation of semantic, lexical, and phonological nodes. The efficiency of the temporary activation process (i.e., how strongly these nodes are activated and how well their activation holds up in the face of passive decay) depends in part on the resting activation strength of semantic and lexical representations (see Figure 11 for an illustration of the impact of resting activation strength on VSTM).

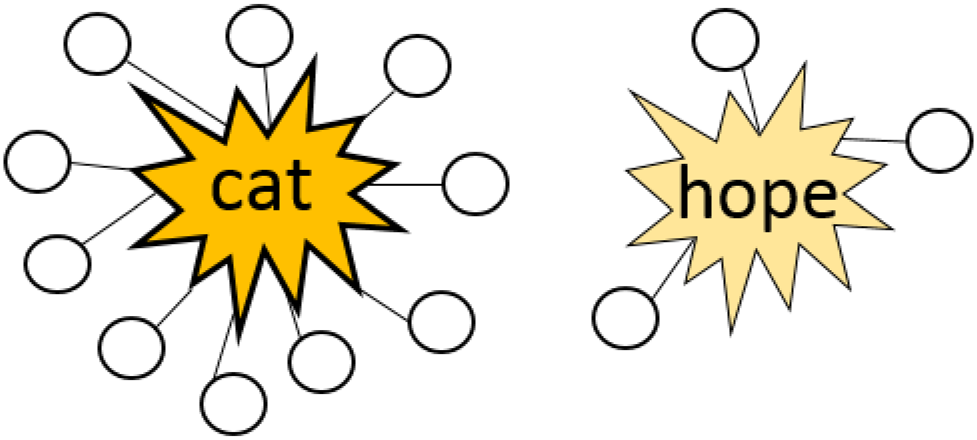


Figure 11. An illustration of high and low resting activation levels. The lemma (or lexical) node for each word is represented in the middle and the word's semantic nodes are represented with circles surrounding the lemma node. The lemma of the word *cat* has a higher resting (i.e., baseline) activation level than the lemma of *hope* because it is a more frequently activated word. Additionally, the highly imageable word *cat* is represented by a much richer semantic network than the word *hope*. The words' resting activation levels impact how strongly these words can be activated and selected when the word is briefly held in VSTM. The lemma for *cat* will be selected more easily, decay back to baseline activation more slowly, and send activation to the lower level phonological nodes more easily than the lemma for *hope* due to *cat*'s higher resting activation level (ideas based on Dell et al., 1997).

Though the main analyses and interpretations involved the relationship between word retrieval impairment type (i.e., P-S scores) and psycholinguistic biases, the lack of significant correlations between S scores and both frequency and imageability bias scores, a result that is inconsistent with the findings of Martin & Saffran (1997) for semantic language comprehension score and imageability bias, deserves mention. It is important to remember that S and P are both raw scores that do not take into account relative impairment, and thus are less useful for the study of links between word retrieval impairment type and susceptibility to psycholinguistic manipulations. Still, the current study demonstrated several significant associations with P scores, so the lack of any significant associations with S scores deserves some exploration. One important difference to take into account is that Martin & Saffran (1997) analyzed imageability biases for the first and second targets of each word pair separately, and only found a link between imageability bias on the *second* word and semantic score (the same thing was true for phonological score). It is possible that a re-analysis of each word in the pair separately would have yielded different results in the current study.

Two other methodological issues that are inherent in picture naming tasks may also have contributed to the null findings between S scores and both frequency and imageability bias scores. One is that in picture naming tasks, the activation of semantic features is initiated in part by looking at the picture to be named. Therefore, semantic nodes receive a large boost due to the presentation of a picture, while the activation of phonological nodes needs to be initiated by the individual completing the task. Such a discrepancy may cause the raw number of semantic errors to be lower than if the individual was attempting to produce the same word in conversation, because he/she could not rely on the assistance of a picture to initiate the activation of semantic nodes, thus making it harder to produce the correct word. A second methodological issue involves the inevitable problem of visual confusions on picture naming tasks. For example, some of the participants called the picture whose target was *bowl* a *cup*, a possible visual confusion due to the somewhat ambiguous picture representing the word *bowl.* Consistent with the PNT rules, this response was coded as a semantic error, though in reality, the response could have been the result of a visual confusion that could just as likely have been made by a neurologically healthy individual. In this case, the raw number of semantic naming errors could have been overestimated. This confounding factor is difficult to avoid even in thoroughly normed and psychometrically sound naming tests such as the PNT.

A much more general and related issue is that each experimental task must be looked at critically in terms of its stimulus and response in the context of the more natural act of interest to the researcher. Picture naming, for example, is a window into the much more natural task of retrieving words in the midst of a conversation. Though some aspects of the laboratory task are similar to this more natural process, such as the required coactivation of various linguistic nodes, other aspects, such as the presentation of a picture and the allowance of 30 seconds to respond, deviate significantly from the usual lack of visual support in a conversation and the much quicker speed at which word production in a typical conversation operates. Though laboratory tasks are an extremely useful way to study language in a carefully controlled manner in order to make conclusions about the processes underlying its breakdown, they should always be analyzed with their limitations and deviations from language processing in a natural environment in mind. Conducting language research with this large caveat in mind becomes doubly critical for clinical researchers who seek to contribute to a knowledge base in order to eventually influence the development of better language assessments and treatments and, ultimately, improve people’s lives.

