Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

11-22-2021 3:00 PM

The Role Of Working Memory And Linguistic Knowledge On Language Performance

Theresa Ai Vy Pham, *The University of Western Ontario*

Supervisor: Archibald, Lisa M. D., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Health and Rehabilitation Sciences
© Theresa Ai Vy Pham 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Speech and Hearing Science Commons

Recommended Citation

Pham, Theresa Ai Vy, "The Role Of Working Memory And Linguistic Knowledge On Language Performance" (2021). *Electronic Thesis and Dissertation Repository*. 8236.
https://ir.lib.uwo.ca/etd/8236

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

The language system is highly flexible and draws on distinct but interconnected cognitive mechanisms, including verbal working memory and long-term linguistic knowledge. Verbal working memory is the ability to manipulate verbal information in mind. Long-term linguistic knowledge refers to our knowledge of the language (i.e., phonology, semantics, syntax), stored in long-term memory. The close interaction between verbal working memory and linguistic knowledge highlights a pressing need to investigate the construct of verbal working memory, its separability and its relationship with linguistic knowledge. To understand the way working memory influences and interacts with language abilities in children and adults, I ask the following questions: Are verbal working memory and language separable constructs? And, does verbal working memory operate within a dynamic network of cognitive systems including the language network? In Chapter 2, I examined whether working memory and linguistic abilities could be teased apart using the same language task, namely a modified Token Test. Indeed, factors related to working memory and linguistic abilities explained performance on our modified Token Test and were differentially related to other language measures. Despite evidence of separability, it must be acknowledged that verbal working memory and language processing are highly intertwined. Chapters 3 and 4 investigated this interrelationship in detail. Specifically, I used experimental tasks to delineate the involvement of phonological and semantic representations in the maintenance of verbal items (words, sentences) in working memory. In Chapter 3, I used a novel word recognition paradigm and found separable phonological and semantic effects on immediate memory, with semantic processing supporting long-term retention. These findings confirmed that both phonological and semantic information were readily activated and accessed when a word is encountered and processed. Chapter 4 further evaluated the interplay between different cognitive processes underlying verbal working memory in the context of sentence recall. Similarly, results supported the idea that multiple representations influence performance, but their contributions differ. Semantic processing was beneficial for both immediate and long-term memory whereas phonological processing had more immediate benefits. Finally, in the concluding chapter, I discuss the importance of these results for models of verbal short-term memory and highlight some potential implications for clinical practice.

Keywords

Verbal working memory, Linguistic knowledge, Phonology, Semantics, Language performance, Language development

Summary for Lay Audience

Language processing is influenced by many cognitive factors including verbal working memory, which is the ability to hold and process some aspect of language in mind, and linguistic knowledge, which is our knowledge about the rules of language. Think about what it means to hold a word or sentence in mind. This means not only knowing the phonological word form (or speech sounds that make up the word) but also having rich linguistic knowledge such as semantic information (or meaning-based information) associated with it. It would be difficult to fully understand what factors are impacting language performance without considering the role of both working memory and linguistic knowledge. In this thesis, I aim to investigate the construct of verbal working memory, its separability and its relationship with long-term linguistic knowledge. Chapter 2 evaluated potential working memory and language demands of different language tools in a group of children. Although children employ both working memory skills and linguistic knowledge during language-based tasks, I found that differential performance across language measures could be revealing of greater reliance on working memory or language skills to support performance. Chapters 3 and 4 examined the relationship between verbal working memory and long-term language knowledge by manipulating linguistic variables specifically. In particular, I focused on how phonological and semantic knowledge contribute to the maintenance of verbal items (words, sentences) in working memory. In Chapter 3, I examined the involvement of phonological and semantic information at the word-level and found that phonological and semantic information were all activated when a word was encountered and processed, with access to semantic knowledge benefiting long-term retention. In Chapter 4, I built on these findings to investigate how phonological, semantic, and attentional mechanisms are contributing to language performance at the sentence-level. Similarly, results supported the idea that sentence recall taps phonological, semantic, and attentional processes interactively but their contributions for immediate and long-term recall differed. Overall, results from this dissertation will help us understand the relationship among working memory processes, linguistic knowledge, and language functioning from a theoretical and clinical perspective.

Co-Authorship Statement

Chapter 2:

Pham, T., Bardell, T. E., Vollebregt, M., Kuiack. A. K., & Archibald., L. M. D. (Accepted). Evaluating the Modified-Shortened Token Test as a working memory and language assessment tool. *Journal of Speech, Language, and Hearing Research*

Chapter 3:

Pham, T. & Archibald., L. M. D. (2021). The role of phonological and semantic representations in verbal short-term memory and delayed retention. *Memory & Cognition*. Advance online publication. https://doi.org/10.3758/s13421-021-01216-8

Chapter 4:

Pham, T. & Archibald., L. M. D. (In prep). Assessing the role of available resources and sentence concreteness on immediate and long-term sentence recall.

Chapters 2 and 3 included in this dissertation are co-authored manuscripts completed by the candidate (the first listed author) under the supervision of Dr. Lisa M. D. Archibald. Chapter 4 is being prepared for submission to a scientific journal. The candidate completed the majority of the work contributing to the manuscripts, including obtaining ethics approval, study design, data collection, statistical analysis, and manuscript preparation and submission. Chapter 2 was conducted in collaboration with Taylor E. Bardell, Meghan Vollebregt, and Alyssa K. Kuiack. The co-author(s) on each manuscript contributed significantly to obtaining ethics approval, securing funding and partnerships for the work, recruitment, data collection, and to the final drafts of the manuscripts.

Acknowledgments

This work would not have been possible without the remarkable support and guidance of many people.

To begin, I wish to acknowledge the land that I have lived and worked on throughout my life. I grew up, learned, and worked in Toronto and completed my graduate studies in London. I acknowledge that Toronto is on the traditional territory of many nations including the Mississaugas of the Credit River, the Anishnabeg, the Chippewa, the Haudenosaunee, the Huron-Wendat, and the Seneca peoples and is now home to many diverse First Nations, Inuit and Métis peoples. I also wish to acknowledge that the University of Western Ontario and the City of London is located on the traditional lands of the Anishinaabek, the Haudenosaunee, the Lūnaapéewak, and the Attawandaron peoples. This land continues to be home to diverse Indigenous peoples (First Nations, Inuit, and Métis) and I am grateful to have the opportunity to live and work on this land.

First and foremost, I owe my deepest gratitude to my advisor, Dr. Lisa Archibald. Thank you for always championing my ideas while at the same time challenging me to think deeply and critically about them. I feel so fortunate to have had the opportunity to learn and grow from your guidance, expertise, and creativity. Lisa, you have nurtured and helped me develop the confidence I needed to grow as a clinician and scientist.

A special thanks to my committee, Drs. Janis Cardy and Marc Joanisse for their helpful thoughts and feedback throughout the development of my dissertation.

I have also been blessed with many incredible mentors earlier in my research career. Drs. Daphna Buchsbaum and Amy Finn, thank you for taking a chance on me as an undergraduate researcher. A career in research did not even cross my mind until I began working in your labs. I am forever grateful to have begun my research career under your mentorship.

I am so thankful to my lab mates, Nicolette Armstrong, Areej Balilah, Taylor Bardell, Caitlin Coughler, Alex Harder, Olivia Daub, Christine Davison, Alyssa Kuiack, Elaine Kwok, Laura Pauls, Rachael Smyth, Meghan Vollebregt; amazing research assistants, Rebecca Day, Katie Flannery, Areej Malik, Nive Varagunan; and the M.Cl.Sc SLP Class of 2020. Thank you for your friendship. You have all kept me motivated and laughing throughout my time of Western.

A warm thank you to all my friends and family who have been my biggest supporters and greatest cheerleaders, and who provided necessary distractions from work. My brother and cousins have especially kept me grounded and very humble during graduate school. And, last but not least, I sincerely thank my parents, Tuan Pham and Hong Dang, for their unconditional love and support. My parents have made countless sacrifices so that I can pursue my dreams. Thank you for always believing in me and for being my source of inspiration. This thesis is dedicated to them.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 Introduction

Language acquisition and language processing must be supported by neurocognitive mechanisms that enable the retention and analysis of language. The two cognitive processes that I will focus on are *working memory*, the ability to briefly hold and manipulate information in mind, and *linguistic knowledge*, our knowledge of the language itself. Word learning, at minimum, involves holding and repeating a novel phonological word form (or speech sounds of the word) and learning the meanings of words (or semantic representations). Working memory for verbal information, or *verbal working memory*, plays a role in establishing such phonological and semantic representations. Additionally, knowing a word involves understanding it's linguistic entailments. Long-term linguistic knowledge refers to language knowledge held in long-term memory including knowing the rules governing speech sounds or word forms (phonological knowledge) and rules governing the meanings of words (semantic knowledge). Recent evidence suggesting a close interaction between verbal working memory and long-term linguistic knowledge has prompted a shift away from more modularized views of specialized verbal storage accessing linguistic knowledge to viewing the language network as a highly flexible system supported by various cognitive processes. This dissertation is aimed at providing further consideration of the way working memory influences and interacts with language abilities. In particular, I will try to answer the following questions: Are verbal working memory and language separable constructs? And, does verbal working memory operate within a dynamic network of cognitive systems including the language network ? Across three studies, I examined the construct of verbal working memory, its separability and its relationship with long-term linguistic representations to answer these questions.

1.1 Terminology

I will be using the following terminology in this dissertation. Immediate memory includes short-term memory and working memory. *Short-term memory* refers to the maintenance of a limited number of items for a short period of time, while *working memory* subsumes short-term memory storage, and additionally includes complex processing functions (e.g., immediate processing, manipulation of information). Some consider the two terms interchangeable (Cowan, 2017). The studies in this dissertation were not designed to systematically distinguish between short-term memory and working memory. Nevertheless, I am specifically interested in *verbal working memory* which involves holding some aspect of language in mind for processing. Henceforth, when discussing my studies, I am referring to verbal short-term memory when only temporary maintenance is involved and verbal working memory when tasks involve both storage and processing.

Long-term memory, on the other hand, refers to a vast store of knowledge. Long-term memory has traditionally been divided into episodic and semantic memory. Episodic memory refers to the ability to recall personal experiences and prior events. Semantic memory involves memory for general knowledge and factual information. I am specifically interested in *long-term linguistic knowledge*, aligning most closely with the conceptualization of semantic memory. The term long-term linguistic knowledge is used in this thesis to specifically refer to language knowledge such as phonological and semantic representations stored in long-term memory. Lastly, memory researchers often distinguish between *item information* (linguistic identity of verbal items) and *order information* (memory for the order of items in a correct sequence). This distinction is made to reflect the suggestion that separable yet interacting cognitive mechanisms support encoding of item and order information (e.g., Farrell & Lewandowsky, 2002; Majerus, 2009).

1.2 Working memory and language: Are they the same or different?

1.2.1 Support for making a distinction

The close interaction between working memory and language has raised questions about whether or not these constructs are, indeed, conceptually distinct from each other. On the one hand, linguistic knowledge impacts verbal working memory performance. For example, consider the repetition of non-words, a primarily immediate memory task requiring the recall of a novel phonological form with no semantic referent. Despite non-word repetition primarily tapping phonology, considerable evidence shows that performance on a non-word repetition task is influenced by knowledge of the language as well. Performance is better when the non-words themselves follow a highly familiar sound sequence (/*vɪdæg*/) than non-words that do not (/*vuŋab*/) (Edwards et al., 2004) and when non-words resemble English words (*fewd*) than non-words that do not (*hyct*) (Gathercole et al., 1995; Ritchie et al., 2015). At the same time, working memory plays an important role in language learning (Gathercole & Baddeley, 1993). The ability to learn a new word relies on phonological storage, especially in the early stages of language acquisition given the relative lack of prior (semantic) knowledge (Gathercole, 2006; Stoel-Gammon, 2011). This link is suggested to continue to support word learning across the lifespan (Gathercole, 2006; Papagno & Vallar, 1995). However, as proficiency in language grows, the link with working memory weakens. For example, correlations with vocabulary learning of a foreign language and working memory decrease over time as reliance on other mechanisms such as existing linguistic knowledge grows (Gathercole & Baddeley, 1989; Masoura & Gathercole, 2005).

Verbal working memory has been found to be a potential constraint on language learning and performance. Specific links between working memory and language learning and processing is perhaps most evident in research involving children with and without language-related disorders. One population in particular is children with a developmental language disorder (DLD), who have a persistent problem learning, understanding, and using language that is not attributed to known biomedical conditions or less rich experience with language (Bishop et al., 2016, 2017). Children with DLD show marked and pervasive deficits on tasks tapping verbal short-term memory such as non-word repetition (Archibald, 2008). As well, there are consistent findings that children with DLD have marked working memory impairments for verbal relative to visuospatial information (Archibald & Gathercole, 2007; Vugs et al., 2013). Visuospatial working memory, then, seems to not be as closely related to language skills as verbal working memory. It would follow that it is verbally mediated working memory that appears to be specifically problematic for children with DLD. The relationship between verbal working memory and language has led to the suggestion that verbal working memory deficits underlie DLD (Montgomery, 2002).

Another body of research, however, suggests that the core linguistic deficits in DLD is not always working memory, but working memory could constrain language processing (Archibald & Joanisse, 2009; Montgomery et al., 2010). Archibald and Joanisse (2009) were the first to identify three groups of children: children with DLD, children with a working memory-based impairment, and children with co-occurring deficits in working memory and language. This was one of the first studies to demonstrate dissociable working memory and language impairments. Similar results were reported for bilingual children by Kapantzoglou et al. (2015), who identified three language ability profiles: low working memory, low grammaticality, and average skills. More recently, Gray et al. (2019) also found that many, but not all, children with language-based disorders such as DLD or dyslexia have co-occurring working memory deficits, suggesting that one domain could be impaired but not the other. In another study demonstrating overlapping but separable influences of working memory and language in children, Noonan et al. (2014) found that children with DLD performed poor overall in a grammaticality judgement task whereas children with co-occurring language and working memory impairments had difficulties judging errors that occurred late in the sentence. Late (vs. early) errors were presumed to impose high working memory demands, thereby, taxing the limited working memory capacity of children with working memory deficits.

The separable influence of working memory on language processing is also evident in children without working memory and/or language deficits. Working memory and language have been distinguished in large-scale factor analysis studies involving an unselected sample of children (e.g., Alloway et al., 2004; Archibald, 2013; also Gray et al., 2019). Relatedly, in a recent study of cognitive predictors of sentence comprehension, the four most salient characteristics of cognitive processing in children with and without DLD were found to be fluid reasoning, controlled attention, complex working memory, and language knowledge (Gillam et al., 2019). Further, complex working memory was found to mediate the relationship between sentence comprehension and the other three cognitive mechanisms (fluid reasoning, controlled attention, and language knowledge). In a subsequent paper, Montgomery et al. (2021) described working memory as having a conduit function such that long-term language knowledge operates through working memory to indirectly influence sentence comprehension. This description is broadly convergent with theories emphasizing a close interaction between verbal working memory and language processing (e.g., Cowan, 1999; Majerus, 2013) because sentence comprehension relies on a subset of activated language knowledge and on attention to maintain items in memory during processing. These findings regarding sentence comprehension could be applied to understand the separable yet intrinsic connections between working memory and language processing more generally.

1.2.2 Challenges for this separation

Despite evidence of separability, it must be acknowledged that working memory and language processing are highly intertwined. It may therefore be as important, and perhaps, more ecologically valid to examine the close relationship between these two factors. For example, one interesting finding from Archibald's (2013) large-scale study was that some language tasks such as the abilities to recall sentences (i.e., Recalling Sentences) or follow directions (i.e., Following Directions) cross-loaded on both immediate memory and language factors. Overlapping factors suggest that the same language task could tap linguistic knowledge primarily (Klem et al., 2015; Polišenská et al., 2015) and working memory secondarily (Archibald, 2013; Riches, 2012).

Marton and Schwartz (2003) examined the interaction between working memory and language comprehension in children with and without DLD. The authors examined the effect of sentence length and syntactic complexity on working memory performance. Arguably, working memory demands are higher for sentences that are longer (vs. shorter) or more complex (vs. simple) because more verbal information needs to be processed. The goal was therefore to assess whether increase in sentence length or syntactic complexity would have similar effects on sentence comprehension. Results revealed reduced performance for all children when sentences were complex regardless of length. This finding suggests that syntactically complex sentences are even more demanding on working memory as compared to increases in sentence length alone. More broadly, this study highlights the interdependence between working memory and language.

Recent suggestions have even considered the idea that verbal working memory performance is driven by language knowledge (Klem et al., 2015; Mainela-Arnold & Evans, 2005). Mainela-Arnold and Evans (2005) reported that differences in working memory capacity between children with and without DLD were no longer significant when language knowledge was controlled. The authors interpreted this finding to support the view that verbal working memory may not be distinct from language representations (Schwering & MacDonald, 2020). Recently, this supposition has also been supported based on inferencing. Mainela-Arnold et al. (2010) reported that semantic knowledge and lexical processing predicted performance on a verbal working memory span task of children with and without DLD. Relatedly, Klem et al. (2015) found that sentence repetition loaded strongly on the unitary language factor rather than being a separate memory construct for typically developing children. These findings challenge the claim that working memory and linguistic knowledge are separable entities.

It may be too that the relationship between verbal working memory and language changes over the course of development (Kidd, 2013). The separability of working memory and language might be most revealing when studied in a developmental population, as the intertwining nature of these two systems is presumably more stable – and harder to disentangle – in language users who have a developed language facility, such as adults. This suggestion is based on findings demonstrating that working memory is an important learning mechanism in early stages of acquiring a language but has a diminishing role with increasing language experience (e.g., Gathercole, 2006; Gathercole & Baddeley, 1989; Masoura & Gathercole, 2005). In adults, linguistic knowledge strongly influences verbal working memory performance (e.g., Acheson et al., 2011; Poirier et al., 2015; Kowialiewski & Majerus, 2018a, 2018b), making this distinction difficult to make. Thus, the possible distinction between working memory and language might be clearer when considered in a developmental context.

1.2.3 Relevant issues in measuring working memory and language

Understanding the influences of working memory and language on performance poses a challenge because the nature of measurements can play a major role in highlighting the separability or connections between these systems. Evidence demonstrating a distinction between working memory and language has primarily come from studies using a battery of tests of language, working memory, and other related cognitive measures. These results all broadly point to the constructs of working memory and language being separable yet highly intertwined (e.g., Archibald, 2013; Gillam et al., 2019). However, one of the challenges in examining verbal working memory and linguistic knowledge is devising one suitable task to provide (separable) estimates of working memory and language skills for the following reasons. First, it would be difficult to separate the contribution of working memory from language entirely given evidence for their interdependence (Mainela-Arnold & Evans, 2005; Marton & Schwartz, 2003). Second, verbal working memory tasks use verbal stimuli (digits, words) and thus are not free from linguistic influences, and many linguistic tasks include a memory component. For instance, working memory demands of language tasks stem from using lengthier and more complex items. This highlights a need to develop a task in which the memory load increases while keeping the linguistic load constant and vice versa. Also, tasks that are considered language measures are used to assess working memory by other researchers. For example, the Token Test was designed to assess language (e.g., De Renzi & Vignolo, 1962; Ellis Weismer & Evans, 1999; Isaki et al., 2008; Schmoeger et al., 2020), but other researchers have used it to assess working memory (e.g., Cohen-Mimran & Sapir, 2007; Mahurin et al., 2006; Bohm et al., 2004), possibly because of requirements to process syntactically complex sentences (tapping working memory).

Interestingly, the Token Test may be one potential task that could be used to evaluate working memory and linguistic demands underlying language processing. In the Token Test, participants are following directions differing in working memory and language load, with systematic increases in length largely separably from changes in syntactic complexity. The test starts with syntactically identical sentences of increasing length and then the last part is similar in length to the longest simple commands but uses complex sentences. Given the nature of the test, I anticipate that the contributions of verbal working memory and linguistic knowledge will differ depending on the cognitive demands embedded in the sentence being processed. Determining the extent to which language performance relies on verbal working memory will also be clinically important. Children employ working memory skills and language knowledge to process sentences. Speech-language pathologists will therefore need to understand why some children perform poorly on language tasks: it is due to working memory and/or language impairments? Further, having a clinical tool to disentangle working memory and language could help speech-language pathologists understand how verbal working memory interacts with language ability in children with DLD. Thus, the aim of Chapter 2 is to understand the involvement of verbal working memory and linguistic skills in a language-specific task, namely, the Token Test.

1.3 Interrelationships between working memory and language

The idea that verbal working memory may be a theoretically separate construct from language should not detract from the highly intertwined nature of these two domains. In fact, a number of verbal working memory theories have been proposed to explain the close interaction between verbal working memory and long-term knowledge. Models of working memory typically differ in their explanations of the connection between short- and long-term memory and involvement of storage components (see Cowan, 2017 for a review), with a recent shift toward a more integrative view of working memory (e.g., Majerus, 2013; Schwering & MacDonald, 2020). Nevertheless, working memory models have been fundamental in advancing our understanding of working memory within the language architecture. I will first review relevant working memory theories and supporting evidence before discussing linguistically-motivated frameworks of verbal working memory.

1.3.1 Working memory models

1.3.1.1 Multicomponent model

An early and still influential model is the multicomponent model of working memory (Baddeley & Hitch, 1974). In the original model, working memory refers to a limited-capacity store composed of specialized short-term storage components working together: the central executive for attentional control, the visuospatial sketchpad for visual information, and the phonological loop for verbal information. Importantly for language maintenance, the phonological loop is further divided into the phonological store (which holds verbal-phonological information) and the articulatory process (which allows for subvocal rehearsal). In fact, recall is reduced when participants are asked to articulate an irrelevant word aloud (known as articulatory suppression), thereby preventing rehearsal (e.g., Baddeley et al., 1975; Larsen & Baddeley, 2003). Information held in working memory is otherwise susceptible to decay (e.g., Baddeley, 2000; Barrouillet & Camos, 2012; Ricker & Cowan, 2014) or interference (e.g., Lewandowsky et al., 2009; Oberauer & Kliegl, 2006) over time. The multicomponent working memory model was revised to acknowledge the contribution of long-term memory to working memory by adding the episodic buffer (Baddeley, 2000). The episodic buffer integrates information from the specialized working memory subcomponents and from long-term memory to support performance. In the most recent model, the episodic buffer interacts directly with the visuospatial sketchpad and phonological loop, and not through the central executive (Baddeley et al., 2011), arguably making the multicomponent model hard to distinguish from other approaches theorizing working memory as the temporary activation of long-term memory (e.g., Cowan, 2005; Oberauer, 2002; also Majerus, 2013; Dell et al., 1997; Schwering & MacDonald, 2020). In fact, Baddeley et al. (2009, p. 440) acknowledges this current way of thinking, “The proposed [episodic buffer] thus made the working memory model more compatible with other approaches such as those of Cowan (2005) and Engle (2002) who have tended to emphasize the executive and integrative aspects of working memory, rather than its subsystems”.

1.3.1.2 **Embedded models**

Integrative accounts of working memory attempt to capture the interaction between working memory and long-term memory without proposing specialized components (e.g., Cowan, 1999; Engle, 2002; Oberauer, 2002). These models conceptualize working memory as the temporary activation of long-term memory representations. For instance, according to Cowan’s embedded processes model (1999, updated 2019), working memory is an activated subset of long-term memory under the focus of attention. Attention plays a pivotal role in holding about four items in the focus of attention. Notably, while activated portions of long-term memory is part of working memory, working memory remains distinct from long-term memory (Cowan, 2017). Similarly, Oberauer (2002) proposed a three-layer memory model: the activated portion of long-term memory, a region within it limited to 3-4 items, and within that, one item in the focus of attention. Overall, these models were better equipped to explain the influence of long-term linguistic knowledge and the role of attention on working memory performance without invoking additional mechanisms, a problem faced by the original multicomponent theory.

1.3.2 Linguistic factors influencing working memory performance

1.3.2.1 Phonological involvement in verbal working memory

One way to understand the close interactions between verbal working memory and language is by observing the influence of linguistic variables. Historical findings have led to the assumption in verbal working memory research that short-term memory codes phonological information whereas long-term memory is associated with semantics (e.g., Baddeley, 1986, 2007). In fact, the phonological loop is a well-defined mechanism for retaining phonological information in working memory. The phonological loop is sometimes even referred to as phonological short-term memory (e.g., Alloway & Gathercole, 2005; Archibald, 2013; Meltzer et al., 2016). This view has largely been based on seminal work reporting a phonological similarity effect in working memory, that is, serial recall is impaired for lists of phonologically similar items as opposed to distinct items (Conrad, 1964, 1965; Conrad & Hull, 1964). Baddeley (1966) extended this work by demonstrating a greater detrimental effect of phonological than semantic similarity on short-term memory, suggesting that verbal short-term memory is largely influenced by phonological effects and minimally by semantic effects. These findings were taken as evidence for phonological coding in verbal working memory and spurred on a large body of work focusing on the major role that phonology plays in verbal working memory.

Traditionally, serial recall, the immediate repetition of items in presented order, has been used widely to study a number of phonological hallmark findings in working memory. Along with the phonological similarity effect, the word length effect (better recall for list of shorter words than longer words; Baddeley et al., 1975) and irrelevant speech effect (poor recall in the presence of background verbal material, Salamé & Baddeley, 1982) all demonstrate the influence of phonological knowledge on working memory. Further, serial recall is poor when rehearsal is prevented such as when individuals engage in articulatory suppression, the concurrent repetition of some irrelevant word or phrase aloud (Baddeley et al., 1975; Larsen & Baddeley, 2003). The articulatory suppression effect indicates that serial recall performance relies on the ability to rehearse phonological information. Finally, phonological processing and rehearsal are considered relatively automatic and important processes to immediate recall (Campoy & Baddeley, 2008; Tehan et al., 2004) and also very effective for storing serial order (vs. item) information, which is often a requirement of serial recall tasks (Gathercole et al., 2001; Romani et al., 2008). This latter point is supported by empirical evidence showing that order information relies on phonological representations and minimally on long-term memory knowledge, whereas item information relies on long-term language knowledge (Majerus, 2009, 2019).

1.3.2.2 **Semantic involvement in verbal working memory**

The focus on phonological representations misses some crucial evidence for the influence of other linguistic representations such as semantic knowledge on the verbal retention of information in working memory. For instance, when processing familiar verbal items (e.g., words, sentences), maintenance and processing relies on not only phonological knowledge but also linguistic knowledge (lexical, semantic, syntactic knowledge) associated with it. Early work by Shulman (1970) and McElree (1996) are some of the first experiments to consider the influence of semantic representations on short-term memory. Shulman (1970) presented participants with a list of 10 words using various presentation rates (range = 350 to 1400 ms) followed by a recognition probe probing for item, homophone, or synonym information. Results revealed that synonym judgements improved with slower presentation rates whereas item and homophone judgements were made more accurately and quickly. McElree (1996), using a similar task in the context of the speed-accuracy trade-off paradigm, presented participants with a 5-word list and participants made slowed or speeded judgements (range = 128 – 3000 ms) about whether the probe word matched, rhymed, or was synonymous with a word on the preceding list. Retrieval dynamics (a speed-accuracy trade-off analysis of response accuracy and timing) were slower for both rhyme and synonym judgements compared to match/item judgements, but nonetheless suggested sufficient access to phonological and semantic information to enable a comparison between the probe and associated list word.

Relatedly, behavioural studies have shown that working memory performance is influenced by language knowledge held in long-term memory. That is, words with richer semantic representations are remembered better as demonstrated by the lexicality effect (better recall for words than non-words; Hulme et al., 1991a), word frequency effect (better recall for high than low frequency words; Hulme et al., 2003), semantic similarity effect (better recall for semantically related than unrelated words in a list; Kowialiewski & Majerus, 2018a; Poirier et al., 2015), sentence superiority effect (better recall for words that form sentences than arbitrary word lists; Baddeley et al., 2009; Brener, 1940), and concreteness effect (better recall for concrete than abstract words; Romani et al., 2008; Walker & Hulme, 1999). Interestingly, the concreteness benefit is also evident at the sentence-level (Meltzer et al., 2016, 2017). Even the repetition of non-words—primarily a phonological task—is facilitated by activation of long-term linguistic knowledge; memory is better for non-words with high than low wordlikeness (Ritchie et al., 2015; Gathercole et al., 1999). Finally, recent studies accounting for the role of semantics have observed semantic influences in order retention (Acheson et al., 2011; Poirier et al., 2015) and when using running-span procedures instead of serial recall (Kowialiewskia & Majerus, 2018a, 2018b).

Hypothesizing that if verbal working memory operates within a network of linguistic knowledge, then semantic effects should influence order retention, Acheson et al. (2011) and Poirier et al. (2015) pioneered studies to directly test this theory. Acheson et al. (2011) used a dual-task paradigm to explore the influence of semantic processing on serial ordering for concrete and non-word lists. Memory is typically better for concrete words than non-words given the additional semantic support available for concrete words. Acheson et al. (2011) used a novel animacy categorization task to impair semantic processing in working memory. It was reasoned that if the concreteness advantage arises because of semantic support, then engaging in a simultaneous animacy categorization task should impair access to semantic resources. In turn, this should selectively reduce recall for concrete words. Indeed, results revealed that their novel animacy categorization task increased item-ordering errors for concrete words but not non-words. Building off Acheson et al.'s work, Poirier et al. (2015) manipulated the semantic relationships of words in a list to examine the effect on serial recall. Word lists were manipulated such that the first three items and the target fifth item were semantically related or unrelated. The authors found that the target fifth item was recalled in earlier positions with related items, which was not found to be due to a grouping strategy. Together, these findings lead to the suggestion that semantic representations impact serial ordering in working memory, possibly because order information is tied to the activation of lexical items within a linguistic network. This interpretation suggests that semantic resources are available in working memory to some extent, and this idea will be further explored in Chapters 3 and 4.

Further, the activation of semantic resources in working memory may be more rapid and automatic than previously suggested. Studies using different paradigms, namely serial recall and running-span tasks, have demonstrated that associated phonological, lexical, and semantic representations may be activated as soon as the verbal stimuli is encountered. Campoy et al. (2015, Experiments 2 and 3), for example, found that the concreteness advantage was maintained in serial recall even with attention-demanding concurrent tasks. The lack of a concurrent task effect indicates that semantic effects might reflect automatic processes instead of strategies such as grouping or generating mental images. However, one general problem with serial recall is that it stresses order processing and taps item processing minimally (Majerus, 2009, 2013). The stressing of phonological maintenance over semantics means that serial recall is strongly tied to phonological processing making it a poor task for examining semantic effects in working memory. The running-span task is another way to study working memory without stressing phonological or semantic representations to a greater degree. In a running-span task, participants process a word list presented at a fast rate and then are cued to recall items unpredictably. Using running-span tasks, Kowialiewski and Majerus (2018a, 2018b) found that semantic effects—lexicality effect, word frequency effect, and semantic similarity—impacted performance, leading to the suggestion that semantic knowledge was accessed very rapidly and automatically. Neuroimaging studies also support the activation of phonological and semantic representations during all stages of a verbal working memory task (Fiebach et al., 2007; Majerus et al., 2010). In general, it seems that unless the role of semantics was explicitly considered in the methodological design, past reliance on serial recall created a bias towards reporting only phonological effects. Hence, one of the goals for Chapter 3 is to develop a novel method for assessing phonological *and* semantic factors underlying verbal working memory without emphasizing one linguistic factor to a greater extent. Further, Chapter 4 takes a different approach to understand the mechanisms underlying verbal working memory. Specifically, I draw inspiration from Acheson et al. (2011) and used concurrent tasks to investigate the interplay between different mechanisms underlying language maintenance.

Converging evidence from neuropsychological studies also support the role of phonological and semantics in verbal working memory. For instance, patients with lesions in the left inferior and middle frontal gyri have difficulties maintaining semantic information in short-term memory, whereas lesions in the inferior parietal areas are associated with short-term phonological deficits (Hamilton et al., 2009). Patients have demonstrated a double dissociation between phonological and semantic maintenance mechanisms. On the one hand, there are patients who have difficulties maintaining phonological information during linguistics tasks but relatively intact abilities to use semantic representations. For instance, some patients who have difficulties repeating single words show a recall advantage for highly imageable words (e.g., Howard & Nickels, 2005; Majerus et al., 2001; N. Martin & Saffran, 1992) or sentences (Baldo et al., 2008). On the other hand, patients with semantic dementia have an impairment restricted to semantic knowledge, while their phonological representations remain intact such that they have normal digit span, show typical phonological effects, and are able to read and repeat words despite not understanding the meaning (e.g., Jefferies et al., 2005; Kertesz et al., 2010). Further, patients with short-term semantic deficits do not show the typical advantage for remembering words over non-words in immediate recall tasks (R.C. Martin et al., 1994, 1999). Crucially, the dissociation in patients' performance implies that multiple representations including phonological and semantics influence performance. By using data from brain-damaged individuals, research in cognitive neuropsychology initiated a linguistic conceptualization of verbal working memory. Theories proposed to account for the influence of language representations to verbal working memory will be discussed in the next section.

1.3.3 Verbal working memory models: Explaining the impact of linguistic knowledge on verbal working memory

1.3.3.1 Redintegration

A crucial limitation of the original multicomponent working memory model was its inability to account for empirical evidence demonstrating the influence of both phonological and semantics in verbal working memory. Although the phonological loop captured storage of verbal information in a phonological format, semantic representations was said to play no role in such brief maintenance. In fact, phonological and semantic processing were largely considered as operating in distinct memory systems. One way this issue has been addressed, whilst keeping long-term memory separate from short-term memory, is by proposing secondary, long-term mechanisms, namely the episodic buffer and redintegration. As previously discussed, the episodic buffer connects long-term memory with short-term stores which would allow linguistic knowledge to influence the content of the phonological loop (Baddeley, 2000). The redintegration hypothesis specifically accounts for the influence of long-term linguistic knowledge on verbal working memory (Hulme et al., 1991b; Schweickert, 1993; Schweickert, Chen, & Poirier, 1999). According to this view, verbal information is first represented phonologically, as proposed by Baddeley and Hitch (1974). Redintegration is not invoked if verbal information remains intact. As information becomes degraded without rehearsal, however, long-term linguistic knowledge is used to reconstruct the degraded representations held in the phonological loop at the moment of recall – this is the process of redintegration. For instance, if the phonological representation of the to-be-remembered word (*instrument*) has been degraded to some extent (*i n s _ r _ m _ n t*), then by accessing and matching this phonological form to long-term knowledge, such as semantic knowledge, the word can be reconstructed or redintegrated. Redintegration theorists account for short-term semantic effects, such as the concreteness effect, in verbal working memory tasks by presuming that redintegration is more accessible for concrete than abstract words due to less degraded short-term memory representations, coupled with richer connections with long-term memory. Abstract words, by comparison, would be difficult to reconstruct due to more degraded short-term memory traces and less rich or poorer connections with long-term memory representations. However, the mechanism of redintegration in general is still not fully understood. Moreover, the redintegration model has been criticized for emphasizing the role of phonology in verbal working memory. Crucial evidence showing that semantic information is available later than phonological information would provide support for the redintegration hypothesis. This will be examined in Chapter 3 of this thesis.

1.3.3.2 **Language-based models**

In alternative perspectives that take a language-based approach to verbal working memory, phonological *and* semantic knowledge are thought to support the maintenance of verbal information at all stages from encoding, to storage, to recall. These models conceptualize verbal working memory as an integrative part of the language system. Based on patient data, N. Martin and colleagues (2004; Dell et al., 1997) and R.C. Martin and colleagues (1999, 2001) have developed models to account for the maintenance of phonological and semantic information in working memory. N. Martin (2004; Dell et al., 1997) proposed that verbal information is maintained via activation within the linguistic system across the different levels of representations (phonological, lexical, semantic). During word processing, temporary activation spreads back and forth between these different levels, and spreading activation in turn supports recall (Figure 1.1). R. C. Martin (1999, 2001), on the other hand, proposed buffers responsible for short-term retention of phonological and semantic representations that are separate from the lexical processing system, though connected to it. However, the challenge of separate components is the requirement of additional buffers or mechanisms to accommodate for integrated representations and integrative aspects of processing, with the field favouring a more parsimonious explanation.

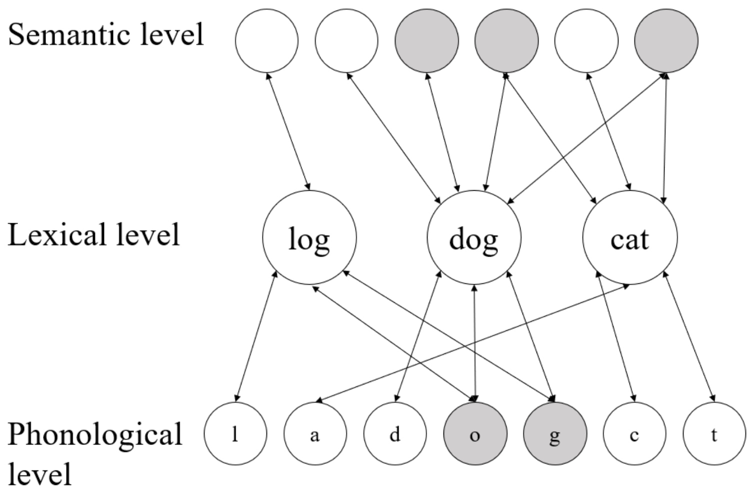

Figure 1.1. Simplified illustration of an interactive model of language production (Dell et al., 1997). Connections are bidirectional. Overlapping phonological and semantic representations are shaded in gray.

Cowan's embedded processes model (1999, updated in 2019), although not a language-based model per se, shares assumptions with linguistically motivated accounts of verbal working memory. According to Cowan's model, working memory is the temporary activation of long-term memory, in line with N. Martin and colleagues (2004; Dell et al., 1997). Attention has a role in Cowan's model, but this model does not specify the interactions between verbal working memory and language activation. Recent verbal working memory models have begun to consider the interrelationships between long-term linguistic knowledge and attention mechanisms with respect to verbal working memory.

Drawing together ideas from early theories, more recent verbal working memory models by Majerus (2013, 2019) and MacDonald and colleagues (2009; Schwering & MacDonald, 2020) provide a detailed account describing the contributions of linguistic factors to language processing. According to the integrative framework proposed by Majerus (2013), short-term language maintenance is achieved by the simultaneous activation of language (phonological and semantic representations), attentional, and serial order processing systems (Figure 1.2). This model, similar to N. Martin and colleagues (2004; Dell et al., 1997), considers that verbal information is the direct activation of phonological and semantic representations within the linguistic system. In particular, dorsal and ventral language processing networks are proposed to provide the neural basis of phonological and semantic representations, respectively. This network supports short non-word and word repetition, for instance. Additionally, this model addresses some of the limitations of previous language-based models by considering the role of domain-general attentional and serial order mechanisms. When verbal retention involves multi-word or sentence repetition, language pathways intervene with bilateral fronto-parietal networks, supporting domain-general attentional and serial order processing. Relatedly, Schwering and MacDonald (2020) advocate for a strongly emergent approach, whereby verbal working memory is not a separate system from language. This approach views the linguistic system itself as responsible for performance on verbal working memory tasks, without the need for separate item and order mechanisms. However, one problem for this emergentist approach is neuroimaging evidence pointing to different neural mechanisms responsible for item and order information (Majerus, 2009). Additionally, although highly intertwined, separable influences on language functioning of working memory and language knowledge in long-term memory have been empirically demonstrated (Archibald & Joanisse, 2009; Gillam et al., 2019; Kapantzoglou et al., 2015; but see Mainela-Arnold & Evans, 2005; Klem et al., 2015) and will be further investigated in this thesis (Chapter 2). In fact, it has been suggested that working memory demands could be reduced by relying on existing language knowledge (Archibald, 2018; Kowialiewski et al., 2020; Montgomery et al, 2021). Parsing verbal information in a meaningful and familiar way would reduce working memory load and leave resources available for integrating more information or processing, for example. Therefore, the close connection between working memory and language could be similar to how Cowan conceptualizes working memory and long-term memory (1999, 2017). That is, verbal working memory is a subset of activated linguistic long-term memory, but verbal working memory does not subsume the linguistic system (also Gillam et al., 2019; Montgomery et al., 2021).

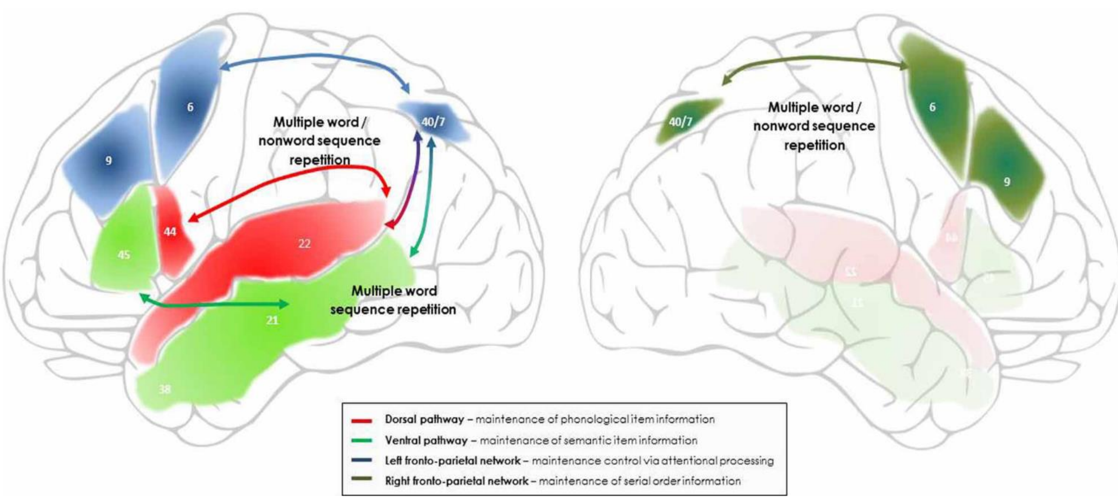

Figure 1.2. Majerus' (2013, 2019) integrative framework of verbal working memory. Short-term storage results from synchronized and flexible recruitment of language (phonological and semantic representations), attentional, and serial order processing systems.

