The nature of linguistic long-term memory effects in verbal working memory

Thèse présentée par

Benjamin Kowialiewski

En vue de l'obtention du titre de
Docteur en Psychologie et Sciences de l'Éducation

Sous la supervision de
Steve Majerus

Psychologie & Neuroscience Cognitives

PsyNCog

Psychology & Neuroscience of Cognition

Acknowledgement

Without the help and support of different persons and organisations, the achievement of this PhD thesis would not have been possible. I'm therefore deeply grateful to them.

First of all, I would like to thank my promoter, Prof. Steve Majerus, who offered me the opportunity to work with him since my Master thesis. Steve Majerus always provided the best supervision I could hope for. His input has been extremely valuable, was present when I had questions and doubts at any stage of my research, and always did so very quickly. Most importantly, because Steve Majerus gives us his trust, I always had a lot of freedom to work in my own way, which is a value I am particularly sensitive to. Thanks to his expertise and methodological rigor, I made more progress than I could ever have imagined.

I am also grateful to the members of my doctoral advisory committee, Prof. Fabienne Collette and Prof. Serge Brédart, for having followed the progress of my PhD research over the last four years.

I also want to thank Prof. Gaën Plancher and Prof. Marc Brysbaert for having accepted to examine this work and for being members of this PhD jury.

This work would have not been possible without the financial support of the Fund for Scientific Research – FNRS nor without the high quality logistical support of the University of Liège.

From a more personal perspective, I would like to thank all my colleagues of our research unit for contributing to a very exciting and stimulating work environment. I also thank the persons who accepted to proofread my manuscript: Coline, Lucie, and Marion. I further thank all the students for their important contribution to this work at the level of data collection, but also all the subjects for their participation to the experiments. I am also very grateful to Simon and Laurens. You were the first persons through which I had my first real research experience during my Master thesis, and it has been really stimulating to work under your supervision. I am glad we still have the opportunity to work together, and I hope we will continue to do so in the forthcoming years.

For their introductory course on a mathematical approach of human cognition, I thank Jacques and Daniel. This has been the most interesting class I ever followed at the university. Thanks to this class, I discovered completely new research perspectives. Importantly, you made mathematics accessible and understandable for me.

I am deeply grateful to my parents and family to accept me as I am. We have always been free to think in our own way, and this freedom has probably been a major

contributing factor to my personality. For their friendship, I thank Adam, Anthony, Guillaume and Maxime. You are the only persons we can laugh about free energy while still drinking beer and playing videogames. You are the smartest guys I know.

<div align="right">
Benjamin Kowialiewski

Liège, March 21, 2019
</div>

"It is, of course, the way of all things. You see, there is only one constant, one universal, it is the only real truth: causality. Action. Reaction. Cause and effect. [...] Beneath our poised appearance, the truth is we are completely out of control. Causality. There is no escape from it, we are forever slaves to it. Our only hope, our only peace is to understand it, to understand the why."

Lana and Lilly Wachowsky (2003)

Table of contents

General introduction

My most compelling recollection of a working memory (WM) situation dates back to when I was almost a teenager. My oldest brother was configuring our home local network in order to solve an internet-related issue we had at that time. I saw him getting more and more frustrated as he tried the simple task of remembering the nine-digit IP address of his computer. Nine digits seems almost like nothing, as compared to the tremendous amount of complex information we process and remember each day of our life. But was my brother really confronted to a simple task after all? All studies in the field of verbal WM would probably answer a straight "no" to this question. It appears that our ability to temporarily maintain verbal information over a short period of time is very limited. Initial sudies pointed to a working memory limitation of about 7±2 digits (G. A. Miller, 1956), but more recent studies rather suggest that the human cognitive system is able to actively hold only around 4 distinct units in mind (Cowan, 2001). Obviously, this is not what our intuition tells us, as we all think being able to maintain much more information at one time. And we are partially right. An important factor that will modulate our ability to maintain information over the short-term is related to our pre-existing knowledge about the information we need to maintain. For example, the digit sequences "0495" or "0479" will probably be very familiar for any citizen living in Belgium, because these prefixes are largely encountered among phone numbers. Accordingly, the maintenance of this type of digit sequence will appear to be easier than the sequence "7059".

Now imagine a different situation where you are ordering some aperitifs in an exquisite restaurant together with four colleagues. The most committed waiter will probably try to hold in mind not only the different beverages that you all have ordered, but also the person that ordered each drink. Furthermore, you could have ordered different drinks: beer, wine, soda, water... or similar drinks: blond beer, brown beer, black beer, amber beer… These two situations will have very different impacts on the waiter's ability to hold the ordered drinks and associated persons in WM. The first situation is likely to make the retention of the drinks more difficult than the second situation where all the drinks can be summarized under the semantic category 'beer' stored in long-term memory. These different examples show that we cannot understand the limits of our WM capacity without considering the long-term memory representations associated with the stimuli that have to be maintained in verbal WM.

The aim of this PhD thesis is to lead to a deeper understanding of the structure and functioning of WM by investigating the role of long-term memory knowledge in

short-term storage of verbal information. Although many studies have shown that long-term memory knowledge support verbal WM, it remains largely unknown where and how these interactions arise. In this thesis, we will try to answer the following questions: Are we able to access long-term memory knowledge in WM because we implement some form of verbal strategies or is verbal knowledge accessed in a direct and obligatory manner? Do the interactions between verbal WM and long-term knowledge reflect the fact that verbal WM is merely an activated part of the language system or do these interactions arise from post-memory reconstructive processes at the stage of recall? Does verbal knowledge interact with all aspects of WM, including the representation of the serial order in which memoranda have been presented, or does it only support the retention of the identity of memoranda? These questions will be answered through different experiments assessing both the cognitive and neural underpinnings of WM-long-term memory interactions.