Specific Aims 1, 2, and 3: Holistic Conclusions

The results interpreted above can be understood most fully by thinking about what all three aims can tell us together about the function of VSTM in word retrieval. The first aim honed in on the specificity of the link between word span length and word retrieval accuracy, a finding that set the stage to explore the process(es) underlying this link. The null results of specific aim 2, which looked at the link between primacy bias in word pair repetition and word retrieval impairment type, were inconsistent with the study's predictions, while the associations between imageability/frequency biases in word pair repetition and word retrieval impairment type (specific aim 3) were mostly consistent with the predictions. While primacy bias likely depends on intricate activation timing, which gives rise to differential activation patterns for words in different positions of a string to be repeated, imageability and frequency biases are likely more dependent on the resting strength of linguistic representations, which allows certain lexical items to transmit and maintain temporary activation more efficiently than others. Both intricate timing and overall strength of linguistic representations are critical to a linguistic STM system modeled through Dell's interactive activation model. The results taken together seem to suggest that, at least in the word retrieval task used in this study, production of words was dependent more on the overall strength of linguistic representations than on the intricate timing of activation transmission. Maybe it is the process of temporary activation of the same linguistic representations that supports both word retrieval and word pair repetition, but the timing of the activation's spread through the network differs for the two tasks classically separately categorized as STM and language tasks. Perhaps it is time to rethink this classic distinction, to accept the idea that access to linguistic representations in word retrieval is dependent on a linguistically-specified temporary activation process, and to let go of the idea that a STM system that is domain-general and/or wholly separable from the process it supports exists.

The Relationship between Language Production and Comprehension Tasks

Though an exploration of links between language production and comprehension processes is not a direct goal of this study, the question of to what extend these processes are related is strongly relevant to the study of the VSTM process underlying the production of words. The previous work of Martin & Saffran (1997) studied the VSTM process underlying the comprehension of sounds and words, and comparing those results with the results of the present study yields some interesting insights into whether the same VSTM system may support both language production and comprehension. While Martin & Saffran (1997) found that both positional (i.e., primacy/recency) and psycholinguistic (i.e., imageability/frequency) biases correlated with comprehension impairment type, the current study yielded significant correlations when looking at links between psycholinguistic biases and word retrieval impairment type only. In the current study, positional biases have been argued to depend on a precisely timed spread of linguistic activation, while imageability and frequency biases have been argued to arise because representations that are more strongly activate at rest can more strongly and efficiently support the activatation of words in online word retrieval. Thus, the biases likely relate to different subcomponents of the VSTM system.

The contrasting findings of the current study and that of Martin & Saffran (1997) provide some insight into the extent to which language production and comprehension processes rely on the same VSTM mechanisms. Language comprehension tasks such as those used to get at language comprehension impairment type (e.g., rhyme judgments, synonymy judgments) likely rely on much more intricate timing between stimulus and response than the word retrieval task used in the current study. This contrast is consistent with the finding that word retrieval impairment type did not correlate with primacy bias in the current study, while comprehension impairment type did correlate with primacy bias (Martin & Saffran, 1997). Along with these results, both the current study and Martin & Saffran (1997) found significant correlations in the predicted direction that linked language impairment type and susceptibility to imageability and frequency manipulations. The fact that the intricate comprehension tasks used in Martin & Saffran (1997) yielded significant correlational results with primacy bias while our word retrieval task did not, and that both yielded significant correlational results with imageability and frequency biases, is a window into the possible subservance of both comprehension and production by a partially joint VSTM system. The collective results suggest that the VSTM link might be in the overall strength of linguistic representations: perhaps both language comprehension and production rely on the temporary activation of the same linguistic representations, but the timing of that activations' spread differs for the two language domains.