Although there are nuances between the language-based models reviewed, importantly, they all emphasize the interaction between different levels of linguistic knowledge and its interrelationship with immediate recall. These models collectively assume that short-term maintenance relies on the fast and direct activation of different levels of representations including phonological and semantics, and that mutual interactions between these levels contribute to maintenance of verbal items in working memory. The concreteness effect, for example, could be explained as the result of concrete words being associated with richer and more distinctive representations at the semantic level, which will more strongly stabilize phonological representations, whereas abstract words will receive weaker stabilizing feedback because of weaker semantic support.

Another consideration is how the interactive nature of these components could support each other. Although prior studies have investigated the interaction between phonological and semantic representations in verbal working memory (Acheson et al., 2010; Nishiyama, 2014), few studies have examined how the different language representations could compensate for each other if there are, in fact, multiple representations influencing performance. The idea is that disruption to phonological processes should lead to greater engagement of semantic processes, for example. Some support for this idea was provided by Nishiyama (2020), who showed that participants adaptively switched from relying on phonological to semantic representations when it was hard to phonologically rehearse words. In fact, articulatory suppression has been found to enhance the advantage for concrete words by weakening phonological representations and thereby encouraging reliance on semantics (Meltzer et al., 2016; Romani et al., 2008). Chapter 4 will test this idea more systematically to understand the interplay between different mechanisms underlying language processing. Sentence repetition is a linguistic task that taps phonological and semantic processes interactively. Further, the advantage for concrete sentences suggest that participants rely on semantic representations in immediate memory. Therefore, I investigated the effect of concurrent tasks on the concreteness advantage in sentence recall. Concurrent tasks were designed to tap a specific process, and minimally imposing on the other processes.

1.3.4 Summary

Despite recent theoretical advancements, experimental investigation of language-based models in healthy participants is still in its infancy at least in part due to methodological limitations. Although we can try to tease apart working memory and language empirically (Chapter 2), in actual fact, most recent work is supporting high interconnections between these two systems especially in adults. If verbal working memory operates within the context of a complex linguistic system, then phonological, semantic, and related cognitive processes should be observed to operate in a very complex, highly interactive way throughout recall from the short- to long-term. Hence, Chapters 3 and 4 are designed to delineate the role of phonological and semantic representations in verbal working memory. Specifically, I asked whether there is direct activation of linguistic knowledge to support maintenance and retention in the context of word recognition (Chapter 3) and sentence recall (Chapter 4). Chapter 3 uses a novel technique to more closely investigate language-based models by minimizing the intervention of redintegration and rehearsal strategies. Chapter 4 takes a different methodological approach to investigate the interplay between different mechanisms—phonological, semantics, and attentional—underlying sentence recall by using concurrent tasks.

1.4 Objectives and Overview

Cognitive abilities that likely contribute to language processing include *working memory* involving simultaneous storage and processing of verbal information and *long-term linguistic knowledge*, which is the ability to activate language content. The intertwining nature of verbal working memory and language processing has led to mixed perspectives; some suggest they are overlapping but separable influences (e.g., Archibald & Joanisse, 2009; Montgomery et al., 2021), while others argue against this distinction (e.g., Mainela-Arnold & Evans, 2005; Schwering & MacDonald, 2020). Nevertheless, there is increasing recognition that language processing is supported by a dynamic cognitive neural network. This highlights a pressing need to investigate the symbiotic relationship between working memory and language learning using robust paradigms that balance opportunities for phonological and semantic influences on working memory. The central objective of this thesis is to examine the role of verbal working memory and linguistic knowledge in children and adult's performance on a variety of verbal tasks, including word recognition, sentence recall, and sentence comprehension. Overall, results from this dissertation will advance theories of verbal working memory as it relates to language processing as well as have the potential to inform clinical practice.

Chapter 2 considers the relational influence of working memory and linguistic knowledge on language performance in a developmental population. Many language tests used clinically by speech-language pathologists appear memory dependent. For example, comprehending sentences of increasing length and syntactic complexity requires the use of working memory, in addition to linguistic skills. Hence, it will be important to understand the contributions of working memory and linguistic abilities to language performance. In Chapter 2, I consider how verbal working memory and linguistic knowledge support verbal abilities in children. Across a series of studies, children completed a version of the Token Test and various measures of working memory and language. Specifically, I evaluated the factor structure of the Token Test and interrelationships between those identified factors and other common working memory and language measures.

Chapter 3 investigates how words are processed in verbal short-term memory. Recent years have seen a shift away from the view that words are processed primarily phonologically in short-term memory, toward a more integrated view. To systematically evaluate how phonological and semantic information are recalled immediately and retained in long-term memory, I use a novel combined probe recognition – running-span paradigm for this investigation. Across two experiments, a list of words was presented sequentially, followed by a probe word probing for phonological or semantic information and delayed memory was also tested. This paradigm allowed me to examine the extent to which linguistic knowledge was readily accessible and retained during word processing even when rehearsal strategies and redintegration processes were prevented.

Chapter 4 further considers the suggestion that verbal working memory and linguistic knowledge are highly intertwined in language tasks and extends this investigation to the context of sentence recall. Sentence recall provides a microcosm for the study of language and working memory skills given the inherent linguistic (sentence) and cognitive (recall) characteristics of the task. In particular, Chapter 4 evaluates the contributions of phonological, semantic, and attentional processing on sentence recall using suppression tasks designed to load on a specific resource. This study would shed light on the interplay between different kinds of resources underlying sentence processing and their respective influence on immediate recall performance and long-term retention.

These three studies were designed to address questions related to cognitive processes that influence and interact with language processing. The findings in this thesis will have the potential to inform theories of verbal working memory, and to begin the work of addressing questions related to verbal working memory and language demands of language tasks at the practical level.

1.5 References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(1), 44-59. https://doi.org/10.1037/a0021205

Acheson, D. J., Postle, B. R., & Macdonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(1),* 17–36. https://doi.org/10.1037/ a0017679

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A.-M. (2004). A structural analysis of working memory and related cognitive skills in young children. Journal of Experimental Child Psychology, 87(2), 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

Archibald, L. M. D. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology, 32(1),* 21-28.

Archibald, L. M. D. (2013). The language, working memory, and other cognitive demands of verbal tasks. *Topics in Language Disorders, 33(3),* 190–207. https://doi.org/10.1097/TLD.0b013e31829dd8af

Archibald, L. M. D., & Gathercole, S. (2007). The complexities of complex memory span: Storage and processing deficits in specific language impairment. *Journal of Memory and Language, 57(2),* 177–194. https://doi.org/10.1016/j.jml.2006.11.004

Archibald, L. M. D., & Joanisse, M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52(4),* 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

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. https://doi.org/10.1080/14640746608400055

Baddeley, A. D. (1986). *Working memory.* Clarendon Press/Oxford University Press.

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

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49(6),* 1393–1400. https://doi.org/10.1016/j.neuropsychologia.2010.12.042

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 647–667). Academic Press

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language, 61(3),* 438–456. https://doi.org/10.1016/j.jml.2009.05.004

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. https://doi.org/10.1016/S0022-5371(75)80045-4

Baldo, J., Klostermann, E., & Dronkers, N. (2008). It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. *Brain and Language, 105(2)*, 134–140. https://doi.org/10.1016/j.bandl.2007.12.007

Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction.* Psychology Press.

Bishop, D.V.M., Snowling, M.J., Thompson, P.A., Greenhalgh, T., & the CATALISE consortium. (2016). CATALISE: a multinational and multidisciplinary Delphi consensus study. 1. Identifying language impairments in children. *PLoS ONE, 11(7):* e 0158753. https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58(10)*, 1068-1080. https://doi.org/10.1111/jcpp.12721

Bohm, B., Smedler, A. C., & Forssberg, H. (2004). Impulse control, working memory and other executive functions in preterm children when starting school. *Acta Pædiatrica*, *93(10)*, 1363–1371. https://doi.org/10.1080/08035250410021379

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*, 467–482. https://doi.org/10.1037/h0061096

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16(4)*, 329–340. https://doi.org/10.1080/09658210701867302

Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology*, *68(4)*, 759-778. https://doi.org/10.1080/17470218.2014.966248

Cohen-Mimran, R., & Sapir, S. (2006). Deficits in working memory in young adults with reading disabilities. *Journal of Communication Disorders*, *40(2)*, 168–183. https://doi.org/10.1016/j.jcomdis.2006.06.006

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning & Verbal Behavior, 4*, 161–169. https://doi.org/10.1016/S0022-5371(65)80015-9

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429–432. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

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, UK: Cambridge University Press. http://dx.doi.org/10.1017/ CBO9781139174909.006

Cowan, N. (2005). *Working memory capacity.* Psychology Press. [Psychology Press and Routledge Classic Edition with new foreword, 2016]

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*, 1158–1170. http://doi.org/10.3758/s13423-016-1191-6

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin, 145(8),* 822–847. https://doi.org/10.1037/bul0000199

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(4),* 801-838. http://doi.org/10.1037/0033-295X.104.4.801

De Renzi, E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasics. *Brain, 85(4),* 665-678. https://doi.org/10.1093/brain/85.4.665

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47(2),* 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Ellis Weismer, S., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42(5),* 1249–1260. https://doi.org/10.1044/jslhr.4205.1249

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11(1),* 19–23. https://doi.org/10.1111/1467-8721.00160

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9(1),* 59–79. http://doi.org/10. 3758/bf03196257

Fiebach, C., Friederici, A., Smith, E., & Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual-semantic representations in verbal working memory. *Journal of Cognitive Neuroscience, 19(12),* 2035–2049. https://doi.org/10.1162/jocn.2007.19.12.2035

Gathercole S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23,* 83–94. https://doi.org/10.3758/BF03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27(4),* 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28,* 200–213. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Erlbaum.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A, 54(1),* 1-30. https://doi.org/10.1080/02724980042000002

Gillam, R., Montgomery, J., Evans, J., & Gillam, S. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: Implications for assessment and treatment. *International Journal of*

Speech Language Pathology, 21(3), 240–251. https://doi.org/10.1080/17549507.2018.1559883

Gray, S., Fox, A., Green, S., Alt, M., Hogan, T., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research, 62(6),* 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Hamilton, A. C., Martin, R. C., & Burton, P. C. (2009). Converging functional magnetic resonance imaging evidence for a role of the left inferior frontal lobe in semantic retention during language comprehension. *Cognitive Neuropsychology, 26,* 685–704. http://doi.org/10.1080/02643291003665688

Howard, D., & Nickels, L. (2005). Separating input and output phonology: semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology, 22(1),* 42–77. https://doi.org/10.1080/02643290342000582

Hulme, C., Maughan, S., & Brown, G. D. A. (1991a). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30(6),* 685-701. https://doi.org/10.1016/0749-596X(91)90032-F

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999b). Think before you speak: Pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25(2),* 447–463. https://doi.org/10.1037/0278-7393.25.2.447

Hulme, C., Stuart, G., Brown, G. D., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language, 49(4),* 500–518. https://doi.org/10.1016/S0749-596X(03)00096-2

Isaki, E., Spaulding, T., & Plante, E. (2008). Contributions of language and memory demands to verbal memory performance in language-learning disabilities. *Journal of Communication Disorders, 41(6),* 512–530. https://doi.org/10.1016/j.jcomdis.2008.03.006

Jefferies, E., Jones, R. W., Bateman, D., & Lambon Ralph, M. A. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology, 22(2),* 183–212. https://doi.org/10.1080/02643290442000068

Kapantzoglou, M., Restrepo, M. A., Gray, S., & Thompson, M. S. (2015). Language Ability Groups in Bilingual Children: A Latent Profile Analysis. *Journal of Speech, Language, and Hearing Research, 58(5),* 1549–1562. https://doi.org/10.1044/2015_JSLHR-L-14-0290

Kertesz, A., Jesso, S., Harciarek, M., Blair, M., &McMonagle, P. (2010). What is semantic dementia? A cohort study of diagnostic features and clinical boundaries. *Archives of Neurology, 67,* 483–489. https://doi.org/10.1001/archneurol.2010.55

Kidd, E. (2013). The role of working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders, 33(3),* 208–223. https://doi.org/10.1097/TLD.0b013e31829d623e

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J. E., and Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science, 18,* 146–154. https://doi.org/10.1111/desc.12202

Kowialiewski, B., & Majerus, S. (2018a). Testing the redintegration hypothesis by a single probe recognition paradigm. *Memory, 26(9),* 1256-1264. https://doi.org/10.1080/09658211.2018.1448420

Kowialiewski, B., & Majerus, S. (2018b). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language, 101*, 64-83. https://doi.org/10.1016/j.jml.2018.03.005

Kowialiewski, B., Lemaire, B., Portrat, S. (2020). Semantic chunks save working memory resources: computational and behavioral evidence. In S. Denison, M. Mack, Y. Xu, & B. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1466-1472). Toronto, ON (virtual): Cognitive Science Society

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? The *Quarterly Journal of Experimental Psychology: Section A, 56(8),* 1249–1268. https://doi.org/10.1080/02724980244000765

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13,* 120–126. http://dx.doi.org/10.1016/j.tics.2008.12.003

Mahurin, R. K., Velligan, D. I., Hazleton, B., Mark Davis, J., Eckert, S., & Miller, A. L. (2006). Trail making test errors and executive function in schizophrenia and depression. *Clinical Neuropsychologist, 20(2),* 271–288. https://doi.org/10.1080/13854040590947498

Mainela-Arnold, E., & Evans, J. (2005). Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research, 48,* 897–909. https://doi.org/10.1044/1092-4388(2005/062)

Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2010). Explaining lexical-semantic deficits in specific language impairment: The role of phonological similarity, working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research, 53*, 1742–1756. https://doi.org/10.1044/1092-4388(2010/08-0198)

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). Psychology Press.

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7,* 357. https://doi.org/10.3389/fnhum.2013.00357

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.04.016

Majerus, S., Lekeu, F., Van der Linden, M., & Salmon, E. (2001a). Deep dysphasia: Further evidence on the relationship between phonological short-term memory and language processing impairments. *Cognitive Neuropsychology, 18,* 385–410. https://doi.org/10.1080/02643290126060

Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., Salmon, E., Seurinck, R., Fias, W., & Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience, 22(11),* 2570–2593. https://doi.org/10.1162/jocn.2009.21378

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46(5),* 1138–1153. https://doi.org/10.1044/1092-4388(2003/089)

Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory*, *13,* 422-429

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition, 24(2),* 173-187. https://doi.org/10.3758/BF03200879

Meltzer, J. A., Kielar, A., Panamsky, L., Links, K. A., Deschamps, T., & Leigh, R. C. (2017). Electrophysiological signatures of phonological and semantic maintenance in sentence repetition. *NeuroImage, 156,* 302–314. https://doi.org/10.1016/j.neuroimage.2017.05.030

Meltzer, J. A., Rose, N. S., Deschamps, T., Leigh, R. C., Panamsky, L., Silberberg, A., Madani, N., Links, K. A. (2016). Semantic and phonological contributions to short- term repetition and long-term cued sentence recall. *Memory & Cognition. 44,* 307–329. https://doi.org/10.3758/s13421-015-0554-y

Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11(1),* 77-91. https://doi.org/10.1044/1058-0360(2002/009)

Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2021). A new memory perspective on the sentence comprehension deficits of school-Age children with developmental language disorder: Implications for theory, assessment, and intervention. *Language, Speech & Hearing Services in Schools, 52(2),* 449–466. https://doi.org/10.1044/2021_LSHSS-20-00128

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: an update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology, 19(1),* 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Martin, N., Fink, R., & Laine, M. (2004). Treatment of word retrieval with contextual priming. *Aphasiology*, *18,* 457-471. https://doi.org/10.1080/02687030444000129

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: evidence from a single case study. *Brain and Language, 43,* 240–274. https://doi.org/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. https://doi.org/10.1080/026432997381402

Nishiyama, R. (2014). Active maintenance of semantic representations. *Psychonomic Bulletin & Review, 21(6),* 1583-1589. https://doi.org/10.3758/s13423-014-0618-1

Nishiyama, R. (2020). Adaptive use of semantic representations and phonological representations in verbal memory maintenance. *Journal of Memory and Language, 111,* 104084. https://doi.org/10.1016/j.jml.2019.104084

Noonan, N. B., Redmond, S. M., & Archibald, L. M. (2014). Contributions of children's linguistic and working memory proficiencies to their judgments of grammaticality. *Journal of Speech, Language, and Hearing Research, 57(3),* 979-989. https://doi.org/10.1044/2014_JSLHR-L-12-0225

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 28, 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*, 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Papagno & Vallar, 1995 Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48A*, 98-107. https://doi.org/10.1080/14640749508401378

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43(3),* 489-499. https://doi.org/10.3758/s13421-014-0470-6

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: what does the task measure? *International Journal of Language & Communication Disorders, 50(1),* 106–118. https://doi.org/10.1111/1460-6984.12126

Martin, R. C., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41(1),* 3–29. https://doi.org/10.1006/jmla.1999.2637

Freedman, M. L., & Martin, R. C. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology, 18*(3), 193–226. https://doi.org/10.1080/02643290042000080

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33(1),* 83-111. https://doi.org/10.1006/jmla.1994.1005

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms: Sentence repetition in children with SLI. *International Journal of Language & Communication Disorders, 47(5),* 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 417–428. http://doi.org/10.1037/a0034301

Ritchie, G., Tolan, G. A., & Tehan, G. (2015). Redintegration, task difficulty, and immediate serial recall tasks. *Canadian Journal of Experimental Psychology, 69(1),* 54–63. https://doi.org/10.1037/cep0000031

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61(2)*, 292-323. https://doi.org/10.1080/17470210601147747

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by irrelevant speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150-164

Schmoeger, M., Deckert, M., Eisenwort, B., Loader, B., Hofmair, A., Auff, E., & Willinger, U. (2020). Evaluating Part V of the German version of the Token Test as a screening of specific language impairment in preschoolers. *Applied Psycholinguistics*, *41*(1), 237–258. https://doi.org/10.1017/S0142716419000493

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21(2),* 168-175. https://doi.org/10.3758/BF03202729

Schweickert, R., Chen, S., & Poirier, M. (1999). Redintegration and the useful lifetime of the verbal memory representation. *International Journal of Psychology, 34(5-6),* 447–453. https://doi.org/10.1080/002075999399800

Schwering, S. C. & Macdonald. M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14,* 68. https://doi.org/10.3389/fnhum.2020.00068

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–508. https://doi.org/10.1016/S0022-5371(70)80093-7

Tehan, G., Fogarty, G., & Ryan, K. (2004). The contribution to immediate serial recall of rehearsal, search speed, access to lexical memory, and phonological coding: An investigation at the construct level. *Memory & Cognition, 32(5),* 711–721. https://doi.org/10.3758/BF03195861

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities, 34(9),* 2586–2597. https://doi.org/10.1016/j.ridd.2013.05.014

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5),* 1256-1271. https://doi.org/10.1037/0278-7393.25.5.1256

Chapter 2

2 Evaluating the Modified-Shortened Token Test as a working memory and language assessment tool

2.1 Introduction

Language acquisition and language processing must be supported by neurocognitive mechanisms that enable the retention and analysis of language. Although conceptualized in different theoretical accounts as either a separable cognitive resource (Baddeley & Hitch, 1974) or an emergent property of language experience (Majerus, 2013; Schwering & MacDonald, 2020), the ability to briefly hold and process information in mind known as working memory has been investigated as a potential constraint on language learning and performance. Evidence for a separable role of working memory in language processing comes from findings distinguishing specific working memory and language impairments (Archibald & Joanisse, 2009) and low grammaticality and low phonological working memory profiles in bilingual children (Kapantzoglou et al., 2015). Indeed, in a recent study of cognitive predictors of sentence comprehension (Gillam et al., 2019), the four most salient characteristics of cognitive processing in children with and without language disorder were found to be fluid reasoning, controlled attention, complex working memory, and language knowledge. Despite evidence of separability, it must be acknowledged that the relationship between working memory and language processing is far from clear cut. For example, children with language-related disorders such as developmental language disorder (DLD) or dyslexia do not always have working memory deficits (Gray et al., 2019). Taken together, this evidence indicates that working memory supports language processing in complex ways. This, in turn, highlights the need for a clinical assessment tool that will assist speech-language pathologists in understanding individual children's performance in relation to working memory and language knowledge. In this paper, we focus on one potential tool, the Token Test, an auditory comprehension test that manipulates the length and linguistic complexity of verbal directions. By drawing on data from a number of studies in our research program, we examine whether the factor structure underlying performance on the Token Test corresponds to separable working memory and language knowledge components, and the extent to which composite scores based on the factor structure relate to subtests of standardized language tests commonly used in the field.

The *Token Test* is widely used to detect receptive language impairments in the context of otherwise relatively intact comprehension in normal communication (De Renzi & Vignolo, 1962). Originally designed to detect auditory comprehension deficits in adults with aphasia, the use of the Token Test has been extended to assess language abilities in children and adolescents (e.g., Cole & Fewell, 1983, Paquier et al., 2007; Gallardo et al., 2011; Fidler et al., 2011). A number of studies have indicated strong to reasonable ranges of internal consistency, inter-rater reliability, and intra-rater reliability for overall score and for individual subtests/parts of different versions of the Token Test (Park et al., 2000; Gallardo et al., 2011; McNeil et al., 2015). Further, various versions of the Token Test continue to be used by researchers and clinicians to measure language skills (e.g., Ellis Weismer et al., 1999; Isaki et al., 2008) or working memory skills (Cohen-Mimran & Sapir, 2007; Bohm et al., 2004). More recently, the Token Test has been evaluated as a potential screening tool for the detection of DLD in preschool children (Schmoeger et al., 2020).

The Token Test assesses comprehension by evaluating the ability to follow verbal directions of increasing length and complexity. Using 20 tokens including 5 small and 5 large circles and squares with equal representation of 5 colours (red, yellow, green, black, white), the respondent manipulates these tokens according to verbal commands. The original Token Test consisted of 62 commands (De Renzi & Vignolo, 1962). To make the tool more efficient for clinical usage, the Token Test has been shortened and revised. Notably, there have been numerous "short" Token Tests developed (e.g., 16-items, Spellacy & Spreen, 1969; 36-items, De Renzi & Faglioni, 1978; 55-items, Arvedson et al. 1985). Here we focus on the *Shortened Token Test* which consists of 36 commands divided into six parts (De Renzi & Faglioni, 1978). It is the only version developed by the original authors as well as being available in many languages (Bastiaanse et al., 2016). The initial parts require pointing to an indicated token with commands increasing in length across the test (e.g., "Touch a green circle"; "Touch a small green circle"; "Touch a green circle and a blue square"; "Touch the small green circle and the large blue square"). The final part (Part 6) incorporates linguistic complexity by requiring sequenced responses and manipulation (e.g., "Put the green square next to the red circle"). Tokens may be displayed in either 2 or 4 rows of 5 tokens depending on the requirements for each part. For example, in the part with the shortest commands such as "Touch a circle", only the two rows of large shapes are displayed. The responses are scored as correct (1 point), correct with repetition (0.5 point), or incorrect (0 point).

Children are likely employing both working memory and language skills to process sentences, but the structure of the Token Test may be suitable for highlighting the contribution of each construct to a greater degree. By both systematically manipulating length and linguistic complexity separately (to some extent), some items may be more sensitive to verbal working memory demands than linguistic abilities in supporting performance and vice versa. In this paper, linguistic complexity was based on word length, the number of declarative clauses, number of phrases, Yngve depth (reflecting the average number of embedded structures in a sentence; Yngve, 1960), readability statistics, age of acquisition of words, concreteness of words, and word frequency. Note that participants are not reading sentences during the Token Test, instead, readability statistics were generated by calculating a score for the reading level of the sentences typed into a document in Microsoft Word. Detailed descriptive information about the linguistic complexity of each Token Test part is provided in the supplemental material (see Table S2.1). Consider, first, the increasing length of commands across the initial parts of the test. The linguistic complexity of the imperative commands across these initial parts (e.g., “Touch [X]” or “Touch [X] and [X]”) is simple and uses a highly constrained vocabulary (i.e., 5 colours, 2 sizes, 2 shapes). As such, processing standard word order and familiar words is done with relative ease. Only the number of content items to be retained increases. For example, “Touch the circle” requires 1 item to be held in mind (i.e., circle) whereas “Touch the small, green circle and the large, blue square” requires 6 items be held in mind. Parts 1 to 3 were indeed the shortest in length, syntactically the simplest, and had little demands on semantics, whereas increasing length through Parts 4 and 5 also resulted in increased sentence depth and some semantic demands, while remaining syntactically identical (Table S2.1 in the supplementary material). It seems likely, then, that the increasing length of the sentences across the initial parts of the Token Test place increasing demands on working memory while imposing a minimal and consistent load on linguistic processing resources. Specifically, the Token Test is likely tapping verbal aspects of working memory (storage and manipulation of verbal information) rather than more general working memory aspects. Indeed, as a myriad of studies have shown, tasks involving verbal storage of digits or non-words load on the same working memory factor (e.g., Alloway et al., 2006; Archibald, 2013).

The final part of the Token Test (Part 6 of the Shortened Token Test), on the other hand, introduces sentences that are grammatically different from the rest of the test. The final part is characterized by a variety of linguistic structures (e.g., adverbial phrases, conditional sentences) and additional vocabulary (e.g., *before, when, all*) whereas the length of the commands (*M* = 9.31 words, *SD* = 1.32) is similar to the longest (but simple) commands from the previous part (*M* = 10, *SD* = 0). Indeed, Part 6 is the only part with sentences consisting of more than one declarative clause and is more difficult in terms of readability score (Part 6, grade 2 level vs. Part 5, grade 1 level; see Table S2.1). Despite the introduction of new vocabulary, we must acknowledge that syntax structure is primarily manipulated in these sentences, with semantic knowledge to a lesser degree. Thus, performance would reflect differences in syntactic knowledge and be limited in terms of addressing differences in semantic knowledge. Although words that are used to make up sentences in real-world contexts are likely to vary much more in meaning than they do in the Token Test, some degree of semantic variation in the Token Test could provide a starting point for us to evaluate semantic knowledge (such as age of acquisition, concreteness of words, frequency of usage) that this tool is tapping. Nevertheless, the final part, then, imposes demands on both working memory and linguistic processing. In fact, the demands may not simply be additive relative to the working memory demands of the simple commands of equivalent length. Remembering a long sentence that is also syntactically complex can be expected to impose an even higher working memory load given the additional language processing needed to assemble a syntactically complex sentence (Magimairaj & Montgomery, 2012; Marton & Schwartz, 2003) as well as to understand the meaning of spatial or abstract words. Indeed, while not their main objective, prior studies have inadvertently demonstrated that the final part in longer versions of the Token Test behaved differently (Gallardo et al., 2011) or did not correlate with the other subtests (McNeil et al., 2015). Thus, the manipulation of length and linguistic complexity in the Shortened Token Test could be mapped onto working memory and language skills respectively and separately, but empirical evidence is needed.

Overall, the Token Test seems to have potentially advantageous properties for providing separable estimates of working memory and language skills in supporting oral language comprehension. Of particular interest as a clinical tool would be a shortened version of the Token Test that could be administered easily and quickly as part of a comprehensive assessment. One problem with De Renzi and Faglioni’s (1978) 6-part Shortened Token Test is the small number of items in each part (Part 1: 7 items; Parts 2-5: 4 items each; Part 6: 13 items) for sampling behaviour. This resulted in poor data variability in our pilot study. To address this limitation, we created the *Modified-Shortened Token Test* by adding 9 more items each to Parts 4 and 5, for a total of 54-items (see Appendix A for examples). Importantly, the modified version had an equal number of items for Parts 4 and 5, the parts with the long and simple commands and Part 6, the part with the long and linguistically complex commands (13 commands each).

2.1.1 The Current Study

The purpose of the present study was to examine the factor structure of the Modified-Shortened Token Test and its relation to other tests of working memory and language by reanalyzing data available from a number of studies in our research program. In Experiment 1, a group of children aged 4- to 7-years-old completed the Modified-Shortened Token Test as part of a larger study on narrative development, which allowed us to examine the constructs underlying this tool. In Experiment 2, a group of kindergarten-aged children completed the Modified-Shortened Token Test as part of a study assessing a board-designed assessment tool. In Experiment 3, participants aged 8- to 17-years-old voluntarily attended an afterschool reading program for children who struggled with reading and spelling. Since data from Experiment 3 was collected simultaneously with the modification of the Token Test, participants completed the original Shortened Token Test as well as other language measures as part of their assessment battery. Experiment 3 is reported in this thesis to preserve transparency. To provide a preliminary validation of the different constructs underlying the Modified-Shortened Token Test, Experiments 2 and 3 assessed correlations between performance on the Token Test and measures of memory and language commonly used by speech-language pathologists.

A comparison between Experiments 2 and 3 may also provide preliminary data on developmental changes of working memory and language. Working memory and language skills are still developing in older children, but these skills may be developing more rapidly in younger children. The greatest increases in working memory capacity occur before mid-adolescence (Cowan et al., 2006; Gathercole, 1999), with the basic structure of working memory in place as early as by age 6 (Gathercole et al., 2004). The developmental time course of different aspects of language knowledge (i.e., syntax, semantics) differs. Syntax skills develop rapidly during early childhood, but the developmental trajectory levels off over time. Indeed, by 4 to 5 years of age, children's sentences include all elements (e.g., adverbial phrases, subordinate clauses) that adults use in their complex sentences (Hoff, 2014). In contrast, semantic knowledge can continue to develop substantially throughout the lifespan (Yee et al., 2017). Differential development could reveal interesting developmental patterns of working memory and linguistic knowledge functioning on language performance. We made the following hypotheses:

Hypothesis 1: The underlying structure of the Modified-Shortened Token Test might be explained better by separate constructs rather than solely assessing receptive language abilities. Specifically, manipulations of length and linguistic complexity in the Modified-Shortened Token Test would tap verbal working memory and language skills, respectively and separately. Heretofore, when discussing our studies, we are referring to verbal working memory unless otherwise stated.

Hypothesis 2: Working memory and language skill composite scores based on the Modified-Shortened Token Test would have convergent validity based on unique correlations with related working memory and language measures commonly used by speech-language pathologists.

Hypothesis 3: Given that working memory and language skills are developing more rapidly in younger (vs. older) children, different patterns of association between working memory and language composites formed from the Token Test and language measures were expected.

2.2 Experiment 1

The aim of Experiment 1 was to determine the factor structure of the Modified-Shortened Token Test. We expected that extracted factors would reflect separate working memory and language constructs underlying this tool.

2.2.1 Methods

2.2.1.1 Participants

We analyzed data from a group of 257 school-aged children from public schools in the southwest region of Ontario, Canada. Children ranged in age from 4- to 7- years old (M_{age} = 5.83, SD_{age} = 0.94; 146 males and 111 females). Children were drawn from the general population and there were no inclusion and exclusion criteria. Although a range of ethnicities resembling the diversity of the population was represented, specific data regarding ethnicities (including language spoken in the home) were not collected. Ethical approval was given by the local ethics committee and school board.

2.2.1.2 Materials

The details of the Modified-Shortened Token Test were discussed in the introduction. Briefly, the Modified-Shortened Token Test consisted of 54-items organized in six parts (Appendix A). As the test progressed, sentences increased in length and linguistic complexity. The child was required to listen to the sentence and then point to the token(s) or carry out simple commands. Responses were scored as 1 for correct, 0.5 for correct with repetition, and 0 for incorrect. The maximum total score for Part 1 is 7, Parts 2 and 3 is 4, and Parts 4, 5 and 6 is 13.

2.2.1.3 **Procedure**

Children were recruited from a larger study examining the efficacy of a narrative retell assessment tool. Children were seen individually by a research assistant (trained undergraduate students or community speech-language pathologists) in a quiet room at the child's school. Children completed a single 30-minute assessment session involving other measures not reported here, and the Modified-Shortened Token Test as the final task.

2.2.1.4 **Data analysis**

Factor analysis. First, the Bartlett's test of sphericity and Kaiser-Meyer-Olkin (KMO) test were performed to evaluate whether the data were suited for factor analysis, which would be indicated by a significant result and a KMO value of greater than 0.6, respectively. The planned factor analysis (if appropriate) included a varimax rotation to enhance the interpretation of the factors, and extraction of factors with eigenvalues greater than 0.7 (Jolliffe, 1972). For each factor, items that loaded highly with values of greater than 0.5 and also secondary loadings with values of greater than 0.3 were retained.

Linguistic parameters. Sentence length was based on number of words and syllables. Syntactic complexity was measured by the number of phrases and Yngve depth. Linguistic trees were generated using the Stanford Core Natural Language Processing website (https://corenlp.run/; Manning et al., 2014). The number of noun, verb, and preposition phrases were counted. Yngve depth also provides a metric of syntactic complexity by accounting for the number of embedded structures in a sentence, with left-branching phrases considered to be more complex than right-branching phrases (Yngve, 1960). Max depth in particular is the deepest (or most embedded) word in the sentence (i.e., highest score; Figure S2.1); this does not correspond to the 'hardest' word. For sentences in the Token Test, this could be thought of as the more adjectives modifying the noun, the more depth the sentence has. For example, *The small green square* has a max depth of 3, while *The green square* has a max depth of 2. Readability scores and reading level were assessed using the "readability statistics" tool in Microsoft Word. Sentences were typed into a Word document and then the tool was used to generate these scores. Semantic knowledge was based on age of acquisition, word concreteness (the degree to which a word can be imagined or perceived through our senses), and word frequency (the level of usage of individual words in spoken language). Only unique words from each part of the Token Test were used for this calculation. We used normative data from Brysbaert and Biemiller (2017) for age of acquisition and Brysbaert, Warriner, and Kuperman (2014) for word concreteness and word frequency.

A 'score' was assigned to each sentence based on the linguistic parameter being measured. 'Scores' refer to word length or max depth calculated for each sentence, for example. For descriptive statistics, a composite score was created by calculating the mean of all ‘scores’ of sentences from the parts that loaded on the same factor. For the statistical analyses, an independent sample *t*-test was used to compare composites. All scores that loaded on the respective factor were treated as an independent sample.

2.2.2 Results and Discussion

Analysis of Descriptive Statistics. Table 2.1 displays the descriptive statistics for each part of our Modified-Shortened Token Test, with corresponding proportion correct for ease of comparison between each part. Children performed near ceiling levels of 99% correct for Parts 1 and 2 and of 94% correct on Part 3. Performance was more variable for the remainder of the test. The mean score for Parts 4, 5, and 6 were 11.41 (SD = 1.76), 8.75 (SD = 3.16), and 9.32 (SD = 2.00), respectively, corresponding to average accuracy rates of 67-88%. Cronbach’s α internal consistency of reliability was 0.62 across all parts and 0.76 for Parts 3-6.

Table 2.1. Descriptive statistics for different parts of the Modified-Shortened Token Test in Experiment 1 (n = 257).

Token test	Mean (SD)	Maximum score	Proportion correct
Part 1	6.97 (.15)	7	.99
Part 2	3.97 (.16)	4	.99
Part 3	3.74 (.60)	4	.94
Part 4	11.41 (1.79)	13	.88
Part 5	8.75 (3.16)	13	.67
Part 6	9.32 (2.00)	13	.72

Exploratory factor analysis. Both a significant Bartlett's test of sphericity ($\chi^2(15) = 285.49$, $p < .001$) and the KMO measure of 0.72 indicated that the implementation of the factor analysis was appropriate. A principal axis factoring approach was performed on the raw scores of the Modified-Shortened Token Test. The analysis yielded three factors with eigenvalues greater than 0.7 (Jolliffe, 1972). The eigenvalue for the first factor was 2.37, 1.16 for the second factor, and .90 for the third factor, accounting for 73.76% of the total variance. Specifically, Factor 1 accounted for 39% of the variance, Factor 2, 19%, and Factor 3, 15%.

The factor loadings are given in Table 2.2, together with the percentage of variance explained in the rotated solution by each factor. Three distinct factors emerged, consistent with the three-factor model found for the 100-item Revised Token Test Spanish version (Gallardo et al., 2011). The first factor showed high loadings with Parts 3, 4, and 5 and minimal secondary loading with Part 6. This pattern reflects commands that increased in length, thereby, reflecting the need to retain more information in verbal working memory. Hence, Factor 1 may be considered a working memory factor. Part 6 loaded highly (loading = 0.89) on Factor 2. Part 6 involves processing commands that are grammatically and semantically different than the rest of the test, suggesting that Factor 2 is a linguistic factor. Finally, Factor 3 was deemed a basic attention factor with Parts 1 and 2 loading on this factor. We used this label to capture the possibility that minimal vigilance might be needed to get familiar with the task and complete even the easy parts. Gallardo et al. (2011) also found that their third factor corresponded to the easiest imperative sentences and made similar interpretations about this factor being related to children adjusting to the task requirements. Also consistent with Gallardo et al. (2011), performance on the easiest parts was very high and yielded little variability in scores as well as having lower reliability, suggesting a more likely alternative explanation that this factor simply identified the easiest parts of the Token Test, and further interpreting in cognitive terms would be unwarranted. Therefore, we do not pursue further analyses with this factor.

Table 2.2. Factor loadings (>.30) for the principal axis factoring with varimax rotation in Experiment 1 (n = 257).

	Factor		
	1	2	3
Part 1			.43
Part 2			.32
Part 3	.57		
Part 4	.68		
Part 5	.84		
Part 6	.35	.89	
% variance explained	39	19	15

Linguistic parameters. A composite score for each linguistic parameter was formed based on the three identified factors. The composite score for the working memory factor was created by averaging the scores for all the items across Parts 3, 4, and 5 – parts that loaded highly on Factor 1. Part 6 was excluded from the working memory composite and these items were used to form the linguistic composite, given its high loading on Factor 2. The composite score for the linguistic factor was created by averaging the scores for all 13 items in Part 6. The composite score for the basic attention factor comprised the averaged scores from Parts 1 and 2 (Table 2.3). The crucial comparison between the working memory and linguistic composites revealed no difference in word length, $t(41) = -1.60$, $p = .12$, but a difference in syllable length, $t(41) = -2.47$, $p = 0.018$, with sentences in the linguistic composite about 1 or 2 syllables longer than the working memory composite. The linguistic composite was more complex than the working memory composite in terms of phrasal complexity, readability statistics, concreteness of words, and age of acquisition of words. The linguistic composite had sentences with significantly more complex phrase structure than the working memory composite, $t(41) = -3.57$, $p < .001$. Readability scores indicated that the linguistic composite (Ease = 94.4; Grade = 2) was one grade level above the working memory composite (Ease = 100; Grade = 0.87). There were also semantic differences. Age of acquisition was higher for words in the linguistic than working memory composite, *t*(56) = -2.24, p = .02. Words in the linguistic composite were also less concrete (i.e., more abstract) than words in the working memory composite, *t*(56) = 2.77, p = .007. Word frequency did not differ, *t*(56) = 0.51, p = .61. Note, however, that there is a high degree of variability in word frequency because ‘the’ and ‘a’ are some of the most common words. In contrast, the Yngve max depth was significantly greater for the working memory than linguistic composite, *t*(41) = 4.78, *p* < .001.

Table 2.3. Composite score of each linguistic parameter for factors identified.

	Basic attention	Working memory	Linguistic
Length:			
Word length	3.82 (0.40)	8.47 (1.68)	9.31 (1.32)
Syllable length	4.27 (0.65	9.67 (1.93)	11.46 (2.70)
Syntax:			
Phrases	2 (0)	3.73 (0.69)	5.23 (1.17)
Max depth	1.82 (0.40)	4.3 (0.70)	3.23 (0.60)
Readability statistics:			
Ease	100	100	94.4
Grade level	0.0	0.87	2.0
Semantics:			
Age of acquisition	3.84 (0.93)	3.94 (0.79)	4.75 (1.86)
Concreteness	3.81(0.86)	3.59 (0.99)	2.85 (1.043)
Word frequency	122 047 (325 763)	173 174 (423 850)	121 963 (318 769)

Evaluating linguistic differences between the working memory and linguistic composites provided additional support for the idea that differential performance across items could reveal relative verbal working memory and language knowledge abilities. Crucially we found that the working memory and linguistic composites only differed in terms of syntactic and semantic processing and not word length. These findings support the assumption that working memory demands are higher for long sentences because more verbal information needs to be processed, captured by the working memory composite. In contrast, the linguistic composite or Part 6, while just as long, was linguistically more challenging. Sentences contained complex phrase structure, reading ease and grade level were lower, and involved semantically harder words, thus imposing a higher language load.

Interestingly, the working memory factor may have had a minimal linguistic load via depth of sentences, indicated by the significant max depth result. This result is likely driven by the high depth score in Part 5, reflecting the combination of using a compound declarative sentence and an additional adjective node specifying size (Part 6 only used large tokens). As an everyday example of increased depth but not linguistic complexity, think of the children's memory game "I'm going on a picnic", in which each person will add to the list and then recite all the items on the list ("*I'm going on a picnic and I'm bringing apples and she's bringing juice and he's bringing a blanket…*"). The sentence structure is simple but by continuously adding items to the list, this will increase depth and inadvertently impose a working memory load.

Following Experiment 1, additional work is needed to evaluate the factor structure of the Modified-Shortened Token Test. Therefore, the goal of Experiments 2 and 3 was to evaluate the relationships between this factor structure with associated measures in separate groups of kindergarten-aged children and older children, respectively.

2.3 Experiment 2

In Experiment 2, we examined the relationship between composite scores formed based on identified factors (from Experiment 1) and other standardized tests of language and working memory. This would provide a preliminary evaluation of external validity of the Modified-Shortened Token Test and its factors using other common measures in speech-language pathology.