Theoretical introduction

Chapter 1

Verbal working memory processes: a general overview

In this first chapter, we are going to briefly describe the main processes currently considered to support the short-term maintenance of information in verbal working memory (WM), before considering more specifically in Chapter 2 the role of linguistic knowledge which is the main focus of this PhD thesis. A first process we consider is subvocal articulatory rehearsal. This is one of the first processes that has been proposed in the WM literature and it is a major component of many older and recent theoretical accounts of verbal WM. The second process we will review is the control and focalization of attention, which is a further core process in many models of WM. The third aspect concerns the processes involved in the retention of serial order information and their distinction from processes involved in the retention of item information. As we will show, these different processes, despite their specificity, cannot be fully understood without considering their interactions with the linguistic system.

Subvocal articulatory rehearsal

One of the most well-known WM processes is the subvocal articulatory rehearsal process and its integration within the phonological loop model proposed by Alan Baddeley (Baddeley & Hitch, 1974; Baddeley, 1986). The phonological loop is composed of two components: a phonological store where phonological information is maintained, and a subvocal articulatory rehearsal process (see **Figure 1.1**). When information is visually presented, it needs to be recoded in phonological codes via the subvocal articulation process in order to be maintained within the phonological store. Auditorily presented items, on the other side, are supposed to gain obligatory access to the phonological store. Over time the phonological traces progressively decay, leading to forgetting in working memory. To counteract the deleterious impact of decay, information can be refreshed, by being reintroduced in the phonological store via the subvocal articulatory rehearsal process, the two components forming an articulatory loop. These components had been proposed on the basis of strong phonological effects observed in WM tasks. First, WM performance is known to decrease as a function of word length (Baddeley et al., 1984, 1975). That is, there is often
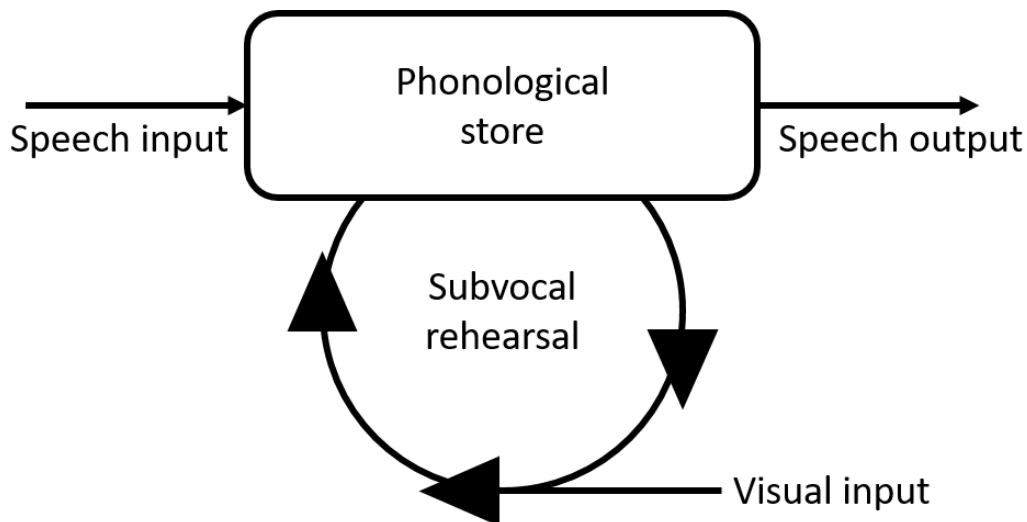
Figure 1.1. The phonological loop model by Baddeley and Hitch (1974).

a WM recall advantage for short, as compared to long words. The phonological loop model accounts for the word length effect by assuming that long words take more time to be rehearsed, leading to a stronger impact of decay over time for non-refreshed items. Additional evidence supporting the phonological loop construct stems from the observation that phonologically similar lists are associated with poorer recall performance as compared to dissimilar lists – the so-called phonological similarity effect (Baddeley, 1966; Levy, 1971). The phonological similarity effect is accounted for by assuming that due to overlapping WM representations and more confusable codes, phonologically similar words are more poorly recalled than dissimilar ones.

A further important finding is the reduction or disappearance of the phonological similarity effect under concurrent articulation. More specifically, the phonological similarity effect disappears when participants are required to repeatedly articulate an irrelevant sound, such as "ba ba ba ba" when encoding the items. Critically, this effect is observed for visually presented sequences, but not for auditorily presented sequences. The continuous repetition of irrelevant sounds leads to suppression of the articulatory rehearsal process, which on turn will prevent the recoding of visually presented memoranda in phonological codes, thereby reducing the impact of phonological similarity for visually presented information. Auditorily presented sequences, on the other side, are supposed to gain obligatory access to the store, even under articulatory suppression, leading to a preserved phonological similarity effect. The situation is a little bit different with the word length effect, which is considered to reflect the process of subvocal articulatory rehearsal, not the properties of the phonological store itself. This assumption is based on the observation of an abolished word length effect under concurrent articulation, regardless of modality of

presentation (Baddeley et al., 1984). Finally, articulatory suppression also leads to generally decreased recall performance in WM, as would be expected if memoranda cannot be refreshed.