Of course, the discussion above rests on the results of two different studies, with two different participant groups, and a study that replicates both the comprehension and production tasks in a single group of participants will yield important additional insight into the similarities and difference in VSTM function in the two domains. Though not analyzed in the current study, comprehension data that includes both semantic and phonological tasks has been collected from 17 of the participants in the current study, and the raw data is included in Appendix IV. Thus, though the conclusion that language comprehension and production rely on a partially shared VSTM system that functions through the activation of shared linguistic representations is premature, the extant data is consistent with this view, and future analyses will provide additional insights into this topic. Additionally, several other studies that have explored the connection between language comprehension and production tasks demonstrated results consistent with the idea of partially connected linguistic networks. Semantic errors in comprehension and production tasks are frequently found to correlate significantly (Nickels, 1997; Nickels & Howard, 1994, p. 297), and the majority of models argue for the reliance of production and comprehension processes on the same semantic representations (N. Martin & Saffran, 2002, p. 108). The nature of the link between phonological representations in production and comprehension tasks is more controversial, with some studies showing significant negative correlations between numbers of phonological errors in production and accuracy scores on phonological comprehension measures (e.g., N. Martin & Saffran, 2002) and others showing no significant correlations between the two types of tasks (e.g., Nickels & Howard, 1995b). The general conclusion seems to be that though a link between phonological representations in production and comprehension likely exists, the relationship is not necessarily a straightforward one. Together, these extant findings are consistent with the current study's suggestion of language production and comprehension systems that are *partially* linked through the temporary activation of the same linguistic representations. The extent to which each representational level (i.e., semantic, lemma, phonological, etc.) is shared in production and comprehension is an important issue that is beyond the scope of this work.

Clinical Implications

Though the current study's work is theoretical in nature, its motivation stems from a desire to eventually influence clinical practice. Though the present work requires replication and extension before it can be applied in clinical settings, its results demonstrate potentially beneficial clinical uses. If a clinician is to understand an individual's aphasia comprehensively, he/she must take into account the linguistic level(s) at which the impairment lies. The analysis of speech errors is one way to tap into the impairment at the production level; however, it is also a very time-consuming and difficult process. The current study's results suggest that a repetition span measure may possibly get at the same impairment. If the clinician were to administer a word pair repetition task and calculate an imageability bias score, might this analysis provide valuable insight into the nature of language breakdown when it is not time- or cost-effective to perform word retrieval error analyses? For example, if a patient demonstrated a large bias towards the correct repetition of highly imageable words, it may mean that his/her word retrieval impairment stems mostly from a breakdown of temporary activation at the phonological level. On the other hand, if a patient demonstrated little or no imageability bias, that may mean that his/her word retrieval impairment stems mostly from a breakdown of temporary activation at the semantic level. Such observations would help guide impairment-level, VSTM-based word retrieval treatments and contribute to a small but growing literature on process-driven treatments of aphasia (for a review, see Salis, et al., 2015). Of course, much more replication and extension work needs to be done before the association between imageability bias and word retrieval impairment type found in this study is clinically applied in this way. Nevertheless, the idea that there may be a more efficient and feasible way to get at the nature of word retrieval breakdown than the time consuming and elusive speech error coding process is an encouraging one.

Limitations and Future Directions

Several methodological considerations and limitations deserve mention. In terms of participant selection, one challenge in all studies of aphasia is to determine the criteria for exclusion of participants with concomitant apraxia of speech. Often, researchers must consider the trade-off between including fewer participants that can more confidently be argued to have a purely aphasic impairment and including a larger sample of participants, some of whom demonstrate a mild-moderate apraxia of speech. In the current study, several participants appeared to have apraxia, and though multiple steps were taken to minimize the influence of their apraxia on the study's results (i.e., counting distortions and schwa insertions as correct and allowing certain distorted substitutions that have been convincingly argued to occur at the motor level of the production process), it is important to consider the possibility that these individuals' motor errors could have been mistaken for phonological errors. All studies of language production that choose to include individuals who are also apraxic must consider this issue, document these patients' motor impairments carefully, and determine ways in which their experimental protocol can minimize their influence, if possible. If the patient has a more severe apraxia of speech that makes it exceedingly difficult to test the process that the experimenter wishes to tap into, they likely need to be excluded. Though I believe that the current study took adequate precautions to minimize the influence of motor speech errors on the study's results, the possibility of its influence must nevertheless be noted.

An important and complex theoretical issue to consider in the current study and in future work is the limitation of using a repetition-based task to measure VSTM. The parsimonious use of interactive activation principles to describe both repetition and word retrieval frames the reasoning behind the current study's predictions and results. Repetition is, in some ways, a more complicated process to describe, because it involves both comprehension and production steps, along with an intermediate step of retaining the presented target before it is repeated. The model of VSTM at the heart of this study centers on the nature of temporary linguistic activation during the intermediate step between listening to a word pair and producing it. Though this intermediate timepoint is most critical to the study of VSTM because it represents a moment during which linguistic representations are purely held in memory and devoid of overt influences of comprehension and production, the model does not account for the complex interactions between this moment and the stimulus and response modalities required by the repetition task (i.e., auditory comprehension and production). While it may be impossible to make the VSTM task purer (e.g., using a pointing rather than a repetition modality to eliminate the language production component) and still ask the questions that are the focus of this study, the limitations of the modality through which a process (in this case, VSTM) is tested should always be considered.