One of the most widely used tool to assess language is the Clinical Evaluations to Language Fundamentals—Fourth Edition (CELF-4; Semel et al., 2003). The purpose of the CELF-4 is to screen for and diagnose language disorders in children. Briefly, there are four core subtests for children ages 5-8: Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences (more details in Methods). Despite primarily assessing language, Archibald (2013) found that some subtests, namely Concepts and Following Directions and Recalling Sentences, were also associated with working memory demands. This is consistent with our assumption that many, if not all, language-based tasks employ both working memory and language skills, but that some verbal items (e.g., increasing length, complexity) may tap one or another construct to a greater degree. Indeed, in these two subtests children are processing spoken sentences of increasing length and complexity, like the Token Test. Unlike the Token Test, on the other hand, the manipulation of length and complexity was not systematic, but sentences did have semantic variability (unique words). Based on these findings, we predict that the all CELF-4 language tasks would correlate with the linguistic composite, with some tasks (i.e., Concepts and Following Directions and Recalling Sentences) requiring verbal working memory skills to a greater degree.

We also had additional measures available that tapped different aspects of language knowledge. The school-board’s tool included a measure of Phonological Awareness (and a Narrative measure reported in the supplemental material). Phonological awareness is said to be a precursor skill to narrative abilities (Farrar et al., 2005), and therefore, we might expect this measure to be associated with language abilities. However, large-scale studies have shown that phonological awareness is not related to language-based weaknesses (e.g., Archibald et al., 2013). The Test of Narrative Language (TNL; Gillam & Pearson 2017) has primarily been used to measure children's ability to understand and tell stories, but could also be used to assess semantic and grammatical knowledge (Gillam et al., 2021). A comparison of the TNL with Part 6 could inform us about the extent to which semantic knowledge was assessed, although we recognize that the manipulation of semantic depth in Part 6 was limited. Finally, the Finger Windows task (Wide Range Assessment of Memory and Learning, Second Edition (WRAML-2); Sheslow & Adams, 2003) was used as a measure of non-verbal working memory. How performance on Token Test correlates with this measure would be indicative of reliance on general cognitive abilities not specific to verbal processing.

2.3.1 Methods

2.3.1.1 Participants

Twenty-four kindergarten-aged children were recruited from a larger study assessing a board-designed assessment tool at two time points five months apart, spring (range = 5;5-6;4) and fall (range = 5;10-6;9). Participants came from public schools in the southern region of Ontario, Canada. Grade and month of birth were reported; other demographic information as well as specific ages were not collected. Ethical approval was given by the local ethics committee and school board.

2.3.1.2 Materials

Modified-Shortened Token Test. The same 54-item Modified-Shortened Token Test was used as in Experiment 1.

Language measures[1]. Each child completed the four core subtests from the CELF-4 (Semel et al., 2003). In the Recalling Sentences subtest, the child immediately repeated the sentence they previously heard verbatim. In the Formulated Sentences subtest, the child was required to make a sentence based on the picture they were shown and word given. In the Concepts and Following Directions subtest, the child listened to the spoken instruction and then pointed to the corresponding picture. In the Word Structure subtest, the child completed a sentence with the grammatically correct word form. Each child completed the board-designed phonological awareness measure. The Phonological Awareness screening measures consisted of 10 tasks including sentence segmentation, syllable blending, syllable segmenting, onset and rime blending, onset and rime segmenting, initial sound correspondences, detecting individual sounds in words (blending), detecting individual sounds in words (segmenting), rhyme recognition, and rhyme production. All items were scored as correct or incorrect for a total score of 42 (each task had 4 items; exception: rhyme production had 6 items). Each child completed the TNL (Gillam & Pearson 2017). They heard stories with and without picture support and then were asked to retell the stories, answer questions, and make up their own stories.

---

[1] We have results from additional measures that overlapped with those reported in the main text or performance was at floor, but we reported them in the Supplementary Materials to preserve transparency

Working memory measure. Each child completed the Finger Windows subtest, a measure of visuospatial working memory, from the WRAML-2 (Sheslow & Adams, 2003). A card with holes was held up and the examiner pointed to a series of holes in turn. The child was then asked to point to the holes in the same order. The sequences became increasingly longer. Items gradually increased in length from sets of 1 to sets of 6 holes. Testing continued until three consecutive errors.

2.3.1.3 **Procedure**

Children were seen individually in a quiet room at the child's school by trained research assistants, who were speech-language pathologists or speech-language pathology graduate students. Testing occurred at two time points, separated by five months. In the spring, children completed a comprehensive battery of standardized tests of oral language and phonological awareness and then our Modified-Shortened Token Test and the finger windows subtest in the fall.

2.3.1.4 **Data analysis**

First proportion correct was calculated for each of the six part of the Token Test. Then scores were averaged across parts that loaded together to form the working memory and linguistic composite corresponding to their respective factor in Experiment 1. The working memory composite was made by averaging proportion correct from Parts 3, 4, and 5. Part 6 was not included in the working memory composite because it loaded highly on a separate factor, the linguistic factor. The linguistic composite was created by computing the proportion correct for all items in Part 6. Despite the small sample size, we conducted partial correlations using Pearson's correlation coefficient. This analysis would allow us to investigate (1) the link between working memory and related measures by controlling for linguistic effects and (2) the link between language and related measures by controlling for working memory effects. Importantly, when the non-parametric measure, Spearman's rank correlation was conducted, results were similar. Bayes Factor (BF_{10}) value are supplemented with standard partial correlations and p-values to quantify the strength of evidence for the correlation. Bayesian analyses were conducted using JASP (JASP Team, 2020). BF_{10} between 3-10 provides substantial evidence and $BF_{10} > 10$ provides strong evidence in favour of an effect (i.e., there is an association between the two variables) than the null hypothesis (JASP Team, 2020; Wagenmakers et al., 2018).

2.3.2 Results and Discussion

Descriptive statistics for each part of our modified Token Test as well as raw test scores across the memory and language measures are shown in Table 2.4. Results from the exploratory Pearson partial correlational analyses is presented in Table 2.5. The working memory composite was only correlated with Recalling Sentences, partial r = 0.45, p = .032, BF = 4.5, and Formulated Sentences, partial r = 0.42, p = .049, BF = 3.3. In contrast, the linguistic composite was only correlated with the Concepts and Following Directions, partial r = 0.58, p = .004, BF = 27, and Word Structure, partial r = 0.46, p = .027, BF = 5.06. Neither of the composites correlated with the Phonological Awareness measure (partial r < .37, p > .11, BF < 2.04, both cases), Narrative Language measure (partial r < .23, p > .23, BF < 0.74, both cases), or the Finger Windows task (partial r < 0.18, p > .42, BF < 0.55, both cases).

Table 2.4. Descriptive statistics for Modified-Shortened Token Test, language, and working memory measures used in Experiment 2 (n = 24).

Measures	Mean (SD)	Proportion correct
Part 1[a]	7.00 (0)	1.00
Part 2[b]	3.92 (.24)	.98
Part 3[b]	3.83 (.32)	.96
Part 4[c]	11.69 (1.33)	.90
Part 5[c]	8.81 (2.75)	.68
Part 6[c]	9.083 (1.95)	.70
Working memory composite		0.85 (0.089)
Linguistic composite		0.70 (0.15)
Language:		
Recalling Sentences	36.04 (14.83)	
Formulated Sentences	17.79 (9.14)	
Concepts and FD	28.08 (10.53)	
Word Structure	18.92 (4.69)	
Test of Narrative Language	53.04 (20.17)	
Phonological Awareness	12.67 (8.27)	
Working memory:		
Finger Windows	17.34 (5.36)	

Note. [a]Maximum score = 7. [b]Maximum score = 4. [c]Maximum score = 13.

Table 2.5. Pearson partial correlations between the identified factors and test measures, with Bayes Factor (BF), for kindergarten-aged children in Experiment 2 (n = 24).

Test measures	Working memory	BF_{10}	Linguistic	BF_{10}
Recalling Sentences	0.45*	4.52	0.36	1.88
Formulated Sentences	0.42*	3.28	0.37	2.044
Concepts and FD	0.29	1.09	0.58**	27.032
Word Structure	0.30	1.17	0.46*	5.058
Test of Narrative Language	0.14	0.46	0.23	0.74
Phonological Awareness	0.34	1.58	0.37	2.044
Finger Windows	-0.27	0.13	0.18	0.55

Note. FD = Following Directions. *$p < .05$; **$p < .01$. BF_{10} between 3-10 provides substantial evidence and $BF_{10} > 10$ provides strong evidence in favour of a correlation over the null hypothesis.

Preliminary evidence from Experiment 2 revealed that each composite showed a selective association with the language measures, though not in the way we had anticipated. Surprisingly, language tasks did not cross-load on both composites, contrasting the results of Archibald (2013). Presumably many language-based tasks tap both verbal working memory and linguistic abilities, but we speculate that these partial correlations are revealing that some tasks may be more sensitive to tax these differential constructs after controlling for specific variances. That is, working memory may contribute to performance on Recalling Sentences and Formulated Sentences above and beyond linguistic variables. It is interesting to note that the two language tasks that correlated with working memory only were language production tasks. However, it is difficult to interpret why Recalling Sentences only correlated with working memory, given that many studies have used sentence recall to index language (Archibald et al., 2013; Frizelle et al., 2019). We speculate that young children may be more sensitive to the memory component of sentence recall, that is having to retain and repeat the sentences verbatim was more taxing. Similarly, formulating sentences required that the child plan a complete, semantically and grammatically correct sentence in mind before responding, potentially imposing a high memory load.

In contrast, the link between language demands and Concepts and Following Directions and Word Structure, after accounting for verbal working memory, suggest that the language characteristics embedded in these tasks influenced young children to a greater extent. The structure of the Concepts and Following Directions task mimics the structure of the Token Test, that is, children are following verbal directions of increasing length and complexity. We did not analyze the structure of this subtest further, but it could be the case that overall performance reflects linguistic abilities generally, whereas performance across items would have tapped working memory and language differently. The Word Structure test was designed to assess morphological and grammatical knowledge (Wiig et al., 2013) and expectedly, primarily imposed a linguistic load for young children.

Notably, the Finger Windows subtest, Phonological Awareness, and the TNL were not correlated with any factors. Regarding the Finger Windows subtest, a test of non-verbal working memory, Schmoeger et al. (2020) also reported that only verbal, and not non-verbal intelligence scores, were correlated with performance on the Token Test for typically developing children. These findings together indicate that performance on the Token Test is specifically related to the verbal domain of working memory. This is also in line with suggestions that verbal working memory specifically supports language processing (Montgomery et al., 2021; Archibald & Gathercole, 2007), rather than more general working memory resources. For instance, children with DLD show considerably more marked and consistent impairments on verbal than visuospatial working memory tasks (Archibald & Gathercole, 2006; Vugs et al., 2013 for a review). Although non-verbal deficits can impact language processing (Vugs et al., 2013), they may not play a central role. Future work is needed to examine the extent to which the Token Test is related to working memory more generally.

These findings suggest that linguistic knowledge captured by the Modified-Shortened Token Test, Phonological Awareness, and TNL likely differs in some way. Phonological Awareness and TNL are both related to narrative skills, whereas the Token Test does not involve this skill. Moreover, the link between phonological awareness and language is not well-established (Archibald et al., 2013) or depends on language abilities (Khan et al., 2021). Khan et al. (2021) found that phonological awareness was correlated with poor but not high language skills. The lack of an association in our study may be due to the fact that children were randomly selected and represented a continuum of language abilities. On the other hand, the TNL has been used to index episodic, lexical-semantic and grammatical aspects of extant language knowledge within long-term memory (Gillam et al., 2019), whereas variation in semantic complexity is more limited in the Token Test. A limitation of the Token Test is that this test may primarily assess syntactic knowledge and minimally semantic knowledge. As such, differences between what each tool was designed to measure might have contributed to the differences in the factor loading. However, all these are tentative interpretations of the results that should be made with caution given the small sample size in Experiment 2.

2.3.3 Developmental patterns in the factor structure

The intriguing finding that language measures were differentially associated with the working memory or linguistic composite motivated us to take a closer at the factor structure of the Token Test in younger children. It could be that some items in the Token Test may tap working memory or language to a greater degree in younger children. Or, that young children would be more sensitive to the working memory and linguistic demands of certain sentences. Given that the Experiment 2 sample was younger (kindergarten-aged) compared to the Experiment 1 sample (kindergarten to grade 2), we reanalyzed the factor structure of the Modified-Shortened Token Test from Experiment 1 with a subset of the original data constrained to only younger-aged children (n = 101; M_{age} = 5;4, range = 4;6 – 5;11). We used only data from Experiment 1 because of the relatively large sample size, whereas a factor analysis would not be warranted with the small sample size of Experiment 2 as well as lack of variability in Part 1. Results were largely similar to the factor analysis of the entire sample, with the same three factors accounting for 75% of the total variance (Table 2.6). The only exception was that Part 5 also loaded on Factor 3 in the younger sample, in addition to Part 6. This suggests that Part 5 predominantly has verbal working memory demands, but for younger children Part 5 also imposed linguistic demands via sentence depth. Older children, then, may be less sensitive to these trivial linguistic demands, resulting in Part 5 not loading on the linguistic factor for the entire sample. This also confirms that Part 6 predominately has linguistic demands.

Table 2.6. Factor loadings (> 0.3) for the principal axis factoring for the younger group of children in Experiment 1 (n = 101) with varimax rotation.

	Factor		
	1	2	3
Part 1		.62	
Part 2		.66	
Part 3	.64		
Part 4	.52		
Part 5	.90		.31
Part 6	.30		.51
% variance explained	39	23	13
Eigenvalue	2.33	1.37	0.79

2.4 Experiment 3

The goal of the third experiment was to further evaluate the relationships between the factor structure from Experiment 1 with other language measures in a group of older children who struggled with reading and spelling. Notably, the original Shortened Token Test was used in Experiment 3. Because this paper relied on secondary data for analysis, data for Experiment 3 was inadvertently collected before the implementation of the

Modified-Shortened Token Test used in Experiments 1 and 2. Nonetheless, we expected older children to perform differently than younger children given developmental increases in working memory and language skills. However, language-related difficulties in reading and spelling may also impact performance.

The language-based tasks we had available for Experiment 3 included Recalling Sentences from the CELF-4, the Test of Word Reading Efficiency–Second Edition (TOWRE-2; Torgensen et al., 1999), and a Definition and Spelling task that were created for the reading program (more details about each task in Methods). In Experiment 2, we found that Recalling Sentences was associated with the working memory composite and could expect the same for this group of children. Children with language-related difficulties may face higher working memory loads just to retain basic verbal information in mind. However, given the plethora of evidence supporting Recalling Sentences as a language task (Archibald, 2013; Frizelle et al., 2019; Klem et al., 2015), we might expect children with language challenges to be more sensitive to the linguistic characteristics of the task. The TOWRE-2 required children to read words and decode non-words, which is most closely related to reading and phonological abilities. Since we found that both the Phonological Awareness and TNL tests did not correlate with any composites in Experiment 2, we might expect similar results here. Finally, it is hard to directly compare the Definition and Spelling tasks to any measures used in Experiment 2. Providing a definition (or using the word in a sentence) is a language production task and we found that such tasks imposed a working memory load to a greater degree. Spelling skills, on the other hand, is tied to knowledge of phonology to spell words, and hence, we predict that this task draws on language knowledge not tapped by the Token Test. Alternatively, the original Token Test was used in Experiment 3 and could lead to poor data variability overall. Correlations may not be detected when variance is not sufficient.

2.4.1 Methods

2.4.1.1 Participants

Twenty-three participants, aged 8-17 years (M_{age} = 11.59, SD_{age} = 2.46), were recruited from an afterschool reading program. The majority of children were 9-years-old (n = 7) or 12-years-old (n = 6), with the remaining age groups having very few participants (n = 1 for age 8, 10, 13, and 17; n = 2 for age 15; n = 3 for age 14). Participants were invited to be part of the research and voluntarily attended the program in southwest Ontario. This study was approved by the local ethics committee.

2.4.1.2 Materials

Shortened Token Test. In this study, participants completed the original 36-item Shortened Token Test. This did not include the additional 9 commands added to parts 4 and 5 used in Experiments 1 and 2.

Language measures[1]. Each child completed a battery of tests including two standardized tasks and two experimental tasks. In the Recalling Sentences subtest of the CELF-4, the child was required to repeat sentences after hearing them (Semel et al., 2003). Each child completed TOWRE–2 (Torgensen et al., 1999). Each child was presented with a list of 108 words or 62 non-words and was asked to read as many printed words (Sight Word Reading) or non-words (Non-word Reading), respectively, as possible within 45 seconds. Words and non-words increased in difficulty from monosyllabic to multisyllabic. Words were counted as correct if they were read accurately within the time limit. The Spelling and Definition tasks consisted of 20 words that were drawn from the reading program; the same words were used for all children. In the Spelling task, the child spelled a list of words. In the Definition task, the child was asked to provide a definition for a given word or use it in a sentence if they could not provide a definition. Examiners did not note whether the child provided a definition or sentence. Items for the Spelling and Definition tasks were scored as correct (1 point) or incorrect (0).

2.4.1.3 Procedure

Some children were seen individually in a quiet room at the local university's clinic or community center, while others were tested in a relatively quiet room with other participants seated near-by. A graduate student in speech-language pathology conducted the assessment battery in one session lasting 1 hour. Each child completed the original Shortened Token Test as well as various tests assessing language skills. Scores from the initial assessment are reported here.

2.4.2 Results and Discussion

Descriptive statistics are provided in Table 2.7. Pearson partial correlations were computed between composites scores based on the factor analysis and language measures (Table 2.8). The working memory composite was not correlated with any language tasks. The linguistic composite was correlated with Recalling Sentences, partial r = 0.43, p = .045, BF = 3.64. BF analysis provided only substantial, not strong, evidence for this

Table 2.7. Descriptive statistics for original Shortened Token Test and language measures used in Experiment 3 (n = 23)

Measures	Mean (SD)	Proportion correct
Part 1[a]	6.98 (0.10)	.99
Part 2[b]	4 (0)	1.00
Part 3[b]	3.85 (0.35)	.96
Part 4[b]	3.93 (0.23)	.98
Part 5[b]	3.54 (0.71)	.89
Part 6[c]	11.02 (1.40)	.85
Working memory composite		0.94 (0.076)
Linguistic composite		0.85 (0.11)
Language:		
Recalling Sentences	62.48 (13.71)	
Word Reading	63.43 (16.13)	
Non-word Reading	31 (14.31)	
Spelling	10.95 (6.37)	
Definition	16 (2.89)	

Note. [a]Maximum score = 7. [b]Maximum score = 4. [c]Maximum score = 13.

Table 2.8. Pearson partial correlations between the identified factors and test measures, with Bayes Factor (BF), for the older group of children in Experiment 3 (n = 23).

Test measures	Working memory	BF_{10}	Linguistic	BF_{10}
Recalling Sentences	0.31	1.26	0.43*	3.64
Sight word Reading	0.012	0.27	0.16	0.50
Non-word Reading	-0.17	0.16	0.26	0.89
Spelling	0.052	0.31	-0.13	0.17
Definition	0.066	0.33	0.16	0.50

Note. **p* < .05; ***p* < .01. BF_{10} between 3-10 provides substantial evidence and BF_{10} > 10 provides strong evidence in favour of a correlation over the null hypothesis.

relationship (Wagenmakers et al., 2018). The two subtests of the TOWRE-2 as well as the Spelling and Definition tasks did not correlate with any composites.

We did not find strong support for Hypothesis 3, but there was preliminary evidence that Recalling Sentences differed between the two groups of children. Recalling Sentences was only related to working memory in the younger group of children, whereas it was related to the linguistic abilities in the older group of children. However, a direct comparison is difficult because the younger group of children in Experiment 2 were drawn from the general population, whereas older children in Experiment 3 were attending a program for children with language difficulties. Differential performance could be tied to both age and language-related difficulties in the older group of children. It could be that sentence recall largely acts as a working memory test for young children or children with good linguistic skills, but indexes language for older children or imposes a language load for children with poorer linguistic abilities. Nonetheless, given key differences between the dataset, it is difficult to draw firm conclusions about the relationships between performance on the Token Test and other language measures. These interpretations are clearly speculative and aims only at stimulating further research.

We again found that tasks relying on phonological knowledge such as reading tasks from the TOWRE-2 and the Spelling task did not correlate with any factors, further indicating that the Token Test is presumably tapping syntactic and semantic language knowledge, but not phonological. The Definition task was not associated with any composites either, despite being a language production task. It might simply be that the word list (e.g., *mixes, myself, unfriendly*) was relatively easy for older children.

A major limitation of Experiment 3, however, concerns the use of the original Token Test and close to ceiling performance for older children. Across all parts, the mean was ≥ 85% and there was little variance in performance. Coupled with developmental factors, the small number of items for each part in the original version may be limiting data variability. In fact, this was the reason for developing the modified version, but unfortunately, data for Experiment 3 was collected beforehand. Only items in Part 6, which formed the linguistic composite, was consistent with the modified version and indeed, we found some emerging evidence, but only at the substantial level, for a relationship between Recalling Sentences and the linguistic factor. Nonetheless, Experiment 3 perhaps makes the case as to why the original version of the Token Test may not be adequate for discriminating between working memory and language abilities, and for use with children with DLD. More research is needed to investigate how the Modified-Shortened Token Test could inform working memory and linguistic influences on language performance in older children.

2.5 General Discussion

The current study investigated the separability of verbal working memory and linguistic skills in performance on our modified version of the Token Test, a comprehension measure manipulating length and complexity of verbal commands. The results demonstrated that children's differential performance across on the Modified-Shortened Token Test could reveal relative verbal working memory and language knowledge abilities as employed in sentence processing. Second, Experiments 2 and 3 provided preliminary evidence for correlations between the factors of our Token Test and language tests commonly used by speech-language pathologists. Although most language tasks likely tap both verbal working memory and language knowledge, we found specific links suggesting that working memory or language constructs had distinct contributions to individual tasks. Overall, our findings provide preliminary evidence that performance across items on the Token Test could be used to differentially tap working memory and language knowledge in children. The results are consistent with prior empirical studies demonstrating overlapping but separable influences on language functioning of both working memory and existing language knowledge (Archibald & Joanisse, 2009, 2012; Noonan et al., 2014)

2.5.1 Factor structure of the modified Token Test

Of particular interest for the present study and consistent with Hypothesis 1 was that performance on the Modified-Shortened Token Test was explained by separate factors. Specifically, the working memory factor included parts that involved a higher memory load via sentence length (Parts 3-6), while the linguistic factor captured sentences involving higher linguistic complexity (Part 6). Processing long sentences imposes a working memory load as there is more verbal information to be processed. However, linguistic demands in Parts 3 to 5 are relatively low given that sentences use standard word order and familiar words. In contrast, Part 6 independently and separately loaded onto the linguistic factor. Not only does Part 6 involve remembering a command that is just as long as the previous part, it may also be sensitive to linguistic abilities because of complex phrase structure and vocabulary, thus imposing a higher load on linguistic processing resources than the working memory factor.

Indeed, we quantified that the linguistic composite had unique linguistic demands compared to the working memory composite. In Part 6, syntax (phrase structure), readability scores, and semantics (age of acquisition, concreteness of words) were more difficult, stemming from a combination of using logical operators (*if-then* conditional statements), logical forms (inferring that *except* means *all but*), prepositions (*with*), and spatial vocabulary (*far away*, *next to*). Therefore, in Part 6, on top of processing a long sentence (tapping working memory), children also had to understand and process different linguistic structures and new vocabulary (tapping language), whereas Parts 3-5 (the working memory composite) placed a constant linguistic load given syntactically identical commands (e.g., “Touch [X] and [X]”). Though note that the Token Test might be assessing differences in syntactic knowledge primarily as the variation in semantic complexity is somewhat limited. That is, the number of unique words in the Token Test is limited (e.g., *red, small, next to, in addition to*) compared to real-world sentences that would vary much more in semantic meaning. Balancing the variation in syntactic and semantic complexity will be an important point to consider in future work in this area. Nevertheless, our factor analysis provides preliminary support for the hypothesis that, in contrast to the original development of the Token Test as a test of oral language comprehension, children’s performance across items could differentially tap working memory or linguistic skills at least separately to some extent.

2.5.2 Relationships between the Token Test and language measures

As the first step to validate the identified factors and evaluate interrelationships between these factors and related language and memory measures, performance by separate groups of children in Experiments 2 and 3 were analyzed. Results from Experiments 2 and 3 provide preliminary evidence that the Token Test was tapping into verbal working memory and language knowledge processing as unique relationships with other measures that also tap into verbal working memory and language knowledge processing emerged, though not in the directions we had predicted (Hypothesis 2). In Experiment 2, kindergarten-aged children were sensitive to demands of memory in the sentence repetition and sentence formulation tasks, whereas language demands were more prevalent in the following directions and word structure tasks. The dissociation between the contributions of working memory and language in young children motivated us to reanalyze the factor structure of the Token Test to assess differential performance across age groups. For kindergarten-aged children, Parts 5 and 6 loaded with the linguistic factor, both parts were quantified to be the most syntactically complex via depth of sentence and phrasal structure, respectively. Very young children were therefore sensitive to linguistics demands that were otherwise trivial to the entire sample.

On the other hand, for older children in Experiment 3, the only correlation that emerged was between the linguistic composite and Recalling Sentences, with BF analysis deeming this evidence to only be substantial. It may be that children who struggled with reading and spelling, or language more generally, were more sensitive to the linguistic demands involved in sentence recall. Sentence recall does not simply require just repeating a series of word, but is also supported by semantic, morphological, and syntactic knowledge (Frizelle et al., 2019; Klem et al., 2015). Although distinct correlations between Experiments 2 and 3 emerged (Hypothesis 3), an important point to consider when interpreting the results of Experiment 3 is that a number of factors may be impacting performance in this older group (e.g., developmental increases, language-related learning difficulties, original Token Test).

Somewhat surprisingly, the results did not directly support our initial assumption that all languages subtests of the CELF-4 would correlate with the linguistic composite and verbal memory to play a role in some subtests. However, partial correlational analyses were used to control for working memory and language factors in respective analyses. The results suggest that each language task taps into at least partially distinct constructs once the associated variance was removed. Another note is that the verbal working memory composite provided better estimates given that it was composed of more Token Test parts (Parts 3, 4, 5) compared to the linguistic composite (Part 6 only). Future work in our lab aims to extend this investigation by constructing a version of the Token Test with better balance between composite scores thereby improving estimates of working memory and language.

2.5.3 Limitations

There are several other limitations and considerations that should be noted. Most notably, the current work represents a secondary analysis of data and was limited by what data was available. For instance, the sample was small in Experiments 2 and 3. Correlational findings must be interpreted with caution and this study needs to be conducted with a larger sample. Second, the wide and older age range in Experiment 3 (ages 8 – 17) meant that we were unable to reanalyze our factor structure to examine developmental patterns in the factor structure. Relatedly, performance on the Modified-Shortened Token Test in older children is needed. Third, the finding that the Finger Windows and other language measures (e.g., narrative tasks) did not correlate with the any composites suggest that additional research is needed to discern what type of specific working memory skills (verbal and nonverbal) and language abilities, respectively, are being reflected by performance on the Modified-Shortened Token Test. Future modifications of the Token Test should systematically vary sentences in terms of syntactic and semantic complexity to better reflect real-world sentences. Finally, since we showed that working memory and language skills could be separated constructs in a sample of children drawn from the general population, it will be important for future work to understand this separation in children with DLD and to examine correlations between these constructs and related cognitive measures in children with and without

DLD. Hence, the overall observations here serve only as pointers for further research with limited implications for immediate practice.

2.5.4 Conclusion

Verbal working memory and language knowledge are highly intertwined. Therefore, it is important for clinicians to understand how verbal working memory and linguistic skills influence language performance. This study demonstrated that the Modified-Shortened Token Test could be one tool to provide separate estimates of verbal working memory and linguistic skills. Further, unique relationships with language measures emerged based on whether the verbal task primarily tapped verbal working memory or language skills. The results highlighted how language tests may pose higher memory or language demands thereby influencing language performance. More broadly, these findings have the potential to inform assessment, and contribute to our understanding of why some children experience language learning difficulties.

2.6 References

Alloway, T., Gathercole, S., & Pickering, S. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77(6),* 1698–1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Archibald, L. M. D. (2013). The language, working memory, and other cognitive demands of verbal tasks. *Topics in Language Disorders, 33(3),* 190–207. https://doi.org/10.1097/TLD.0b013e31829dd8af

Archibald, L., & Gathercole, S. (2007). The complexities of complex memory span: Storage and processing deficits in specific language impairment. *Journal of Memory and Language, 57(2),* 177–194. https://doi.org/10.1016/j.jml.2006.11.004

Archibald, L. M. D., & Joanisse, M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52(4),* 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Archibald, L. M. D., Oram Cardy, J., Joanisse, M. F., & Ansari, D. (2013). Language, reading, and math learning profiles in an epidemiological sample of school age children. *PloS One, 8(10),* e77463–e77463. https://doi.org/10.1371/journal.pone.0077463

Arvedson, C.J., McNeil, M., and West, T. (1985). Prediction of Revised Token Test overall, subtest, and linguistic unit scores by two shortened versions. *Clinical Aphasiology, 15,* 57-63.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 647–667). Academic Press.

Bastiaanse, R., Raaijmakers, S., Satoer, D., & Visch-Brink, E. (2016). The Multilingual Token Test. *Aphasiology, 30(4),* 508–508. https://doi.org/10.1080/02687038.2015.1121710

Bohm, B., Smedler, A. C., & Forssberg, H. (2004). Impulse control, working memory and other executive functions in preterm children when starting school. *Acta Pædiatrica, 93(10),* 1363–1371. https://doi.org/10.1080/08035250410021379

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods, 49(4),* 1520–1523. https://doi.org/10.3758/s13428-016-0811-4

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46(3),* 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cohen-Mimran, R., & Sapir, S. (2006). Deficits in working memory in young adults with reading disabilities. *Journal of Communication Disorders, 40(2),* 168–183. https://doi.org/10.1016/j.jcomdis.2006.06.006

Cole, K., & Fewell, R. (1983). A quick language screening test for young children: The Token Test. *Journal of Psychoeducational Assessment, 1 (2),* 149–153. https://doi.org/10.1177/073428298300100205

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology, 42(6),* 1089–1102. https://doi.org/10.1037/0012-1649.42.6.1089

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token test. *Cortex, 14(1),* 41-9.

De Renzi, E., &Vignolo, L. A. (1962). The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain, 85(4),* 665-678. https://doi.org/10.1093/brain/85.4.665

Ellis Weismer, S., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42(5),* 1249–1260. https://doi.org/10.1044/jslhr.4205.1249

Fidler, Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. American Journal of Speech-Language Pathology, 20(1), 2–13. https://doi.org/10.1044/1058-0360(2010/09-0096)

Frizelle, P., Thompson, P., Duta, M., & Bishop, D. V. M. (2019). Assessing Children's Understanding of Complex Syntax: A Comparison of Two Methods. *Language Learning, 69(2),* 255–291. https://doi.org/10.1111/lang.12332

Gallardo, G., Guardia, J., Villasenor, T., & McNeil, M. R. (2011). Psychometric data for the Revised Token Test in normally developing Mexican children ages 4-12 Years. *Archives of Clinical Neuropsychology, 26(3),* 225–234. https://doi.org/10.1093/arclin/acr018

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3(11),* 410–419. https://doi.org/10.1016/S1364-6613(99)01388-1

Gillam, R., & Pearson, N. (2017). *Test of Narrative Language (2nd ed.).* Pro-Ed.

Gillam, R., Montgomery, J., Evans, J., & Gillam, S. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: Implications for assessment and treatment. *International Journal of Speech Language Pathology, 21(3),* 240–251. https://doi.org/10.1080/17549507.2018.1559883

Gray, S., Fox, A., Green, S., Alt, M., Hogan, T., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research, 62(6),* 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Isaki, E., Spaulding, T., & Plante, E. (2008). Contributions of language and memory demands to verbal memory performance in language-learning disabilities. *Journal of Communication Disorders, 41(6),* 512–530. https://doi.org/10.1016/j.jcomdis.2008.03.006

JASP Team. (2020). JASP (Version 0.14.1) [Computer software].

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis, I. Artificial data. *Applied Statistics, 21,* 160–173

Kapantzoglou, M., Restrepo, M., Gray, S., & Thompson, M. (2015). Language Ability Groups in Bilingual Children: A Latent Profile Analysis. *Journal of Speech, Language, and Hearing Research, 58*(5), 1549–1562. https://doi.org/10.1044/2015_JSLHR-L-14-0290

Magimairaj, B. M., & Montgomery, J. W. (2012). Children's verbal working memory: Role of processing complexity in predicting spoken sentence comprehension.

Journal of Speech, Language, and Hearing Research, 55, 669-682. https://doi.org/10.1044/1092-4388(2011/11-0111)

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7,* 357. https://doi.org/10.3389/fnhum.2013.00357

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60)

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46(5),* 1138–1153. https://doi.org/10.1044/1092-4388(2003/089)

McNeil, M., Pratt, S., Szuminsky, N., Sung, J., Fossett, T., Fassbinder, W., & Lim, K. (2015). Reliability and validity of the computerized Revised Token Test: comparison of reading and listening versions in persons with and without aphasia. *Journal of Speech, Language, and Hearing Research, 58(2),* 311–324. https://doi.org/10.1044/2015_JSLHR-L-13-0030

Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2021). A new memory perspective on the sentence comprehension deficits of school-Age children with developmental language disorder: Implications for theory, assessment, and intervention. *Language, Speech & Hearing Services in Schools, 52(2),* 449–466. https://doi.org/10.1044/2021_LSHSS-20-00128

Park, G. H., McNeil, M. R., & Tompkins, C. A. (2000). Reliability of the Five-Item Revised Token Test for individuals with aphasia. *Aphasiology, 14(5-6),* 527–535. https://doi.org/10.1080/026870300401298

Paquier, P., Van Mourik, M., Van Dongen, H., Catsman-Berrevoets, C., Creten, W., &Van Borsel, J. (2007). Normative data of 300 Dutch-speaking children on the Token Test. *Aphasiology, 23 (4),* 427–437.

Schmoeger, M., Deckert, M., Eisenwort, B., Loader, B., Hofmair, A., Auff, E., & Willinger, U. (2020). Evaluating Part V of the German version of the Token Test as a screening of specific language impairment in preschoolers. *Applied Psycholinguistics*, *41*(1), 237–258. https://doi.org/10.1017/S0142716419000493

Schwering, S. C. & Macdonald. M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14*, 68. https://doi.org/10.3389/fnhum.2020.00068

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, Fourth Edition (CELF-4).* Psychological Corporation.

Sheslow, D., & Adams, W. (2003). *Wide range assessment of memory and learning (2nd ed.).* Wide Range, Inc.

Spellacy, F. J., & Spreen, O. (1969). A short form of the Token Test. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *5*(4), 390-397. https://doi.org/10.1016/S0010-9452(69)80015-8

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary.* American Chemical Society.

Torgensen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Pro-Ed.

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities, 34(9),* 2586–2597. https://doi.org/10.1016/j.ridd.2013.05.014

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., …Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review, 25,* 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wiig, E. H., Semel, E., Secord, W. A. (2013). *Test Objectives and Descriptions for CELF-5: Clinical evaluation of language fundamentals, Fifth Edition*. Pearson. Retrieved from https://www.pearsonclinical.ca/content/dam/school/global/clinical/canada/programs/celf/CELF-5-objectives-descriptions_cdn_lr.pdf

Yee, E., Jones, M. N., & McRae, K. (2017). Semantic Memory. In J. T. Wixted & S. Thompson-Schill (Eds.), *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th Edition, Volume 3: Language and Thought). New York: Wiley.

Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society, 104*, 444–466

2.7 Supplemental Material

2.7.1 Supplementary Experiment 1

2.7.1.1 Methods

Linguistic parameters

Length. We analyzed the length of a sentence by counting the number of words and number of syllables.

Syntactic complexity. Syntactic complexity was defined by number of clauses and Yngve depth (Yngve, 1960). Linguistic trees were generated using the Stanford Core Natural Language Processing website (https://corenlp.run/; Manning et al., 2014). The number of clauses was calculated in two ways: 1) declarative clauses which was defined as the number of S nodes in the linguistic tree and 2) phrasal nodes (XPs) in the linguistic tree were also analyzed.

Another way to measure syntactic complexity is by using Yngve depth to evaluate the tree depth (reflecting the average number of embedded structures in a sentence). We computed the max and total Yngve depth of each sentence. Scores are assigned by giving a score of 0 to the rightmost branch under a given node and then increasing the score of each branch by 1 going from right to left. The total Yngve depth of each word is the sum of all the branches that connect that word to the root node. The max Yngve depth is the word in the sentence with the most depth (i.e., highest score) and the total Yngve depth is the sum of depth over all words. For example, the sentence, “In addition to touching the yellow circle, touch the black circle” has a max depth of 3 and total depth of 18 (Figure

S2.1). Total Yngve depth was not reported in the main text because it is similar to our sentence length measure, but is provided in this online supplement.

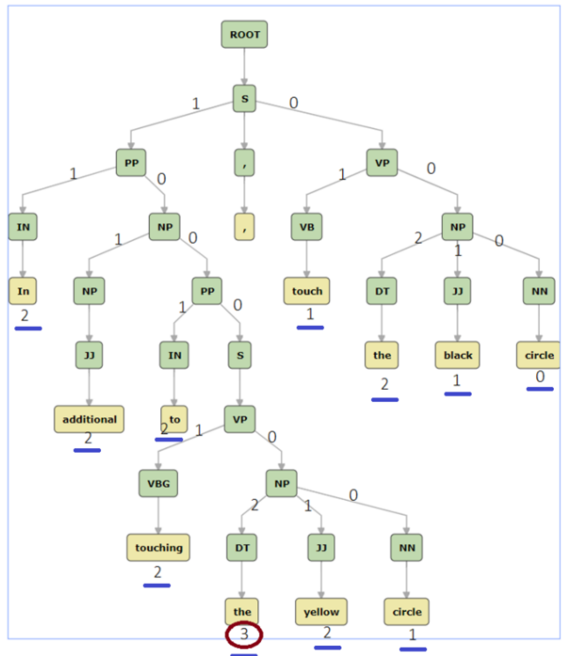

Figure S2.1. A linguistic tree illustrating the calculation of Yngve max depth (score circled in red = 3) and total depth (sum of all underlined scores = 18).

Readability. We used the “readability statistics” tool that is available in Microsoft Word to estimate the reading level for each part of the Modified-Shortened Token Test. Note that participants are not reading sentences during the Token Test, instead, readability statistics were generated by computing a score for the reading level of the sentences typed into a document in Microsoft Word. Readability was measured by Flesch Reading Ease, the higher the score, the easier it is to understand, and Flesch-Kincaid Grade Level, which determines the minimum level of education required for the reader to understand the text.

Table S2.1. Descriptive statistics for linguistic measures for each part of the Modified-Shortened Token Test.

	Part 1	Part 2	Part 3
Length:			
Word length	3.71 (0.49)	4 (0)	5 (0)
Syllable length	4 (0.57)	4.75 (0.5)	5.75 (0.5)
Syntax:			
Declarative clause	1 (0)	1 (0)	1 (0)
Phrasal nodes	2 (0)	2 (0)	2 (0)
Max depth	1.71 (0.49)	2 (0)	3 (0)
Total depth	3.43 (0.98)	4 (0)	7 (0)
Readability statistics:			
Ease	100	100	100
Grade level	0.0	*0.0*	0.0
Semantics:			
Age of acquisition	3.8 (0.95)	*3.89 (0.95)*	4 (0.96)
Concreteness	3.86 (0.82)	*3.76 (0.96)*	*3.67 (0.87)*
Word frequency	113 158 (93 811)	134 270 (366 460)	108 264 (327 806)
	Part 4	**Part 5**	**Part 6**
Length:			
Word length	8 (0)	10 (0)	9.31 (1.32)
Syllable length	9.31 (0.63)	11.23 (0.83)	11.46 (2.70)
Syntax:			
Declarative clause	1 (0)	1 (0)	1.31 (0.48)
Phrasal nodes	4 (0)	4 (0)	5.23 (1.17)
Max depth	4 (0)	5 (0)	3.23 (0.60)
Total depth	14 (0)	22 (0)	14.31 (3.40)
Readability statistics:			
Ease	100	100	94.4
Grade level	1.1	*1.5*	2.0
Semantics:			
Age of acquisition	3.86 (0.69)	3.96 (0.76)	4.75 (1.86)
Concreteness	3.58 (1.13)	3.55 (1.03)	2.85 (1.04)
Word frequency	222 521 (497 442)	186 142 (457 906)	121 963 (318 769)

2.7.1.2 Results and Discussion

Descriptive statistics for the linguistic parameters calculated for each part of the Modified-Shortened Token Test are presented in Table S2.1.

Based on the three factors identified in Experiment 1, a composite score was formed based on proportion items correct for relevant sections and then correlations were formed. There was no significant difference between the working memory and linguistic composites with respect to total Yngve depth, $t(41) = 1.38$, $p = .18$ (Table S2.2). Total Yngve depth is similar to sentence length, which also did not differ between the two composites, as reported in the main text.

Table S2.2. Composite score of total depth for factors identified.

Composite	Syntactic complexity Total depth
Basic attention	3.64 (0.81)
Working memory	17.4 (5.03)
Linguistic	14.31 (3.40)

2.7.2 Supplementary Experiment 2

2.7.2.1 Methods

Language measure. Each child completed the board-designed *DDSB narrative language* measure. The child listened to a story with corresponding images and then was asked to retell the story, answer comprehension and vocabulary questions, and share a personal narrative about a similar event. All other available test measures are described fully in the main text.