However, recent (and less recent) evidence raises doubts about a causal role of subvocal articulatory rehearsal in WM maintenance. For instance, even under very fast encoding conditions leaving little room for phonological recoding, word length and phonological similarity effects can still be observed (Coltheart, 1999; Page & Norris, 1998). Furthermore, the deleterious effect of articulatory suppression has little cumulative effects over time; that is, if participants are required to repeatedly utter the same sound such as "the – the – the – the…", WM performance stays relatively stable if a longer interval is embedded between encoding and recall (Lewandowsky, Geiger, & Oberauer, 2008; Lewandowsky & Oberauer, 2015). Longer retention intervals should logically lead to a more dramatic impact of decay. Conversely, varying the uttered distractor (e.g. "Monday – Tuesday – Wednesday…") has a much stronger impact on WM recall performance. These results have been explained as showing that the articulatory suppression effect reflects an interference effect rather than the effect of memory decay (Gupta & MacWhinney, 1995).

The effects supporting the phonological loop framework also appear to be strategy-dependent. Studies have shown that when participants adopt a phonological coding strategy, phonological similarity and word length effects can be observed, even under concurrent articulation (Hanley & Bakopoulou, 2003; Logie, Della Sala, Laiacona, Chalmers, & Wyn, 1996). However, when participants are explicitly instructed to use a semantic coding strategy, these effects are abolished, or are strongly reduced (Campoy & Baddeley, 2008; Hanley & Bakopoulou, 2003). Furthermore, the test-rested reliability of these effects is relatively weak (Logie et al., 1996).

Most critically for the research question of this PhD thesis, a set of studies has also shown that these effects may at least be partially driven by interactions with linguistic knowledge stored in the language system. Indeed, most studies assessing the word length effect actually manipulated, unintentionnally, the impact of a specific psycholinguistic variable, namely neighborhood density (Jalbert, Neath, Bireta, & Surprenant, 2011). This variable refers to the number of phonological (lexical) neighbors associated with a target word. For instance, the word CAT has the phonological neighbors FAT, BAT, RAT, and MAT. Words with many neighbors are generally better recalled in WM tasks than words with a small number of neighbors, as we will see in more detail during Chapter 2. The problem with the word length effect is that short words typically have a larger number of neighbors in the linguistic knowledge base than long words: hence, when not explicitly controlling the

neighborhood dimension, the word length effect is actually confounded with the neighborhood density effect. When short and long words are equated for neighborhood density, the word length effect disappears. However, when stimuli are equated for word length, the neighborhood density effect is preserved (Guitard, Gabel, Saint-Aubin, Surprenant, & Neath, 2018; Jalbert, Neath, Bireta, et al., 2011; Jalbert, Neath, & Surprenant, 2011). These results have been recently replicated in a study controlling a large number of psycholinguistic dimensions (Guitard, Gabel, et al., 2018). Furthermore, under articulatory suppression, the neighborhood density effect disappears (Jalbert, Neath, & Surprenant, 2011), just as the word length does. These results clearly show the importance of considering the interactions with linguistic long-term memory knowledge when exploring verbal WM, as some of its most important benchmark effects can be traced down to these interactions.

Note that the phonological loop model has changed substantially over time, with the addition of a central executive (Baddeley & Logie, 1999) and an episodic buffer (Baddeley, 2000), and consideration of its interactions with the linguistic system (Baddeley, Gathercole, & Papagno, 1998) as we will see in Chapter 3.

Control and focalization of attention

Attention is a fundamental cognitive function allowing for the selection of currently relevant information in a WM situation or in any other task-related cognitive situation (Corbetta & Shulman, 2002). Two different types of attention are generally distinguished: controlled and automatic attention. The former refers to our capacity to intentionally select relevant information, while the latter refers to the automatic attentional capture provoked by – most of the time – unexpected external stimuli. Controlled attention is often studied using simple location tasks involving participants to perform spatial judgement over visually presented objects (e.g. left or right). When the location of an object is pre-cued by a briefly presented arrow (i.e. congruent trials), response times are enhanced when compared to trials where the arrow does not match the to-be-presented object (i.e. incongruent trials) (Shulman et al., 2008). This very simple example shows that the human cognitive system is able to select relevant information for an ongoing task. Accordingly, attention can be viewed as a system whose function is to prioritize relevant information. Automatic attention, on the other side, is often studied by presenting rare and unexpected stimuli during a task. In these conditions, participants automatically (unintentionally) redirect their attention toward the unexpected stimulus (Corbetta, Patel, & Shulman, 2008; G. L. Shulman et al., 2010). Controlled and automatic attention have shown to be two antagonist processes and recruiting two distinct neuroanatomic networks: controlled attention is supported by

the dorsal attentional network, composed of the frontal eye field and the intraparietal sulcus, while automatic attention is supported by the ventral attentional network, composed of the ventral frontal cortex and the temporal-parietal junction.

Attention and WM are currently considered to be two highly related constructs. Previous studies have shown a strong relationship between performance in WM and attentional controlled tasks, with high WM span individuals also performing at high levels during attentional tasks (Engle, 2002; Kane, Bleckley, Conway, & Engle, 2001). It has also been shown that items held in WM automatically bias attention towards information in visual search tasks (Dowd, Pearson, & Egner, 2015; Greene & Soto, 2014; Mallett & Lewis-Peacock, 2018). Recent studies have shown the involvement of the two major attentional systems also in WM tasks (Corbetta et al., 2008; Majerus et al., 2016; Majerus, Attout, et al., 2012; Wen, Yao, Liu, & Ding, 2012). Like in attentional tasks, the dorsal (controlled) and ventral (automatic) attentional networks appear to play antagonist roles during WM tasks, with increased involvement of the dorsal attentional network during high load WM conditions and resulting in the suppression of the ventral attention network. This suppression leads to inattentional blindness effects in WM: the higher the WM load, the lower the probability that an unexpected distractor stimulus is perceived during a WM task (Majerus, Attout, et al., 2012; Matsuyoshi et al., 2010; Todd et al., 2005).