In terms of task design considerations, both the word retrieval and word pair repetition tasks may benefit from some modification in future studies. To make the word retrieval task more consistent with the very rapid timeframe of language production in natural settings, several changes can be made. One possible alteration is the decrease of the allotted response time from 30 seconds to a shorter period. Such a task might better approximate the rapid spread of activation required to produce a word in conversation, and may possibly yield results more comparable with that of Martin & Saffran's (1997) language comprehension-VSTM association study. Another possibility might be to allow the picture to be named to only be present for a brief period of time (as is done in TALSA subtest 7, N. Martin, 2012) to require more active generation of features at the conceptual-semantic level of processing, thus more closely simulating conversational word retrieval.

To maximize the possible future clinical utility of the word pair repetition task, a future methodological recommendation may be to re-analyze the existing data to determine how many trials it takes to accurately estimate imageability and frequency bias scores. In the current study, participants completed 240 total trials of words of varying imageability and frequency before bias scores were calculated. The task took an average of about 30 minutes to complete (not inluding the break given at the halfway point), and many participants demonstrated fatigue when completing this task. Even if results of future replications were to be consistent with the idea that the word pair repetition task is a reliable way to get at word retrieval impairment type in individuals with aphasia, a task of this length would not be practical in a typical clinical setting with limited assessment time. If fewer trials might yield essentially the same imageability and frequency bias scores as seen with the full set of trials used in this study, the potential clinical utility of this measure would increase significantly.

Last but not least, an exciting future direction of this work is the examination of associations between the word pair (i.e., VSTM) task used in the current study and both production and comprehension impairment types in the same group of participants. Though several theoretically and clinically interesting similarities between the current study and that of Martin & Saffran (1997), which focused on comprehension, were noted, clearer conclusions about the function of VSTM in language production and comprehension could be made if both word retrieval and comprehension measures were administered to the same group of participants. Semantic (i.e., word-to-picture matching and synonymy judgments) and phonologic (i.e., minimal pair discrimination and rhyming judgments) comprehension measures have been collected for 17 of the current study's participants, allowing for a more precise future exploration of the shared and divergent processes supporting language production and comprehension.

Conclusion to the Dissertation

A desire to understand the function of VSTM in word retrieval, a process that is ubiquitously disrupted in aphasia, was the heartbeat of this dissertation. Process-driven investigations of language breakdown are important because they tap into the mechanisms that underlie the linguistic impairments, and these mechanisms must be understood in order to effectively evaluate and treat individuals with aphasia. Theoretically, the current study demonstrated some interesting links between word retrieval impairment type (i.e., relatively more semantic or phonological) and word pair repetition (i.e., a measure of VSTM), supportive of the idea that a linguistically-specified STM system underlies word retrieval. Though, at the present moment, the study adds primarily to the theoretical realm of knowledge, it was designed with eventual clinical application in mind. Thus, anomia, the most omnipresent symptom of aphasia, was chosen as the focus, with the hope to potentially determine a more efficient way to tap into the nature of word retrieval breakdown that might more easily provide some of the valuable information that speech error analysis has proven to provide. It is my strong belief that investigations such as this one, which are primarily theoretical in nature, must consider their potential impact at the clinical level during the inception of the work rather than only after the results of the study have been interpreted. If all theoretical aphasia research had at its heart the collective ultimate goal of informing clinical practice, I believe we would get to the difficult step of clinical application more rapidly. Persistent clinically-driven investigations of processes (e.g, STM and attention) that aid access to language representations, processes deemed "nonlinguistic" in the past but that have been repeatedly shown to be anything but, have the potential to inform and influence clinical practice, eventually leading to more impairment-specific and generalizable treatments for aphasia.

References

Albert, M. L. (1976). Short-term memory and aphasia. *Brain and Language, 3*, 28-33. doi: 10.1016/0093-934X(76)90003-1

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64-88. doi:10.1016/j.jml.2006.02.002

Aphasia Fact Sheet (2015). from http://www.aphasia.org/aphasia-resources/aphasia-factsheet/

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 89-195). New York: Academic Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A., & Hitch, G. J. (1974). Working Memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). San Diego: Academic Press.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365. doi:10.1080/14640746608400055

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology, 20*, 249-264. doi:10.1080/14640746808400159

Baddeley, A. D. (1986a). The articulatory loop. *Working memory* (pp. 75-107). Oxford: Oxford University Press.

Baddeley, A. D. (1986b). Beyond the modal model. *Working memory* (pp. 18-29). Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*, 485-493. doi: 10.1037/0894-4105.8.4.485

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589. doi: 10.1016/S0022-5371(75)80045-4

Ballard, K. J., Granier, J. P., & Robin, D. A. (2000). Understanding the nature of apraxia of speech: Theory, analysis, and treatment. *Aphasiology, 14,* 969-995. doi: 10.1080/02687030050156575

Bastiaanse, R., Wieling, M., & Wolthuis, N. (2015). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology*. doi: 10.1080/02687038.2015.1100709

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology, 58*, 667-692. doi: 10.1080/02724980443000142

Benson, D. F. (1979). *Aphasia, alexia and agraphia.* New York: Churchill Livingstone.