2.7.2.2 Results and Discussion

A composite score was formed for each factor identified in Experiment 1 and then correlated with the board-designed narrative measure. Pearson partial correlations of the board-designed DDSB narrative measure with the working memory composite (controlling for language) and with the linguistic composite (controlling for working memory) were not significant, partial $r = .003$, $p = 1$, BF = 0.26, and partial $r = .11$, $p = .65$, BF = 0.40, respectively. Similarly, the Test of Narrative Language (TNL; Gillam &

Pearson 2017) was not associated with any composites either, as reported in the main text.

2.7.3 Supplementary Experiment 3

2.7.3.1 Methods

Language measures. Each child also completed nonstandardized tasks including two reading tasks, an identifying affixes task, and a morphological task. Each child read a List of words drawn from the reading program and a Passage from the DIBELS Oral Reading Fluency (DIBELS ORF) corresponding to their grade level (Good et al., 2007). The Identifying Affixes task used the same list of 20 words as the Spelling and Definition tasks described in the main text. In this task, however, each child analyzed and identified the prefixes and suffixes of each word. Each child completed two tests of morphological structure, with 30 items on each test (Carlisle, 2000). In the Decomposition task, the child was provided with a morphologically derived word and then a sentence context requiring the child to provide the morphological base word form (The word is *driver*. The sentence is: *Children are too young to* ___.). In the Derivation task, each child is provided with a word and then a sentence context that required the child to provide the morphological derived word form (The word is *farm*. The sentence is: *My uncle is a* ___.). Scores from the two tests were combined for a total score. All other language measures are described fully in the main text.

2.7.3.2 Results and Discussion

Pearson partial correlational analyses revealed that these additional test measures did not correlated with either the linguistic or working memory composite. Full results are provided in Table S2.3. Finding that the two reading tasks did not correlate with the identified factors substantiates the presumption that the Token Test is not likely tapping language knowledge related to phonology or narrative abilities, as discussed in the main text. Nevertheless, the presence of floor effects across most of the language tasks presented in the supplement coupled with highly accurate performance on the original Token Test with little variability, might have limited our ability to find differential correlations.

Table S2.3. Partial correlations between the identified factors and additional test measures, with Bayes Factor (BF), for the older group of children in Experiment 3 (n = 23).

Test measures	Working memory	BF	Linguistic	BF
Reading list	0.043	0.30	0.022	0.28
Reading passage	-0.003	0.26	0.22	0.70
Identifying affixes	-0.056	0.22	-0.24	0.13
Morphological awareness	0.19	0.59	0.12	0.41

2.7.4 Supplementary References

Carlisle, J. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading & Writing, 12(3),* 169–190. https://doi.org/10.1023/A:1008131926604

Gillam, R., & Pearson, N. (2017). *Test of narrative language (2nd ed.).* Pro-Ed.

Good, R. H., & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th ed.).* Eugene, OR: Institute for the Development of Education Achievement.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60)

Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society, 104*, 444–466.

Chapter 3

3 The role of phonological and semantic representations in verbal short-term memory and delayed retention

3.1 Introduction

Knowing a familiar word entails representing the word form via phonological representations as well as having enriched linguistic knowledge (lexical, semantic, syntactic knowledge) associated with it. Yet, traditional accounts of verbal short-term memory suggest a distinction between language processes involved in short- and long-term memory tasks. In particular, rapid encoding of phonological representations has been attributed to short-term memory processes, whereas activation of associated semantic knowledge has been considered independently related to episodic long-term memory processes or strategy use. In contrast, alternative explanations based on language-based models of verbal short-term memory conceptualize verbal short-term memory as the activation of different levels of linguistic knowledge (phonological, lexical, semantic) within the language system (Majerus, 2013; Schwering & MacDonald, 2020). As such, language-based models allow for direct activation of phonological and semantic information during language processing. Serial recall, the immediate repetition of items in presented order, is one of the most common measures of short-term memory, but is best suited for investigating the role of phonological processing (Campoy & Baddeley, 2008) and rehearsal (Tan & Ward, 2008) in keeping verbal items active. In contrast, there is a relative dearth of paradigms for investigating semantic processing in verbal short-term memory. Thus, the current study aimed to reconcile traditional phonological effects with the growing literature of semantics effects in verbal short-term memory by employing a new paradigm to directly compare the retention of phonological and semantic information in verbal short-term memory as well as long-term impacts.

The notion that verbal short-term memory is largely influenced by phonological effects and minimally by semantic effects comes from a long history. Seminal work by Baddeley (1966) found a greater detrimental effect of phonological similarity (poorer recall for lists that are phonologically similar as opposed to distinct) than semantic similarity on short-term memory, which was taken as evidence for phonological coding in verbal short-term memory. This finding spurred on a large body of research focusing on phonological properties and rehearsal mechanisms in verbal short-term memory, including the word length effect (better recall for lists of shorter than longer words) and articulatory suppression (Baddeley et al., 1975). Some researchers continue to use the terms phonological short-term memory and verbal short-term memory interchangeably (e.g., Papagno & Cecchetto, 2019).

However, it is now well established that semantic knowledge influences verbal short-term memory as well. Evidence comes from the findings that words with richer semantic representations are remembered better: the lexicality effect (better recall for words than non-words; Hulme et al., 1991) and the concreteness effect (better recall for concrete than abstract words; Romani et al., 2008; Walker & Hulme, 1999). Neuropsychological studies have been instrumental in demonstrating the interaction between language processing and verbal short-term memory. On verbal short-term memory tasks, some patients present with phonological deficits but intact semantic effects (difficulties repeating words, but not sentences, Baldo et al., 2008; semantic support despite impaired rehearsal, Howard & Nickels, 2005), while other patients have difficulties maintaining semantic but not phonological information (diminished lexicality effects, Jefferies et al., 2005; N. Martin et al., 1996; R. C. Martin et al., 1994). Additionally, behavioural data (Nishiyama, 2014, 2018) and neuroimaging studies (Fiebach et al., 2007) have provided corroborating evidence supporting unique phonological and semantic contributions throughout all stages of a verbal short-term memory task. Explaining this interaction between verbal short-term memory and the linguistic system has produced different theoretical positions, two receiving the most research attention are the redintegration hypothesis (Gathercole et al., 2001; Hulme et al., 1997; Schweikert, 1993) and language-based models (Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999; Schwering & MacDonald, 2020).

The redintegration hypothesis assumes a two-part process to recall. According to this view, processing verbal information first relies on forming a phonological representation of the item and then semantic knowledge stored in long-term memory is accessed and used at retrieval to reconstruct or “clean up” degraded phonological traces (Gathercole et al., 2001; Hulme et al., 1997; Schweikert, 1993). Some argue that redintegration can also take place during rehearsal (Hulme et al., 1999) or maintenance (Barrouillet & Camos, 2015) and not only at recall. Nevertheless, the distinction between phonological short-term and semantic long-term processes is inherent in this theory, with the redintegration mechanism as a potential account for the influence of semantic knowledge in immediate recall tasks. More recently however, there has been a shift away from viewing semantic contributions through this stepwise perspective toward a more integrated view of linguistic knowledge and short-term memory.

Language-based models offer a more parsimonious explanation for the interaction between language processing and verbal short-term memory. These current models account for the influence of semantic knowledge on immediate memory by assuming that activation occurs within the linguistic system (N. Martin & Saffran, 1997) or a dedicated buffer for short-term semantic maintenance (semantic short-term memory, R. C. Martin et al. 1999; conceptual short-term memory, Potter, 2012). Despite their differences, language-based models collectively assume that semantic representations are maintained along with phonological representations when verbal items are encountered and processed (for review, see Majerus, 2013, and Schwering & MacDonald, 2020). Motivated by this line of reasoning, we conducted the present study to systematically explore whether phonological and semantic information, while interactive, could have independent effects in verbal short-term memory using novel techniques.

Verbal short-term memory has typically been investigated by using serial recall tasks. Serial recall has been used to study the efficiency of phonological encoding (Campoy & Baddeley, 2008), the effect of list length (Grenfell-Essam & Ward, 2012), and rehearsal in immediate memory (Tan & Ward, 2008). However, serial recall stresses order over item information, and thereby taps linguistic knowledge minimally (Majerus, 2009, 2013). Phonological effects might have been inevitable with serial recall because phonological coding is very effective for storing serial order (Romani et al., 2008; Gathercole et al., 2001). Indeed, with rapid presentation rates, participants can easily encode serial order via phonological processes, while semantic encoding appears to be less optimal (Campoy & Baddeley, 2008; Campoy et al., 2015, Experiment 1). Semantic processing, conversely, then, is more engaged when processing item information and when verbal items are familiar and meaningful. Notably, when semantics have been considered in the methodology, they have been found to impact serial recall (Acheson et al., 2011; Poirier et al., 2015).

A further way in which results from serial recall studies have lent themselves to interpretations related to short-term and long-term memory has been in examining phonological effects using short lists (e.g., Campoy & Baddeley, 2008; Tan & Ward, 2008) and semantic effects using long lists (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014). Theorists argue that, given the limits of short-term memory, short list lengths can be kept within the focus of attention (Cowan, 2001), while longer lists would likely involve episodic long-term memory to support recall. Thus, as list length increases, there should be less involvement of short-term memory processes. Indeed, adults and children spontaneously use rehearsal to aid serial recall of short lists (McGilly & Siegler, 1989; Tan & Ward, 2008), and it has been suggested that this strategy is abandoned at longer list lengths as cognitive load increases (e.g., Baddeley & Larsen, 2007). Rehearsal at short list lengths seems to be reflected in reaction time such that reaction time linearly increases up to about four to six items (serial recall, Vergauwe et al., 2014; recognition task, Rypma & Gabrieli, 2001; Sternberg, 1966). Others report flatter reaction times in recognition tasks, especially with long list lengths (Burrows & Okada, 1975), indicating that memory search may occur in parallel rather than in a serial fashion (Townsend & Fific, 2004; see Cowan, 2001, for theoretical discussion). This latter notion is broadly in line with language-based models, according to which access to both phonological and semantic representations would be available throughout encoding and maintenance, without the need for rehearsal or redintegration processes. Therefore, the current study seeks to better understand whether there is evidence for semantic effects, in addition to phonological effects, in verbal short-term memory when the role of list length and strategies such as rehearsal are considered.

Given the limitations of serial recall paradigms, techniques more suitable for the investigation of phonological and semantic retention capacities are needed. Such tasks should avoid stressing serial order and thereby phonological encoding. Tasks that may be particularly well suited are the rhyme (or homophone) and synonym probe-recognition paradigms (R. C. Martin, et al., 1994; McElree, 1996). Item probe-recognition task (primarily testing item information) is in contrast to the serial order or list probe-recognition task, which was designed to primarily test order information and requires serial rehearsal (Henson et al., 2003; Murdock, 1976). Rhyme (or homophone) and synonym probe tasks are typically studied separately. Participants would first process a list of words followed by a probe word. In the rhyme (or homophone) probe task, participants must decide if the probe word rhymes with (or sounds the same as) a word on the list. In the synonym probe task, participants must decide if the probe word is synonymous with a word on the list. Since these two tasks have primarily been used to show a dissociation between phonological and semantic short-term memory in patients (R. C. Martin & He, 2004; R. C. Martin, et al., 1994), it is important to also investigate phonological and semantic mechanisms in healthy adults, and only few studies on this topic exist.

Although addressing a different theoretical question than the current work, an early study by McElree (1996) provided preliminary evidence for phonological and semantic effects in healthy adults. Participants studied a list of five words followed by a recognition probe probing for item, rhyme, or synonym information. Participants answered more quickly or slowly (range = 0.128 – 3 s) depending on the condition. Results revealed slower retrieval dynamics (a measure of speed-accuracy trade-off) for rhyme and synonym judgements compared to item judgements, except there was a recency effect across all conditions. In turn, rhyme and synonym judgements did not differ. The data were interpreted to indicate sufficient access to phonological and semantic information to enable a comparison with the probe (see also Shulman, 1970). This interpretation, however, should be made with caution due to a very small sample (n = 4), use of closed sets – which could encourage phonological representations, and rhyme and synonym probe stimuli not being perfectly comparable as synonym probes were multisyllabic words (car-automobile) while rhyme probes were shorter words (car-far). Using a modified recognition task paradigm, Nishiyama (2014) investigated the separability of phonological and semantic representations in working memory in healthy adults. Participants studied ten-word lists using either a phonological (rehearsal) or a semantic strategy (focus on the meaning) while completing a concurrent task impairing phonological (articulatory suppression) or semantic processing (finger-tapping). At test, participants had to choose the target word that was a homophone or synonym for one of the items on the list and were tested on all ten words. Results revealed that articulatory suppression impaired homophone judgements, whereas the finger-tapping task impaired synonym judgements, indicating distinct representations in healthy adults. However, task instructions could have favoured phonological or semantic encoding, and the use of long list lengths (ten words) could have involved episodic long-term memory in supporting recall. Further, the influence of semantics was found only in an indirect way – tapping impaired semantic via attentional load (Ruchkin et al., 2003). We addressed these issues by not prompting a particular strategy in order to necessitate engagement of both phonological and semantic processing across both short and long lists lengths.

The running-span task may be another way to study the content of short-term memory (and mitigate involvement of episodic long-term memory) as it minimizes opportunities for strategy use (e.g., Cowan, 2001; Pollack et al., 1959). In a typical task, participants process items presented at a fast presentation rate and then are cued to recall the most recent *n*-items unpredictably (where *n* is the number of items they are asked to recall). The fast presentation rate requires constant updating and serves to necessitate attention to each item, while the unpredictable list length prevents the use of strategic encoding such as the use of rehearsal or grouping (Bunting et al., 2006; Cowan, 2001). Rehearsal is actually detrimental to performance (Hockey, 1973). Similarly, research in visual working memory has shown that conceptual knowledge can be activated rapidly and without rehearsal using the Rapid Serial Visual Presentation procedures (Potter, 2012).

Kowialiewski and Majerus (2018a) developed a novel recognition variant of the running-span task to more closely evaluate the direct activation of semantic knowledge as well as rapidity of such processing in verbal short-term memory. Participants studied a list of words and nonwords presented at a fast rate of 2 items/s with list length varying from 11 to 14 items. A probe word appeared after the word list and participants had 1750 ms to decide if the probe word matched one of the items in the word list. This speeded response further served to prevent redintegration during retrieval. Findings revealed a lexicality effect, with better and faster performance for real words compared to nonwords, even in a task that minimized rehearsal and redintegration processes. The results were interpreted to be consistent with language-based models. Semantic knowledge that was activated when a word (vs. nonword) was encountered served to stabilize phonological representations. Task conditions could, however, have favoured semantic encoding; participants were only being tested on semantic knowledge (lexicality effect) and with long list lengths (11–14 items). Further, item or matching judgements behave differently than judgements based on phonological and semantic information, indicated by McElree (1996). Thus, the current study used a similar but modified procedure that required accessing phonological and semantic information in short-term memory across various list lengths and processing times.

Finally, little is known about how information processed in verbal short-term memory can help or hinder long-term memory encoding. On the one hand, phonological processing or rehearsal can support short-term retention, but phonological information decays rapidly (Baddeley, 2012), making it ineffective for long-term retention (Craik & Lockhart, 1972; Gallo et al., 2008). Semantic processing, on the other hand, is a process that involves deeper processing, leading to encoding of more contextually unique features and making it less susceptible to forgetting (Gallo et al., 2008). Related to the methodology adopted here, the probe word in the probe recognition task may act as a cue itself to re-activate relevant information. Studies in the visual working memory domain have found that when a cue (arrow) is presented after displaying the to-be-remembered items and before the probe item, this reactivates relevant information *already* in working memory and benefits retention for cued items (e.g., Berryhill et al., 2012). For word processing, by consequence of verbal short-term memory being emergent from the language network, probing for reactivation of semantic (vs. phonological) representations may promote encoding into long-term memory.

Taken together, results from the aforementioned studies using novel paradigms to study semantic effects in verbal short-term memory provide complementary data to well-established work on phonological effects (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014). However, past studies on short-term phonological effects differ from these more recent studies on semantic effects in many respects (e.g., serial recall vs. probe recognition; letters, numbers vs. words, sentences; short vs. long lists), making a direct comparison difficult. Therefore, our aim was to use one common paradigm to minimize methodological differences in our investigation of the retention of phonological and semantic information in verbal short-term memory and long-term impacts.

3.1.1 The Current Study

The present study aimed to address methodological issues with respect to testing of order information, instructions encouraging semantic or phonological encoding, the use of long list lengths favoring semantic encoding, and the possibility of post-list processes and rehearsal strategies. In contrast to Nishiyama (2014, 2018), participants were not instructed in advance on how to encode or maintain the word list. Our approach also extends work by Kowialiewski and Majerus (2018a) by directly tapping semantic (synonym judgement) and phonological information (rhyme judgement). Short and long list lengths were also used. Across two studies, participants studied a list of words that varied in length from three to 11 items. Following the word list, participants were cued to make an immediate rhyme or synonym judgement on a probe word appearing after the cue. Critically, this cue was not known in advance, meaning that participants had to engage in both phonological and semantic processing. After a 10-minute delay, participants completed a surprise delayed recognition test to assess long-term retention. In Experiment 1 we used a modified version of the rhyme and synonym probe-recognition tasks, and this paradigm was further modified in Experiment 2 by incorporating aspects of the running-span procedure to minimize strategy use and redintegration processes.

The main goal was to develop a new methodology that would be more suitable for assessing language representations underlying maintenance of verbal items than paradigms used in earlier studies (e.g., serial recall; Kowialiewski & Majerus, 2018a; Nishiyama, 2014). We drew inspiration from the item probe-recognition and running-span tasks because these paradigms do not stress phonological or semantic maintenance to a greater degree nor should they primarily involve rehearsal or serial order retention (Henson et al., 2003; Murdock, 1976). We hypothesized that the emergence of any phonological and semantic effects would reflect rapid access in short-term memory and further indicate that semantic effects do not require post-encoding reconstruction processes, aligning with language-based models. In contrast, if semantic representations rely on post-list processing, that is, activation of semantic knowledge to reconstruct incomplete phonological representations, then we expect to observe subtle semantic effects across the two experiments. Specifically, with the rapidity of Experiment 2, we expect that semantic knowledge would not be accessed within the allotted time.

A secondary goal was to examine performance across short and long list lengths. There may be an effect of list length in that long list lengths with increased load would impair performance regardless of probe. However, at a given list length, we hypothesized comparable phonological and semantic performance that would indicate direct activation, and thus, the ability to access linguistic knowledge regardless of list length. If, however, phonological and semantic effects reflect short- and long-term memory processes, respectively, then different advantages should emerge across list lengths. Short list lengths could easily be maintained using rehearsal, which would convey an advantage for phonological over semantic probe, but rehearsal would be less efficient as list length increases. This should also be reflected in reaction time analyses; reaction time should increase linearly with increased load (Vergauwe et al., 2014). If semantics only affect episodic long-term memory, then accuracy and reaction times for the semantic relative to phonological probes should be much worse for the short list lengths, but perhaps be better for longer list lengths.

Finally, short- and long-term retention for probed items were examined. The novelty of this task relied on the use of the probe word to reactivate relevant phonological or semantic information that would induce phonological or semantic processing, respectively. Meaningful processing should differentiate words more than focusing on sounds. As such, memory is expected to be better for words probed with a semantic than a phonological cue.

3.2 Experiment 1

Experiment 1 served to demonstrate the use of our modified probe recognition paradigm as a verbal short-term memory task that could tap phonological and semantic information. The paradigm was revised to address the limitations of previous work by (a) requiring both phonological and semantic processing (probe type was a within-subject variable), (b) no instructed encoding or maintenance strategy (probing occurred after the word list), and (c) presenting short and long list lengths. Further, a surprise delayed recognition test was added to compare short- and long-term retention.

3.2.1 Methods

Experiment 1 was not pre-registered. Experiment 2 was pre-registered: https://osf.io/ms7k3. The data and analysis code for both experiments are available on the Open Science Framework: https://osf.io/zye6a.

3.2.1.1 Participants

We recruited 31 participants (15 females; M_{age} = 19.55 years; SD_{age} = 2.01) who were proficient or native English speakers. Three additional participants were excluded because they did not follow directions or understand the instructions. Informed written consent was obtained for all participants. Ethical approval was provided by the University of Western Ontario's research ethics committee.

3.2.1.2 Materials

We selected a total of 350 monosyllabic words (nouns, verbs, adjectives) using the SUBTLEX norms (Brysbaert & New, 2009). Words were of medium lexical frequency, with a mean word frequency of 73.70 per million (*SD* = 39.60). They had a mean concreteness rating of 3.70 (*SD* = 0.93), with 5 being the most concrete (Brysbaert et al., 2014). Lists of three, five, seven, nine, and 11 words were generated by random selection without replacement, so that each word appeared only once during the experiment. There were ten lists at each length; matched for word frequency and concreteness rating. List lengths were presented in an ascending order (trials within list lengths were presented randomly). Within each length, participants made a rhyme judgement on five trials (rhyme probe) and a synonym judgement on five trials (synonym probe). Further, within those five trials, three of them were matching probes (affirmative responses) and two of them were nonmatching probes (negative responses).

For matching trials, one word within each list was paired with a rhyming or a synonymous word using an open set of monosyllabic words as well. Rhyming words were orthographically both similar (e.g., *hat* – *cat*) and dissimilar (e.g., *note* – *throat*). A monosyllabic synonymous word (e.g., *hat* – *cap*) was obtained from the norms of Nelson et al. (1998). The mean forward-associative strength was 0.27 (*SD* = 0.25).

3.2.1.3 Procedure

Participants were tested individually, seated in front of a 14-in laptop, using the PsychoPy 3.1 software (Peirce et al., 2019). The paradigm is shown in Figure 3.1. Each trial began with a fixation cross at the center of the screen for 1000 ms. Each word from the word list was presented at the top-center of the screen one at a time for 1000 ms with an interstimulus interval of 1000 ms. After the word list, participants received a cue word, “rhymes” or “means”, in capitalized letters at the center of the screen, followed by the probe word at the bottom-center of the screen. Participants were instructed to press the key labelled “Yes” or “No” in response to whether the probe word had rhymed or meant something similar to an item on the word list (depending on the cue). For example, in Figure 3.1, if the probe word *cap* appears after the cue RHYMES, then the participant must decide if “cap” rhymed with any words from the list. Participants would press “No” since there is no rhyming word on the list. If the cue was MEANS, then participants would press “Yes” because “cap” is a synonym for the word list *hat*. Participants were required to make their response as quickly and as accurately as possible. Participants completed practice trials at list length four. After completing the immediate probe recognition task, participants performed a nonverbal task for 10 minutes, then they completed a surprise delayed recognition test.

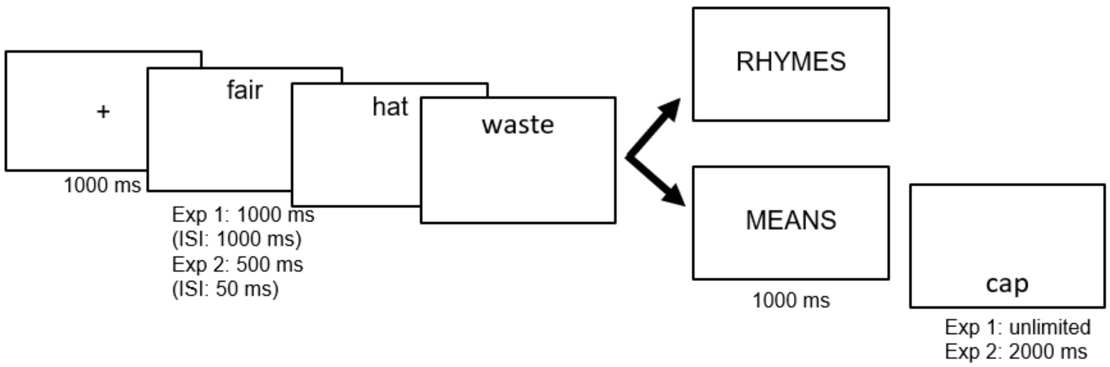

Figure 3.1. Schematic of the experimental design. After a word list, a single cue appeared. The cue was either ‘rhymes’ or ‘means’. After the cue, a probe word appeared, and participants were instructed to indicate whether the probe word was associated with an item on the list based on the cue. The timing intervals were the same in both experiments unless otherwise indicated. *Note*: If this was a synonym probe trial, then the word “hat” would be presented in the delayed recognition test.

For the delayed test, the 30 list words that had a matching probe word (i.e., given the match *hat* (list word) – *cap* (probe word), the word *hat* would be tested) and 30 new words were presented individually on the computer. The new words were matched on word frequency and concreteness rating. Participants responded to each word by pressing the key labelled “Old” for old words or “New” for new words. The word was considered “old” if it was a word from the word list, while a “new” word meant that it was never presented in the experiment, neither as list nor probe word. The word remained on the screen until participants made a decision.

3.2.1.4 Data analysis

Recognition accuracy was analyzed using d’ to remove response bias. Based on signal detection theory, d' is a measure of sensitivity that accounts for false alarms when measuring proportion of hits (Stanislaw & Todorov, 1999). Hits reflect correctly identifying that the probe word rhymed or was synonymous with a word from the list, whereas false alarms were acceptances even when there were no associations. High d’ indicates high sensitivity, or more accurate performance (fewer misses or false alarms). The effective limit is 4.65, for a hit rate of .99 and false alarm of .01 (Macmillan & Creelman, 2005). Zero d’ indicates a lack of sensitivity, or chance-level performance. The typical range of d’ for yes-no paradigms is 0.5–2.5, corresponding to about 60-90% accuracy.

We use a Bayesian analysis approach to analyze recognition accuracy (d’) and reaction time (Wagenmakers et al., 2018). For each model, Bayes Factor or BF_{10} is used to evaluate the strength of evidence for the alternative model (H_1) against a null model (H_0). If support for the model was ambiguous, we ran an analysis of *specific effects* to untangle this ambiguity and report inclusion BF (BF_{incl}) based on all models (van den Bergh et al., 2020). We used the following classification scheme to interpret BF: BF < 1 provides no evidence, BF between 1 and 3 provides anecdotal evidence, BF between 3 and 10 provides substantial evidence, BF between 10 and 30 provides strong evidence, BF between 30 and 100 provides very strong evidence, and BF > 100 provides decisive evidence (Jeffreys, 1961; Wagenmakers et al., 2018). Delayed data were analyzed using

Bayesian Wilcoxon Signed-Rank Test due to the small sample size[2]. Bayesian analyses were conducted using JASP (JASP Team, 2020). Immediate and delayed data were submitted to separate analyses given the differences between the tasks, and visual inspection was used to compare performance.

3.2.2 Results

Preliminary analyses. Orthographically dissimilar rhyme pairs may be more difficult than similar pairs so we first verified that this manipulation did not unintentionally affect performance, and indeed there was no evidence for a difference, $BF_{10} = 0.61$. These conditions were collapsed in all remaining analyses.

Accuracy. Performance for each probe across list length is depicted in Figure 3.2 and descriptive statistics are provided in Table 3.1. A Bayesian repeated-measures ANOVA with probe and list length as within-subject factors provided anecdotal evidence in favour of the full model containing both main effects and the interaction term (BF_{10} = 1.03e+8) preferred by a factor of 1.47 over the second best model containing both main effects (BF_{10} = 7.03e+7). We ran an analysis of specific effects in order to untangle the ambiguous evidence. This revealed decisive evidence for an effect of probe with better accuracy for synonym than rhyme probes, BF_{incl} = 1.57e+4. There was also decisive evidence for an effect of list length, BF_{incl} = 3.58e+4. Follow-up Bayesian t-tests indicated that performance was best at list length 3 compared to all other lengths ($BF_{10} > 29.80$, all cases), anecdotal evidence for 9 vs 11 ($BF_{10} = 1.75$), and no evidence for

[2] Due to experimenter error, only 13 of the 31 participants completed the delayed task in Experiment 1.

remaining comparisons ($BF_{10} < 0.34$, all cases). Finally, substantial evidence supported the interaction between probe and list length, $BF_{incl} = 5.48$. Follow-up Bayesian t-tests showed better performance for the synonym than rhyme probe at list length 3 (BF_{10} = 115.96) as well as at list length 5 and 9 ($BF_{10} > 12.12$); however, no evidence for a difference at list length 7 or 11 ($BF_{10} < 0.22$, both cases). These results indicate that

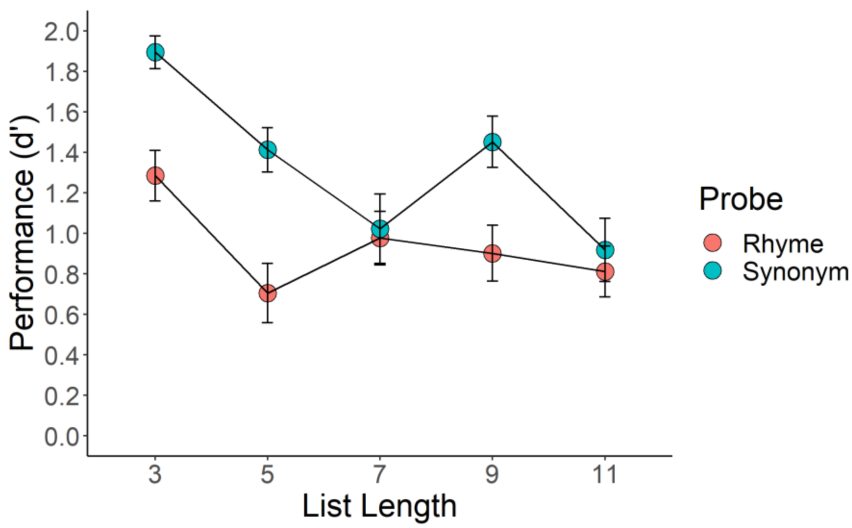

Figure 3.2. Participants' accuracy (d') across list lengths for both the rhyme and synonym probes in Experiment 1. Error bars are the standard error of d-prime.

Table 3.1. Proportion of hit responses as a function of probe type and list length and proportions for hits and false alarms for Experiment 1

List length	Rhyme	Synonym
3***	0.73 (0.32)	0.94 (0.13)
5**	0.70 (0.30)	0.78 (0.22)
7	0.65 (0.24)	0.73 (0.26)
9**	0.65 (0.21)	0.80 (0.25)
11	0.76 (0.21)	0.67 (0.29)
Hits	0.70 (0.26)	0.78 (0.25)
False alarms	0.24 (0.35)	0.14 (0.37)

Note. Standard deviations are in parenthesis. Probe effect, where Rhyme < Synonym: ** = strong evidence; *** = decisive evidence synonym judgements were just as good as – or even better than – rhyme judgements across most list lengths. Additionally, we analyzed each probe separately via a Bayesian repeated-measures ANOVA. There was decisive evidence supporting differences across list lengths for the synonym probe, BF_{10} = 4.96e+4, but no reliable evidence for the rhyme probe, BF_{10} = 1.11. For the synonym probe, list length 3 differed from all other lengths (BF_{10} > 3.83, all cases), strong evidence between 9 vs 11 (BF_{10} = 14.77), and anecdotal evidence between 5 vs 7, 5 vs 11, and 7 vs 9 (BF_{10} > 1.96, all cases).

Reaction time. Only correct trials were included in reaction time analyses and data from three participants were removed due to insufficient data for analysis (i.e., no correct trials in a given list length). A Bayesian repeated-measures ANOVA with probe and list length as within-subject factors revealed that the model with the effect of list length only was the best model (BF_{10} = 2.54e+5) favoured over the second best model including both main effects (BF_{10} = 3.92e+4) by a factor of 6.47 (see Figure 3.3). Further comparisons using Bayesian t-tests revealed decisive evidence that list length 3 was faster than all other lists (BF_{10} > 110.32, all cases), anecdotal evidence for 5 vs 7 and 5 vs 11 (BF_{10} > 2.50, both cases), and no support for remaining comparisons (BF_{10} < 0.48, all cases).

Delayed data. Although the difference in task precludes direct comparison, the immediate data are provided to contrast with delayed performance in Figure 3.4. There was substantial support for a difference between recognition of words previously probed by rhyme and synonym cue, BF_{10} = 8.48. Semantically processed words were recognized better than phonologically processed words.

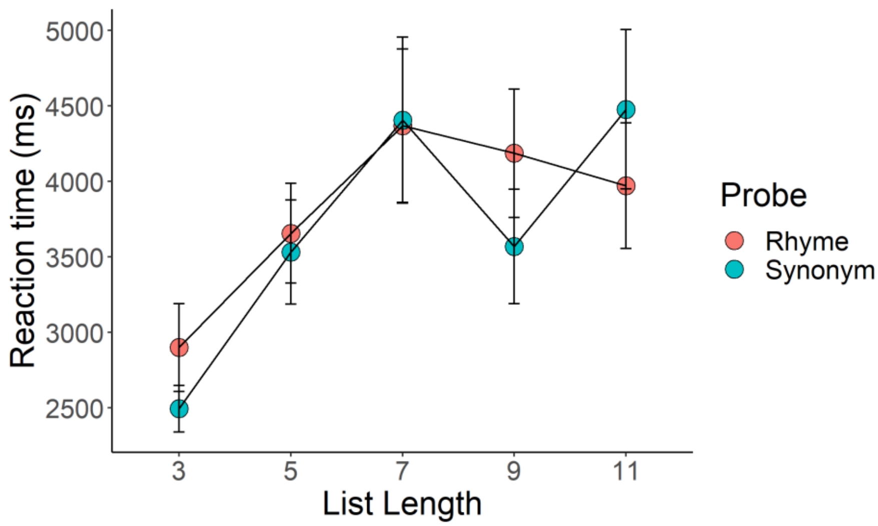

Figure 3.3. Mean reaction time across list lengths for both the rhyme and synonym probe in Experiment 1. Error bars are the standard error of the mean.

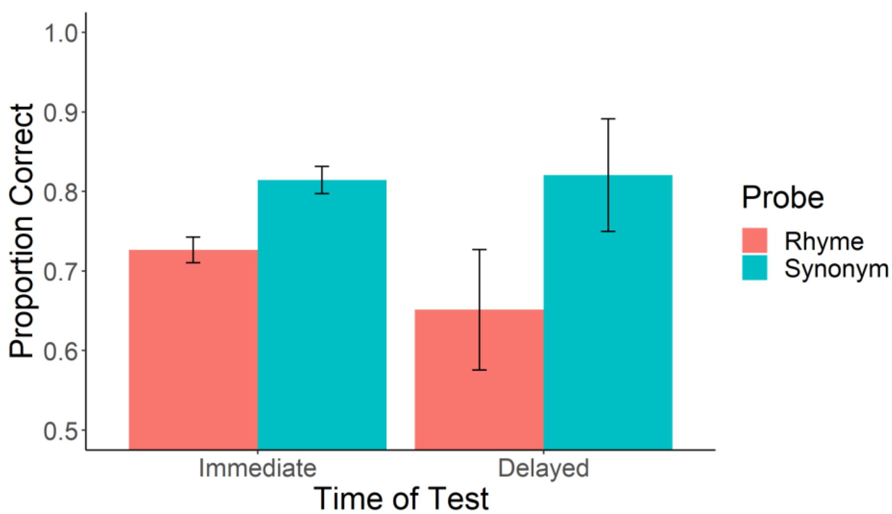

Figure 3.4. Mean proportion correct for rhyme/synonym probe across short- and long-term retention in Experiment 1. Error bars are the standard error of the mean.

3.2.3 Discussion

Using a modified probe recognition paradigm, Experiment 1 demonstrated that verbal items immediately activated phonological and semantic knowledge. Performance was better for synonym than rhyme judgements, and this advantage was consistent until the longest list length. Importantly, words probed semantically had an advantage not only in long list lengths exceeding the capacity of short-term memory (list length 9), but also when items were within the focus of attention (list lengths 3 and 5). Semantic effects in short lists minimize the contribution of episodic long-term memory processes, and addresses a limitation from prior work investigating semantics effects with only long lists (e.g., Nishiyama, 2014; Kowialiewski & Majerus, 2018a). Reaction time did not increase linearly. Instead, in both accuracy and reaction time analyses, there was only decisive evidence for a difference between the shortest list length (3) with longer list lengths; there were no reliable evidence for differences among long list lengths. Finally, delayed data revealed that, similar to immediate recognition, participants were better at remembering words that were probed with a semantic than a phonological cue.

Of note, performance on rhyme probes at list length 3 may not have been at ceiling because vigilance could be low in tasks that seem trivially easy (Thomson et al., 2015); practice trials using list length 4 were at ceiling (rhyme: M = .90, SD = 0.20; synonym: M = 0.87, SD = 0.22). An alternative explanation is that participants found the synonym task more challenging (e.g., it is less clear whether two words are synonymous than rhyme with each other) and thus chose to focus more attention on semantic access and maintenance given the relatively slow presentation rate, causing a decrement in performance for rhyme probes. Another potential indication of a trade-off between task difficulty and accuracy comes from inspecting the performance patterns across lists 7, 9, and 11 for synonym probes. Although there was no evidence or, at best, anecdotal evidence for differences in accuracy or reaction time in direct comparisons of these lengths, visual inspection of the data revealed somewhat slower but more accurate responses for list length 9, which could reflect a shift to increased vigilance, and a relative 'giving up' of the task with list length 11.

One limitation of Experiment 1 is that the timing intervals could potentially introduce confounding effects. In our initial conceptualization of this study, we were drawing on traditional serial recall tasks where a presentation rate of one item/s is common (e.g., Baddeley, 1966; Nishiyama, 2014; Poirier et al., 2015). However, longer times could have given participants time to make the strategic choice to focus attention on semantic relations as well as time for elaborative processing and engaging episodic long-term memory. Probe effects in delayed retention too could be attributed to stabilization from tapping long-term memory. We addressed this limitation in Experiment 2 by using a fast-encoding, running-span procedure to prevent rehearsal and grouping strategies (Bunting et al., 2006; Cowan, 2001). Words were presented at a rate of about two items/s (500 ms on, 50 ms off). Further, we started at list length 4 so that the task did not seem trivial and only used orthographically similar rhyme pairs. In this way, Experiment 2 was designed to investigate whether access to phonological and semantic information in verbal short-term memory occurs rapidly, in the absence of strategy use and redintegration processes. If short-term effects arise by consequence of the linguistic system directly supporting verbal short-term memory maintenance rather than being attributed to redintegration processes, then we should expect similar phonological and semantic effects, even under conditions of fast encoding.

3.3 Experiment 2

In Experiment 2, we re-examined phonological and semantic processing in verbal short-term memory by incorporating a running-span procedure. This meant manipulating the presentation rate of the probe recognition task to be very fast and ending lists unpredictably to minimize opportunities for strategy use and redintegration. Experiment 2 was revised by (a) presenting words visually and auditorily at a rate of one every 550 ms, (b) using list lengths 4, 6, 8, and 10, and (c) selecting only orthographically similar rhyme pairs.

3.3.1 Methods

3.3.1.1 Participants

We recruited 30 new participants (19 females; M_{age} = 18.33 years; SD_{age} = 0.61). None had participated in Experiment 1.

3.3.1.2 Materials

We selected 280 monosyllabic words from Experiment 1 ($M_{word\ frequency}$ = 74.90 per million, $SD_{word\ frequency}$ = 37; $M_{concreteness}$ = 3.76, $SD_{concreteness}$ = 0.95). From this set, words were further divided into ten trials for list lengths 4, 6, 8, and 10. Trials were presented in random order so that the list ended unpredictably. Trials varied by probe type (i.e., rhyme or synonym) and whether there was a match or not in the same way as was done in Experiment 1. Rhyme pairs were orthographically similar (e.g., *hat – cat*). The mean forward-associative strength for synonym pairs was 0.45 (*SD* = 0.23).

To control the duration in this experiment, audio recordings of each word were recorded by the first author, a native English speaker. Audacity was used to remove background noise and adjust the duration of each word to 500 ms without altering the pitch.

3.3.1.3 Procedure

In this experiment, the running-span procedure was incorporated into the probe recognition task (from Experiment 1). Stimuli were presented both visually and auditorily. A fixation cross appeared at the center of the screen for 1000 ms to begin each trial. Words from the list were presented sequentially for 500 ms with an interstimulus interval of 50 ms. After the word list, participants were cued to make a yes-no decision on the probe word to indicate whether it had rhymed or was synonymous with a word on the list. Participants had 2000 ms to respond, further limiting the use of strategies and redintegration processes (Experiment 1 had no time restrictions). When participants did not respond in time, the word “FASTER” would appear on the screen reminding them to respond faster and the trial was not repeated. Participants completed a practice before the experiment. After the immediate task and a 10-minute delay, participants performed a surprise delayed recognition test.

For the delayed test, the 24 old and 24 new words were presented both visually and auditorily, one at a time, in the center of the computer screen. For each word, participants had to determine whether the word was old or new.

3.3.1.4 Data analysis

Accuracy and reaction time were analyzed in the same way as in Experiment 1.

3.3.2 Results

Accuracy. Performance accuracy on the immediate test is displayed in Figure 3.5 and provided in Table 3.2. Data from one participant was removed due to insufficient data for analysis. We further excluded 26 rhyme trials (2.24% of the data) and 40 synonym trials (3.45%) where participants did not respond within the allotted time. A Bayesian repeated-measures ANOVA on recognition accuracy scores (d') with probe and list length as within-subject factors provided anecdotal evidence in favour of the full model (BF_{10} = 4.75e+5) preferred by a factor of 1.48 over the second best model containing the list length effect only (BF_{10} = 3.20e+5). Given the ambiguous results, we ran an analysis of specific effect. There was no reliable evidence for an effect of probe, BF_{incl} = 1.19, suggesting that there was no evidence for the probe effect, nor evidence to state that there is no effect[3]. There was a list length effect, BF_{incl} = 4.73e+5, characterized by decisive support for a difference between 4 vs 8 and 6 vs 8 (BF_{10} > 832.41, both cases), strong support for 4 vs 10 (BF_{10} = 13.44), and substantial support for 6 vs 10 (BF_{10} = 7.40). Finally, there was substantial evidence in favour of the interaction, BF_{incl} = 4.72. Follow-up comparisons using Bayesian t-tests indicated only anecdotal evidence for a probe effect at list lengths 6 and 8 (BF_{10} > 1.27, both cases) and no evidence at list lengths 4 and 10 (BF_{10} < 0.84, both cases). This indicates that, at minimum, immediate recognition of semantic information was just as good as recognition of phonological information across all list lengths. Instead, the interaction was due to decisive evidence supporting differences across list lengths for rhyme probes, BF_{10} = 1.44e+7, but no evidence for synonym probes, BF_{10} = 0.49. A comparison for rhyme judgements across list lengths revealed decisive support for a difference between 4 vs 8 and 6 vs 8 (BF_{10} >

---

[3] As an alternative, and to untangle this ambiguity further, we computed inclusion probabilities for "matched" models only (van den Bergh et al., 2020) and found no evidence for the main effect of probe, BF_{incl} = 0.32. In fact, there is 3.13 times more evidence for the null model than a model including the effect of probe.