A number of WM models include attention as a central process for WM maintenance. This is for example the case for the embedded-process model of WM proposed by Cowan (Cowan, 1995, 1999, 2001), illustrated in **Figure 1.2**. Critically, these models also include sensory and long-term knowledge bases as attention is considered to be deployed on representations activated in these bases. The embedded processes framework by Cowan considers that WM emerges from three main components: (a) latent knowledge stored in long-term memory, (b) the long-term memory representations that are temporarily activated and (c) the representations that are kept within the focus of attention with a capacity limit of 3 to 5 chunks (Cowan, 2001). Hence, for this theoretical framework, attention is an obligatory process in order to fully access and activate long-term memory knowledge, and this activation in long-term memory is considered to be the representational basis for WM maintenance. In this framework, activated long-term memory representations are supposed to progressively decay over time, and can be directly accessed and refreshed via the focus of attention, unless they are degraded up to a point that they can no longer be retrieved. This assumption is supported by recent multivariate neuroimaging studies showing that WM content can be reliably decoded in sensory-processing regions when items are actively maintained in the focus of attention. However, once attention is

Chunk

Focus of attention

Activated long-term memory
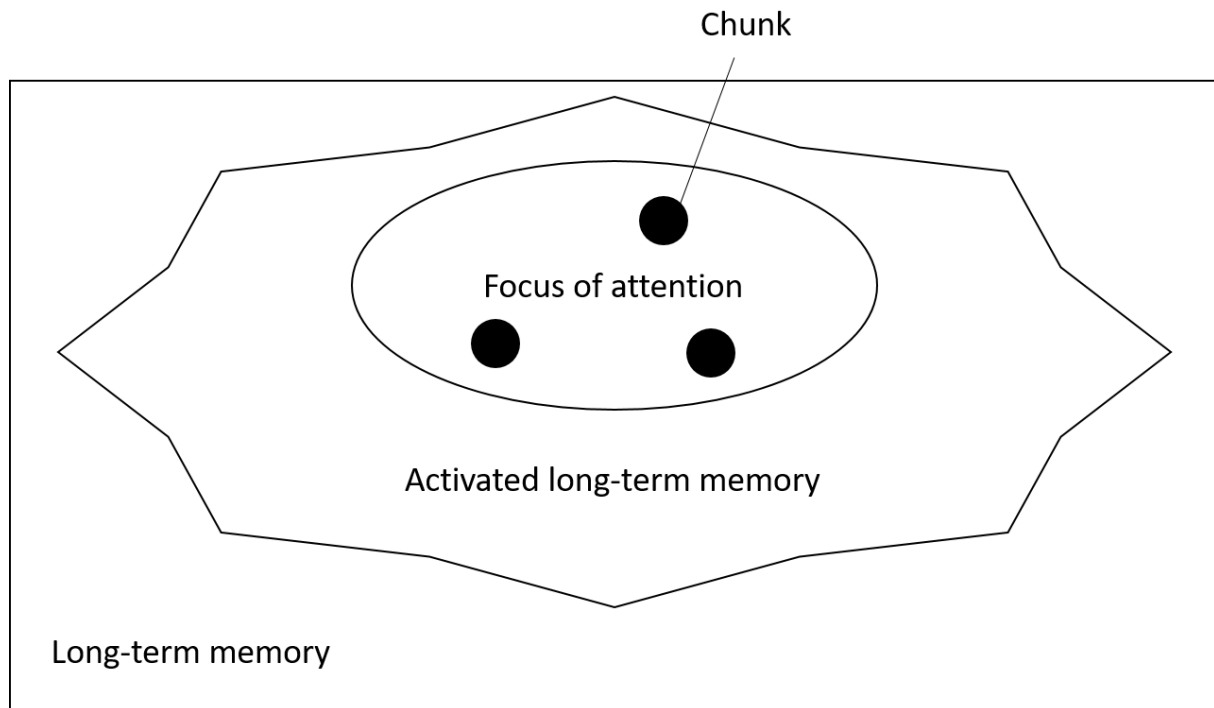
Long-term memory

Figure 1.2 - Cowan's embedded processes model of WM. Three different component can be distinguished: (a) latent knowledge stored in long-term memory, (b) the part of long-term memory currently activated and (c) the focus of attention, holding a limited number of chunks.

redirected toward an irrelevant stimulus, the original neural signature is rapidly lost (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2011; Lewis-Peacock & Postle, 2012). These results suggest that WM content may be directly represented within sensory-related regions.

Other theoretical accounts make different assumptions about the role of attention in WM. For instance, Oberauer (2002) (see also Oberauer & Hein, 2012) dissociates the "broad" from the "narrow" focus of attention, the broad focus being a region of direct access where items, activated in long-term memory, stay in a highly accessible state, while only the narrow focus is supposed to hold items that are fully and consciously attended. The Time-Based Resource Sharing model, on the other side, considers that WM maintenance is performed via an attentional refreshing mechanism that restores the constantly decaying WM representations (Barrouillet, Bernardin, & Camos, 2004). For Barrouillet and Camos, WM maintenance is constrained by this balance between decay and refreshing: when more time is available between two processing steps, more refreshing attempts can be performed in order to restore the degraded WM representations. Note that in certain versions of this framework, items are not

10

supposed to be directly activated in long-term memory, contrary to the embedded-processes model.