Benson, D. F. (1988). Anomia in Aphasia. *Aphasiology, 2*, 229-235. doi: 10.1080/02687038808248915

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*, 122-125. doi: 10.3758/BF03200768

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*, 467-482. doi: 10.1037/H0061096

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods, 41,* 977-990. doi: 10.3758/BRM.41.4.977

Cahana-Amitay, D., & Albert, M. L. (2015). Neuroscience of aphasia recovery: the concept of neural multifunctionality. *Current Neurology and Neuroscience Reports, 15,* doi: 10.1007/s11910-015-0568-7

Caplan, D., Waters, G., & Howard, D. (2012). Slave systems in verbal short-term memory. *Aphasiology, 26*, 279-316. doi:10.1080/02687038.2011.642795

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex, 26,* 95-122. doi: 10.1016/S0010-9452(13)80077-9

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition*, *37*, 205-223. doi: 10.1006/brcg.1997.0970

Chang, A. (2015). Computer Program to Calculate Sample Size Requirement for Estimating the Correlation Coefficient. StatsToDo, 2014, from http://www.statstodo.com/SSizCorr_Pgm.php#

Chang, D. F. (1995). Opthalmic examination. In D. Vaughan, T. Asbury & P. Riordan-Eva (Eds.), *General Ophthalmology*. Norwalk: Appleton & Lange.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2): Laurence Erlbaum Associates.

Coltheart, M. (1981). The Mrc Psycholinguistic Database. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 33*(Nov), 497-505. doi: 10.1080/14640748108400805

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1283-1296. doi: 10.1037/0278-7393.26.5.1283

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language, 94*, 94-103. doi: 10.1016/j.bandl.2004.12.002

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*, 163-191. doi: 10.1037/0033-2909.104.2.163

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge: Cambridge University Press.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185. doi: 10.1017/S0140525X01003922

Crosson, B. (2000). Systems that support language processes: verbal working memory. In S. E. Nadeau, L. J. Gonzalez Rothi & B. Crosson (Eds.), *Aphasia and Language: Theory to Practice* (pp. 399-418). New York: The Guilford Press.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 345-361. doi: 10.1037/0278-7393.25.2.345

Darley, F. L. (1982). Aphasia. Philadelphia: W.B. Saunders.

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex, 11*, 341-354. doi: 10.1016/S0010-9452(75)80026-8

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283-321. doi:10.1037/0033-295X.93.3.283

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56,* 490-520. doi: 10.1016/j.jml.2006.05.007

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42,* 287-314. doi: 10.1016/0010-0277(92)90046-K

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104,* 801-838. doi: 10.1037//0033-295X.104.4.801

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition, 102,* 219-260. doi: 10.1016/j.cognition.2005.12.010

Goodglass, H., Kaplan, E., & Barressi, B. (2000). *Boston Diagnostic Aphasia Examination* (3 ed.). San Antonio: Pearson.

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: spontaneous and structured contexts. *Brain and Language, 82,* 113-145. doi: 10.1016/S0093-934X(02)00001-9

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language, 38,* 313-338. doi: 10.1006/jmla.1997.2547

Gvion, A., & Friedmann, N. (2012). Phonological short-term memory in conduction aphasia. *Aphasiology, 26,* 579-614. doi:10.1080/02687038.2011.643759

Howard, D., Swinburn, K., & Porter, G. (2010). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. *Aphasiology, 24*, 56-74. doi: 10.1080/02687030802453202

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685-701. doi: 10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, M., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 23*, 1217-1232. doi: 10.1037/0278-7393.23.5.1217

Itoh, M., Sasanuma, S., & Ushijima, T. (1979). Velar movements during speech in a patient with apraxia of speech. *Brain and Language, 7*, 227-239. doi: 10.1016/0093-934X(79)90019-1

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning Memory and Cognition, 24,* 1256-1274. doi: 10.1037//0278-7393.24.5.1256

Kasselimis, D. S., Simos, P. G., Economou, A., Peppas, C., Evdokimidis, I., & Potagas, C. (2013). Are memory deficits dependent on the presence of aphasia in left brain damaged patients? *Neuropsychologia, 51*, 1773-1776. doi: 10.1016/j.neuropsychologia.2013.06.003

Kay, J., Coltheart, M., & Lesser, R. (1992). *Psycholinguistic Assessments of Language Processing in Aphasia*. Hove: Psychology Press.

Kendall, D., del Toro, C., Nadeau, S. E., Johnson, J., Rosenbek, J. and Velozo, C. (2010). The development of a standardized assessment of phonology in aphasia. Poster presented at the Clinical Aphasiology Conference, Isle of Palm, SC.