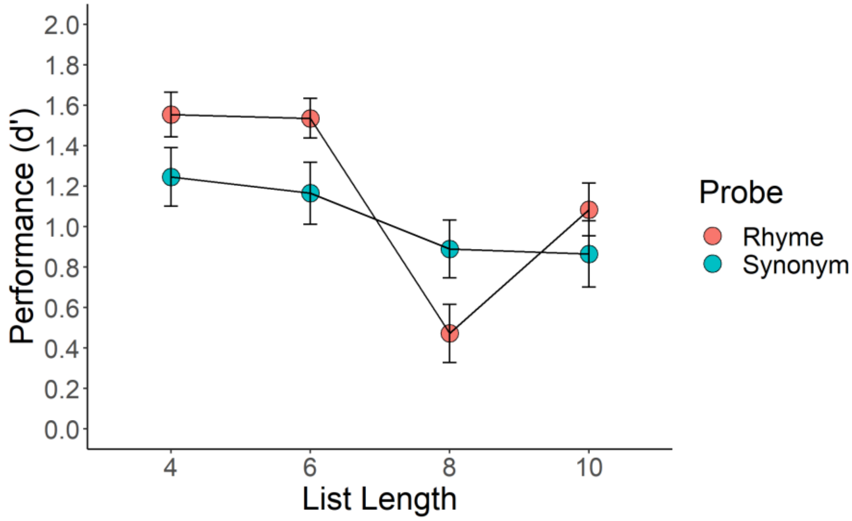

Figure 3.5. Participants' accuracy (d') across list lengths for both the rhyme and synonym probe in Experiment 2. Error bars are the standard error of d-prime.

Table 3.2. Proportion of hit responses as a function of probe type and list length and proportions for hits and false alarms for Experiment 2

List length	Rhyme	Synonym
4	0.83 (0.25)	0.72 (0.31)
6*	0.86 (0.19)	0.73 (0.25)
8*	0.70 (0.28)	0.75 (0.27)
10	0.75 (0.25)	0.64 (0.27)
Hits	0.79 (0.25)	0.71 (0.28)
False alarms	0.21 (0.32)	0.18 (0.31)

Note. Standard deviations are in parenthesis. Probe effect: * = anecdotal evidence

6850.63, both cases), substantial support for 6 vs 10 and 8 vs 10 (BF_{10} > 4.71, both cases), and anecdotal support for 4 vs 10 (BF_{10} = 2.95).

Reaction time. Only correct trials were included in reaction time analyses. Data from two participants were removed due to insufficient data for analysis. We conducted a Bayesian repeated-measures ANOVA to assess reaction time performance (Figure 3.6). The best model included the main effect of probe only (BF_{10} = 7.66e+6), which was

12.50 times more likely than the second best model with the two main effects (BF_{10} = 6.13e+5). Participants were faster when making rhyme (M = 1007.85, SD = 221.41) than synonym judgements (M = 1135.65, SD = 247.93).

Delayed data. Figure 3.7 contrasts phonological and semantic performance at immediate and delayed testing. There was substantial evidence for a difference between words probed semantically and phonologically, BF_{10} = 3.17. Delayed recognition again showed an advantage for semantically processed words, even though immediate recognition was similar.

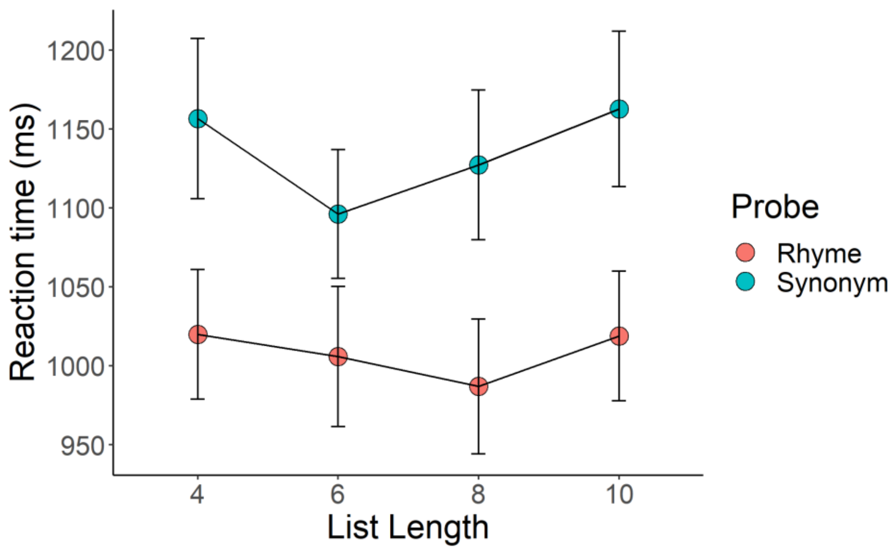

Figure 3.6. Mean reaction time across list lengths for both the rhyme and synonym probe in Experiment 2. Error bars are the standard error of the mean.

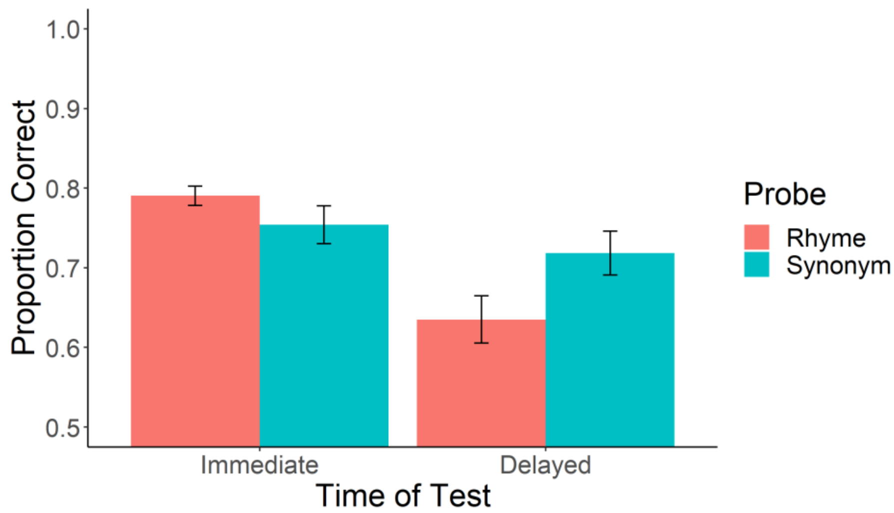

Figure 3.7. Mean proportion correct for rhyme/synonym probe across short- and long-term retention in Experiment 2. Error bars are the standard error of the mean.

3.3.3 Discussion

Our aim in Experiment 2 was to use a running-span probe task to investigate whether short-term phonological and semantic effects were evident in a task that prevented strategic encoding and redintegration processes. Indeed, we found that there appears to be similar access to both phonological and semantic information as participants were equally accurate at making rhyme and synonym judgements, indicated by no evidence for a probe effect (unlike Experiment 1). Further support came from demonstrating that there was no evidence or, at best, anecdotal evidence in support of a difference between probe type at any given list length. Therefore, at minimum, our results imply that semantic information was activated as readily and rapidly as phonological information in verbal short-term memory, consistent with language-based models. Interestingly, despite no immediate advantage of probing with a rhyme or synonym cue, there was substantial evidence supporting better memory for words processed semantically than phonologically after a brief delay.

Two somewhat surprising results were (a) substantial evidence in support for a difference between list lengths 8 and 10 in the rhyme probe condition and (b) that there was a main effect of probe in the reaction time data. First, improved accuracy between list lengths 8 and 10 could be attributed to inadvertently probing for rhymes across more positions in list length 8 (serial positions 1, 4-8 were probed) than list length 10 (serial positions 5-8 were probed), resulting in more degraded representations in list length 8. Since we did not systematically vary serial position, further analyzes would not be appropriate. But in Experiment 1 there was also some evidence that performance increased after a dip between list lengths 7 and 9 for synonym pairs, indicating a potential shift in processing. We elaborate on this point in the *General Discussion*. Second, although synonym judgements were associated with slower responses, this did not correspond to more accurate responses, suggesting it was not due to a speed-accuracy trade-off. Instead, it likely takes longer for participants to determine if words are synonymous with each other than if they had rhymed because synonym judgements are inherently more difficult.

3.4 General Discussion

The purpose of the study was to use our modified verbal short-term memory task to tap phonological and semantic information as well as to compare short- and long-term retention. Importantly, the methodology adopted across two experiments, particularly the running-span version of Experiment 2, avoids an emphasis on order information and the employment of rehearsal, both of which favor phonological coding. In Experiment 1, accuracy was better for synonym than rhyme judgements. Further, words probed with a synonym (vs rhyme) cue were better retained after a delay. However, given the slower presentation rate in Experiment 1, findings of semantic effects could have been attributed to contributions from long-term memory via redintegration rather than immediate co-activation of linguistic knowledge. Experiment 2 was designed to rule out this possibility by preventing redintegration and strategic processes. Indeed, in Experiment 2, under a running span task, similar activation of phonological and semantic knowledge was indicated by accuracy being equally good for both the rhyme and synonym probes, confirmed by no evidence for a probe effect. Despite initial similar levels of activation, delayed retention was better for words probed with a semantic than phonological cue. Taken together, phonological and semantic effects were evident across both studies suggesting that linguistic knowledge was rapidly and readily available to support word list retention.

3.4.1 Phonological and semantic representations in short-term memory

This study used a modified probe recognition task coupled with the running-span task in Experiment 2 to assess phonological and semantic information across list lengths. Results revealed automatic and rapid access to phonological and semantic representations in a task that minimized redintegration and strategic processes, with slow encoding leading to an immediate semantic advantage. As discussed, synonym judgements are more difficult than rhyme judgments even without short-term memory demands, leading participants to focus their attention on semantic information when they had time to make such strategic choices (Experiment 1) and resulting in longer reaction times (Experiment 2). Further, semantic effects in particular do not require post-encoding processes such as redintegration, consistent with prior work demonstrating the influence of semantics in verbal short-term memory (e.g., Kowialiewski & Majerus, 2018a, 2018b). Although we cannot completely rule out the existence of redintegration processes, it is difficult to reconcile the rapidity of semantic processing and lack of a phonological advantage with redintegration theories. However, it could be the case that more traditional serial recall tasks or certain procedures requiring serial order and phonological processing may benefit from redintegration. Nonetheless, our study more broadly implies that different levels of representations (phonological and semantic knowledge have been the focus here) influence verbal short-term memory and support retention of verbal information.

What is particularly novel in these results is that we demonstrated access to phonological and semantic representations using a relatively novel recognition variant of the running-span task. Procedures used in prior work including serial recall, short list lengths, and task instructions could have encouraged phonological representations and minimized semantic access (e.g., Nishiyama, 2014; Tan & Ward, 2008; but see McElree, 1996). It was therefore important to adopt a paradigm that would capture verbal short-term memory processes in the absence of rehearsal and redintegration (Kowialiewski & Majerus, 2018a), strategic encoding (Bunting et al., 2006; Cowan, 2001), and without primarily relying on serial order retention (Henson et al., 2003; Murdock, 1976). Results revealed that the same verbal item can engage both phonological and semantic processing when the task does not emphasize certain codes, even under fast encoding conditions. Given that this methodological approach may be better suited for tapping phonological and semantic information simultaneously, this work needs to be replicated and extended to further understand phonological and semantic access and maintenance. Future work can directly compare well-established phonological (e.g., phonological similarity, word length) and semantic effects (e.g., lexicality, concreteness). For instance, Kowialiewski and Majerus (2018b) studied a variety of semantic effects (lexicality, frequency, semantic similarity, and imageability) and found that most of these effects, except imageability, continued to emerge even in a running-span task. Phonological effects, however, were not investigated. Nishiyama (2013) studied both phonological and semantic effects using word frequency and word imageability, respectively. Although results were interpreted to support the separability of phonological and semantic representations, using word frequency to index phonology limits this interpretation because word frequency is typically viewed as a semantic variable.

3.4.2 Phonological and semantic effects across list lengths

A secondary goal was to investigate phonological and semantic effects across list lengths. Prior work often used short lists to examine phonological effects (e.g., Tan & Ward, 2008) and long lists to examine semantic effects (e.g., Kowialiewski & Majerus, 2018a; Nishiyama, 2014), with some arguing that long list lengths likely tap episodic long-term memory. We found that although there was an overall effect of list length in that accuracy decreased with longer lengths, synonym judgements were for the most part better than phonological judgements in Experiment 1 and performance was comparable at any list length in Experiment 2. Importantly, semantic maintenance was observed not only at long list lengths (replicating prior work; Kowialiewski & Majerus, 2018a; Nishiyama, 2014), but also short list lengths, contrasting with the dominant view of semantics only affecting episodic long-term memory. McElree (1996) also found comparable rhyme and synonym judgements using five-word lists. Furthermore, reaction time did not increase linearly with list lengths (except a difference between list length 3 vs. longer lengths). This finding, coupled with the lack of a phonological advantage even at short lists, suggests that rehearsal or phonological processes was not the primary mechanism being used to process the word list. In fact, there was a semantic advantage in Experiment 1, while rehearsal was prevented in Experiment 2. Instead, the results can be accommodated by the predictions of language-based models. That is, upon hearing a word, maintenance would rely on the direct activation of phonological and semantic representations simultaneously within the linguistic system. It would follow that participants can automatically and rapidly activate phonological and semantic knowledge needed for processing, without the need to rehearse and this activation is unaffected by list length manipulations (see also Potter, 2012).

Nevertheless, substantial support for the interaction in both experiments suggests a potential shift in processing with increased load (i.e., longer list lengths). In particular, after an initial decrease, accuracy increased between list lengths 7-8 and 9-10, but it was not due to a speed-accuracy trade-off. Instead, there could have been a trade-off between allocating attentional processes to semantic (and phonological) access and accommodating the increased load. As a result, this likely minimized the semantic advantage at length 7 in Experiment 1 and caused a decrement to rhyme judgements at length 8 in Experiment 2 (which is in line with Experiment 1). The subsequent increase at lengths 9 and 10, respectively, might be due to paying more attention when the task got even harder, leading to better performance. However, we need to interpret this trend with caution as there was only anecdotal support for a difference between 7 and 9 for synonym whereas substantial evidence supported a difference between 8 and 10 for rhyme. Increased performance for rhyming pairs may be due to a number of factors (e.g., serial position probed, visual plus auditory presentation). Additionally, we did not measure individual differences in attention allocation which further limits our interpretations. Future studies, for instance, could evaluate general attentional demands on performance by using a dual-task that primarily imposes an attentional load while not chiefly tapping phonological or semantic processing. Results would shed light on the role that attentional processes play in list length findings and its interrelationships with language processing in verbal short-term memory more broadly.

3.4.3 Semantic processing facilitates retention

We also examined how probing for phonological and semantic information already in verbal short-term memory benefited delayed retention. Although linguistic knowledge may be directly activated in short-term memory, after a brief delay, focusing on semantic (vs. phonological) information led to better long-term retention. Interestingly, this was even the case in Experiment 2 despite no initial probe advantage. To understand how short-term processing impacts subsequent memory, we integrate ideas from language-based models and from the distinctiveness hypothesis.

According to language-based models, different levels of linguistic representations (including phonological and semantic) are actively maintained in verbal short-term memory (Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999; Schwering & MacDonald, 2020). On the other hand, creating distinctive representations via immediate semantic processing allows participants to encode more unique features associated with studied words (Gallo et al., 2008). Through these frameworks, when cued to make a synonym judgement, semantic processing is associated with richer and more distinctive representations at the semantic level, which serves to stabilize phonological representations through mutual interactions between the different levels (e.g., Majerus, 2013; N. Martin & Saffran, 1997), and subsequently enhance memory (i.e., less susceptible to forgetting). In contrast, information activated at the semantic level by a rhyme cue will be less rich given the lack of distinctiveness of rhyming pairs that share a limited number of surface features. Thus, phonological processing contributes less to short-term maintenance and would be suboptimal for promoting retention. The results for delayed performance are an interesting adjunct to the short-term memory findings, indicating that semantic processing during short-term memory benefitted retention.

3.4.4 Conclusions

In sum, it is clear that verbal short-term memory does not operate in isolation, but within the context of a complex linguistic system. In particular, when a word is encountered and processed, verbal short-term memory has access to and relies on both phonological and semantic representations for maintenance and retention. Further, focusing on semantic information in short-term memory leads to better long-term memory, even if there seems to be no immediate advantage. Our novel finding of separable phonological and semantic effects on both short- and long-term retention contrast with the dominant view that phonological and semantic effects reflect short- and long-term memory processes, respectively. More broadly, the results are consistent with viewing verbal short-term memory as an emergent property of the language processing system with rapid access to available word-related knowledge.

3.5 References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(1), 44-59. https://doi.org/10.1037/a0021205

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63(1),* 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

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

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. https://doi.org/10.1016/S0022-5371(75)80045-4

Baddeley, A. D., & Larsen, J. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *The Quarterly Journal of Experimental Psychology*, *60(4)*, 497–504. https://doi.org/10.1080/17470210601147572

Baldo, J., Klostermann, E., & Dronkers, N. (2008). It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. *Brain and Language, 105(2),* 134–140. https://doi.org/10.1016/j.bandl.2007.12.007

Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction.* Psychology Press.

Berryhill, M. E., Richmond, L. L., Shay, C. S., & Olson, I. R. (2012). Shifting attention among working memory representations: Testing cue type, awareness, and strategic control. *The Quarterly Journal of Experimental Psychology, 65(3),* 426-438. https://doi.org/10.1080/17470218.2011.604786

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41(4)*, 977-990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46(3),* 904-911. https://doi.org/10.3758/s13428-013-0403-5

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology, 59(10),* 1691–1700. https://doi.org/10.1080/17470210600848402

Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science, 188(4192),* 1031–1033. https://doi.org/10.1126/science.188.4192.1031

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16(4),* 329–340. https://doi.org/10.1080/09658210701867302

Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology, 68(4),* 759-778. https://doi.org/10.1080/17470218.2014.966248

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24(1),* 87–114. https://doi.org/10.1017/S0140525X01003922

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684

Fiebach, C., Friederici, A., Smith, E., & Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual-semantic representations in verbal working memory. *Journal of Cognitive Neuroscience, 19(12),* 2035–2049. https://doi.org/10.1162/jocn.2007.19.12.2035

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58(4),* 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A, 54(1),* 1-30. https://doi.org/10.1080/02724980042000002

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67(1),* 106–148. https://doi.org/10.1016/j.jml.2012.04.004

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A, 56(8),* 1307-1334. https://doi.org/10.1080/02724980244000747

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology, 25(1),* 104-111. https://doi.org/10.1080/14640747308400328

Howard, D., & Nickels, L. (2005). Separating input and output phonology: semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology, 22(1),* 42–77. https://doi.org/10.1080/02643290342000582

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(6),* 685-701. https://doi.org/10.1016/0749-596X(91)90032-F

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search, and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25(2),* 447–463. https://doi.org/10.1037/0278-7393.25.2.447

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & 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(5),* 1217.

JASP Team. (2020). JASP (Version 0.14.1) [Computer software].

Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology*, *22(2)*, 183–212. https://doi.org/10.1080/02643290442000068

Jeffreys, H. (1961). *The theory of probability.* Oxford University Press.

Kowialiewski, B., & Majerus, S. (2018a). Testing the redintegration hypothesis by a single probe recognition paradigm. *Memory*, *26(9)*, 1256-1264. https://doi.org/10.1080/09658211.2018.1448420

Kowialiewski, B., & Majerus, S. (2018b). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language*, *101*, 64-83. https://doi.org/10.1016/j.jml.2018.03.005

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.).* Lawrence Erlbaum Associates Publishers.

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). Psychology Press.

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7,* 357. https://doi.org/10.3389/fnhum.2013.00357

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

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

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*, *89(1)*, 76-82. https://doi.org/10.1016/S0093-934X(03)00300-6

Martin, R. C., Lesch, M., & Bartha, M. (1999). Independence of Input and Output Phonology in Word Processing and Short-Term Memory. *Journal of Memory and Language*, *41(1)*, 3–29. https://doi.org/10.1006/jmla.1999.2637

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, *33(1)*, 83-111. https://doi.org/10.1006/jmla.1994.1005

McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition*, *24(2)*, 173-187. https://doi.org/10.3758/BF03200879

McGilly, K., & Siegler, R. S. (1989). How children choose among serial recall strategies. *Child Development*, *60*, *172–182*. https://doi.org/10.2307/1131083

Murdock, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, *105(2)*, 191–216. https://doi.org/10.1037/0096-3445.105.2.191

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.* http://www.usf.edu/FreeAssociation/

Nishiyama, R. (2013). Dissociative contributions of semantic and lexical-phonological information to immediate recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39(2),* 642-648. https://doi.org/10.1037/a0029160

Nishiyama, R. (2014). Active maintenance of semantic representations. *Psychonomic Bulletin & Review, 21(6),* 1583-1589. https://doi.org/10.3758/s13423-014-0618-1

Nishiyama, R. (2018). Separability of active semantic and phonological maintenance in verbal working memory. *PloS One*, *13(3),* e0193808. https://doi.org/10.1371/journal.pone.0193808

Papagno, C., & Cecchetto, C. (2019). Is STM involved in sentence comprehension? *Cortex*, *112*, 80–90. https://doi.org/10.1016/j.cortex.2018.08.028

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51(1)*, 195-203, https://doi.org/10.3758/s13428-018-01193-y

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition*, *43(3)*, 489-499. https://doi.org/10.3758/s13421-014-0470-6

Pollack, I., Johnson, I. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*, 137–146.

Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Frontiers in Psychology*, *3*, 113–113. https://doi.org/10.3389/fpsyg.2012.00113

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, *61(2)*, 292-323. https://doi.org/10.1080/17470210601147747

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *The Behavioral and Brain Sciences*, *26(6)*, 709-728. https://doi.org/10.1017/S0140525X03000165

Rypma, B., & Gabrieli, J. D. (2001). Functional neuroimaging of short-term memory: The neural mechanisms of mental storage. *The Behavioral and Brain Sciences, 24(1)*, 143–144. https://doi.org/10.1017/S0140525X01503921

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21(2),* 168-175. https://doi.org/10.3758/BF03202729

Schwering, S. C. & Macdonald. M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14,* 68. https://doi.org/10.3389/fnhum.2020.00068

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–508. https://doi.org/10.1016/S0022-5371(70)80093-7

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31(1),* 137-149. https://doi.org/10.3758/BF03207704

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153,* 652–654. https://doi.org/10.1126/science.153.3736.652

Tan, L. & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15(3),* 535-542. https://doi.org/10.3758/PBR.15.3.535

Thomson, D., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science, 10(1),* 82–96. https://doi.org/10.1177/1745691614556681

Townsend, J. T., & Fifić, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics, 66(6),* 953–962. https://doi.org/10.3758/BF03194987

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E., Derks, K .. & Wagenmakers, E. (2020). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique, 120,* 73-96. https://doi.org/10.3917/anpsy1.201.0073

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 40(4),* 1072–1095. https://doi.org/10.1037/a0035779

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., …Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review, 25,* 58–76. https://doi.org/10.3758/s13423-017-1323-7

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5),* 1256-1271. https://doi.org/10.1037/0278-7393.25.5.1256

Chapter 4

4 The role of available processes and sentence concreteness on immediate and long-term sentence recall

4.1 Introduction

Verbal working memory refers to a limited-capacity system that is responsible for retaining verbal information over short periods of time. It is now well established that verbal working memory is also influenced by long-term linguistic knowledge. According to the integrative framework proposed by Majerus (2013, 2019), short-term language maintenance is achieved by the simultaneous activation of language (phonological and semantic representations), attentional, and serial order processing systems (*also see* Schwering & MacDonald, 2020). Theories taking this language-based approach to verbal working memory hypothesize that sentence recall, for instance, taps these processes simultaneously, but little work has investigated the integrative operation of these mechanisms using a common paradigm. This highlights a pressing need to evaluate this integrative framework for sentence recall. The purpose of the present study was to systematically evaluate the respective contributions of phonological, semantic, and attentional processing on sentence recall using selective suppression tasks.

Sentence recall is a multi-faceted linguistic task involving phonological processes for rehearsal, semantic knowledge for comprehension, and attentional support to integrate incoming words and recall the sentence. Perhaps because of the varying demands, some researchers argue that sentence recall is an index of verbal memory (Alloway & Gathercole, 2005), while others see it as a measure of language skills (Klem et al., 2015). Alternatively, the language-based framework takes a more integrative view of the multiple cognitive processes involved in verbal maintenance by postulating that sentence repetition is achieved by simultaneously activating the language processing system (phonological and semantic representations), encoding order information, all with support from attentional resources (Majerus, 2013, 2019). Motivating the current study is how sentence maintenance is theorized to rely on the interaction between phonological representations, semantic representations, and attentional resources to support recall. We argue that these mechanisms are jointly involved in sentence recall, but may be engaged differently. For example, immediate recall requires verbatim repetition and retention of the linguistic form and structure of the sentence, placing high demands on phonological processing (e.g., Alloway & Gathercole, 2005; Caramazza et al., 1981). Despite the high phonological demands in immediate memory, numerous studies have shown that long-term memory knowledge, such as semantic knowledge, also supports short-term maintenance of verbal information. This is supported by data showing that concrete words are recalled better than abstract words (Romani et al., 2008; Walker & Hulme, 1999) and even sentences (Meltzer et al., 2016), an effect called the concreteness effect. The concreteness effect is thought to reflect the use of semantic representations in verbal working memory. Relatedly, using intrusion paradigms, Schweppe et al. (2011) showed that phonological and semantic lure words interfered with recall, leading to the suggestion that phonological and semantic representations contribute to sentence recall.

Delayed recall, on the other hand, might draw on recalling the gist of the sentence, supported by semantics (Potter & Lombardi, 1990). Delayed recall is assumed to depend mainly on semantic information because phonological information decays rather quickly without rehearsal, thereby limiting its contribution to delayed recall

(Rummer & Engelkamp, 2003). For instance, findings by Polišenská et al. (2014, 2015) suggest that delayed recall may rely relatively more on semantics and pragmatics and immediate recall on phonology and morphosyntax.

Beyond phonological and semantic knowledge, general attentional processes support maintenance by focusing attention on the verbal information. As well, when verbal information is less familiar such as abstract sentences, maintenance may require more attentional resources. According to Barrouillet et al. (2004), speed of processing is associated with cognitive load. Abstract words are processed slower than concrete words (e.g., Paivio, 2007; Schwanenflugel, 1991), imposing a higher cognitive load and requiring more attention. Thus, processing load and concreteness can be manipulated to impose differing demands on phonological, semantic, and attentional mechanisms supporting immediate and delayed sentence recall.

Given that sentence recall taps various processes jointly, it can be expected that the introduction of a suppression task should interfere with selective resources, and may even shift reliance to other (available) resources. In a dual-task paradigm, participants perform a primary task (sentence recall) alone or with a secondary task (suppression task) with the latter being designed to heavily tax a specific resource. Based on the rationale that working memory has a limited capacity, the completion of two tasks requiring access to a common underlying resource is expected to impair performance on the primary task (Baddeley, 2003; Barrouillet et al., 2004). Indeed, a range of suppression tasks have been used to impair verbal performance (e.g., Larsen & Baddeley, 2003; Parker & Dagnall, 2009; Rende et al., 2002). However, few studies have examined the extent to which other available processes could be engaged to support recall if, indeed, multiple representations influence performance. For example, Nishiyama (2020) demonstrate that individuals can adaptively switch to relying on semantics to support serial recall when articulatory suppression disrupts phonological representations (also Romani et al., 2008; Meltzer et al., 2016). Therefore, one way to better understand how multiple interacting mechanisms including phonological, semantic, and attentional can independently support maintenance of verbal information would be to use a range of suppression tasks specifically designed to engage one or another of these supports.

Disrupting phonological processing in verbal working memory is straightforward. Articulatory suppression, the continuous repetition of an irrelevant word, is widely used to specifically interfere with phonological processing by preventing rehearsal (Baddeley, 1975). This task has been shown to disrupt verbal recall more than a concurrent visuospatial task (e.g., Alloway et al., 2010; Vergauwe et al., 2010). Interestingly, articulatory suppression has been found to reduce immediate but not delayed recall (Camos & Portrat, 2015). Moreover, articulatory suppression would prevent phonologically-based verbatim rehearsal without placing any appreciable demands on semantic resources. Meltzer et al. (2016) and Romani et al. (2008) directly tested this assumption and demonstrated an enhanced concreteness advantage under articulatory suppression indicating that reduced phonological processing allowed for maximal semantic processing.

Although there is an established body of research on using dual-task paradigms to disrupt phonological processing, to date, research using a similar method to block semantic processing during a verbal working memory task is sparse. In one of the first studies to demonstrate distinct processing in healthy adults, however, Nishiyama (2014) used articulatory suppression (repeating a word aloud) to disrupt phonological processing and attentional suppression (tapping the “0” key) to disrupt semantic processing, with the latter based on the assumption that attention underlies semantic maintenance (Ruchkin et al., 2003). Participants processed a word list phonologically (focus on the sounds of the word) or semantically (focus on the meaning of the word) under articulatory or attentional suppression. Importantly, results revealed that articulatory suppression interfered with phonological processing but not semantic processing, whereas imposing an attentional load interfered with semantic processing but not phonological processing. While these findings provide preliminary evidence for phonological and semantic representations in immediate memory, one methodological problem was using the finger-tapping task to suppress semantic processing. The finger-tapping task is not an established method for disrupting semantic processing. In fact, it was reasoned to impose an attentional load and arguably, attention has its own role in supporting recall.

Until recently it has been difficult to assess the contribution of semantic processing to verbal working memory with a task that selectively disrupts semantic processing. A potential method developed by Acheson et al. (2011) is the semantic categorization task. Participants were asked to recall concrete and nonword lists while completing a concurrent task involving judging whether or not an animal picture was a dog. Concrete words are thought to be represented by more semantic features (Paivio, 1986, 2007; Plaut & Shallice, 1993) or having richer and more semantic support (Romani et al., 2008; Walker & Hulme, 1999; also Schwanenflugel & Shoben, 1983). More broadly, the size of the concreteness effect would reflect the extent to which long-term linguistic knowledge affects short-term maintenance, a relationship used to index available semantic resources in the present study. Acheson et al. (2011) reasoned that the concurrent semantic suppression task would require accessing visual semantic features, making it difficult to rely on semantic representations to support recall of concrete words but not nonwords because nonwords lack semantic support. Although the concurrent semantic task did result in more item order errors for lists of concrete words but not nonwords, the concreteness advantage remained robust in all other analyses despite semantic interference. It could be the case that deciding if the image was or was not a dog using a limited set of images (i.e., only animals) was relatively easy and imposed only a low load on semantic resources. Thus, the semantic manipulation might have been too weak to eliminate the concreteness advantage in recall altogether. Nevertheless, following this line of reasoning, we designed two semantic suppression tasks to engage semantic representations while involving phonological processing only minimally: (1) an animacy categorization task (participants judged the animacy of images using an open set of images), and (2) a semantic relatedness task (participants judged semantic relations between objects). Critically, we hypothesized that our bespoke semantic tasks involved semantic cues that were more salient and reliable, and would thereby load more heavily on semantics than Acheson et al.'s (2011) original dog judgement task.

Another consideration is how attentional mechanisms contribute to sentence recall. Sentence recall requires attention to keep the sentences in the focus of attention for later repetition. We also consider how abstract sentences, imposing a higher cognitive load, might require more attentional resources than concrete sentences. As well, there might be some inherent attentionally demanding processes involved in the articulatory and semantic suppression tasks. Hence, a comparison with a non-verbal secondary task expected to impose an attentional load while not chiefly tapping phonological or semantic processing would provide an evaluation of the general attentional demands on performance. Simple finger-tapping[4] has been used as a control condition for articulatory suppression (Meltzer et al., 2016; Emerson & Miyake, 2003) and used to suppress semantic processing via taxing attention (Nishiyama, 2014). Simple tapping is thus expected to disrupt attentional processing while having little impact on phonological processing and may have some impact on semantic processing. This should lead to greater reliance on phonological rehearsal. Rehearsal requires only minimal attention and occurs spontaneously in the absence of specific instructions (Dunlosky & Kane, 2007; Turley-Ames & Whitfield, 2003). Drawing these different lines of research together, we conducted the current study to systematically investigate how imposing additional demands on phonological, semantic, and general attention processes would influence sentence recall.

In addition to immediate recall, it is also important to understand how sentential semantics and available resources impact long-term retention. We previously alluded to the idea that delayed recall relies on knowing the gist of the sentence, highlighting potentially a greater role of semantics relative to phonology. This would suggest that concrete sentences, being more familiar and meaningful, would be remembered better than abstract sentences in both the short- and long-term. What is transferred into long-term memory could also be informative about the processing that takes place during a

---

[4] Complex finger-tapping, which requires participants to tap in a syncopated rhythm or in a clockwise sequence, has also been used to disrupt attentional processes. We did not use complex tapping because it has been shown to disrupt verbal recall (Larsen & Baddeley, 2003) and wanted to avoid this confound.

verbal working memory task. For instance, articulatory suppression reduces phonological resources and should shift engagement to available semantic resources, which would benefit long-term recall (Craik & Lockhart, 1975; Gallo et al., 2012). Interestingly, Meltzer et al. (2016) found that this was the case. Articulatory suppression impaired immediate sentence recall, but resulted in better delayed recall (see also Rose et al., 2012; 2014). Conversely, semantic suppression reduces semantic resources but leaves phonological resources available; though phonological information is not expected to contribute to long-term recall (Rummer & Engelkamp, 2003; Potter & Lombardi, 1990). Finally, although tapping may generally occupy attention during sentence processing, participants could readily engage in rehearsal but this type of shallowing processing would be suboptimal for subsequent memory (Craik & Lockhart, 1975; Gallo et al., 2012). Thus, long-term retention may depend on processes involved during initial repetition, making it important to assess long-term sentence recall in relation to resources accessed during immediate recall.

4.1.1 The Current Study

The purpose of the present study was to understand the associations between phonological, semantic, and general attentional mechanisms underlying sentence recall. Additionally, we used the concreteness effect to examine the influence of semantic knowledge to working memory performance. To this end, we modified the novel paradigm developed by Meltzer et al. (2016) such that participants immediately recalled concrete and abstract sentences while engaging in one of the suppression tasks during a 12-s retention period (Experiment 1) or concurrently (Experiment 2). Long-term recall was tested 24-hours later in both experiments. The suppression tasks were designed to selectively tap phonological, semantic, or general attentional processes. For articulatory suppression, participants articulated six nonwords aloud to continuously impose a phonological processing load. There is no well-established task for interfering with semantic processing. We therefore imposed a semantic load in Experiment 1 by using an animacy categorization task and in Experiment 2 with the animacy task and a more demanding, semantic relatedness task. In Experiment 1, tapping was used to impose an attentional load, with little impact on phonological resources but some impact on available semantic resources. In Experiment 2, a no-suppression control condition was included.

We made the following predictions. Semantic representations were hypothesized to influence immediate and long-term sentence recall, evidenced by the concreteness effect: better and faster recall for concrete than abstract sentences. Importantly, we hypothesized that processes underlying sentence processing jointly affect recall. That is, when a certain process is blocked via suppression task, participants would invoke some other available process to support sentence recall. This led to a number of specific predictions for each suppression task. First, articulatory suppression would make it difficult to use phonological representations, shifting reliance to semantic representations during sentence recall. Thus, immediate recall should be less accurate and slow overall but with an enhanced concreteness effect. Long-term recall should be relatively good, however, because maximal semantic processing afforded by articulatory suppression would benefit long-term retention. Second, both semantic tasks were hypothesized to selectively disrupt semantic maintenance. As a result, immediate recall should be reduced and slowed, but better and faster than under articulatory suppression due to the availability of phonological resources. There should also be little (if any) concreteness effect given that a secondary semantic task would reduce semantic resources. Long-term recall would also be reduced compared to the articulatory suppression condition due to the lack of initial semantic engagement. Finally, the tapping task was hypothesized to impose an attentional load. We expected that participants would freely engage in verbatim rehearsal to support immediate recall. Thus, initial recall should be accurate and fast overall. If abstract sentences rely on attentional resources, then we might expect a larger impact on abstract than concrete sentences. Further, if tapping imposed less of a load on semantics than semantic suppression, then we expected that it would not have as much impact on the concreteness effect as a task that specifically engaged semantics. Long-term recall should be relatively poor compared to the other conditions because repeating sentences verbatim would be suboptimal for long-term memory transfer. Alternatively, it is possible that the semantic suppression task may be worst overall in the long-term if it limits the use of semantic information.

4.2 Experiment 1

In Experiment 1, we aimed to gain a more comprehensive understanding of phonological, semantic, and attentional contributions to sentence recall. We modified the Meltzer et al. (2016) paradigm by including three suppression tasks (i.e., articulatory suppression, semantic categorization, and tapping) as well as examining longer term retention (24 hours later). These three tasks were chosen because of their putative involvement of phonological, semantic, and general attentional resources, respectively, and minimally imposing on the other processes. The way in which they differentially influence performance should inform us about linguistic and cognitive processes underlying sentence recall.

4.2.1 Methods

Experiment 1 was pre-registered: https://osf.io/at8re. Experiment 2 was not pre-registered. The experiment materials (sentences, picture stimuli), data, and the analyses scripts for both experiments are available via the Open Science Framework at https://osf.io/5a2p6.

4.2.1.1 Participants

We recruited 41 participants who were proficient or native speakers of English, ranging in age from 17 to 40 years (M_{age} = 19.85; SD_{age} = 3.97, 27 female). Second/foreign languages known to participants included Arabic, Bengali, French, Chinese, Greek, Gujarati, Korean, Spanish, Telugu, Tamil, Urdu, and Vietnamese. A range of ethnicities resembling the diversity of the population was represented. Participants were recruited from the undergraduate psychology research pool and received a course credit or monetary compensation for their participation. Informed written consent was obtained for all participants. The study was approved by the institutional research ethics board for human subjects at the local university.

4.2.1.2 Materials

Sentences. The present study included 96 sentences adopted from Meltzer et al. (2016; for more details)[5]. There were 48 concrete and 48 abstract sentences. Concrete and abstract sentences provided an index of semantic resources available. Concrete sentences contained highly imageable concepts affording rich visual imagery. For instance, “The boy sneaked himself some chocolate-chip cookies while his mother was away”. Abstract sentences referred to abstract concepts such as qualities, feelings, and abstract nouns that did not easily evoke a mental image. For instance, “The doctor retired because there was very little demand for his services”. Concrete sentences had an average rating of 4.58 (SD = 0.64) and abstract sentences had an average rating of 1.15 (SD = 0.26), on a scale from 1 (least imaginable) to 5 (most imaginable). Decisive evidence confirmed a concreteness difference between the sentence sets (BF_{10} = 1.11e+51). Sentences were 10 to 16 words in length and were pre-recorded by a female speaker. Sentences were otherwise matched on a number of linguistic parameters including length, word frequency, syntactic complexity, and predictability (see Meltzer et al., 2016 for more details).

Pictures. For the semantic categorization task, a subset of 86 images were selected from the Snodgrass and Vanderwart (1980) picture set. Images were chosen such that they were clearly living or nonliving items and could easily be represented pictorially; foods and plants were excluded. Pictures were black and white line drawing of a single animal or inanimate object against a white background.

---

[5] We thank Jed A. Meltzer for sharing his stimuli with us.

4.2.1.3 Procedure

Participants were tested individually. The experiment was written using PsychoPy 1.90.3 (Peirce et al., 2019). Each trial of the immediate recall task proceeded as follows (see Figure 4.1A). First, a fixation cross appeared at the center of the screen and then a speaker image appeared to cue participants to listen to the auditory sentence being presented. Following the sentence, a cue word was displayed for 1 s to indicate which suppression task they would perform during the delay. This was followed by an 11 s phase (for a 12-s delay total) during which participants performed one of the following suppression tasks: articulatory suppression, animacy categorization, or tapping. There was a total of 32 sentences for each suppression condition, with 16 being concrete and 16 being abstract sentences. Sentences were randomly assigned to each suppression condition and differed for each participant.

Participants engaged in one of the three suppression tasks. In the articulatory suppression condition, after an initial “WORD” cue, participants were instructed to say aloud the nonword that appeared on the screen. There were six different nonwords (babataka, jujumupu, sosohopo, totoboko, riritidi, dadalara). Each word was displayed for 1500 ms with an interstimulus interval (ISI) of 500 ms. In the animacy categorization condition, after an initial “PICTURE” cue, participants were instructed to make animacy judgements about a series of pictures. Participants made an animacy judgement (is this picture a living thing or not?) by pressing the key labelled “Yes” for yes/living and the key labelled “No” for no/non-living. Participants were explicitly instructed to make a decision about the picture’s animacy as quickly and as accurately as possible, and to not label the image. Each picture remained on screen until a response (ISI of 500 ms), which triggered the appearance of the next image. Reaction time was recorded from the onset of the stimuli. In the tapping condition, after an initial "TAP" cue, participants were instructed to press the "T" key each time the letter "T" appeared on the screen. Participants tapped about twice per second (200 ms on, 500 ms off).