To sum up, it appears that attention and WM are two highly related cognitive functions sharing many properties. Most critically, for several of the attentional WM accounts, attention cannot operate in WM tasks without interacting with representations stored and temporarily activated in linguistic, visual and other long-term memory systems.

Processes involved in the retention of serial order information

Another fundamental process in WM is the retention of serial order information, in addition to item identity. During the short-term maintenance of a verbal sequence, the serial order of to-be-remembered items is arbitrary, as in "hand – lake – road" and needs to be represented in some form, in addition to the representation of the items and their characteristics. By considering the specific processes involved in the retention of serial order information, we will again show the importance of considering the interactions with linguistic knowledge in verbal WM. Indeed, several studies have shown that item and serial order information are not equal as regards the influence of language knowledge. Different psycholinguistic variables have been shown to impact the ability to recall item identity, as we will show in detail in Chapter 2. For example, word lists lead to consistently higher verbal WM performance than nonword lists (Brener, 1940). This however appears to be only the case for the recall of item information (i.e., when scoring recall independently of the serial position in which items are recalled); the different psycholinguistic effects discussed in Chapter 2 typically do not lead to a specific benefit for the recall of serial position information (i.e., serial order errors do not decrease for memoranda with richer long-term memory representations) (Campoy, Castellà, Provencio, Hitch, & Baddeley, 2014; Hulme et al., 1997; Hulme, Stuart, Brown, & Morin, 2003; Poirier & Saint-Aubin, 2005; Romani, McAlpine, & Martin, 2008; Walker & Hulme, 1999; but see Allen & Hulme, 2006; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Tse & Altarriba, 2007). And even when there is an impact of psycholinguistic variables on serial order recall, the impact is generally more subtle than what is observed for item recall (Jefferies, Frankish, & Lambon Ralph, 2006a; Tse, Li, & Altarriba, 2011).

Hence, the processes involved in the retention of serial order and item information can already be distinguished based on the differential impact of linguistic knowledge on serial order and item retention performance. Other lines of evidence further support this item-order dissociation. First, several studies have shown that item and serial order WM processes are not equally affected by interfering tasks. Serial

order recognition performance is indeed more strongly impacted by rhythmic and articulatory interfering tasks than is item recognition, and this has been observed across the verbal and musical domains (Gorin, Kowialiewski, & Majerus, 2016; Henson, Hartley, Burgess, Hitch, & Flude, 2003). This cross-domain dissociation of item and serial order WM processes further suggests the existence of domain-general mechanisms supporting the processing of serial order information (Hurlstone, Hitch, & Baddeley, 2014). Second, studies performed in brain-injured populations have shown double-dissociations between the processing of item identity and serial order information. For instance, patients with specific impairment to semantic knowledge show poorer recall performance at the item level, while their recall for serial order information stays relatively unaffected (Majerus, Van der Linden, Poncelet, & Metz-Lutz, 2004; Patterson, Graham, & Hodges, 1994). A fine-grained analysis of WM profiles in brain injured patients with verbal WM deficits has recently revealed a strong heterogeneity as regards item versus serial order WM impairment, with patients showing either specific impairment for the short-term retention of item information, specific impairment for the retention of serial order information, or mixed profiles (Majerus, Attout, Artielle, & Van der Kaa, 2015). Dissociations between item and serial order retention processes have also been observed in neurodevelopmental disorders such as dyslexia (Leavitt, Mendoza-Halliday, & Martinez-Trujillo, 2017; Martinez Perez, Majerus, & Mahot, 2012; Martinez Perez, Steve, & Martine, 2013) dyscalculia (Attout, Fias, Salmon, & Majerus, 2014) and Down syndrome (Brock & Jarrold, 2005).

Finally, neuroimaging studies are also informative about the item-order dissociation. Using a procedure requiring either the maintenance of item or serial order information, a stronger recruitment of the right intraparietal sulcus has been observed for the maintenance of serial order information as compared to the maintenance of item information (Majerus et al., 2010). Conversely, the processing of verbal item information seems to recruit to a higher extent sensory-specific neural regions, such as the superior temporal gyrus involved in phonological and lexico-semantic linguistic processes. The right intraparietal sulcus appears also to be less recruited in populations presenting selective serial order WM deficits (Martinez Perez, Poncelet, Salmon, & Majerus, 2015). Even within the linguistic system, dissociations between item and serial order processing can be observed. Kalm & Norris (2014) observed distinct neural patterns in the dorsal language network (superior temporal gyrus, supramarginal gyrus, and inferior prefrontal cortex) that allowed to decode the memorization of the serial order of nonwords as compared to their identity. This has also been confirmed by a microelectrode stimulation study. Papagno et al. (2017)
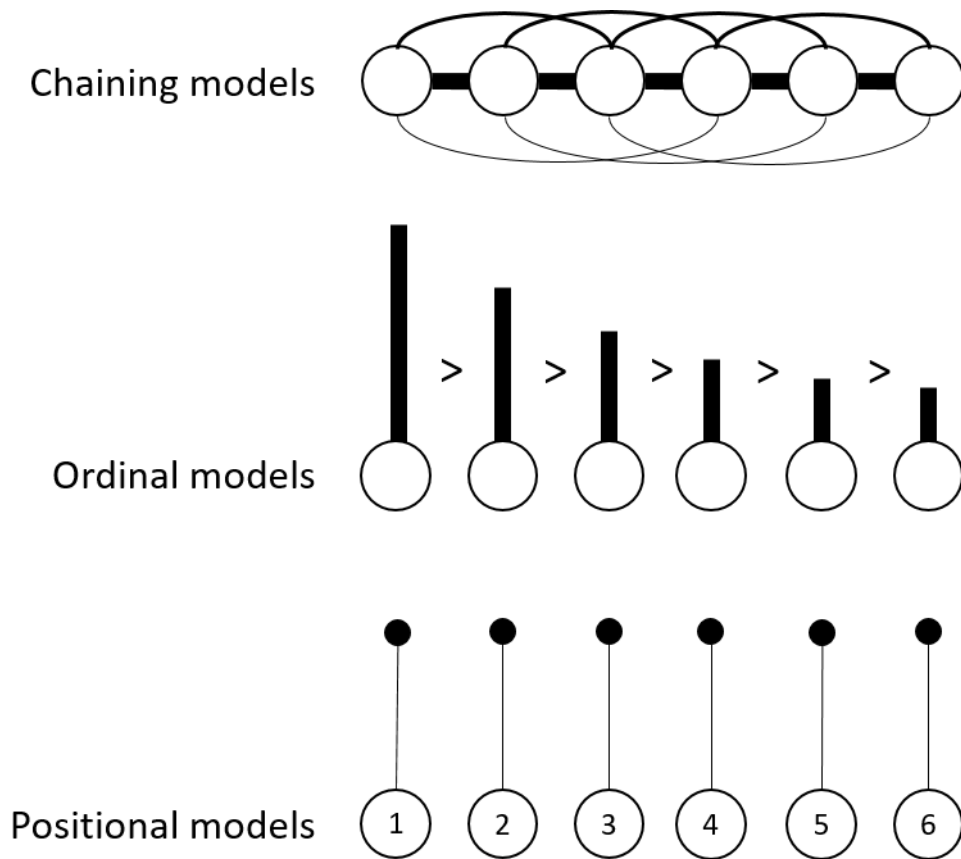
Figure 1.3 - The three family of models accounting for serial order effects.