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

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25*, 463-492. doi: 10.1080/02643290701674851

Koenigs, M., Acheson, D. J., Barbey, A. K., Solomon, J., Postle, B. R., & Grafman, J. (2011). Areas of left perisylvian cortex mediate auditory-verbal short-term memory. *Neuropsychologia, 49*, 3612-3619. doi:10.1016/j.neuropsychologia.2011.09.013

Kucera, H., & Francis, W. N. (1967). Providence: Brown University Press.

Landis, J. R., & Koch, G. G. (1977). The Measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174. doi: 10.2307/2529310

Lang, C. J. G., & Quitz, A. (2012). Verbal and nonverbal memory impairment in aphasia. *Journal of Neurology, 259*, 1655-1661. doi:10.1007/s00415-011-6394-1

Laures-Gore, J., Marshall, R. S., & Verner, E. (2011). Performance of individuals with left hemisphere stroke and aphasia and individuals with right brain damage on forward and backward digit span tasks. *Aphasiology, 25*, 43-56. doi: 10.1080/02687031003714426

Leff, A. P., Schofield, T. M., Crinion, J. T., Seghier, M. L., Grogan, A., Green, D. W., & Price, C. J. (2009). The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. *Brain, 132,* 3401-3410. doi: 10.1093/brain/awp273

Madigan, S. (1980). The serial position curve in immediate serial-recall. *Bulletin of the Psychonomic Society, 15*, 335-338. doi: 10.3758/BF03334550

Martin, N. (2000). Word processing and verbal short-term memory: How are they connected and why do we want to know? *Brain and Language, 71*, 149-153. doi: 10.1006/brln.1999.2237

Martin, N. (2012). Managing communication deficits associated with memory disorders. In R. Peach & L. Shapiro (Eds.), *Cognition and Acquired Language Disorders: A Process-Oriented Approach* (pp. 275-297). St. Louis: Mosby.

Martin, N., & Ayala, J. (2004). Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. *Brain and Language, 89*, 464-483. doi: 10.1016/j.bandl.2003.12.004

Martin, N., Ayala, J., & Saffran, E. M. (2002). Lexical influences on serial position effects in verbal STM span in aphasia. *Brain and Language, 83*, 92-95.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language, 47*, 609-660. doi: 10.1006/brln.1994.1061

Martin, N., & Gupta, P. (2004). Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology, 21*, 213-228. doi:10.1080/02643290342000447

Martin, N., & Saffran, E. M. (1990). Repetition and verbal STM in transcortical sensory aphasia: a case study. *Brain and Language, 39*, 254-288. doi: 10.1016/0093-934X(90)90014-8

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: evidence from a single case study. *Brain and Language, 43*, 240-274. doi: 10.1016/0093-934X(92)90130-7

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641-682. doi: 10.1080/026432997381402

Martin, N., & Saffran, E. M. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology, 16*, 107-150. doi: 10.1080/02687040143000447

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language, 52*, 83-113. doi: 10.1006/brln.1996.0005

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 3-29. doi: 10.1006/jmla.1999.2637

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*, 346-362. doi: 10.1016/j.cognition.2011.11.003

Mauszycki, S. C., Dromey, C., & Wambaugh, J. L. (2007). Variability in apraxia of speech: A perceptual, acoustic, and kinematic analysis of stop consonants. *Journal of Medical Speech-Language Pathology, 15*, 223-242. doi: 10.1080/02687030903438516

McNeil, M. R., Odell, K., & Tseng, C. H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical Aphasiology, 20*, 21-39.

McNeil, M. R., & Pratt, S. R. (2001). Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. *Aphasiology, 15*, 901-911. doi: 10.1080/02687040143000276

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97. doi: 10.1037/h0043158

Navarette, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*, 359-377. doi: 10.1016/j.jml.2005.05.001

Neath, I., & Suprenant, A. M. (2003). Working Memory. *Human Memory* (pp. 76-78). Belmont: Wadswirg.

Nickels, L. (1997). Semantic Deficits and Semantic Errors. *Spoken Word Production and its Breakdown in Aphasia* (pp. 115-116). East Sussex: Psychology Press Ltd.

Nickels, L., & Howard, D. (1994). A frequent occurence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology, 11*, 289-320. doi: 10.1080/02643299408251977

Nickels, L., & Howard, D. (1995a). Aphasic naming: What matters? *Neuropsychologia, 33*, 1281-1303. doi: 10.1016/0028-3932(95)00102-9

Nickels, L., & Howard, D. (1995b). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex, 31*, 209-237. doi: 10.1016/S0010-9452(13)80360-7

Odell, K., McNeil, M. R., Rosenbek, J. C., & Hunter, L. (1990). Perceptual characteristics of consonant production by apraxic speakers. *Journal of Speech and Hearing Disorders, 55*, 345-359. doi: 10.1044/jshd.5502.345

Paivio, A., Yuille, J., & Madigan, S. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph, 76,* 1-25. doi: 10.1037/h0025327

Potagas, C., Kasselimis, D., & Evdokimidis, I. (2011). Short-term and working memory impairments in aphasia. *Neuropsychologia, 49*, 2874-2878. doi:10.1016/j.neuropsychologia.2011.06.013

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*, 460-499. doi: 10.1037/0033-295X.107.3.460

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford Psychologists.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology, 24*, 121-133.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1019-1034. doi: 10.1037/0278-7393.28.6.1019

Rosenbek, J. C., LaPointe, L. L., & Wertz, R. T. (1989). Nosology: What aphasia is and what it is not. *Aphasia: A clinical approach* (pp. 34-54). Austin: PRO-ED.