After the suppression task, recall was probed by an empty speech bubble appearing on the screen. Participants were instructed to recall the sentence verbatim (i.e., exactly as they heard it). Responses were recorded by an attached microphone for later transcription and scoring. Once they were finished, participants proceeded to the next trial by pressing the "space" bar, and a new trial began starting with the fixation cross. There was a 20 s time limit, after which the next trial would begin automatically. Participants practiced each suppression task once independently, without the sentence recall task. They then practiced three trials of the full task before beginning the experimental trials. A short break was taken after every 20 trials.

Participants completed the long-term cued recall task approximately 24-hours later (Figure 4.1B). On each trial, participants received two cue words (the subject and main verb) from a sentence that they previously heard in the initial session. They were asked to recall the sentence based on these two cued words. Sentences were presented in a randomized order. Participants had 30 s to recall each sentence, after which the next trial would begin automatically. Participants had three practice trials.

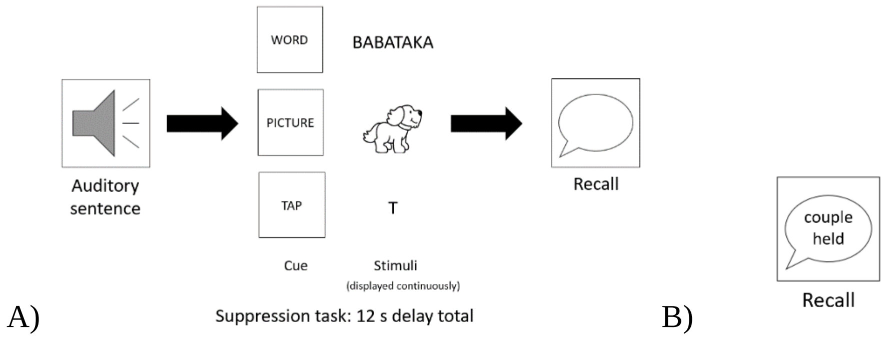

Figure 4.1. Schematic of Experiment 1. A) Immediate recall. After hearing an auditory sentence, a cue appeared. The cue was either 'word' (articulatory suppression), 'picture' (animacy categorization), or 'tap' and then participants began the suppression task. After the suppression task, participants were required to recall the sentence they previously heard. B) Long-term cued recall. After 24 hours, participants completed a cued-recall task, in which they were presented with two words (subject and verb) from the sentence that they previously heard and were asked to recall the sentence.

4.2.1.4 **Sentence transcription and scoring**

We closely followed the analysis of recall accuracy as described by Meltzer et al. (2016). Audio recordings were manually transcribed by the first author. For both immediate and long-term recall, a verbatim score was computed for each sentence. A conditional score was additionally computed for long-term data. To assess reliability, a researcher blind to the purpose of the experiment re-transcribed data from 21 participants (51%). The intra-class correlation coefficient (ICC) was used to evaluate reliability for the verbatim and conditional scores. ICC values between 0.75 and 0.90 are considered good and above 0.90, excellent (Koo & Li, 2016). There was excellent agreement across all conditions: immediate verbatim = 0.990, long-term verbatim = 0.988, and long-term conditional = 0.984.

Verbatim score. The strict verbatim score required that participants recalled the exact word including grammatical inflections; otherwise, the word would be scored as incorrect. However, credit was given even if the exact word was recalled in a different order. The number of exact words recalled was compared to the number of words in the target sentence to calculate proportion correct, ranging from 0 to 1. See Table 4.1 for examples of verbatim scoring.

Table 4.1. Examples of verbatim scoring

Target sentence (length)	Recalled sentence	Verbatim transcript (# exact words)	Verbatim score
Rainbows have coloured the sky this week during the ongoing thunderstorms (11)	The **rainbows coloured the** blue skies **during the** continuous **thunderstorms**	Rainbows coloured the during the thunderstorms (6)	6/11 = .55
The internet distracts many students form their studies and harms productivity (11)	**The internet harms productivity** by distracting **students from their studies**	The internet harms productivity students from their studies (8)	8/11 = .73

Conditional score. For long-term recall, the verbatim score might reflect not recalling words that were not immediately recalled and forgetting during the 24-hour delay. A conditional score was therefore computed to evaluate long-term recall relative to how much was initially remembered. The number of exact words recalled in the long-term was compared to the number of exact words immediately recalled. Note, it was not a word-for-word comparison so different words recalled could form the score (see Table 4.2). In rare cases, this could include recalling more words in the long-term than short-term. In even rarer cases, there was no response immediately whereas some of the sentence was recalled in the long-term. For both of these rare instances, we set proportion correct to a maximum of 1. These rare cases occurred in 6% of the data.

Conditional recall = # words in long-term verbatim recall / # words in immediate verbatim recall

Table 4.2. Examples comparing verbatim and conditional scoring

Target sentence (length)	Immediate response (# exact words)	Long-term response (# exact words)	Verbatim score	Conditional score
This miracle inspired the townspeople to a higher level of faith (11)	The **miracle inspired townspeople to a** new **level of faith** (8)	The **miracle inspired** others to believe in something **higher** (3)	Immediate: 8/11 = 0.73 Long-term: 3/11 = 0.27	3/8 = 0.38
Sausages can burn if left to sizzle on the barbecue for too long (13)	**Sausages burn** when **left to sizzle on the** grill **for too long** (10)	The **sausages burn if left on the** grill **for** a **long** time (8)	Immediate: 10/13 = 0.77 Long-term: 8/13 = 0.62	8/10 = 0.80
The angry housewife locked the bedroom door so her husband could not enter (13)	**The** wife was **angry** at the **husband** (3)	**The angry housewife locked the door** on **her husband** (8)	Immediate: 3/13 = 0.23 Long-term: 8/13 = 0.62	8/3 = 2.67; set to 1

4.2.1.5 **Statistical analysis**

Accuracy and response time data were submitted to separate Bayesian repeated-measures ANOVAs and follow-up Bayesian t-tests. Each model returns a Bayes Factor or BF_{10} value. BF_{10} is used to evaluate the strength of evidence for the alternative model (H_1) against a null model (H_0). If support for the model was ambiguous, we ran an analysis of *specific effects* to untangle this ambiguity and report inclusion BF (BF_{incl}) based on all models (van den Bergh et al., 2020). We used the following classification scheme to interpret BF: BF < 1 provides no evidence, BF between 1 and 3 provides anecdotal evidence, BF between 3 and 10 provides substantial evidence, BF between 10 and 30 provides strong evidence, BF between 30 and 100 provides very strong evidence, and BF > 100 provides decisive evidence (Jeffreys, 1961; Wagenmakers et al., 2018). Bayesian analyses were conducted using the BayesFactor (Morey & Rouder, 2018) and bayestestR packages (Makowski et al., 2019) for R. Immediate and long-term data were submitted to separate analyses given the differences between the tasks.

4.2.2 Results

Preliminary analyses. We first verified that participants were completing the animacy categorization task. On average, participants were categorizing 11.98 images on each trial (*SD* = 2.48, range = 6 – 17). Participants' response time was compared to normative naming data from Snodgrass and Yuditsky (1996) to determine whether participants were classifying based on semantic knowledge (animacy) rather than phonological (naming the images). On correct trials only, there was decisive support that response time (*M* = 559.31 ms, *SD* = 25.80) was about 315 ms faster than it would have taken participants to verbally name the images aloud (M_{vocal} = 865.00 ms, SD_{vocal} = 163.43), BF_{10} = 1.79e+22, and even when compared to subvocal responses, naming the images in their mind, faster by about 204 ms ($M_{subvocal}$ = 754.83 ms, $SD_{subvocal}$ = 168.95), BF_{10} = 2.30e+12. In fact, participants were making animacy decisions in as little as 176 ms and were therefore unlikely to be accessing the phonological word form corresponding to the image. Nevertheless, animacy judgment accuracy was high (*M* = 0.94, *SD* = 0.028), indicating participants were accessing semantic information.

4.2.2.1 Immediate recall performance

Accuracy. In Figure 4.2A, immediate recall (top row) reveals an increase in scores across the three suppression conditions, with a larger concreteness effect under articulatory suppression compared to the other conditions. Results across conditions are shown in Table 4.3. These observations were substantiated by submitting immediate verbatim scores to a 3 (suppression task) x 2 (concreteness effect) Bayesian repeated-measures ANOVA. The results provided only anecdotal evidence in favour of a model with both effects of concreteness and suppression task (BF_{10} = 4.50e+34) preferred by a factor of 1.33 over the second best model including both effects and the interaction term (BF_{10} = 3.39e+34). Given the ambiguous results, we ran an analysis of specific effects. This analysis revealed decisive evidence for an effect of sentence concreteness, BF_{incl} = 2.49e+12, with accuracy higher for concrete (*M* = 0.80, *SD* = 0.097) than abstract sentences (*M* = 0.74, *SD* = 0.13). There was also decisive evidence for the effect of suppression task, BF_{incl} = 3.18e+28. Follow-up comparisons using Bayesian t-tests showed decisive evidence that articulatory suppression (*M* = 0.69, *SD* = 0.12) reduced immediate recall compared to the other tasks (BF > 5.57e+13, both cases) and substantial evidence that recall was better for the tapping than animacy task (*M* = 0.82, *SD* = 0.088 and *M* = 0.79, *SD* = 0.099, respectively, BF_{10} = 5.46). Given substantial evidence supporting the interaction, BF_{incl} = 3.014, we explored this interaction using Bayesian t-tests. Decisive evidence supported better recall for concrete than abstract sentences across all tasks (BF_{10} > 2310, all cases) but the effect size was greater for articulatory suppression (*d* = 1.10) as compared to remaining conditions (*d* < 0.82, both cases). This result indicates an enhanced concreteness effect with articulatory suppression specifically. Analysis of each sentence separately revealed for concrete sentences, differences across all conditions (with BF_{10} ranging between 1.30 and 6.38e+8), and for abstract sentences, differences between articulatory suppression and both remaining conditions (BF_{10} ranging between 4.33e+6 and 1.15e+8) but not between animacy and tapping (BF_{10} = 0.96).

Response time. We also ran a Bayesian repeated-measures ANOVA on reaction time with sentence type and suppression task as within-subject factors. The results provided only anecdotal evidence in favour of a model with both effects of sentence concreteness and suppression task (BF_{10} = 2.088e+23) preferred by a factor of 1.76 over the second best model including both effects and the interaction term (BF_{10} = 1.18e+23). We ran an analysis of specific effects in order to untangle the ambiguous evidence. This confirmed an effect of concreteness, BF_{incl} = 2863.62, indicating that response time was faster for concrete (*M* = 8.15 s, *SD* = 1.80) than abstract sentences (*M* = 8.61 s, *SD* = 2.064). Decisive evidence also supported an effect of suppression task, BF_{incl} = 4.73e+21. Follow-up Bayesian t-tests revealed decisive support for slower response time with articulatory suppression (*M* = 9.21 s, *SD* = 2.11) compared to the animacy and tapping tasks (*M* = 8.068 s, *SD* = 1.83 and *M* = 7.85 s, *SD* = 1.62, respectively, BF_{10} > 3.87e+12,

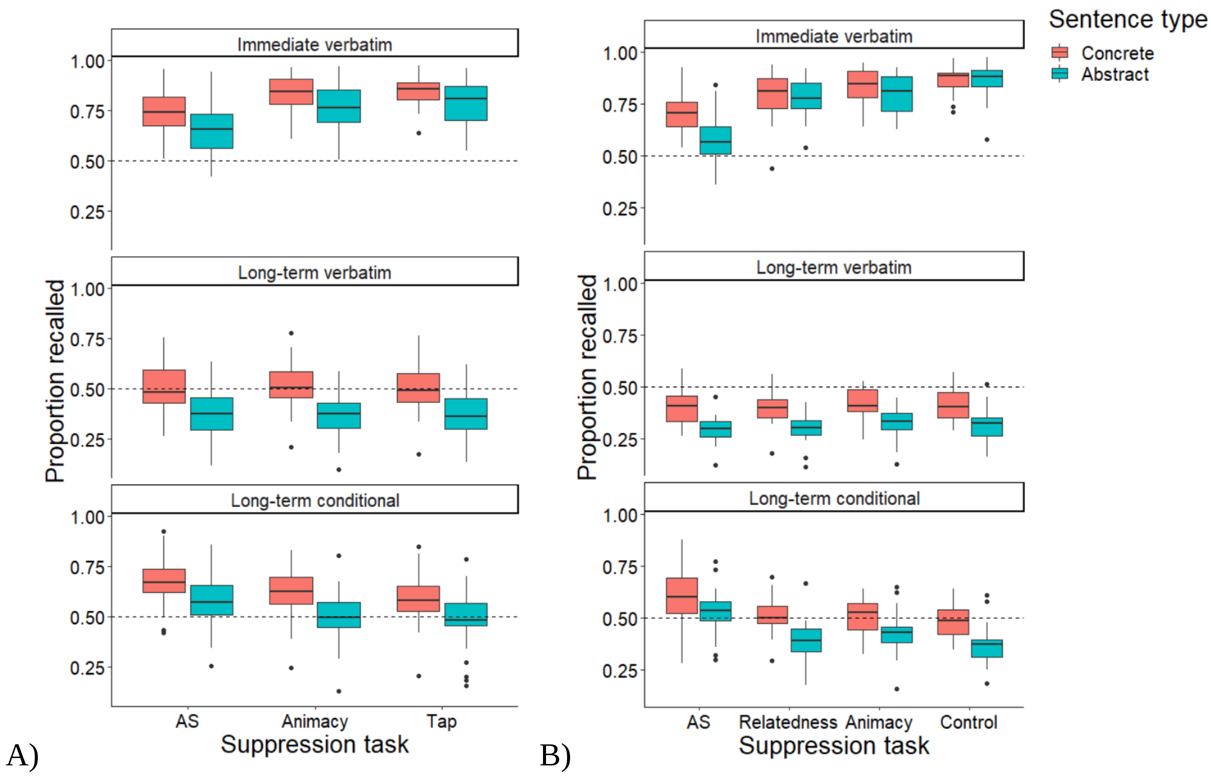

Figure 4.2. Boxplots for words recalled accurately for abstract and concrete sentences for immediate and long-term recall in A) Experiment 1 and B) Experiment 2. Top: Immediate verbatim. Middle: Long-term verbatim. Bottom: Long-term conditional. AS = articulatory suppression.

Table 4.3. Mean proportion of words correctly recalled (and standard deviations) across the suppression conditions for immediate and long-term recall in Experiment 1.

	AS	Animacy	Tap
Immediate verbatim:			
Concrete	0.74 (0.10)	0.83 (0.086)	0.84 (0.068)
Abstract	0.65 (0.12)	0.77 (0.10)	0.79 (0.099)
Long-term verbatim:			
Concrete	0.51 (0.12)	0.51 (0.12)	0.51 (0.12)
Abstract	0.38 (0.10)	0.38 (0.11)	0.38 (0.12)
Long-term conditional:			
Concrete	0.68 (0.12)	0.62 (0.12)	0.60 (0.13)
Abstract	0.58 (0.13)	0.50 (0.12)	0.49 (0.14)

Note: AS = articulatory suppression.

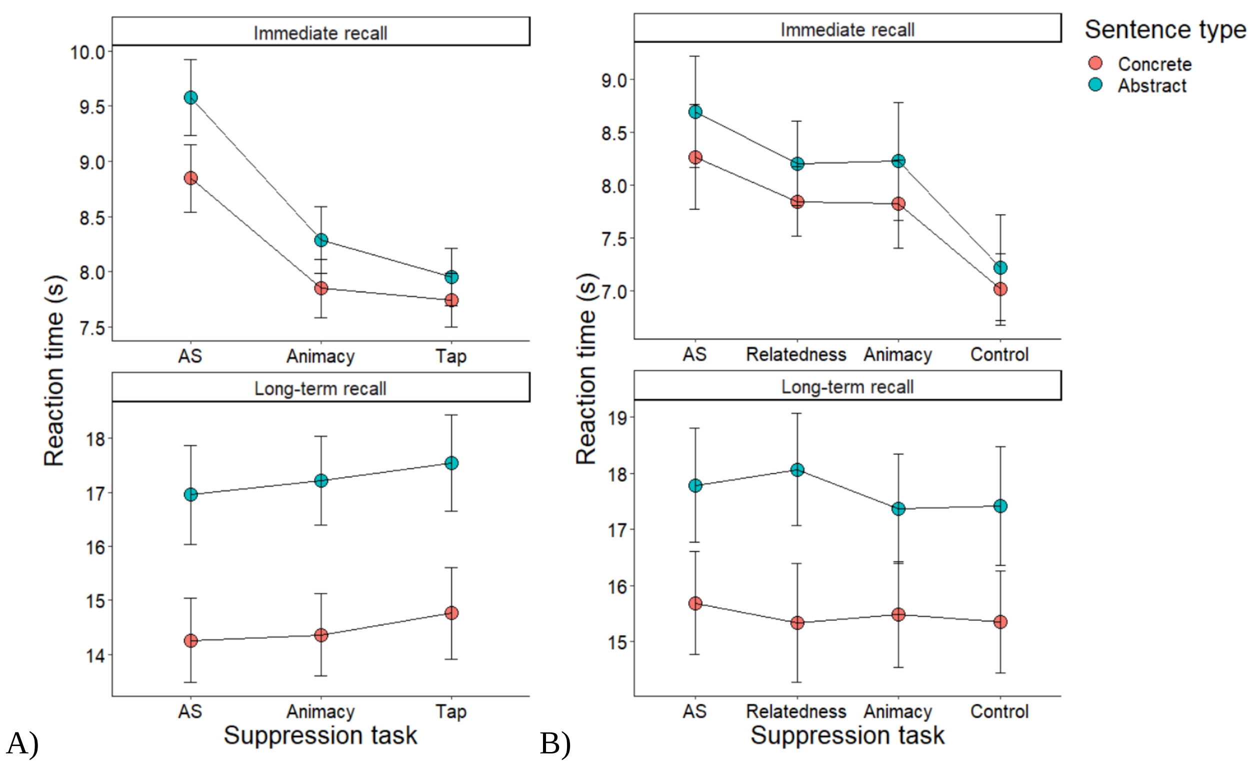

Figure 4.3. Mean reaction time for abstract and concrete sentences in the immediate and long-term recall task in A) Experiment 1 and B) Experiment 2. AS = articulatory suppression. Error bars are the standard error of the mean.

both cases) whereas there was anecdotal support for faster response time with tapping than animacy categorization (BF_{10} = 1.21). Finally, support for the presence of an interaction was only anecdotal, BF_{incl} = 2.68. See Figure 4.3A.

4.2.2.2 **Long-term recall performance**

Accuracy. We first assessed long-term verbatim scores using a Bayesian repeated-measures ANOVA. The model with the effect of concreteness only resulted in the highest BF value (BF_{10} = 7.094e+43) and was 23.72 times more likely compared to the second best model containing the two main effects (BF_{10} = 2.99e+42). Concrete sentences (*M* = 0.51, *SD* = 0.12) were recalled better than abstract sentences (*M* = 0.38, *SD* = 0.11) after 24-hours.

We next performed a Bayesian repeated-measures ANOVA on long-term conditional scores to assess long-term recall relative to immediate recall. There was substantial support for the model containing the two main effects (BF_{10} = 1.22 e+30), this model being preferred over the full model (BF_{10} = 1.94e+29) by a factor of 6.98. This supported the presence of a concreteness effect as accuracy was higher for concrete (*M* = 0.63, *SD* = 0.12) than abstract (*M* = 0.52, *SD* = 0.13) sentences. The main effect of suppression task was characterized by better recall with articulatory suppression (*M* = 0.63, *SD* = 0.13) compared to animacy judgements (*M* = 0.56, *SD* = 0.13, BF_{10} = 3.28e+7) and tapping (*M* = 0.55, *SD* = 0.14, BF_{10} = 5.96e+6) but no difference between the latter two tasks (BF_{10} = 0.23). Notably, the effect of suppression task was opposite to the pattern for immediate recall. Conditional scores (Figure 4.2A, bottom row) illustrate that long-term recall was best following articulatory suppression and that there was a concreteness effect across all conditions.

Response time. A Bayesian repeated-measures ANOVA revealed that the model with the main effect of sentence type (BF_{10} = 1.88e+21) was preferred with substantial evidence, or 4.39 times over the second best model containing the two main effects only (BF_{10} = 4.29e+20). Concrete sentences (*M* = 14.46 s, *SD* = 5.064) were associated with faster response times than abstract sentences (*M* = 17.24 s, *SD* = 5.58). See Figure 4.3A.

4.2.3 Discussion

Experiment 1 conceptually replicates the patterns of results from Meltzer et al. (2016, 2017) using a Bayesian analysis approach but crucially, extends them in important ways (i.e., inclusion of semantic suppression, assessing longer term recall, analysis of response time). There was a concreteness effect across immediate and long-term recall as concrete sentences were recalled better and faster than abstract sentences. The effect of suppression condition differed for immediate and long-term recall. For immediate recall, performance was best under tapping (least phonological disruption), followed by animacy judgements, and impaired under articulatory suppression (most phonological disruption). In contrast, long-term recall was better with articulatory suppression than the animacy and tapping tasks. The interaction between the concreteness effect and suppression task at short-term only could help explain these contrasting results. Under articulatory suppression, there was a very large and robust concreteness effect, while this effect was reduced (and not different) in both the animacy and tapping conditions. It would follow that articulatory suppression, which reduced phonological resources, increased reliance on (available) semantics at initial recall leading to an enhanced concreteness effect and subsequent benefits to long-term memory. These results suggest that semantic resources (e.g., semantic engagement, sentence concreteness) are beneficial for both immediate and long-term memory whereas phonological resources have more immediate benefits.

There was some evidence for a difference between the animacy categorization and tapping tasks hypothesized to tax semantic and general attentional resources, respectively. Substantial and anecdotal evidence supported higher accuracy and faster response with tapping than animacy judgements, respectively. These results provide suggestive evidence that the animacy task is more demanding than the tapping task, presumably because it is disrupting a linguistic mechanism specifically. These two tasks did not show a difference in effect on concreteness advantage nor long-term recall, however. In particular, although animacy categorization reduced the concreteness effect relative to articulatory suppression, this reduction was equivalent to that found in the tapping condition. Since both the animacy and tapping tasks exerted similar impacts on the concreteness effect, this could be interpreted to suggest that the advantage for concrete sentences is not only supported by long-term semantic knowledge, but may be also supported by general attentional resources (Nishiyama, 2014; Ruchkin et al., 2003). Another explanation is related to the demands imposed by both tasks. Simple tapping may impose a negligible attentional load, explaining the relatively good and fast recall overall. When conceptualizing a novel semantic suppression task, we chose animacy categorization because animacy is a highly salient and reliable semantic cue (Culbertson et al., 2017; Pham et al., 2020), and anticipated that it would have loaded more heavily on semantics than Acheson's et al. (2011) original task. However, animacy may also impose a negligible cognitive load resulting in its virtually automatic calculation (Pham et al., 2020). Response time data shows that participants recognized animacy with incredible speed – in as little as 176 ms, but there was no evidence of a speed-accuracy trade-off as accuracy was at 94%. This suggests that the task may have been relatively easy. Therefore, in Experiment 2, we additionally included a more demanding, semantic suppression task: evaluating the semantic relation between objects. Participants had to actively decide whether two objects were semantically related (e.g., used together, have a similar purpose). Pilot data indicated that judging semantic relatedness was more demanding and difficult than animacy categorization.

Other changes to Experiment 2 were as follows. First, participants engaged in a concurrent suppression task simultaneously with sentence presentation instead of during a delay period, in keeping with how dual-task paradigms have traditionally been studied. Second, articulatory and semantic suppression conditions were contrasted with a control condition without any secondary task at all in order to more rigorously evaluate how the concreteness effect changes when concurrent tasks tap linguistic knowledge specifically.

4.3 Experiment 2

Experiment 2 sought to replicate the key findings in Experiment 1 while addressing some possible limitations. In Experiment 1, it could be argued that animacy categorization was relatively easy, imposing a negligible load on semantics. Therefore, in addition to the animacy categorization task, we designed a semantic relatedness task to increase the level of semantics required by the task. Participants had to evaluate the semantic relation between objects. We also used a no-suppression condition as a baseline instead of an alternative suppression task, such as finger-tapping, to more directly examine phonological and semantic supports to sentence recall.

4.3.1 Methods

4.3.1.1 Participants

Twenty-one new participants were recruited aged between 18 and 21 (M_{age} = 18.60; SD_{age} = 1.03, 10 female). Five additional participants were excluded, four due to technical problems and one did not complete the experiment. None had participated in Experiment 1.

4.3.1.2 Materials

The sentences and pictures for the animacy categorization task were identical to those of Experiment 1.

For the semantic relatedness task, we selected 214 object pairs from the pool of pairs of related objects database which consists of images with norms for semantic relatedness (Kovalenko et al., 2012). There were 106 pairs of semantically related objects and 108 pairs of unrelated objects. As described in Kovalenko et al. (2012), two objects were related when they (a) are used together (hammer-nail), (b) serve the same purpose (glass-cup), (c) often occur in the same situation (needle-stethoscope), or (d) are from the same basic category (dog-cat). By contrast, objects that are visually similar (orange-basketball) were classified as not related. We chose object pairs that were judged to be high (M = 0.84, SD = 0.049) and low (M = 0.094, SD = 0.072) in semantic relatedness, and decisive evidence supported this difference, BF_{10} = 3.74e+164.

4.3.1.3 Procedure

The experiment was constructed with PsychoPy v2020.1.3 (Peirce et al., 2019) and hosted on Pavlovia (example of Experiment 2 procedure can be accessed at https://osf.io/kdf95). In Experiment 2, participants heard the sentence only or while concurrently engaging in an articulatory or semantic suppression task (Figure 4.4). The 96 sentences were randomly divided into 24 sentences for each suppression condition, with 12 being concrete and 12 being abstract sentences. The experiment was separated into two blocks, and the order of blocks was counterbalanced across participants. One block consisted of all 24 trials of the animacy categorization task and half, or 12 trials each, of the articulatory suppression and control conditions, while the other blocked consisted of all 24 trials of the semantic relatedness task and the remaining 12 trials for each of the articulatory suppression and control tasks. This means that half of the participants completed the animacy task first and the remaining participants completed the semantic relatedness task first, but all participants completed 24 trials of each condition. The trials within blocks were presented in random order.

At the beginning of each trial, participants saw a cue to indicate which suppression task they would perform concurrently with sentence presentation. Both the suppression task and auditory sentence began after the cue. The articulatory suppression and animacy categorization conditions followed the same general procedure as Experiment 1, except that it was done concurrently. In the semantic relatedness condition, participants saw a pair of images on the screen and had to decide whether the images were semantically related according to the criteria described above. Participants were instructed to answer as quickly and as accurately as possible. Reaction time was recorded from the onset of the stimuli. Once participants made their decision, another image would appear, and the task continued until the end of the sentence. There was no ISI between images for the semantic tasks. In the control condition, participants did nothing and listened to the sentence only. Other aspects of the procedure were the same as those in Experiment 1. Recall was tested immediately and long-term recall was tested about 24 hours later.

4.3.1.4 Data analysis

Sentence and statistical analyses were analyzed in the same way as in Experiment 1. Two researchers who were blind to the experimental conditions re-transcribed data from all participants to assess reliability. Using ICC to evaluate reliability, there was excellent reliability across the condition: immediate verbatim = 0.993, long-term verbatim = 0.962, and long-term conditional = 0.939.

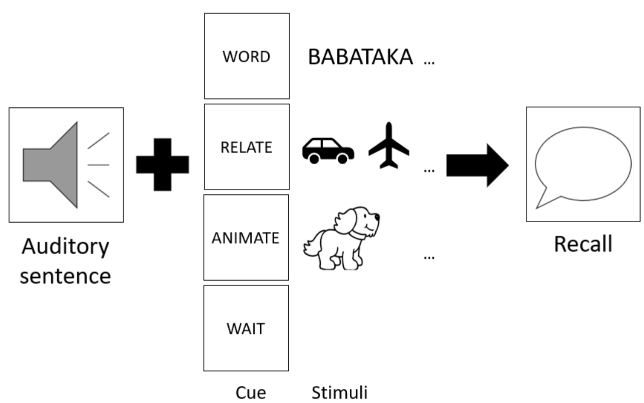

Figure 4.4. Schematic of the immediate recall procedure only of Experiment 2. After an initial cue word, both the suppression task and auditory sentence began. After the suppression task and sentence ended, participants were required to recall the sentence they previously heard. Cued long-term recall (24 hours later) was the same (not shown here).

4.3.2 Results

Preliminary analyses. We first verified that participants performed the animacy and semantic relatedness tasks. On average, participants were categorizing 8.82 images (*SD* = 1.14, range = 3 – 12) in the animacy task and 5.18 object pairs (*SD* = 0.76, range = 2 – 8) in the semantic relatedness task. Reaction time was faster in the animacy task (*M* = 668.79 ms, *SD* = 50.99) than the semantic relatedness task (*M* = 1240.94 ms, *SD* = 224.17), BF_{10} = 5.11e+7. For the animacy task, reaction time was 197 ms faster compared to normative vocal naming data (BF_{10} = 4.36e+11) and 86 ms faster compared to subvocal naming data (BF_{10} = 174.17), similar to Experiment 1. Nevertheless, accuracy was high in both the animacy (*M* = 0.98, *SD* = 0.020) and semantic relatedness task (*M* = 0.94, *SD* = 0.11), with decisive evidence supporting this difference, BF_{10} = 2318.28.

4.3.2.1 Immediate recall performance

Accuracy. We analyzed immediate verbatim accuracy scores (Table 4.4) using a 2 (sentence concreteness) x 4 (suppression task) Bayesian repeated-measures ANOVA. The best model was the full model (BF_{10} = 7.76e+31) and this model was about 7793.83 times better than the model with the two main effects (BF_{10} = 9.96e+27). Recall accuracy was higher for concrete (*M* = 0.80, *SD* = 0.11) than abstract (*M* = 0.75, *SD* = 0.15) sentences, confirming the concreteness effect. The results also indicated a main effect of suppression task, with decisive evidence supporting impaired recall under articulatory suppression (*M* = 0.65, *SD* = 0.13, BF_{10} > 7.31e+6, all cases) and best recall in the control condition (*M* = 0.86, *SD* = 0.077, BF_{10} > 120.70, all cases), with the semantic tasks intermediate. There was substantial evidence for a difference between the semantic tasks

(BF_{10} = 5.87) as accuracy was higher with animacy (*M* = 0.82, *SD* = 0.091) than semantic relatedness judgements (*M* = 0.78, *SD* = 0.10). We analyzed the interaction with Bayesian t-tests, revealing decisive evidence for the concreteness advantage with articulatory suppression (*d* = 1.85, BF_{10} = 2.91e+5) and only anecdotal evidence with animacy judgements (*d* = 0.55, BF_{10} = 2.87); however, no evidence for the semantic relatedness or control conditions (*d* = 0.087, BF_{10} = 0.24 and *d* = 0.029, BF_{10} = 0.23, respectively). This means that concrete sentences had a marked advantage over abstract sentences in the articulatory suppression condition only and this advantage was minimized or absent in the remaining conditions. We also analyzed each sentence separately. For concrete sentences, differences only emerged between articulatory suppression when compared to all other conditions ($BF_{10} > 11.78$, all cases) and no reliable difference between remaining comparisons ($BF_{10} < 3$, all cases). For abstract sentences, accuracy increased across conditions ($BF > 395.58$, all cases), from articulatory suppression (worst), to the semantic tasks, to the control (best). The semantic tasks did not differ (BF_{10} = 0.57). Thus, the absent concreteness effect in the control condition was likely due to better recall for abstract sentences compared to dual-task conditions. As can be seen in Figure 4.2B (top row), across conditions, accuracy increased while the gap between concrete and abstract sentences decreased, and was very high in the control condition.

Table 4.4. Mean proportion of words correctly recalled (and standard deviations) across the suppression conditions for immediate and long-term recall in Experiment 2.

	AS	Relatedness	Animacy	Control
Immediate verbatim:				
Concrete	0.71 (0.10)	0.79 (0.12)	0.84 (0.088)	0.87 (0.071)
Abstract	0.58 (0.12)	0.78 (0.094)	0.80 (0.092)	0.86 (0.085)
Long-term verbatim:				
Concrete	0.41 (0.092)	0.39 (0.076)	0.42 (0.077)	0.41 (.087)
Abstract	0.29 (0.068)	0.29 (0.068)	0.32 (0.076)	0.32 (0.084)
Long-term conditional:				
Concrete	0.59 (0.13)	0.51 (0.096)	0.51 (0.092)	0.49 (0.088)
Abstract	0.53 (0.12)	0.40 (0.10)	0.43 (0.11)	0.37 (0.10)

Note: AS = articulatory suppression.

Response time. A similar analysis was conducted on response times. A Bayesian repeated-measures ANOVA provided anecdotal evidence in favour of a model with both main effects (BF_{10} = 5.089e+5) preferred by a factor of 1.91 over the second best model containing the effect of suppression task only (BF_{10} = 2.66e+5). An analysis of specific effects provided decisive evidence for the concreteness effect, BF_{incl} = 2.53e+5, with faster response time for concrete (*M* = 8.085 s, *SD* = 2.30) than abstract sentences (*M* = 7.74 s, *SD* = 1.86). There was only anecdotal evidence for the effect of suppression task, BF_{incl} = 1.39, but a visual inspection of Figure 4.3B suggests that the control condition was fastest and the remaining dual-task conditions were similar to each other.

4.3.2.2 Long-term recall performance

Accuracy. We first assessed long-term verbatim scores using a Bayesian repeated-measures ANOVA with sentence concreteness and suppression task as within-subject factors. There was anecdotal evidence in favour of the model with the concreteness effect only (BF_{10} = 1.26e+21), preferred over the second best model with both the main effects (BF_{10} = 4.41e+20) by a factor of 2.84. An analysis of specific effects confirmed the presence of a concreteness effect, BF_{incl} = 1.028e+21, but no evidence for the suppression task effect, BF_{incl} = 0.26. Concrete sentences (*M* = 0.41, *SD* = 0.082) were remembered better than abstract sentences (*M* = 0.31, SD = 0.074).

We next examined performance using long-term conditional scores (Figure 4.2B, bottom row). A Bayesian repeated-measures ANOVA indicated that the best model containing the two main effects (BF_{10} = 2.93e+18) was only ambiguously preferred by a factor of 2.44 over the next best model including an interaction term (BF_{10} = 1.20e+18). An analysis of specific effects provided decisive evidence for the concreteness effect, BF_{incl} = 1.064e+11, and for the effect of suppression task, BF_{incl} = 8.86e+10. Concrete sentences (*M* = 0.52, *SD* = 0.11) were remembered better than abstract sentences (*M* = 0.43, *SD* = 0.12). The effect of suppression task was in the opposite direction of immediate recall. Decisive evidence supported better recall with articulatory suppression (*M* = 0.56, *SD* = 0.13) compared to remaining conditions (BF_{10} > 1.40e+4, all cases) and substantial evidence for better recall with the animacy than control condition (*M* = 0.47, *SD* = 0.11 and *M* = 0.43, *SD* = 0.11, respectively, BF_{10} = 4.57). There was no difference between the two semantic tasks (BF_{10} = 0.29) nor between the semantic relatedness (*M* = 0.46, *SD* = 0.11) and control conditions (BF_{10} = 0.79). There was only anecdotal evidence for the presence of an interaction term, BF_{incl} = 1.70.

Response time. A Bayesian repeated-measures ANOVA on response time revealed that the model with the highest BF was the model with the concreteness effect only (BF_{10} = 1.045e+11) and was 19.31 times more likely as compared to the second best model containing both main effects (BF_{10} = 5.42e+9). There was a concreteness effect as response times were faster for concrete than abstract sentences (*M* = 15.46 s, *SD* = 4.31 and *M* = 17.66 s, *SD* = 4.57, respectively). See Figure 4.3B

4.3.3 Discussion

Experiment 2 replicated the key findings of Experiment 1, but with the suppression task happening concurrently rather than during a 12-s delay. Concrete sentences were recalled better and faster than abstract sentences overall. The main effect of suppression task indicated that dual-task demand had a negative impact on performance overall. Recall was impeded and slow with articulatory suppression, accurate and slow with both semantic tasks, and highly accurate and fast in the control condition. Further, a comparison between the semantic tasks revealed that semantic relatedness judgements (more demanding) indeed impaired accuracy more than animacy judgements (less demanding) providing evidence that the two semantic tasks differed in their load. There was a crucial interaction between sentence concreteness and suppression task in immediate recall. Importantly, the concreteness effect was decisively supported under articulatory suppression only indicating that semantic resources were recruited to support recall when phonological resources were unavailable, similar to Experiment 1. In contrast, the immediate effects of semantic suppression as reflected by an impaired concreteness effect indicate that participants were less reliant on the semantic characteristics of the sentence when such resources were taxed. Results overall provided some evidence of graded differences in semantic suppression; the higher the semantic load imposed by a semantic task, the smaller the concreteness effect and reduced recall. The absent concreteness effect in the control condition, on the other hand, could be attributed to high recall overall. With regards to long-term recall, recall was again better and faster for concrete than abstract sentences. Similar to Experiment 1, the same interesting finding regarding the suppression tasks was observed, that of better recall for sentences in the articulatory suppression condition compared to all other conditions. Thus, although articulatory suppression impaired immediate recall, this led to less forgetting in the long term, consistent with the findings of Experiment 1 as well as Meltzer et al. (2016, 2017).

4.4 General Discussion

Our goal was to examine how phonological, semantic, and general attentional resources interactively support immediate and long-term recall of concrete and abstract sentences. In Experiment 1, participants immediately recalled sentences after completing one of three suppression tasks (articulatory suppression, animacy categorization, tapping), and again 24 hours later (long-term recall). In Experiment 2, participants concurrently engaged in one of the suppression tasks (articulatory suppression, animacy categorization, semantic relatedness, no-suppression control) happening with sentence presentation prior to recall. Long-term recall was also tested 24-hours later. Strikingly, the results were similar across both studies. Immediate and long-term recall were better and faster for concrete than abstract sentences overall due to semantic information available to support recall. The effect of suppression task was as follows. Compared to the other conditions, immediate recall was most accurate and fastest in the control condition and least accurate and slowest with articulatory suppression, with the remaining conditions intermediate. In Experiment 1, tapping was associated with higher accuracy and anecdotally faster performance than animacy categorization. In Experiment 2, animacy categorization was associated with higher accuracy but similar response time compared to semantic relatedness judgements. Further, the concreteness effect was large and robust for articulatory suppression only in immediate recall. The concreteness effect decreased with the animacy and tapping tasks in Experiment 1 and there was no reliable evidence for the concreteness effect with both semantic tasks and the control condition in Experiment 2. Finally, compared to the remaining conditions, articulatory suppression resulted in higher long-term scores relative to immediate recall.

4.4.1 Shifting from phonological to semantic processes

Recently, language-based models of verbal working memory have advanced our understanding about the nature of language processing. Language processing is supported by various neurocognitive mechanisms such as phonological and semantic information during sentence repetition, all coordinated by attentional mechanisms (Majerus, 2013, 2019). We considered this integrative framework in our investigation of how phonological, semantic, and general attentional processes were affected by interfering tasks. Prior research has focused on the immediate detrimental effect of concurrent tasks on performance, with little work investigating if other processes would be engaged and sufficient to support recall as well as long-term impacts. Therefore, the effects of various concurrent tasks were examined in the present study to understand the mechanisms underlying sentence recall. Across both studies, articulatory suppression, a concurrent task known to tax phonological representations, impaired immediate recall (low accuracy and slow response time). However, articulatory suppression, by disrupting phonological processes, afforded opportunities for greater reliance on semantics, as revealed in two ways. First, the enhanced concreteness effect for articulatory suppression only in immediate recall suggests a shift from phonological representations to semantics, and that semantic knowledge (indexed by concreteness effect) can protect against the impact of articulatory interference. This result is consistent with previous findings of enhanced concreteness effects under articulatory suppression (Meltzer et al., 2016; Romani et al., 2008; also Polišenská et al., 2014) as well as new evidence demonstrating the protective effect of semantic knowledge against an interfering secondary task (Kowialiewski & Majerus, 2020). Second, we found higher conditional long-term scores under articulatory suppression compared to the other conditions indicating an increased reliance on semantics at initial recall with subsequent benefits. Congruent with other studies, articulatory suppression disrupted immediate recall, but not delayed recall (Camos & Portrat, 2015; Meltzer et al., 2016, 2017).

A potential concern may be that articulatory suppression was associated with less forgetting in the long-term due to an artefact of the conditional scoring method, and not a result of task manipulation. That is, because articulatory suppression led to fewer exact words immediately recalled, the denominator would be small (compared to the numerator) resulting in large changes to the proportion. However, this was not the case for abstract sentences; immediate recall was impaired and long-term conditional recall was even more impaired. Nevertheless, we explored this possibility by combining both studies and constraining the dataset to highly accurate performance in immediate recall (≥ 80% accuracy) across collapsed suppression conditions (articulatory suppression, semantic suppression [animacy and semantic relatedness tasks], control [tapping and control tasks]) to boost power. In both cases, the model with the effects of concreteness and suppression task (verbatim: BF_{10} = 6.58e+30; conditional: BF_{10} = 2.32e+31) was preferred over the second best model by a factor of 11.24 and 8.92, respectively.

Crucially, for both long-term verbatim *and* conditional scores, decisive evidence (BF_{10} > 153) supported better recall under articulatory suppression compared to most conditions and substantial evidence (BF_{10} = 8.90) for articulatory suppression vs. control using verbatim scores, despite high performance for all conditions in immediate recall. It is also interesting to note that although we did not expect participants to recall more in the long-term relative to immediate recall, there were 310 trials (6% of the total data) where this was the case, with 182 of those trials (59%) from the articulatory suppression condition. Nevertheless, the suggestive evidence that articulatory suppression led to improved conditional long-term recall may be difficult to interpret because recall score could reflect an effect of immediate recall, long-term recall, or both in some complex combination. Further studies evaluating this scoring method and the potential benefit of articulatory suppression in general are needed.