showed that stimulation of the supramarginal gyrus disrupted the processing of serial order information for a digit span task (Papagno et al., 2017), while the stimulation of Broca's area disrupted to a larger extent recall of item information. Overall, these studies suggest that distinct neural cortices, within and outside linguistic cortices, encode item and serial order information in WM tasks.

More generally, the retention of serial order information is characterized by a series of benchmark phenomena, which are at the basis of many recent models of serial order WM (Hurlstone et al., 2014; Oberauer et al., 2018). First, WM performance, when scored as a function of serial position, displays a strong primacy and a somewhat reduced recency effect, with recall performance progressively decreasing across serial position, and a slight recall advantage for the last presented item. Second, serial order errors in WM tasks are characterized by a *locality constraint* (Henson, 1996): when a serial order error occurs (i.e. recalling an item at a wrong serial position), it more often involves adjacent as compared to distant displacements, giving rise to a *transposition gradient*. Third, another important benchmark effect typically observed in studies assessing serial order recall is the temporal grouping effect. When memoranda are

temporally grouped via the insertion of one or several pauses between items (e.g., ABC – *pause* – DEF), WM performance increases (Hartley, Hurlstone, & Hitch, 2016; Henson, 1996; Hitch, Burgess, Towse, & Culpin, 1996), and an increase of *interposition* errors occurs. These errors involve the displacement of items from one group to another, but keeping the same within-group position. For instance, transposing **B** instead of **E** in "A**B**C – *pause* – DEF" constitutes an interposition error. More generally, temporal parameters have been shown to be crucial for WM, for instance in order to boost refreshing of WM representations during maintenance (Plancher, Bernard, Tillmann, Neuroscience, & Bernard, 2018).

These benchmark phenomena have been accounted for by a range of mechanisms and models. The primacy effect is often explained by an output interference mechanism. As participants recall the items over the different positions, WM representations at the item and serial order level get blurry, leading to increasingly poorer recall performance as recall progresses (Cowan et al., 1992; Cowan, Saults, Elliott, & Moreno, 2002). Other have postulated the existence of a primacy gradient, according to which each item is successively encoded with decreasing activation strength. One possible reason for this could be that each item gradually receives less attentional resources across serial position (Oberauer, 2003) or because items positioned at the beginning of the list are rehearsed more often (Tan & Ward, 2008). At the moment of recall, items are recalled according to their activation level, with the most activated item being selected for recall. Since items at the beginning of the list receive more initial activation, their probability to be recalled also increases.

The active maintenance of serial order information has been accounted for more specifically by several types of serial order WM models (see **Figure 1.3**). According to chaining models, maintenance of serial order involves the creation of strong associations between adjacent items, but also between more distant items, with association strength decreasing as the distance between items increases. At the moment of recall, each item correctly recalled serves as a cue to recall the directly following item (Lewandowsky & Murdock, 1989; Wickelgren, 1969). The primacy gradient (Page & Norris, 1998) described in the previous paragraph belongs to a family of models called ordinal models, which assume that items are coded according to their relative activation. In the primacy model, the most activated items at the moment of recall is selected in priority; since items at the beginning of the encoding list receive more initial activation, they are also recalled first, as compared to items at the end of the list. Positional models, on the other side, assume that each item is associated to a specific position thanks to mechanisms coding for positional information. For instance, in a connectionist model of the phonological loop (Burgess & Hitch, 1999, 2006; Hitch

et al., 1996), each item is associated with partially overlapping positional markers. These associations are created in memory via Hebbian learning, and recall is performed by reactivating these positional markers. Hence, this model assumes the existence of direct associations between items and their serial position, although the item level remains relatively unspecified.

These evidence suggest that the serial order processing is coded via specific codes, and these codes appear to be distinct from those involved in the processing of item identity. As we will see, the linguistic system is a fundamental component for the processing of item identity.