Ryan, W. J. (1982). *The nurse and the communicatively impaired adult*. New York: Springer Publishing Company.

Saffran, E. M. M., N. (1990). Neuropsychological evidence for lexical involvement in short-term memory. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge: Cambridge University Press.

Salis, C., Kelly, H., & Code, C. (2015). Assessment and treatment of short-term and working memory impairments in stroke aphasia: a practical tutorial. *International Journal of Language and Communication Disorders, 50*, 721-736. doi:10.1111/1460-6984.12172

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language, 54*, 199-227. doi:10.1016/j.jml.2005.10.002

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language, 54*, 228-264. doi:10.1016/j.jml.2005.10.001

Shelton, J. R., Martin, R. C., & Yaffee, L. S. (1992). Investigating a verbal short-term memory deficit and its consequences for language processing. In D. I. Magnolin (Ed.), *Cognitive neuropsychology in clinical practice* (pp. 131-167). New York: Oxford University Press.

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*. New York: Psychology Press.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychol Bull, 133*, 1038-1066. doi: 10.1037/0033-2909.133.6.1038

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 1.4 [Data File] Available from http://www.iphod.com

Verhaegen, C., Piertot, F., & Poncelet, M. (2013). Dissociable components of phonological and lexical-semantic short-term memory and their relation to impaired word production in aphasia. *Cognitive Neuropsychology, 30*, 544-563. doi:10.1080/02643294.2013.876398

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*, 422-488. doi:10.1016/j.cogpsych.2003.09.001

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*, 1256-1271. doi: 10.1037/0278-7393.25.5.1256

Watkins, M. J. (1972). Locus of the modality effect in free-recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 644-648. doi: 10.1016/S0022-5371(72)80048-3

Watkins, M. J., & Watkins, O. C. (1973). Postcategorical status of the modality effect in serial recall. *Journal of Experimental Psychology, 99*, 226-230. doi: 10.1037/H0034651

Wilshire, C. E., Keall, L. M., & O'Donnell, D. J. (2010). Semantic contributions to immediate serial recall: Evidence from two contrasting aphasic individuals. *Neurocase, 16*, 331-351. doi: 10.1080/13554791003620256

Appendix I. Summary of error coding guidelines, adapted from the Philadelphia Naming Test (Roach, Schwartz, Martin, Grewal, & Brecher, 1996)

Error Coding Step	Rules
1) Determine whether the response that constitutes first complete attempt. This is the response that will be coded.	**1)** Minimal acceptable response is a consonant-vowel or vowel-consonant production. Schwas do not count as vowels. **2)** Self-interrupted responses are not considered the first complete attempt. **3)** Response has to meet one of the following criterions: a) Produced with downward or upward intonation with or without a pause before next production. b) Produced with level intonation with noticeable (~1s) pause before next production.
2) Determine whether the response is phonologically similar to the target.	**1)** Response has to meet one of the following criterions: *a*) Shares stressed vowel, initial, or final phonemes *b*) Shares two or more phonemes at any position *c*) Shares one or more phonemes at corresponding syllable and word positions (aligned from left to right) **2)** The additions/exceptions to the rule above are: *a*) Shwa phonemes and plural morphemes are not considered when determining whether the error is phonologically similar to target. *b*) Consonant clusters are considered one unit when determining syllable position. If a phoneme that is part of a cluster is in a corresponding syllable position with a target’s phoneme, the error is considered to be phonologically similar to the target.

3a) If response is phonologically similar to the target, determine if it should be coded as a phonological, formal, or mixed error. **3b)** If response is not phonologically similar, determine if it should be coded as a semantic error or placed in the "other errors" category.	**1) Phonological errors** are phonologically similar nonword errors that are not semantically related to the target *(rules for determining semantic similarity are included below, see #4).* **2) Formal errors** are phonologically similar real word errors that are not semantically related to the target. **3) Mixed errors** are phonologically similar real word errors that are also semantically related to the target. **4) Semantic errors** are real word noun errors that are related to the target in one of the following ways: a) synonym (e.g., dog → canine) b) category coordinate (e.g., dog → cat) c) superordinate (e.g., dog → animal) d) subordinate (e.g., dog → puppy) e) associated (e.g., dog → leash) f) diminutives (e.g., dog → doggie) **5) Other errors** are responses that do not fit into the four categories listed above.

Appendix II. Simple linear regression results for main analyses (a-c) and posthoc analyses (d-g). Significant findings are starred.