The idea that sentence processing draws on different levels of linguistic representations simultaneously is interesting to consider. Some researchers reason that articulatory suppression blocks access to language and interpret their findings accordingly (e.g., Dymarska et al., 2021). However, our results indicate that not all linguistic information was blocked with articulatory suppression. The presumed shift from phonological to semantic processing suggests that articulatory suppression specifically interfered with the phonological aspect while having little effects on other linguistic mechanisms such as semantics. In fact, sentence processing inadvertently engaged available semantic processes because articulatory suppression prevented phonological processing, resulting in an enhanced concreteness effect and better long-term performance. Similarly, Nishiyama (2020) showed that participants adaptively shifted from relying on phonological to semantic representations in serial recall when it was hard to rehearse words (Experiment 1) or when they were instructed to use a semantic strategy (Experiment 2). The mutual interactions between different levels of linguistic knowledge aligns with recent theories conceptualizing working memory as an emergent property from long-term linguistic knowledge (e.g., Majerus, 2013, 2019; Martin & Gupta, 2004). This calls for future studies to assess both immediate and long-term effects of articulatory suppression.

4.4.2 The influence of semantic knowledge on verbal working memory

The role of semantic resources was examined through the concreteness effect and the impact of our novel semantic suppression tasks on the concreteness effect. Consistent with prior work (e.g., Walker & Hulme, 1999; Romani et al., 2008; Meltzer et al., 2016, 2017), semantic knowledge influenced verbal working memory: concrete sentences were recalled better and faster than abstract sentences in the short- and long-term. Importantly, the extent to which semantic resources were available to support the concreteness effect was demonstrated through the immediate impact of semantic suppression. The concreteness effect was minimized and absent during animacy categorization and semantic relatedness judgements, respectively, indicating semantic resources were unavailable to support the advantage afforded by concrete sentences. Moreover, unique effects on immediate accuracy and response time data were observed for the two semantic tasks when compared to other dual-tasks. The two semantic tasks collectively resulted in better immediate performance than articulatory suppression (better recall, slower-to-comparable speed), performance was not as good as the tapping task (poorer accuracy, somewhat slower speed), and impaired compared to control trials (reduced recall and speed). The distinct impact of articulatory, semantic (both animacy and relatedness judgements), and attentional suppression on performance highlights the fact that each task was likely tapping a different resource underlying sentence recall.

We compare our findings to those obtained by Acheson et al. (2011). In that study, the original dog judgement task – presumed to disrupt semantic processing – only had limited effects on recall of concrete words. While there was an impact on item ordering errors, their semantic task did not reduce the concreteness advantage in recall in remaining analyses. We reasoned that this may be attributed to instructing participants to make relatively easy categorical judgements of pictures. Indeed, our findings revealed that the more demanding the concurrent semantic task, the more it reduced accuracy and the concreteness effect in immediate recall. Importantly, our method is also advantageous compared to prior work using finger-tapping to indirectly tap semantics (Nishiyama, 2014). We have provided a method to more directly tap semantics and in fact found that the animacy and tapping tasks differed from each other in Experiment 1. Immediate recall was better (substantial evidence) and somewhat faster (anecdotal evidence) with animacy categorization than tapping. Additionally, we were able to demonstrate a dissociation between phonological and semantic representations using tasks imposing a phonological and semantic load, respectively and separately. Taken together, the novelty of the animacy and semantic relatedness judgements, requiring participants to recruit semantic resources and phonological minimally, validate these tasks as potential semantic suppression tasks. Having a task engaging semantics to parallel with articulatory suppression (imposing phonological demands) and tapping (imposing attentional demands) would allow future studies to systematically examine the relative contributions of each process to different verbal tasks.

One particular finding that warrants further discussion is the unexpected finding that animacy categorization and tapping had similar effects on the concreteness effect in Experiment 1, despite the former task additionally tapping semantic knowledge. Given that animacy can be recognized automatically (Pham et al., 2020), Experiment 2 included a more demanding semantic relatedness task and indeed we found that accuracy was higher with animacy than semantic relatedness judgements, confirming that the two tasks differed in their load. Crucially however, both semantic tasks resulted in no evidence or at best, anecdotal evidence for the concreteness effect. Why then did the animacy categorization task have a greater effect on the concreteness effect in Experiment 2 than Experiment 1? To reconcile this, we suggest that *additional time* before immediate recall impacted involvement of semantic processing. In Experiment 1, participants performed the suppression task during a 12-s delay period leaving semantic resources available at the time of sentence presentation for encoding. This would have allowed greater semantic processing of the sentences supporting the advantage for concrete sentences. In contrast, participants would not have time to engage semantics in Experiment 2 given the simultaneous occurrence of sentence presentation and suppression task. This might also explain similar results found in the tapping task. The concreteness effect emerging in a control condition by Meltzer et al. (2017) using a similar paradigm, in contrast to the results here, can also be explained by the fact that there was a 5.5 s delay period before recall, hence, allowing opportunities for semantic processing. Relatedly, other work in verbal short-term memory has found increasing reliance on semantics with additional time (Pham & Archibald, 2021; Polišenská et al., 2014). Nevertheless, along with the semantic relatedness task, the animacy categorization task (despite its automaticity) has the potential to be used to selectively engage semantics, with concurrent presentation imposing an even higher semantic load.

4.4.3 The role of attentional processes and rehearsal

Another finding of the current study was that without a high cognitive load, participants were likely to repeat sentences verbatim, but such superficial processing was not beneficial for long-term memory, in line with prior work (Craik & Lockhart, 1975; Gallo et al., 2012). We found that when participants were free to rehearse, as was the case in the tapping and control conditions, recall was fast and accurate. However, maintaining surface level phonological representations was not beneficial to long-term recall. Both the tapping and control conditions, along with the semantic tasks, hindered long-term recall more than articulatory suppression.

Beyond the influence of rehearsal, phonological representations appeared to directly contribute to the maintenance of verbal information in working memory. Unsurprisingly, when rehearsal was available and linguistic mechanisms were not disrupted (tapping and control conditions), then there was a negligible impact on immediate recall accuracy. In contrast, when rehearsal was blocked via articulatory suppression, performance was impaired. However, immediate recall was relatively good under semantic suppression even though participants would not have been able to rehearse as freely as in the tapping or control tasks while phonological resources were more available than with articulatory suppression. Accordingly, graded differences in immediate recall emerged in that performance was best for the control condition, followed closely by tapping, intermediate for both semantic tasks, and worst for articulatory suppression. This suggests that phonological knowledge can support recall, even with more limited opportunities for rehearsal. Indeed, phonological coding has been deemed an important contributor to immediate serial recall, over and above the effect of rehearsal (Tehan et al., 2004). This result is also consistent with predictions of the language-based models, according to which multiple types of interactive representations such as phonological and semantic can be activated rapidly to support recall (Majerus, 2013, 2019; Martin & Gupta, 2004). Recent empirical work has also demonstrated that verbal information directly activates corresponding representations within the linguistic system upon encounter, without the need for rehearsal (Pham & Archibald, 2021). Thus, direct activation of phonological representations could support verbal maintenance when rehearsal is prevented and semantics is taxed, as was the case under semantic suppression. However, the lack of immediate semantic engagement with phonological encoding – whether through rehearsal (tapping and control conditions) or direct activation (semantic tasks) – impaired long-term remembering. Future research on the relation between verbal working memory and long-term memory would be helpful in understanding the exact strategies and processes influencing immediate recall and subsequent memory.

4.4.4 Conclusion

We present a novel perspective that highlights the relational influences of phonological, semantic, and attentional mechanisms on sentence recall, aligning with recent theories that considers short-term maintenance of verbal working memory as emergent from neurocognitive mechanisms (Majerus, 2013, 2019). This study is the first to systematically investigate the interplay between phonological, semantic, and general attentional resources on a linguistic task such as sentence recall. Our findings suggest that immediate recall tends to rely on verbatim repetition unless such phonological processes are blocked or the sentential content engages semantics, while long-term recall benefits from engaging semantics either intentionally (sentence concreteness) or incidentally (maximizing semantic processing by blocking phonological). Further, we provide a novel way to study the influence of semantic knowledge in working memory. We conclude that sentence recall is a multi-faceted task relying on language skills as well as verbal memory skills to support processing.

4.5 References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 44-59. https://doi.org/10.1037/a0021205

Alloway, T. P, Kerr, I., & Langheinrich, T. (2010). The effect of articulatory suppression and manual tapping on serial recall. *European Journal of Cognitive Psychology, 22(2),* 297-305. https://doi.org/10.1080/09541440902793731

Alloway, T. P., & Gathercole, S. E. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology, 17,* 207-220. https://doi.org/10.1080/09541440440000005

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36,* 189-208. https://doi.org/10.1016/s0021-9924(03)00019-4

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589. https://doi.org/10.1016/S0022-5371(75)80045-4

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83-100. https://doi.org/10.1037/0096-3445.133.1.83

Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review, 22(4),* 1029–1034. https://doi.org/10.3758/s13423-014-0772-5

Caramazza, A., Basili, A. G., Koller, J. J., & Berndt, R. S. (1981). An investigation of repetition and language processing in a case of conduction aphasia. *Brain and Language, 14,* 235-271. https://doi.org/10.1016/0093-934x(81)90078-x

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684

Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language, 92,* 343–358. https://doi.org/ 10.1016/j.jml.2016.08.001

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology, 60(9),* 1227–1245. https://doi.org/10.1080/17470210600926075

Dymarska, A., Connell, L., & Banks, B. (2021, March 4). *Linguistic bootstrapping allows more real-world object concepts to be held in mind.* https://doi.org/10.31234/osf.io/5v6de

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: a dual-task investigation. *Journal of Memory and Language, 48,* 148–168. https://doi.org/10.1016/S0749-596X(02)00511-9

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58(4),* 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J.- E., and Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science, 18,* 146–154. https://doi.org/10.1111/desc.12202

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15(2),* 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects (POPORO) for investigating visual semantic integration: Behavioral and electrophysiological validation. *Brain Topography, 25(3),* 272-284. https://doi.org/10.1007/s10548-011-0216-8

Kowialiewski, B., & Majerus, S. (2020). The varying nature of semantic effects in working memory. *Cognition, 202,* 104278. https://doi.org/10.1016/j.cognition.2020.104278

Larsen, J., & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology, 56A,* 1249–1268. https://doi.org/10.1080/02724980244000765

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7,* 357. https://doi.org/10.3389/fnhum.2013.00357

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112,* 122–133. https://doi.org/10.1016/j.cor-tex.2018.04.016

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software, 4(40),* 1541. https://doi.org/10.21105/joss.01541

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. https://doi.org/10.1080/02643290342000447

Meltzer, J. A., Kielar, A., Panamsky, L., Links, K. A., Deschamps, T., & Leigh, R. C. (2017). Electrophysiological signatures of phonological and semantic maintenance in sentence repetition. *NeuroImage, 156,* 302–314. https://doi.org/10.1016/j.neuroimage.2017.05.030

Meltzer, J.A., Rose, N.S., Deschamps, T., Leigh, R.C., Panamsky, L., Silberberg, A., Madani, N., Links, K.A. (2016). Semantic and phonological contributions to short-term repetition and long-term cued sentence recall. *Memory & Cognition. 44,* 307–329. https://doi.org/10.3758/s13421-015-0554-y

Morey, R. D., & Rouder, J. N. (2018). BayesFactor (version 0.9.12-4.2) [Computer software].

Nishiyama, R. (2014). Active maintenance of semantic representations. *Psychonomic Bulletin & Review, 21(6),* 1583-1589. https://doi.org/10.3758/s13423-014-0618-1

Nishiyama, R. (2020). Adaptive use of semantic representations and phonological representations in verbal memory maintenance. *Journal of Memory and Language*, *111*, 104084. https://doi.org/10.1016/j.jml.2019.104084

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Erlbaum

Paivio, A. (1986). *Mental representations*. Oxford University Press.

Parker, A., & Dagnall, N. (2009). Concreteness effects revisited: The influence of dynamic visual noise on memory for concrete and abstract words. *Memory, 17,* 397–410. https://doi.org/10.1080/09658210902802967

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51(1),* 195-203, https://doi.org/10.3758/s13428-018-01193-y

Pham, T. & Archibald., L. M. D. (2021). The role of phonological and semantic representations in verbal short-term memory and delayed retention. *Memory & Cognition.* Advance online publication. https://doi.org/10.3758/s13421-021-01216-8

Pham, T., Kang, J. H., Johnson, A., & Archibald, L. M. D. (2020). Feature-focusing constraints on implicit learning of function word and meaning associations. *Applied Psycholinguistics, 41(2),* 401–426. https://doi.org/10.1017/S0142716420000041

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377-500. https://doi.org/10.1080/02643299308253469

Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: what does the task measure? *International Journal of Language & Communication Disorders, 50(1),* 106–118. https://doi.org/10.1111/1460-6984.12126

Polišenská, K., Chiat, S., Comer, A., & McKenzie, K. (2014). Semantic effects in sentence recall: The contribution of immediate vs delayed recall in language assessment. *Journal of Communication Disorders, 52,* 65–77. https://doi.org/10.1016/j.jcomdis.2014.08.002

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Rende, B., Ramsberger, G., & Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology, 16(3),* 309–321. https://doi.org/10.1037/0894-4105.16.3.309

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61(2),* 292-323. https://doi.org/10.1080/17470210601147747

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1019-1029. https://doi.org/10.1037/a0026976

Rose, N. S., Buchsbaum, B. R., & Craik, F. I. M. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition, 42, 689-700.* https://doi.org/10.3758/s13421-014- 0398-x

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *The Behavioral and Brain Sciences, 26(6),* 709-728. https://doi.org/10.1017/S0140525X03000165

Rummer, R., & Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *The Quarterly Journal of Experimental Psychology, 56A,* 83–95. https://doi.org/10.1080/02724980244000279

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223–250). Erlbaum.

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

Schweppe, J., Rummer, R., Bormann, T., and Martin, R. C. (2011). Semantic and phonological information in sentence recall: experimental evidence and a neuropsychological single-case study. *Cognitive Neuropsychology. 28,* 521–545. https://doi.org/ 10.1080/02643294.2012.689759

Schwering, S. C. & Macdonald. M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14,* 68. https://doi.org/10.3389/fnhum.2020.00068

Snodgrass, J. G., & Vanderwart, M. (1980). Standardized set of 260 pictures—norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Tehan, G., Fogarty, G., & Ryan, K. (2004). The contribution to immediate serial recall of rehearsal, search speed, access to lexical memory, and phonological coding: An investigation at the construct level. *Memory & Cognition, 32(5),* 711–721. https://doi.org/10.3758/BF03195861

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49(4),* 446– 468. https://doi.org/10.1016/S0749-596X(03)00095-0

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science, 21,* 384–390. https://doi.org/10.1177/0956797610361340

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., …Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review, 25,* 58–76. https://doi.org/10.3758/s13423-017-1323-7

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5),* 1256-1271. https://doi.org/10.1037/0278-7393.25.5.1256

4.6 Supplemental Materials

4.6.1 Supplementary Experiment 1

4.6.1.1 Methods

Gist score. The verbatim scoring method reported in the main text was a conservative approach to analyzing sentence recall. Even the conditional scoring relied on the exact words being recalled. This strict criterion could have missed some crucial evidence indicating that participants remembered the gist of a sentence but not the exact wording. Scoring gist recall could be highly subjective though. We therefore used a machine learning algorithm to objectively estimate the degree of similarity between the target sentence and given answer, assessed by cosine similarity. Words in the sentence were first converted into vectors so that words can be represented numerically. We created the word vectors based on the Global Vectors for Word Representation (GloVe) model (Pennington et al., 2014). Specifically, we used the Wikipedia 2014 + Gigaword 5 (50d) pre-trained vectors. The cosine similarity was then computed between the target sentence and recalled sentence, with cosine similarity ranging from 0 (the two sentences have very different meanings) to 1 (the two sentences have very similar meanings) (Figure S4.1 for example). Cosine similarity can be thought of similarly to the standard correlation (Clark, 2018). To enhance interpretation of cosine similarity, we converted cosine distance to a percentage (Sieg, 2018). See Figure S4.1 for an illustration and Table S4.1 for examples.

```
Sentence 1 (Correct answer) = "Sausages can burn if left to sizzle on the barbecue for too long"
Sentence 2 (Given answer) = "When left on a grill for too long the sausages burnt"

Word Embedding method with a cosine distance assesses that our two sentences are similar to 92.85%
```

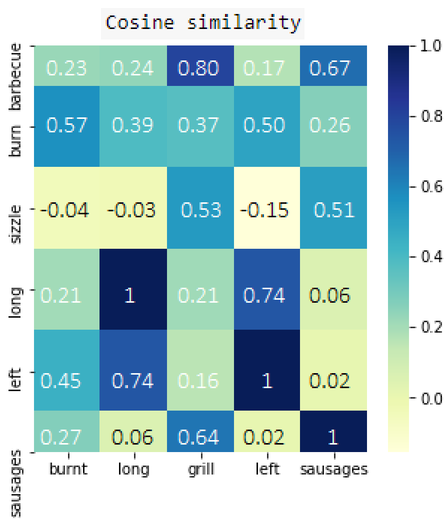

Figure S4.1. Heatmap of co-occurrence matrix representing the occurrence of one word with the other words.

Table S4.1. A comparison of scoring methods

Target sentence (length)	Recalled sentence	Verbatim transcript (exact words)	Verbatim score	Cosine similarity
Sausages can burn if left to sizzle on the barbecue for too long (13)	When **left on** a grill **for too long** the **sausages** burnt	left on for too long sausages (6)	6/13 = .46	92.85%
The internet distracts many students from their studies and harms productivity (11)	Social media is distracting **and** impacts work	and (1)	1/11 = .091	86.67%

Error analysis. We also examined the types of errors that were produced. The errors were classified as order changes, grammatical substitutions, semantic substitutions, phonological substitutions, unrelated additions, and open-class omissions (see Table S4.2 for examples). Order changes were instances where the correct word or phrase was exchanged for another in the sentence. Substitutions were classified as errors when the whole word was replaced by a similar function word or a word that changed in grammatical inflection (plurality, tense, contraction; grammatical substitutions); a content word that was similar in meaning (semantic substitutions); a word that sounded similar (phonological substitutions). Unrelated additions were content words recalled that were not in the original sentence. Omissions were instances when a content word was not recalled. For both short- and long-term performances, we counted the total number of errors in each category for each sentence.

Table S4.2. Example of each error type; error italicized in each case and omissions marked with a strikethrough

Error type	Example of error
Order changes:	
Yeast serves as an increasingly important model for the study of life	*For the study of life* yeast serves as an increasingly important model
Grammatical substitutions:	
A tornado destroyed the farmers house and scattered the pieces over the fields	A tornado destroyed the farmers house *scattering* the pieces over the *field*
Semantic substitutions:	
Retrievers will sniff the ground obsessively until they discover the source of a smell	*Dogs* will *smell* the ground *excessively* until they discover the source of a smell
Phonological substitutions:	
Stocks have declined in value recently as confidence in the market has fallen	Stocks have *decreased* in value recently as confidence in the market has *faded*
Unrelated additions:	
The potato chips crunched loudly in the teeth of the kids in the movie theatre	The *popcorn* and chips crunched loudly in the teeth of the *adults watching* in the theatre
Open-class omissions:	
The senator argued strongly in favour of the bill until it was finally passed	The senator argued ~~strongly~~ in favour of the bill until it was ~~finally~~ passed

4.6.1.2 Results

Immediate performance

Gist accuracy. Gist recall assessed by cosine similarity was very high overall, as shown in Table S4.3 and Figure S4.2A (top panel). A 2 (sentence: concrete, abstract) x 3 (suppression: articulatory suppression, animacy task, tapping) Bayesian repeated-measures ANOVA on cosine similarity provided anecdotal evidence in favour of the full model (BF_{10} = 1.041e+12) preferred by a factor of 1.30 over the second best model containing the effect of suppression task only (BF_{10} = 8.014e+11). An analysis of specific effect provided anecdotal evidence for the concreteness effect, BF_{incl} = 1.49 and decisive evidence supporting the main effect of suppression task, BF_{incl} = 1.094e+12. For the suppression task effect, follow-up Bayesian T-Tests revealed that compared to articulatory suppression (*M* = 93.17, *SD* = 5.74), recall was better with animacy judgements (*M* = 96.77, *SD* = 2.78, BF_{10}= 7.56e+11) and tapping (*M* = 97.19, *SD* = 2.74, BF_{10}= 1.11e+17), though recall was high across the three conditions. There was also substantial evidence for better recall following tapping than animacy judgements (BF_{10}= 10.35). Finally, there was only anecdotal evidence for the presence of an interaction term, BF_{incl} = 2.67.

Error. We examined the occurrence of each error category using separate 2 (sentence type) x 3 (suppression task) repeated-measures Bayesian ANOVAs (Figure S4.3A). Phonological substitutions were rare in all conditions, and in fact, results supported the null model (BF_{10} = 0.13 or there is 7.69 times more evidence for not making phonological errors). Order changes were also low and there was anecdotal evidence for the model with the main effect of sentence only (BF_{10} = 2.11) over the second best model including both main effects (BF_{10} = 0.82) by a factor of 2.58. An analysis of specific effect found no evidence for a main effect of sentence type (BF_{incl} = 0.091) nor suppression task (BF_{incl} = 0.057) however.

For both semantic substitutions and errors of omissions, the model with both main effects (semantic: BF_{10}= 2.61e+14; omissions: BF_{10}= 1.29e+39) was the best model. Abstract sentences showed more semantic substitutions and omissions than concrete sentences. For semantic substitutions, articulatory suppression showed the most semantic errors ($BF_{10} > 1.52e+8$, both cases) while animacy and tapping tasks did not differ ($BF_{10} = 0.12$). Omissions were also most common under articulatory suppression ($BF_{10} > 1.20e+17$, both cases) and more words were being omitted with animacy judgements than tapping at the anecdotal level ($BF_{10} = 2.55$).

Finally, the best model for both grammatical substitutions and unrelated additions was the full model, $BF_{10} = 5.77e+4$ and $BF_{10} = 1.37e+22$, respectively. Abstract sentences showed more grammatical substitutions and additions than concrete sentences. For grammatical substitutions, the effect of suppression was as follows. Articulatory suppression resulted in more grammatical errors than tapping ($BF_{10} = 20.079$) but no evidence for a difference in remaining comparisons ($BF_{10} < 1.85$, all cases). The interaction was explored using Bayesian T-Tests, showing evidence for the reverse concreteness effect in the animacy ($BF_{10} = 20.33$) and tapping conditions ($BF_{10} = 2704.79$) but not with articulatory suppression ($BF_{10} = 0.21$). That is, abstract sentences resulted in more grammatical errors than concrete sentences with the animacy and tapping tasks but no grammatical effects under articulatory suppression. For additions, the effect of suppression was such that articulatory suppression showed the most additions ($BF_{10} > 4.90e+4$, all cases) while animacy and tapping tasks did not differ ($BF_{10} = 0.50$). We explored the interaction and found that the articulatory suppression condition resulted in disproportionately more additions than remaining conditions, especially for abstract sentences.

Long-term performance

Gist accuracy. The result of a Bayesian repeated-measures ANOVA provided strong evidence in favour of a model with the concreteness effect only (BF_{10} = 1.11e+5) and this was preferred over the second best model containing both main effects (BF_{10} = 6576.22) by a factor of 16.88. Concrete sentences (*M* = 83.62, *SD* = 10.61) were remembered better than abstract sentences (*M* = 79.16, *SD* = 13.95), confirming the concreteness effect.

A Bayesian repeated-measures ANOVA on conditional cosine similarity resulted in similar findings. The model containing the concreteness effect (BF_{10} = 4.61e+4) was substantially preferred over the second best model with the two main effects (BF_{10} = 1.35e+4) by a factor of 3.42. Again, concrete sentences (*M* = 85.99, *SD* = 10.84) were remembered better than abstract sentences (*M* = 81.61, *SD* = 14.61). Figure S4.2A (lower panels) illustrates the results of long-term gist recall.

Table S4.3. Cosine similarity in percentage with word embedding method using the GloVe model (and standard deviations) across the suppression conditions for immediate and long-term recall in Experiment 1.

	AS	Animacy	Tap
Immediate gist:			
Concrete	94.31 (4.57)	96.74 (3.07)	97.33 (2.72)
Abstract	92.03 (6.58)	96.80 (2.49)	97.05 (2.78)
Long-term gist:			
Concrete	83.42 (10.55)	84.30 (10.98)	83.13 (10.52)
Abstract	79.28 (12.83)	79.35 (13.054)	78.85 (16.069)
Long-term conditional gist:			
Concrete	86.72 (10.87)	86.43 (11.04)	84.82 (10.78)
Abstract	82.86 (13.078)	81.42 (13.24)	80.55 (16.44)

Note: AS = articulatory suppression.

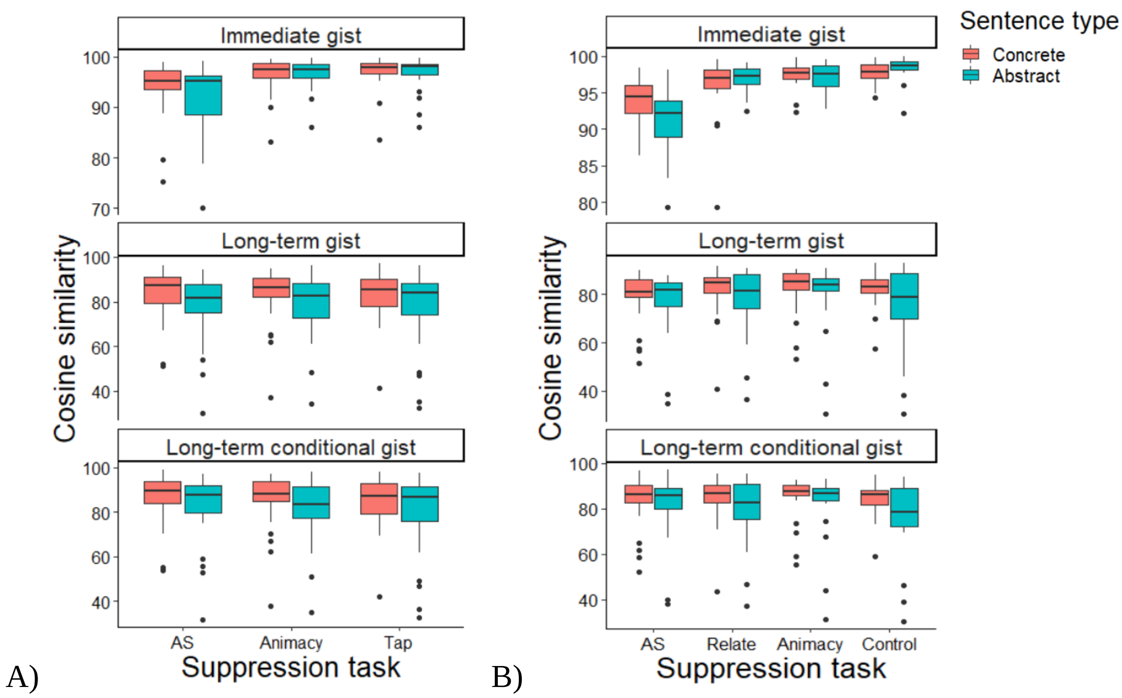

Figure S4.2. Boxplots for cosine similarity across experimental factors for immediate and long-term recall in A) Experiment 1 and B) Experiment 2. Top: Immediate recall. Middle: Long-term recall. Bottom: Long-term conditional recall. AS = articulatory suppression.

Error. Similar to short-term error analyses, we conducted a Bayesian repeated-measures ANOVA for each type of error committed during long-term recall (Figure S4.3B). Semantic substitutions, grammatical substitutions, and order changes were not affected by any experimental factors, with results supporting the null model ($BF_{10} < 0.64$, all cases).

The model with the main effect of sentence type only was the best model for additions (BF_{10} = 6.059e+28), omissions (BF_{10} = 1.031e+33), and phonological substitutions (BF_{10} = 3.83), preferred by a factor of 5.33, 18.57, and 22.53, respectively, over the second best model. Abstract sentences resulted in more unrelated additions and

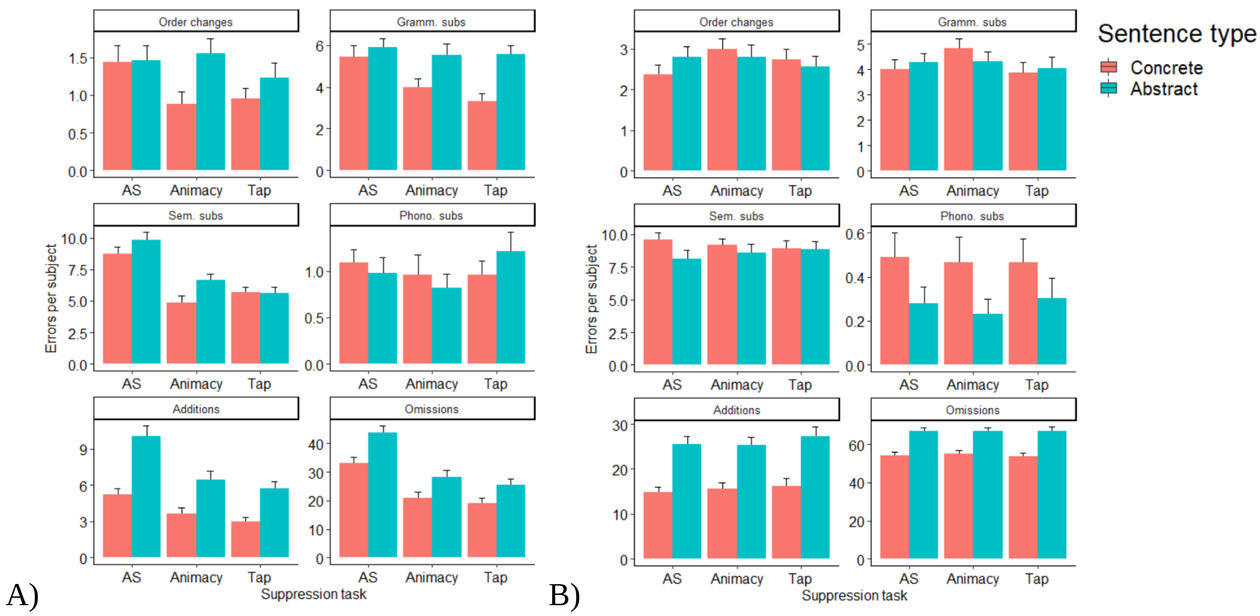

Figure S4.3. The panels show the average number of errors made by each participant across conditions in A) immediate recall and B) long-term recall in Experiment 1. AS = articulatory suppression. Error bars are the standard error of the mean.

omissions than concrete sentences. In contrast, concrete sentences showed more phonological substitutions than abstract sentences.

4.6.1.3 **Discussion**

An analysis of gist accuracy using cosine similarity revealed that performance was relatively good overall. Even after 24 hours, accuracy was above 80% suggesting that participants were maintaining the gist meaning of the sentences. For immediate recall, there was only decisive evidence for an effect of suppression task whereas evidence for the concreteness effect and interaction term was at the anecdotal level (the full model was supported with verbatim recall, as reported in the main text). Exploration of the suppression effect revealed that recall was best with tapping, then animacy categorization, and impaired under articulatory suppression, though note recall was high overall. Nevertheless, this is the same pattern we found with verbatim scores reported in the main text. Cosine similarity for both long-term and conditional recall resulted in support for the concreteness effect as concrete sentences were remembered better than abstract sentences. In contrast to the verbatim results reported in the main text, articulatory suppression did not result in better retention compared to the remaining conditions.

Additionally, an analysis of the error types indicated differences between the experimental manipulations. Abstract sentences showed more semantic substitutions, grammatical substitutions, omissions, and additions than concrete sentences in immediate recall. In short-term recall only, semantic substitutions, grammatical substitutions, and omissions were the most common under articulatory suppression compared to remaining conditions. This pattern might in fact strengthen the idea that there is a shift from phonological to semantic resources under articulatory suppression in that participants are omitting target words and replacing them with synonyms and substitutions. For long-term recall, abstract sentences resulted in more omissions and additions than concrete sentences. Interestingly, concrete sentences resulted in more phonological substitutions than abstract sentences in long-term recall, suggesting some phonological contributions to long-term recall. But this observation should be interpreted with caution as phonological errors were extremely rare.

4.6.2 Supplementary Experiment 2

4.6.2.1 Methods

The same gist score and error analyses were used as in Experiment 1.

4.6.2.2 Results

Immediate performance

Gist accuracy. Table S4.4 and Figure S4.2B (top panel) display the results of gist recall (cosine similarity) across conditions in Experiment 2. A 2 (sentence: concrete, abstract) x 4 (suppression: articulatory suppression, relatedness task, animacy task, control) Bayesian repeated-measures ANOVA on cosine similarity provided very strong support for the full model (BF_{10} = 4.060e+19) preferred by a factor of 57.18 over the second best model containing the effect of suppression task only (BF_{10} = 7.10e+17). Recall was better for concrete (*M* = 96.20, *SD* = 3.21) than abstract sentences (*M* = 95.81, *SD* = 4.063). The model also supported a main effect of suppression task, with recall impaired under articulatory suppression (M = 92.45, SD = 4.32, BF_{10} > 2.95e+4, all cases) and best in the control condition (M = 98.00, SD = 1.55, BF_{10} > 5.43, all cases), with the semantic tasks intermediate. There was no reliable evidence that the semantic tasks differed (relate: *M* = 96.30, *SD* = 3.36 and animacy: *M* = 97.26, *SD* = 1.82, SD = 3.36, BF_{10} = 1.089). We explored the interaction using Bayesian T-Tests to examine the concreteness effect within each suppression task. There was strong evidence for a concreteness effect with articulatory suppression (BF_{10} = 17.032) and this concreteness advantage was reduced in the remaining conditions, with evidence ranging from anecdotal evidence (control: BF_{10} = 1.39) to no evidence (relatedness: BF_{10} = 0.54 and animacy: BF_{10} = 0.31). This means that concrete sentences had a marked advantage over abstract sentences in the articulatory suppression condition only and this concreteness advantage was absent with a semantic load. We then explored the interaction by analyzing concrete and abstract sentences separately. For concrete sentences, recall was better in all conditions when compared to articulatory suppression, with BF_{10} ranging from 2.32 to 1.17e+4. There was also anecdotal evidence for better recall in the animacy and control conditions when compared to relatedness judgements ($BF_{10} > 1.018$, both cases). The animacy and control tasks did not differ ($BF_{10} = 0.28$). For abstract sentences, recall was the most impaired under articulatory suppression ($BF_{10} > 3505.97$, all cases) and best in the control condition ($BF_{10} > 17.67$, all cases), with the semantic tasks intermediate and no evidence for a difference ($BF_{10} = 0.32$).

Error score. We examined the occurrence of each error type using separate 2 (sentence) x 4 (suppression) repeated-measures Bayesian ANOVAs (Figure S4.4A). Grammatical and phonological substitutions were rare in all conditions, and in fact, results supported the null model ($BF_{10} < 1$, both cases).

For both order changes and semantic substitutions, substantial evidence supported the model including the effect of suppression only ($BF_{10} = 30.29$ and $BF_{10} = 5.59e+5$, respectively) over the second best model including both main effects. For order changes, the control condition resulted in the least amount of order changes ($BF_{10} > 8.59$, all cases); remaining comparisons did not differ ($BF_{10} < 0.39$, all cases). For semantic substitutions, articulatory suppression showed the most semantic errors ($BF_{10} > 63.70$, all cases); remaining comparisons did not differ ($BF < 0.33$. all cases).

For unrelated additions, substantial evidence supported the model with both main effects ($BF_{10} = 4.84e+9$) over the full model (4.84e+9) by a factor of 8.38. There were more additions for abstract than concrete sentences. The main effect of suppression was explored using Bayesian T-Tests. Articulatory suppression was found to result in the most additions compared to other conditions ($BF_{10} > 362.17$, all cases) as well as semantic relatedness resulting in more additions than the control task ($BF_{10} = 11.23$). There were no other differences ($BF_{10} < .69$, all cases).

Finally, for omissions, very strong evidence supported the full model ($BF_{10} =$ 9.44e+24) over the second best model including both main effects ($BF_{10} =$ 1.50e+23) by a factor of 62.99. Abstract sentences resulted in more omissions than concrete sentences. The pattern of the main effect of suppression task was as follows. Articulatory suppression showed the most omissions ($BF_{10} >$ 4.96e+5, all cases), while omissions were uncommon in the control condition ($BF_{10} > 12.38$, all cases), with both semantic tasks intermediate. There were more omissions with relatedness than animacy judgements ($BF_{10} = 13.86$). The interaction was explored using Bayesian T-Tests, showing that abstract sentences led to more omissions than concrete sentences only with articulatory suppression ($BF_{10} = 5727.36$) and animacy judgements ($BF_{10} = 4.023$).

Long-term performance.

Gist accuracy. A Bayesian ANOVA was performed on cosine similarity and results revealed that the best model that accounted for the data included the concreteness effect only ($BF_{10} = 86.98$), preferred by a factor of 4.081 over the second best model containing both main effects. Concrete sentences ($M = 80.89$, $SD = 10.32$) were remembered better than abstract sentences ($M = 77.086$, $SD = 15.38$).

A similar Bayesian ANOVA was performed on conditional cosine similarity. The model with the highest BF was also the model with the concreteness effect ($BF_{10} =$ 139.94) and was 7.85 times more likely than the second best model including both main effects (BF_{10} = 17.84). Concrete sentences (*M* = 83.72, *SD* = 10.47) were remembered better than abstract sentences (*M* = 79.65, *SD* = 15.94). See Figure S4.2B (lower panels).

Error score. Separate Bayesian ANOVAs were conducted for each error category for long-term performance (Figure S4.4B). Order changes, phonological substitutions, and semantic substitutions were not affected by the experimental manipulations and results indicated evidence for a null effect (BF_{10} < 0.49, both cases). For grammatical substitutions, there was ambiguous evidence in favour of the model with suppression task only (BF_{10} = 1.31) but an analysis of specific effect revealed no evidence for an effect of suppression (BF_{incl} = 0.89).

For additions and omissions, the best model included the main effect of sentence only (BF_{10} > 3.33e+7, both cases), substantially preferred over the second best model. Abstract sentences led to more errors of additions and omissions than concrete sentences.

Table S4.4. Cosine similarity in percentage with word embedding method using the GloVe model (and standard deviations) across the suppression conditions for immediate and long-term recall in Experiment 2.

	AS	Relate	Animacy	Control
Immediate gist:				
Concrete	93.98 (2.96)	95.74 (4.43)	97.41 (1.79)	97.67 (1.43)
Abstract	90.92 (4.97)	96.86 (1.70)	97.12 (1.87)	98.33 (1.63)
Long-term gist:				
Concrete	78.20 (11.70)	81.28 (11.29)	81.93 (10.40)	82.15 (7.65)
Abstract	76.36 (14.50)	77.79 (14.90)	78.89 (15.25)	75.30 (17.58)
Long-term conditional gist:				
Concrete	82.69 (12.61)	84.22 (11.16)	84.014 (10.57)	83.96 (7.62)
Abstract	81.16 (15.52)	80.036 (15.39)	81.061 (15.62)	76.33 (17.78)

Note: AS = articulatory suppression.

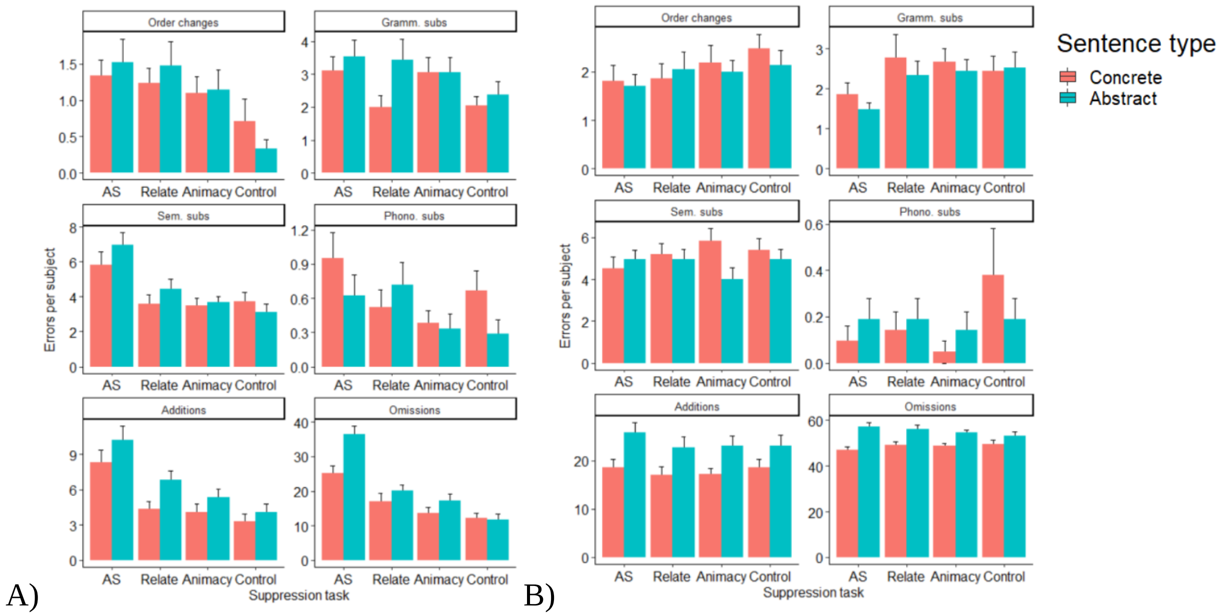

Figure S4.4. The panels show the average number of errors made by each participant across conditions in A) immediate recall and B) long-term recall in Experiment 2. AS = articulatory suppression. Error bars are the standard error of the mean.