Chapter summary

In this chapter, we went through several important processes influencing WM maintenance. Subvocal articulatory rehearsal is one of those processes, and is supposed to be responsible for WM maintenance by refreshing the decaying phonological traces via an articulatory process. As we have shown, the word length effect, a hallmark effect considered to arise from these articulatory refreshing processes, can also be attributed to the intervention of lexical linguistic knowledge. Next, we reviewed the role of attentional processes in WM, by showing again the importance of long-term memory knowledge for understanding the links between attention and WM. Finally, we examined the fundamental distinction of item and serial order processing levels in WM, by highlighting again the importance of linguistic knowledge which is considered to be a major determinant of item storage capacities. Hence, examining more deeply the interactions between the linguistic system and WM appears to be a fundamental aspect for advancing our understanding of the concept of WM.

Chapter 2

The influence of linguistic knowledge in working memory

In Chapter 1, we have seen that processes defining several core aspects of WM cannot be fully understood without considering the interactions between verbal WM and linguistic knowledge stored in the language system. In this chapter, we will review in detail the empirical data that show, in a more or less direct manner, the dependency of verbal WM abilities on access to various levels of linguistic knowledge, from sublexical phonological to semantic levels of linguistic representations. These data will mirror the initial neuropsychological findings that initiated a linguistic conceptualization of verbal WM: N. Martin and colleagues reported several patients with language impairment and associated verbal WM deficits and who showed a close association between error types in linguistic and verbal WM tasks (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; N. Martin & Saffran, 1992). Subsequent reviews showed that the vast majority of patients presenting with so-called specific verbal WM impairment had actually a history of language impairment (Majerus, 2009). The present chapter will describe in detail the empirical data indicating that verbal WM, or at least part of it, is grounded in the linguistic system.

The impact of sublexical phonological knowledge on verbal WM

During language acquisition, we rapidly acquire the sublexical/phonological regularities of the language we are exposed to, and these regularities become part of our linguistic long-term knowledge base. Several studies have demonstrated the impact of phonological long-term knowledge structures on WM processing. More specifically, the regularity at which phonological segments occur in a language has been shown to influence WM performance. This effect is called the *phonotactic probability* or *phonotactic frequency* effect. For example, in French, the diphone /ra/ appears more frequently than the diphone /Sn/. In linguistic tasks, nonwords associated with high phonotactic frequency structures are usually associated with faster decision times, such as in matching tasks requiring participants to judge whether two auditorily presented items are identical (Vitevitch & Luce, 1999). The same effect

is also observed in WM tasks, with increased recall performance for nonwords composed of high frequency diphones (Gathercole, Frankish, Pickering, & Peaker, 1999; Majerus & Van der Linden, 2003). Furthermore, this phonotactic knowledge can be updated very quickly and the newly acquired phonotactic knowledge will also have an immediate effect on verbal WM peformance. This has been demonstrated using an incidental learning paradigm in which young children and adults were passively exposed to a continuous phonological sequence whose phoneme successions were governed by an artificially phonotactic grammar (Majerus, Van der Linden, Mulder, Meulemans, & Peters, 2004). After the incidental learning phase (via simple passive listening while carrying out a complex drawing task), the participants were presented with a nonword repetition task consisting of nonwords of increasing length constructed with the regularities of the artificial phonotactic grammar or with different regularities. Participants showed a reliable recall advantage for 'legal' nonwords as compared to 'illegal' nonwords according to the artificial phonotactic grammar. This effect can be observed with both word and nonword syllables, except for digit syllables (Majerus, Martinez, & Oberauer, 2012).

Overall, these studies on sublexical phonological long-term memory effects suggest that WM performance is constrained by sublexical knowledge structures embedded in the language system.

The impact of lexical knowledge on verbal WM

One of the strongest psycholonguistic effects observed on WM is the lexicality effect, as defined by increased WM recall performance for words versus nonwords (Besner & Davelaar, 1982; Brener, 1940).It reflects the impact of our previously acquired lexical knowledge on WM performance. One may argue that words might be better recalled because they are composed of diphone segments appearing more frequently in natural language processing, but it has been shown that this effect remains after controlling for phonotactic probability structures (Gathercole et al., 1999; Majerus & Van der Linden, 2003). It might also be argued that recall performance for nonwords decreases because they take longer to articulate, and hence to rehearse. However, Hulme et al. (1991) were able to rule out this possibility, by showing that recall performance still differs between words and nonwords when controlling for articulation rate. More specifically, recall performance increases regularly as articulation rate increases for both words and nonwords, but at the same time words and nonwords have different intercepts. The lexicality effect has also been observed using a novel word learning paradigm. In a first phase, participants had to learn auditorily presented new words. During a second phase, participants had to recall

either these newly acquired words, or nonwords that had never been presented during the learning phase. A recall advantage was observed for these newly acquired words (N. Savill, Ellis, & Jefferies, 2016; N. Savill, Metcalfe, Ellis, & Jefferies, 2015), as compared to completely new nonwords. In addition, these studies showed that when the to-be-learnt words were associated with semantic knowledge during the learning phase (i.e. such as pictures of novel objects corresponding to the words), WM performance further increased for this class of words. This result further shows the importance of semantic knowledge in WM, in addition to purely lexical knowledge, as we will see in the next section.