Main Analyses

a. *Predictor*: Word pointing span length *Outcome*: PNT accuracy

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
PNT accuracy	*.54*	*.52*	*25.46(1,22)**				
Intercept				117.67	(6.63)	17.75	*
Word Span Length				34.17	(6.77)	5.05	*

b. *Predictor*: Frequency Bias *Outcome*: P score

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
P Score	*.17*	*.13*	*4.34(1,22)**				
Intercept				8.38	(2.16)	3.89	*
Frequency Bias				4.59	(2.20)	2.08	*

c. *Predictor*: Imageability Bias *Outcome*: P-S score

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
P-S Score	*.18*	*.15*	*4.96(1,22)*				
Intercept				-1.33	(2.62)	-0.51	
Imageability Bias				5.95	(2.67)	2.23	*

Posthoc Analyses

d. *Predictor*: Frequency Bias *Outcome*: PA score

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
PA Score	*.22*	*.19*	*6.35(1, 22)**				
Intercept				14.13	(3.09)	4.58	*
Frequency Bias				7.94	(3.15)	2.52	*

e. *Predictor*: Imageability Bias *Outcome*: PA score

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
PA Score	*.20*	*.17*	*6.35 (1, 22)**				
Intercept				14.13	(3.13)	4.52	*
Imageability Bias				7.57	(3.19)	2.37	*

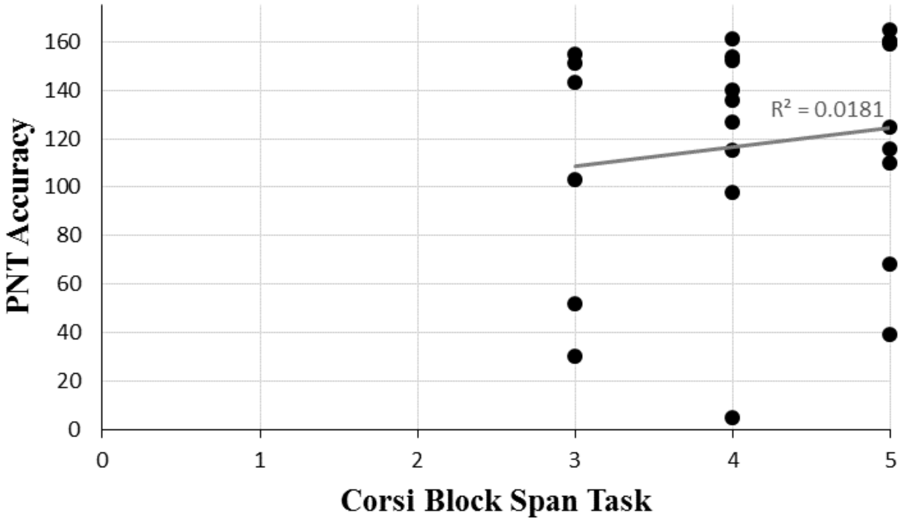
f. *Predictor*: Frequency Bias *Outcome*: PA-S score

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
PA-S Score	*.17*	*.13*	*4.56(1, 22)**				
Intercept				4.42	(3.58)	1.23	
Frequency Bias				7.82	(3.66)	2.14	*

g. *Predictor*: Imageability Bias *Outcome*: PA-S score

	Standard Regression						
	R^2_{total}	R^2_{Adj}	F_{total}	*b*	*(SE)*	*t*	
PA-S Score	*.24*	*.20*	*6.81(1, 22)**				
Intercept				4.42	(3.44)	1.28	
Imageability Bias				9.17	(3.51)	2.61	*

Appendix III. Correlation graph: Insignificant relationship between PNT word retrieval accuracy and nonverbal span (i.e., Corsi Block Span Task)



Appendix IV. Semantic and phonological comprehension data. Participant numbers correspond to those listed in the tables and figures. The highest possible score for each test is listed in parentheses. Comprehension data was collected for 17 out of 24 participants.

Participant ID	**PALPA Spoken Word-Picture Matching (40)**	**PALPA Auditory Synonym Judgments (60)**	**PALPA Minimal Pairs (72)**	**SAPA Nonword Rhyming (22)**
01	40	57	57	19
02	32	39	42	17
03	40	49	60	19
04	39	55	67	14
05	39	47	69	19
06	29	34	42	16
07	40	55	64	20
08	40	56	70	22
09	39	46	67	16
11	35	45	62	14
12	40	59	64	16
14	39	50	68	19
17	38	60	61	19
19	39	51	70	22
20	39	52	60	18
21	40	52	69	18
22	38	55	68	17
AVE	38.00	50.71	62.35	17.94
SD	3.12	6.94	8.62	2.33

PALPA = Psycholinguistic Assessments of Language in Aphasia (Kay, Coltheart, & Lesser, 1992)
SAPA = Standardized Assessment of Phonology in Aphasia (Kendall et al., 2010)