4.6.2.3 **Discussion**

Gist recall was again at a level above chance for both immediate (M > 90%) and long-term recall (M > 75%) suggesting that participants were maintaining the overall gist of the sentence during recall. Immediate recall provided support for the full model, similar to the results reported in the main text. Concrete sentences were recalled better than abstract sentences. There was a clear effect of suppression task in that recall was best for the control condition, intermediate for both semantic tasks (which did not differ), and worst for articulatory suppression. The crucial interaction revealed that the concreteness effect was large and robust for articulatory suppression only, minimized under the control condition, and completely absent with both semantic suppression tasks. With regards to long-term recall, recall was better for concrete than abstract sentences.

Similar to Experiment 1 but unlike verbatim recall, long-term gist recall was unaffected by the suppression task.

The pattern of errors was similar to the observations of Experiment 1. Abstract sentences showed more semantic substitutions, omissions, and additions than concrete sentences in immediate recall. In short-term recall only, control trials showed the least amount of order changes and omissions whereas articulatory suppression showed the most semantic errors, omissions, and additions. We again found that words were more likely to be replaced with synonyms under articulatory suppression compared to the other conditions. This could reflect the tendency for participants to try to maintain the gist of sentences, which led them to make more semantic substitutions. Subsequently, this type of semantic engagement could help explain better long-term verbatim recall for sentences previously processing under articulatory suppression. For long-term recall, abstract sentences resulted in more omissions and additions than concrete sentences.

4.6.3 Supplementary References

Clark, M. (2018, September 9). *An Introduction to Text Processing and Analysis with R*. GitHub. https://m-clark.github.io/text-analysis-with-R/word-embeddings.html

Pennington J., Socher R. and Manning C.D. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-2014).* Association for Computational Linguistics

Sieg, A. (2018, July 4). *Text Similarities: Estimate the degree of similarity between two texts.* Medium. https://medium.com/@adriensieg/text-similarities-da019229c894

Chapter 5

5 Conclusion

The close connection between working memory and language has made it difficult to draw a clear distinction between these two systems. One of the possible reasons could be attributed to the lack of theoretical consensus on the construct of verbal working memory itself. Different theoretical accounts have conceptualized verbal working memory either as a separable cognitive resource (Baddeley & Hitch, 1974) or as an emergent property of the language (Majerus, 2013; Schwering & MacDonald, 2020). This thesis examined both of these conceptualizations of verbal working memory with the aim of developing a better understanding of verbal working memory overall. First, recognizing that many verbal tasks tap working memory and linguistic abilities, I examined whether these abilities could be teased apart in the same language task. Second, the relationships between verbal working memory and linguistic knowledge (phonological and semantic representations specifically) were explored in detail using different verbal tasks. This chapter provides a summary of the main findings and key implications from Chapters 2, 3, and 4, and makes recommendations for future research.

5.1 Relevant Findings

5.1.1 Working memory and linguistic contributions to language performance

There is a growing recognition that both working memory skills and language knowledge, stored in long-term memory, facilitate language processing. However, differentiating working memory abilities from language knowledge is difficult as both may be utilized in any language-based task. The interrelationships of these cognitive constructs were supported by studies demonstrating their interdependence (Mainela-Arnold & Evans, 2005; Marton & Schwartz, 2003). Nevertheless, there is currently no clinical tool that provides separable estimates of working memory and language skills. Therefore, I was interested in addressing this challenge by devising a task that has the potential to disentangle the contributions of verbal working memory and linguistic knowledge to language performance.

Chapter 2 evaluated the suitability of a modified Token Test as one potential tool to capture relative verbal working memory and language knowledge abilities employed in sentence processing. The Token Test is structured such that verbal directions increase in length but are syntactically identical in the initial parts, whereas verbal directions are just as long but linguistically complex (different sentence structure and new vocabulary) in the last part. I found that differential performance across items within the Modified-Shortened Token Test revealed individual differences in verbal working memory and language skills. In particular, long and simple sentences tapped working memory only, whereas long and linguistically complex sentences were more sensitive to linguistic demands. Moreover, unique relationships emerged between these identified factors and other language measures. Specifically, the working memory factor of the Token Test was correlated with languages tasks (e.g., Recalling Sentences) that may have memory demands related to those evaluated by the Token Test, whereas the linguistic factor of the Token Test was associated with tasks (e.g., Concepts and Following Directions) that might have had linguistic characteristics captured by the Token Test. However, this distinction was far from clear cut as Recalling Sentences was correlated with the linguistic component and not working memory in Experiment 3. Although these results are preliminary and determining the precise contribution of each construct to language performance remains a challenge, these finding are encouraging inasmuch as they suggest that some language-based tasks tapped verbal working memory and others, language knowledge.

One clear implication from these findings is that language and memory demands are inherent in language assessment tasks yet can independently contribute to language performance. The findings in Chapter 2 are consistent with prior work demonstrating that measures used in language batteries appear to be memory dependent at least to some extent (Mainela-Arnold & Evans, 2005; Archibald, 2013). The ability to understand and carry out verbal directions, such as "Before you touch the red circle, touch the green square", involves working memory in a linguistic task. Successful performance relies on the child transforming the order of instruction because of the adverbial clause *'before you touch…'* (perform second instruction first, then first instruction), in addition to storing the items in order (4 items: red, circle, green, square) during oral presentation. Despite the intertwining nature of verbal tasks, Chapter 2 is the first study of its kind, showing that the relative contributions of working memory and language could be disentangled using one language tool in children. This is also consistent with prior research using multiple measures to show that working memory and language are separable domains (e.g., Archibald, 2013; Archibald & Joanisse, 2009, Kapantzoglou et al., 2015). Performance on the modified Token Test could be used to capture individual differences in working memory and language skills. It could be that the independent influence of working memory and language emerged in my study because the participants were young children (ages 4 to 7) whose language skills are still developing. In contrast, these factors could become less distinct with developed language facilities. During early stages of development, language processes likely place fairly high demands on working memory. The linguistic demands of complex sentences might be high for young children with less well-developed language skills but a trivial load for adults who are sophisticated language users. In the Token Test, for example, adults would likely not process even the two-part syntactically complex instruction with much difficulty. Instead, standard word order and familiar words could be retained and processed with relative ease and automaticity. Further work is needed to examine age-related differences in working memory and language underlying language performance.

Clinically, the modified Token Test designed in Chapter 2 has the potential to be used as a tool to assess working memory and linguistic skills. Although there are limited implications for immediate practice, providing a novel perspective on the clinical utility of the Token Test is the first step towards assisting speech-language pathologists in understanding the extent to which working memory and linguistic factors influence linguistic behaviours. Verbal directions in the Token Test are similar to instructions children hear in school and at home, “Before you eat, wash your hands”. If a child has difficulties with understanding and carrying out this command, it could be due to a number of factors. The language load may be high given the need to understand various linguistic rules and meaning. Syntactic knowledge would be required to understand that the adverbial clause *“before you eat”* modifies the verb *“wash”* by describing the condition when the action occurs. Semantic knowledge is also important for deciphering the meaning of different words used to process different sentences in real-life contexts. The memory load may be high because of the requirement to transform the word order to carry out this command. Further, children with language-related disorders, such as developmental language disorder (DLD), may face higher working memory loads just to retain basic verbal information in mind. Therefore, speech-language pathologists need to understand whether poor language performance could reflect an effect of impaired language, impaired working memory, or a complex combination of both. Future work is needed to determine whether the Modified-Shortened Token Test could be used as a valid tool for identifying working memory constraints that could be secondary to a language impairment itself.

5.1.2 Phonological and semantic contributions to language performance

The results from Chapter 2 suggest that working memory and language are separable domains, yet also speak to the integration of working memory and language skills. The contribution of phonological and semantic knowledge to verbal working memory performance demonstrates this close association between working memory and linguistic knowledge. However, prior work has focused on how verbal working memory performance is largely influenced by phonological representations and minimally by semantic knowledge (Baddeley, 1966). This has led to an assumption in verbal working memory research that phonological factors influence short-term memory, whereas semantic factors influence long-term memory (e.g., Baddeley, 1986; 2007). More recently, there is a growing body of literature demonstrating that semantic knowledge supports working memory performance. For example, recall is better for concrete words than abstract words (the concreteness effect) presented in lists (Romani et al., 2008; Walker & Hulme, 1999) or sentences (Meltzer et al., 2016, 2017). Language-based models of working memory have been proposed to explain the influence of language knowledge on verbal working memory performance (e.g., Majerus, 2013; N. Martin et al., 2004; R. C. Martin et al., 1999; Schwering & MacDonald, 2020). According to these integrative models, short-term maintenance of verbal information relies on distinct, but highly interconnected cognitive processes including phonological and semantic representations (Majerus, 2013; N. Martin et al., 2004). Therefore, Chapters 3 and 4 explored phonological, semantic, and other cognitive supports in the processing of words and sentences in working memory, respectively.

Chapter 3 investigated the retention of phonological and semantic information in verbal short-term memory. In this study, I developed a novel task to examine the mechanisms underlying verbal short-term memory by combining the probe recognition and running-span paradigms. Participants processed a list of words and then had to decide whether a probe word rhymed or was synonymous with any items on the list in Experiment 1. In Experiment 2, word lists were presented rapidly and ended unpredictably to prevent rehearsal strategies and redintegration processes. Across both studies, results revealed that synonym judgements were just as good – or even better than – rhyme judgements, with semantic processing supporting long-term retention. The immediacy of semantic activation and lack of phonological advantage provides evidence against redintegration. Instead, consistent with language-based models of verbal working memory, there was direct and rapid activation of semantic knowledge, in addition to phonological representations, as soon as a word was encountered.

Chapter 4 built on the findings reported in Chapter 3 by extending the investigation to the context of sentence recall and using a different approach to examine the close interaction between verbal working memory and linguistic long-term memory. According to the integrative framework for verbal working memory (Majerus, 2013, 2019), sentence repetition is supported by phonological, semantic, and attentional resources interactively. This leads to the idea that if one process is blocked via an interference task, the other (available) processes would be relied on to a greater extent to support performance. Across two experiments, various suppression tasks were designed to selectively require phonological, semantic, and attentional processing. To suppress phonological and attentional resources, articulatory suppression and finger-tapping, respectively, have been widely used in the field. Notably, to address the dearth of paradigms for interfering with semantic processing, I developed a novel method to suppress semantic processing. I found that participants shifted between phonological and semantic processes depending on which resource was available to support either immediate or long-term sentence recall. In particular, phonological processing and rehearsal had immediate benefits, whereas semantic engagement was beneficial to both immediate and delayed recall. Across both studies, semantic engagement was evident in two ways: (1) by shifting reliance to semantic processes when phonological representations were disrupted and, (2) relying on sentence concreteness. These findings highlight the integrative principles underlying sentence maintenance in working memory and moreover, the flexibility of the language system to adjust to task demands.

Taken together, the results of Chapters 3 and 4 have implications for theories of verbal working memory and language processing. Instead of making the distinction between phonological short-term and semantic long-term processes, the findings presented in Chapters 3 and 4 support the view of verbal short-term memory as operating more dynamically within the context of a complex linguistic system. In particular, I found that as soon as verbal information (words, sentences) needs to be maintained, verbal working memory is supported by the direct activation of language representations throughout encoding and maintenance, and without relying on rehearsal or redintegration processes. Chapter 4 further showed that verbal information is supported by different processes simultaneously, aligning with an interactive activation account of language maintenance. For example, when articulatory suppression disrupted phonological representations, participants relied on the semantic resources available to support sentence recall, leading to larger concreteness effect and better long-term conditional recall. Overall, these findings support the recent shift away from associating verbal working memory with a specific phonological short-term store toward viewing verbal working memory as operating within the context of a linguistic system.

Although there are no direct clinical implications, the language-motivated accounts of working memory could be extended to how we think about language learning. It has been suggested that language processing proceeds in a ‘good enough’ fashion (Ferreira et al., 2002) such that representations are not fully activated but sufficient for the task at hand without imposing unnecessarily large cognitive loads on the system. According to this account, limited language representations may be activated in language processing tasks. For instance, phonological representations are said to primarily support new word learning because of the need to maintain unfamiliar sound patterns (Baddeley et al., 1998; Gathercole & Baddeley, 1993; Gathercole, 2006), whereas recall of sentences is said to be accurate because of reliance on the representation of the meaning of the sentence rather than retaining the sentence as a string of words (Potter & Lombardi, 1990). However, the results of Chapters 3 and 4 point to a novel perspective. Verbal information – whether it be words or sentences – does have direct access to different representational levels (phonological, lexical, and semantic) within the linguistic system. Since linguistic behaviours reflect the interaction of distinct representations supporting phonology and semantics, language processing can readily and flexibly shift should interference or demands change. It would follow from this that explicit engagement of both phonological and semantic representations during language processing tasks could improve language learning. While phonological representations are important for retaining new verbal information such as novel words, phonological maintenance becomes more stable when novel word forms are associated with semantic representations. For example, Savill et al. (2017) showed that adults recalled new words better when they were linked to semantic knowledge compared to when only the phonological word form was available. Similarly, Benham and Goffman (2020) demonstrated that although children with DLD had difficulties acquiring new phonological word forms, having semantic knowledge associated with these word forms led to more stabilized phonological sequences. For sentence recall, instead of attributing accurate recall to either surface-level, phonological representations (Schweppe et al., 2015) or semantic engagement (Potter & Lombardi, 1990), more recent work is recognizing the role that both types of representations have on sentence processing (Schweppe et al., 2011; Polišenská et al., 2014; also Chapter 4 of this thesis). Therefore, speech-language pathologists who use sentence recall as part of a language assessment should consider how performance could reflect the relative strengths in phonology and semantics as well as memory, and look for parallels with other language assessment measures. Nevertheless, these interpretations are tentative and call for further investigation.

Finally, Chapters 3 and 4 indicate that semantic processing during language processing more generally has long-term effects. This was even the case in Chapter 3; semantic engagement led to better long-term retention, despite initial similar levels of phonological and semantic activation. It would follow from this that phonological processing including rehearsal supports repetition and immediate performance but is a poor strategy for promoting long-term learning, whereas strategies that engage learners with semantic knowledge and meaning promotes long-term retention. Thus, efforts should be made to invoke semantic processing even if it might impose an initial cognitive load. As well, as we begin to understand how linguistic behaviours manifest immediately and in the long-term, the extent to which incorporating delayed recall tasks into practice, which is not generally done clinically, could be evaluated.

5.2 Verbal working memory and linguistic knowledge: Separable yet interacting systems

At first glance, the findings across the three experimental chapters might appear to contradict each other. Whereas Chapter 2 points to a separability between verbal working memory and language in children, Chapters 3 and 4 highlight the close relationship between verbal working memory and language processing in adults. However, I propose that verbal working memory and linguistic knowledge depend on separable yet interacting systems, allowing the relationship between verbal working memory and language to change over development or depending on task demands. First, verbal working memory and linguistic knowledge might be more readily separable in children with developing language skills. Although many language-based tasks employ both working memory and language skills, I found that performance could be differentiated depending on whether verbal items posed a relatively higher memory or language demand. At the same time, I also know from prior work that many verbal tasks require the integration of both working memory and language skills (e.g., Archibald, 2013; Anderson, 2011). Verbal tasks in Chapter 2 might not have shown this integration because partial correlations could have mitigated the effects of overlapping working memory and linguistic contributions to highlight specific links. Nonetheless, connecting my findings to those of prior work leads me to speculate that when language skills are developing, performance on verbal tasks requires support from both verbal working memory and language, separately to some extent, but when language skills are more developed, performance on verbal tasks depends directly on stable long-term representations rather than working memory alone. This latter supposition is akin to theories that view working memory as an activated subset of long-term memory but maintains a distinction between both systems (Cowan, 1999; Montgomery et al., 2021). Empirical studies also support the idea that linguistic knowledge strongly influences verbal working memory performance in adults (e.g., Acheson et al., 2011; Poirier et al., 2015; Kowialiewski & Majerus, 2018a, 2018b; also Chapters 3 and 4 of this thesis). Therefore, as language facility grows, verbal tasks place low demands on working memory functioning, making this distinction harder. Future studies should systematically evaluate the nature of this relationship across development.

Second, due to the interactive nature of these components, the nature of measurements or task demands can play a major role in highlighting the separability or connections between working memory and language skills. Chapter 2 highlighted this separability. Children's differential performance across items on the Token Test likely captured differences in verbal working memory and language knowledge, and in turn, these two constructs were differentially linked to specific language subtests. In contrast, the connection between verbal working memory and language processing were highlighted in Chapters 3 and 4 with respect to interactivity between semantic and phonological systems. Chapter 3 demonstrated that phonological and semantic representations were rapidly accessed in short-term memory, and that semantic activation was not the result of redintegration processes. Chapter 4 found that sentence recall taps phonological, semantic, and attentional processes jointly, but each process uniquely contributed to short- and long-term memory. It would follow from these findings, that relative weakness in one domain, should lead to greater engagement in the other domain. An extension from the Chapter 2 findings would be the speculation that sentence recall imposes a memory load for children with good language abilities, but taps language to a greater extent when language abilities are poor. And, although findings in Chapter 4 are specific to linguistic processes, results revealed that when task conditions made it difficult to use phonological representations to support sentence recall, processing shifted to using available semantic representations, corroborating the findings of recent work (Meltzer et al., 2016, 2017; Nishiyama, 2020). Taken together, the suggestion that verbal working memory and language are separable domains, but highly interconnected are not mutually exclusive accounts. The language system seems to be highly flexible, relying on various cognitive mechanisms differently depending on the task and circumstances.

Future work aimed at understanding verbal working memory as emergent from the language system will be a fruitful area of research.

5.3 Directions for Future Research

Recent theories and evidence have challenged our understanding about the nature of the relationship between verbal working memory, long-term linguistic knowledge, and language processing, making this an interesting area for future work. Across the three chapters, I modified measures or developed relatively novel techniques to evaluate verbal working memory, its separability and its relationship with linguistic knowledge. To determine if working memory and language could be separated, I refined the Token Test and used this tool to measure performance in Chapter 2, whereas previous work have used a battery of measures. To examine the different mechanisms underlying verbal working memory, I used a combined probe recognition and running-span paradigm in Chapter 3 and novel semantic suppression tasks in Chapter 4, whereas the majority of studies in this area have used the serial recall paradigm, with little work engaging semantics specifically. Although the results presented in this thesis are promising, further studies should systematically compare these novel methods to more standard measurements to validate the measurement approach taken in this thesis. This would help address the replicability of these results and the applicability of these findings to theoretical accounts of verbal working memory and language. Additionally, the feasibility of the modified Token Test (Chapter 2) and delayed sentence recall (Chapter 4) as potential clinical tools should be considered in future clinical research.

Another intriguing area for future research is the relationship between working memory and language development in children. The language-based models of verbal working memory that framed this thesis have not been tested in children, despite the increasing number of studies in healthy adults and brain-injured patients. Understanding the basic science of language processing in a relatively stable and intact system, like in adults, has been helpful in understanding developing language and disordered processing. The work conducted in adults will lend insight into the possible neurocognitive mechanisms that enable the retention and analysis of language in children. For example, I predict that in the early stages of word learning, children may rely on phonological rather than semantic learning mechanisms to a greater extent given their relative lack of prior knowledge, but over time, become less reliant on phonological processes (e.g., Gathercole, 2006; Stoel-Gammon, 2011). This shift from phonology to semantic might happen even earlier, as recent work suggests that children as young as 2-years-old organize words into semantic networks to support new word learning, similar to adults (Borovsky et al., 2016). Children with DLD can also rely on semantic information to stabilize phonological word forms, despite a weakness in phonological representations in this population (Benham & Goffman, 2020). Future studies assessing the interactivity between phonological and semantic systems underlying verbal working memory in children will contribute to our understanding of the word learning process, and how to best support acquisition in typical and atypical development.

Finally, it is important for future work to consider domain-general attention and serial order processes, in addition to linguistic mechanisms underpinning language processing. Although all theories of verbal working memory consider the role of domain-general components, this area has received less research attention. I did find some indirect evidence, that of attention being minimally required to complete even the easy parts of the modified Token Test (Chapter 2) and short list lengths (Chapter 3), whereas Chapter 4 assessed sentence recall under an attentional load. Further, Chapter 3 speculated on the role of attention switching in the probe recognition – running-span task. Nevertheless, domain-general attentional and serial order processes are said to more directly contribute to the capacity to support verbal information in working memory, in addition to domain-specific linguistic knowledge (Majerus, 2013; Schwering & MacDonald, 2020). Domain-general attentional processes support both item and order information by keeping items active in the focus of attention and keeping information in the exact order. Moreover, serial order is an inherent part of language learning and processing. For example, learning a novel phonological word form or the ability to repeat non-words requires maintaining and repeating the correct sound sequence. In fact, some studies have shown that children and adults with better serial order maintenance capacities have larger vocabularies and learn new words faster (e.g., Leclercq & Majerus, 2010; Majerus & Boukebza, 2013; Majerus et al., 2008). Future studies will need to consider the multiple components of this integrative framework simultaneously in word learning paradigms and language studies. In particular, more work aimed at specifically understanding the role of serial order retention and attentional processes in supporting language maintenance is warranted.

5.4 Conclusion

The growing recognition that verbal tasks tap working memory and language skills interactively calls for a need to consider the influences of multiple representations on performance. This thesis provides a first step in discerning the relationship between verbal working memory, long-term linguistic knowledge, and language processing.

Results presented in this dissertation align with the view that the language network is a highly flexible system, supported by distinct but interconnected linguistic (phonological, semantic, and syntactic knowledge) and domain-general mechanisms (attention and serial order processes) operating in both the short and long term. Overall, the findings of this thesis point to the importance of understanding the specific role of various neurocognitive mechanisms underpinning language processing, which, in turn, can provide theoretical insights and inform clinical applications.

5.5 References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(1), 44-59. https://doi.org/10.1037/a0021205

Anderson, M. A. (2011). *Memory and language: Evidence of relationships from three studies of school-age children.* (Unpublished Doctoral thesis). Western Michigan University, Michigan, USA. Retrieved from https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=1343&context=dissertations

Archibald, L. M. D., & Joanisse, M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52(4),* 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Archibald, L. M. D. (2013). The language, working memory, and other cognitive demands of verbal tasks. *Topics in Language Disorders, 33(3),* 190–207. https://doi.org/10.1097/TLD.0b013e31829dd8af

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 647–667). Academic Press

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105(1),* 158–173. https://doi.org/10.1037/0033-295X.105.1.158

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

Baddeley, A. D. (1986). *Working memory.* Clarendon Press/Oxford University Press.

Benham, S., & Goffman, L. (2020). Lexical-semantic cues induce sound pattern stability in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63(12),* 4109–4126. https://doi.org/10.1044/2020_JSLHR-20-00244v

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child Development, 87(6),* 1893–1908. https://doi.org/10.1111/cdev.12554

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, UK: Cambridge University Press. http://dx.doi.org/10.1017/ CBO9781139174909.006

Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15. https://doi.org/10.1111/1467-8721.00158

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Erlbaum.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27(4),* 513–543. https://doi.org/10.1017/S0142716406060383

Kapantzoglou, M., Restrepo, M. A., Gray, S., & Thompson, M. S. (2015). Language Ability Groups in Bilingual Children: A Latent Profile Analysis. *Journal of Speech, Language, and Hearing Research, 58(5),* 1549–1562. https://doi.org/10.1044/2015_JSLHR-L-14-0290

Kowialiewski, B., & Majerus, S. (2018a). Testing the redintegration hypothesis by a single probe recognition paradigm. *Memory, 26(9),* 1256-1264. https://doi.org/10.1080/09658211.2018.1448420

Kowialiewski, B., & Majerus, S. (2018b). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language, 101*, 64-83. https://doi.org/10.1016/j.jml.2018.03.005

Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology, 46(2),* 417–427. https://doi.org/10.1037/a0018540

Mainela-Arnold, E., & Evans, J. (2005). Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research, 48,* 897–909. https://doi.org/10.1044/1092-4388(2005/062)

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: New evidence from a novel word learning paradigm. *Journal of Experimental Child Psychology, 116(4),* 811–828. https://doi.org/10.1016/j.jecp.2013.07.014

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107(2),* 395–419. https://doi.org/10.1016/j.cognition.2007.10.003

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7,* 357. https://doi.org/10.3389/fnhum.2013.00357

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112,* 122–133. https://doi.org/10.1016/j.cortex.2018.04.016

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46(5),* 1138–1153. https://doi.org/10.1044/1092-4388(2003/089)

Meltzer, J. A., Kielar, A., Panamsky, L., Links, K. A., Deschamps, T., & Leigh, R. C. (2017). Electrophysiological signatures of phonological and semantic maintenance in sentence repetition. *NeuroImage, 156,* 302–314. https://doi.org/10.1016/j.neuroimage.2017.05.030

Meltzer, J. A., Rose, N. S., Deschamps, T., Leigh, R. C., Panamsky, L., Silberberg, A., Madani, N., Links, K. A. (2016). Semantic and phonological contributions to short- term repetition and long-term cued sentence recall. *Memory & Cognition. 44*, 307–329. https://doi.org/10.3758/s13421-015-0554-y

Montgomery, J. W., Gillam, R. B., & Evans, J. L. (2021). A new memory perspective on the sentence comprehension deficits of school-Age children with developmental language disorder: Implications for theory, assessment, and intervention.

Language, Speech & Hearing Services in Schools, 52(2), 449–466. https://doi.org/10.1044/2021_LSHSS-20-00128

Martin, N., Fink, R., & Laine, M. (2004). Treatment of word retrieval with contextual priming. *Aphasiology, 18,* 457-471. https://doi.org/10.1080/02687030444000129

Nishiyama, R. (2020). Adaptive use of semantic representations and phonological representations in verbal memory maintenance. *Journal of Memory and Language, 111,* 104084. https://doi.org/10.1016/j.jml.2019.104084

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43(3),* 489-499. https://doi.org/10.3758/s13421-014-0470-6

Polišenská, K., Chiat, S., Comer, A., & McKenzie, K. (2014). Semantic effects in sentence recall: The contribution of immediate vs delayed recall in language assessment. *Journal of Communication Disorders, 52,* 65–77. https://doi.org/10.1016/j.jcomdis.2014.08.002

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29,* 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Martin, R. C., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41(1),* 3–29. https://doi.org/10.1006/jmla.1999.2637

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61(2),* 292-323. https://doi.org/10.1080/17470210601147747

Savill, N., Ellis, A. W., & Jefferies, E. (2017). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia, 98,* 85–97. https://doi.org/10.1016/j.neuropsychologia.2016.03.006

Schweppe, J., Rummer, R., Bormann, T., & Martin, R. C. (2011). Semantic and phonological information in sentence recall: Converging psycholinguistic and neuropsychological evidence. *Cognitive Neuropsychology, 28(8),* 521–545. https://doi.org/10.1080/02643294.2012.689759

Schweppe, J., Barth, S., Ketzer-Nöltge, A., & Rummer, R. (2015). Does verbatim sentence recall underestimate the language competence of near-native speakers? *Frontiers in Psychology, 6,* 63–63. https://doi.org/10.3389/fpsyg.2015.00063

Schwering, S. C. & Macdonald. M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14,* 68. https://doi.org/10.3389/fnhum.2020.00068

Stoel-Gammon C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language, 38(1),* 1–34. https://doi.org/10.1017/S0305000910000425

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5),* 1256-1271. https://doi.org/10.1037/0278-7393.25.5.1256

6 Appendices

Appendix A. Examples of the verbal directions used across the different parts of the original and Modified-Shortened Token Test.

Subtest	Example of command in subtest
1 (all tokens)	Touch a circle
2 (large only)	Touch a yellow square
3 (all tokens)	Touch a small white circle
4 (large only)	Touch the red circle and the green square
5 (all tokens)	Touch the large white circle and the small green square
6 (large only)	Touch the black circle with the red square
	If there is a blue circle, touch the red square
	Touch all the circles except the green one

Appendix B. Open Practices Statement

Chapter 3:

Experiment 1 was not pre-registered

Experiment 2 was pre-registered: https://osf.io/ms7k3

The data and analysis code for both experiments are available on the Open Science Framework: https://osf.io/zye6a

Chapter 4:

Experiment 1 was pre-registered: https://osf.io/at8re

Experiment 2 was not pre-registered

The experiment materials (sentences, picture stimuli), data, and the analyses scripts for both experiments are available via the Open Science Framework at https://osf.io/5a2p6

Example of Experiment 2 procedure can be accessed at https://osf.io/kdf95

Appendix C. Copyright permission for Chapter 3

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 16, 2021

This Agreement between Ms. Theresa Pham ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5150861235680
License date	Sep 16, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Memory & Cognition
Licensed Content Title	The role of phonological and semantic representations in verbal short-term memory and delayed retention
Licensed Content Author	Theresa Pham et al
Licensed Content Date	Aug 2, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	full article/chapter
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	The role of working memory and linguistic knowledge on language performance
Institution name	University of Western Ontario
Expected presentation date	Nov 2021
Order reference number	103758
Requestor Location	Ms. Theresa Pham [redacted] Attn: Ms. Theresa Pham
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material

Appendix D. Ethics approval for the studies described in Chapter 2

Experiment 1:

Western Research

Date: 1 November 2019

To: Dr. Lisa Archibald

Project ID: 112343

Study Title: Language Based Literacy Intervention: Program Evaluation

Application Type: NMREB Amendment Form

Review Type: Delegated

Full Board Reporting Date: 06/Dec/2019

Date Approval Issued: 01/Nov/2019 10:14

REB Approval Expiry Date: 20/Aug/2020

Dear Dr. Lisa Archibald,

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
RevisedTT_October2019	Paper Survey	15/Sep/2019	

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted], Research Ethics Officer on behalf of [redacted], NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Experiment 2:

Western Research

Date: 24 September 2018

To: Lisa Archibald

Project ID: 109664

Study Title: A language and literacy kindergarten screening tool

Application Type: NMREB Amendment Form

Review Type: Delegated

Full Board Reporting Date: 05/Oct/2018

Date Approval Issued: 24/Sep/2018 16:37

REB Approval Expiry Date: 10/Oct/2019

Dear Lisa Archibald,

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
CELF_DataCollectionForm&Images	Other Data Collection Instruments	06/Jul/2018	1
EmailRecruitment_ExtensionStudy2018	Recruitment Materials	13/Aug/2018	1
InPersonRecruitment_ExtensionStudy2018	Recruitment Materials	10/Sep/2018	1
LOI_ExtensionStudy2018_CLEAN	Written Consent/Assent	13/Aug/2018	1
SharingContactInformationForm	Other Data Collection Instruments	10/Sep/2018	1
Shortened Token Test_DataCollection Form&Image	Other Data Collection Instruments	06/Jul/2018	1
TelephoneRecruitment_ExtensionStudy2018	Recruitment Materials	13/Aug/2018	1
TNL_DataCollectionForm&Images	Other Data Collection Instruments	06/Jul/2018	1
Updated Narrative Retell Measure	Other Data Collection Instruments	18/Sep/2018	1
UWO Cover Letter_Amendment Clean	Written Consent/Assent	06/Jul/2018	1
WesternProtocol_ExtensionStudy2018_CLEAN	Protocol	10/Sep/2018	1
Word Associations Subtest_CELF-4	Other Data Collection Instruments	18/Sep/2018	1
WRAML2_DataColletionForm&Images	Other Data Collection Instruments	06/Jul/2018	1

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted] Research Ethics Officer on behalf of [redacted] NMREB Vice-Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Experiment 3:

Date: 23 July 2018

To: Lisa Archibald

Project ID: 106896

Study Title: Wise Words - A word building project

Application Type: NMREB Amendment Form

Review Type: Delegated

Full Board Reporting Date: August 3 2018

Date Approval Issued: 23/Jul/2018

REB Approval Expiry Date: 27/Jul/2019

Dear Lisa Archibald,

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
10157_WiseWords_RevisionSept2016_July2018	Protocol	16/Jul/2018	
LOI_WiseWords_July2018	Written Consent/Assent	06/Jul/2018	2.0
MA Assessment Task Revised	Other Data Collection Instruments		
ReadingFluencyAssessment	Other Data Collection Instruments		
RecruitmentLetter_RevJuly2018	Recruiting Advertisements	06/Jul/2018	1
ShortenedTokenTest	Other Data Collection Instruments		
SpellingDefn	Other Data Collection Instruments		

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted] Research Ethics Officer on behalf of [redacted] NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix E. Ethics approval for the studies described in Chapters 3 and 4

Date: 14 December 2017

To: Dr. Lisa Archibald

Project ID: 110440

Study Title: The effects of linguistic knowledge and memory systems on language learning

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 12/Jan/2018

Date Approval Issued: 14/Dec/2017 16:57

REB Approval Expiry Date: 14/Dec/2018

Dear Dr. Lisa Archibald

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
consent-ethics_clean	Written Consent/Assent	04/Dec/2017	
debrief-ethics	Debriefing document	19/Oct/2017	1
demongrahics-language-survey_clean	Online Survey	30/Nov/2017	
email-survey	Online Survey	30/Nov/2017	
sampleitems-ethics	Other Data Collection Instruments		
sona-ethics_clean	Recruitment Materials	04/Dec/2017	

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted] Research Ethics Officer on behalf of [redacted], NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Curriculum Vitae

Name: Theresa Ai Vy Pham

Education

2017-Present — Ph.D. Health and Rehabilitation Sciences (in progress)
University of Western Ontario
Advisor: Dr. Lisa M. D. Archibald
Dissertation title: *The role of working memory and linguistic knowledge on language performance*

2017-Present — M.Cl.Sc. Speech-Language Pathology (in progress)
University of Western Ontario

2012-2016 — B.Sc Psychology and Linguistics
University of Toronto

Academic Honours and Awards

2020 — Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Scholarships – Doctoral Award Nominee (not awarded)

2020 — Travel Award (CA$250), Society of Graduate Students, University of Western Ontario

2019-20 — Ontario Graduate Scholarship (CA$15 000), University of Western Ontario

2019 — Best Talk and Travel Award (CA$360), PsyLinCS-University of Toronto Mississauga

2019 — National Institute of Health Student Travel Award (US$700), Symposium on Research in Child Language Disorders

2018-19 — Faculty of Health Science Travel Award, University of Western Ontario (awarded annually)

2017-18 — Social Sciences and Humanities Research Council (SSHRC) – Canada Graduate Scholarships (CA$17 500), University of Western Ontario

2017-18 — Ontario Graduate Scholarship (CA$15 000), University of Western Ontario (declined for SSHRC)

2017-18	Ontario Graduate Scholarship (CA$15 000), University of Toronto (declined for SSHRC)
2017	The Henry Rogers Memorial Scholarship Fund (CA$1390), Department of Linguistics, University of Toronto
2017	The College Silver Medals, University of Toronto
2013-15	St Michael's College In-Course Scholarship (CA$1500), University of Toronto (awarded annually)

Peer-Reviewed Publications

Pham, T., Bardell, T. E., Vollebregt, M., Kuiack. A. K., & Archibald., L. M. D. (Accepted). Evaluating the Modified-Shortened Token Test as a working memory and language assessment tool. *Journal of Speech, Language, and Hearing Research*

Pham, T., & Archibald, L. (2021). The role of phonological and semantic representations in verbal short-term memory and delayed retention. *Memory & Cognition*. Advance online publication. https://doi.org/10.3758/s13421-021-01216-8

Pham, T., & Buchsbaum, D. (2020). Children's use of majority information is influenced by pragmatic inferences and task domain. *Developmental Psychology, 56(2),* 312–323. https://doi.org/10.1037/dev0000857

Pham, T., Kang, J., Johnson, A., & Archibald., L. M. D. (2020). Feature-focusing constraints on implicit learning of function word and meaning associations. *Applied Psycholinguistics, 41(2),* 401-426. https://doi.org/10.1017/S0142716420000041

Pham, T., Hu, J. C., & Buchsbaum, D. (2017). Pragmatics influence children's use of majority information. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2894-2899). Cognitive Science Society.

Conference Presentations

Talks

Pham, T., & Archibald., L. M. D. (submitted, April 2022). *The modified Token Test: A new way to assess working memory and linguistic skills in young children.* 2022 SAC Speech-Language Pathology Conference. Speech-Language & Audiology Canada (SAC), Vancouver, BC, Canada.

Pham, T., & Archibald., L. M. D. (virtual, June 2021). *Digit Span and Test of Integrated Language and Literacy Skills: Understanding why children score higher on the digit span backward than forward task.* Theory to Therapy: SLP & Audiology Graduate Student Conference. University of Toronto, Toronto, ON, Canada.

Pham, T., Flannery, K., & Archibald., L. (September 2019). *Disruption of verbal short-term memory: Evidence for phonological and semantic levels of representation.* PsyLinCS-UTM. Toronto, ON. **Winner of best talk.**

Pham, T., Hu, J. C., & Buchsbaum, D. (May 2018*). Selective Social Learning: It's All About the Context.* Developmental 2018. Brock University, Niagara, ON, Canada.

Posters

Pham, T., Malik, A., Varagunan, N., & Archibald., L. M. D. (virtual, November 2021). *Immediate and long-term sentence recall: The benefits of blocking rehearsal and concrete sentences.* Psychonomic Society 62nd Annual Meeting. San Diego, CA, USA

Pham, T., Malik, A., Varagunan, N., & Archibald., L. M. D. (virtual, June 2021). *How available resources and sentence concreteness differentially support immediate and long-term sentence recall*. Canadian Society for Brain, Behaviour and Cognitive Science (CSBBCS/SCSCCC). Concordia and McGill Universities, Montreal, QB, Canada

Pham, T., & Archibald., L. M. D. (virtual, June 2021). *Evaluating the Modified-Shortened Token Test as a working memory and language assessment tool*. Symposium on Research in Child Language Disorders. Madison, WI, USA.

Pham, T., Bardell, T. E., Vollebregt, M., Kuiack. A. K., & Archibald., L. M. D. (virtual, June 2021). *Are children performing better on the digit span backward than forward task? An exploratory analysis.* Symposium on Research in Child Language Disorders. Madison, WI, USA.

Pham, T., & Archibald, L. M. D. (July 2020). *Disruption of verbal short-term memory: Evidence for phonological and semantic levels of representation.* Cognitive Science Society. *Withdrew due to COVID-19 outbreak.*

Tandoc, M., Nadendla, B., Shen, C., Otsubo, K., **Pham, T.,** Finn, A. S. (October 2019). *Kids learn what they can't ignore: Developmental differences in the processing of distractors.* Cognitive Development Society. Louisville, KY.

Pham, T. & Archibald., L. M. D. (July 2019). *Phonological and semantic processing in short-term memory* Cognitive Science Society. Montreal, QB, Canada

Pham, T., & Archibald., L. M. D. (June 2019). *A language-based account for semantic encoding in verbal short-term memory.* Symposium on Research in Child Language Disorders. Madison, WI, USA.

Dubois, M., Hang, N., Jeganathan, L., Lim, D., **Pham, T.,** Finn, A. (April 2019). *Younger isn't better: Broader attention does not facilitate learning peripheral information.* Society for Research in Child Development Biennial Meeting, Baltimore, Maryland, USA. **Nominated for best poster.**

Pham, T., Kang, J., Johnson, A., & Archibald, L. (June 2018). *Implicit learning of semantic information depends on contextual cues*. Symposium on Research in Child Language Disorders. Madison, WI, USA.

Dubois, M., **Pham, T**., Lim, D., Finn, A. S. (March 2018) *Limited attention facilitates learning of peripheral information in children.* Cognitive Neuroscience Society, Boston, MA, USA.

Pham, T., Hu, J. C., & Buchsbaum, D. (July 2017). *Pragmatics influence children's use of majority information*. Cognitive Science Society. London, UK

Pham, T., Hu, J. C., & Buchsbaum, D. (April 2017). *Children's conformity: effects of Pragmatic knowledge on majority influence.* Society for Research in Child Development Biennial Meeting, Austin, Texas, U.S.A.

Pham, T., Hu, J. C., & Buchsbaum, D. (November 2016). *The influence of counterfactuals on children's conformity during causal reasoning.* Interdisciplinary Workshop on Counterfactual Reasoning, University of Toronto, Toronto, ON, Canada

Manuscript in prep

Pham, T. & Archibald., L. M. D. (In prep). Assessing the role of available resources and sentence concreteness on immediate and long-term sentence recall.

Professional Experience

2021	Student Clinician (Speech-Language Pathology), Fern Speech and Language Services
2020-Present	Digital Content Coordinator, Ontario Association for Families of Children with Communication Disorders
2020-21	Student Mentor for Wise Words/Leaps to Literacy (Speech-Language Pathology), University of Western Ontario
2020	Student Clinician (Speech-Language Pathology), Bluewater Health Hospital
2020	Teaching Assistant for Speech Science and Aural Rehabilitation, School of Communication Sciences and Disorders, University of Western Ontario
2019-20	Student Clinician (Speech-Language Pathology), Parkwood Institute
2019	Student Clinician (Speech-Language Pathology), H.A. Leeper Speech & Hearing Clinic
2016-17	Lab Manager to Dr. Amy Finn, Learning and Neural Development Lab, Department of Psychology, University of Toronto
2015-17	Lab Manager to Dr. Daphna Buchsbaum, Computational Cognitive Development Lab, Department of Psychology, University of Toronto

Professional Service

2018-Present	Conference reviewer, Cognitive Science Society
2021	Speaker, Second-Year Learning Communities, Department of Psychology, University of Toronto

2020	Ad-hoc reviewer, Autism & Developmental Language Impairments
2019-20	Peer Mentor for Peer Mentorship Program, Department of Psychology, University of Toronto
2019	Community Events Coordinator (student volunteer), Autism Ontario London Chapter