The influence of lexical knowledge on WM performance has also been demonstrated via other psycholinguistic variables, such as lexical frequency. As words are more frequent in everyday language, recall performance increases when presented in WM tasks (Watkins & Watkins, 1977). This effect does not only reflect the higher accessibility of lexical representations for frequent items in the language system, but may also reflect stronger inter-item associations; frequent words are not only more frequent at the individual level, but also appear together more frequently (Landauer & Dumais, 1997). Indeed, Hulme et al. (2003) compared recall performance for pure lists of high and low frequency words, versus lists composed of alternating high and low frequency words (i.e. HLHLHL or LHLHLH, with the letter "H" and "L" refering to high and low frequency words, respectively). If the lexical frequency effect stems from the increased accessibility of individual item representations, we would expect to observe a sawtooth pattern of performance across serial positions for alternating lists, with higher performance each time a frequent word appears in the list. Instead, the authors observed that performance across serial positions followed the classical shape, with overall recall performance being at intermediate levels for alternating lists, when compared to pure lists of high and low frequency words. Hulme et al. concluded from these observations that the recall advantage for high over low frequency words does not stem from the higher lexical frequency of individual words, but instead from increased inter-item associations for frequent words (Landauer & Dumais, 1997). Although Hulme et al. initially showed that recall performance did not differ when high and low frequency words were equated at the level of inter-item associations, other authors were able to show lexical frequency effects in WM at the individual item level (Poirier & Saint-Aubin, 2005; Tse & Altarriba, 2007).

A final lexical factor influencing WM recall performance is the neighborhood density effect, or neighborhood size effect we already mentioned in Chapter 1 when discussing the word length effect. The neighborhood dimension refers to the number of phonological neighbors a target word is associated with. Two words are considered

to be phonological (or lexical) neighbors if they share all but one phoneme, via deletetion, substitution or addition of one phoneme in one of the words. For instance, "cat" and "bat" are considered to be phonological neighbors, as they differ by a single phoneme via substitution of the first phoneme. In WM tasks, lists of words with many neighbors are better recalled as compared to lists of words with fewer neighbors (Roodenrys et al., 2002). Roodenrys and Hinton (2002) also showed this effect with nonwords, while also controlling for phonotactic frequency as the latter variable is typically highly correlated with neighborhood density. Conversely, they also observed that, when controlling for neighborhood density, the impact of phonotactic frequency was virtually absent, raising doubts about a sublexical impact on WM performance. Another study, using more thouroughly controlled sets of stimuli, however observed independent effects of neighborhood density and phonotactic frequency on verbal WM performance (Thorn & Frankish, 2005).

Therefore, it appears that our linguistic system impacts WM performance at the lexical level in a very complex manner, either via the presence of more or less strong and connected lexical representations that can support verbal WM performance, or via the influence of neighboring lexical representations of a target word.

The impact of semantic knowledge on verbal WM

Finally, WM performance appears to be also influenced by semantic knowledge. One of the first studies highlighting this effect is the study by Bourassa and Besner (1994) showing that WM performance differs between high and low imageability words, and this after controlling for other linguistic dimensions such as word frequency and word length. The imageability dimension refers to the ease with which words evoke a mental image (Tyler, Moss, Galpin, & Voice, 2002). This dimension is highly correlated with the concreteness dimension, which refers to the degree to which words not only evoke visual images but also other sensory experiences such as sounds, tactile sensations, smells, etc… Although both variables are highly related, they do not perfectly match. For instance, the word "dragon" may be highly imageable, but may have a smaller concreteness score, since one has never touched a dragon or had a direct sensory experience with a dragon. One of the most comprehensive assessments of the imageability variable and its impact on WM tasks has been performed by Romani et al. (2008). They showed that this effect was observed in immediate serial recall tasks with and without articulatory suppression using different list lengths, and this across several paradigms such as serial order recognition, free recall, and serial order reconstruction. This effect has also been observed for both short and long words (Walker & Hulme, 1999). Interestingly, this effect has recently been manipulated under

different presentation rates, and it has been found that high imageability words were better recalled for slow (i.e. 2 seconds for each item) than for faster (i.e. 1 second for each item) presentation rates, with presentation rate having no impact on the recall of abstract words (Campoy et al., 2015). This observation has been interpreted as reflecting the fact that the process of forming a mental image for high imageability words takes a certain amount of time during WM processing. However, a recent study compared recall performance for high and low imageability words with dynamic visual noise being presented during encoding. In this paradigm, random and constantly changing patterns of white and black squares were presented to participants on the computer screen in order to diminish the vividness of mental images evoked by the to-be-remembered stimuli (Baddeley & Andrade, 2000; Quinn & McConnel, 1996). It follows that if the imageability effect stems from the creation of mental images, then the imageability effect should be abolished under dynamic visual noise. Instead, the imageability effect appeared to be preserved in these conditions (Castellà & Campoy, 2018; Chubala, Surprenant, Neath, & Quinlan, 2018). Interestingly, Chubala et al. also found that the effect was preserved under dynamic visual noise when immediate serial recall was required; as soon as a recognition task was used, the effect was abolished. This result suggests that the imageability effect is sensitive to the recall test conditions.

Another major semantic effect observed in WM tasks is the semantic similarity effect. This effect was initially described by Poirier and Saint-Aubin (1995). They observed that WM recall performance was increased when the words in the WM lists were related at the semantic level such as "leaf – tree – branch" as opposed to "hand – cloud – chair" (Monnier & Bonthoux, 2011; Neale & Tehan, 2007; Saint-Aubin & Poirier, 1999a; Tse, 2009; Tse et al., 2011). One study compared different types of semantic relatedness, such as thematic and taxonomic relationship (Tse, 2009). Two concepts are considered to share a thematic relationship if they are defined by an external relationship that may unify them witin a coherent context. This is the case for instance for the words "car" and "garage", which are frequently used in the context of driving, despite the fact that they share minimal physical properties. The taxonomic relationship, on the other side, is more specifically defined by a broader hierarchical concept. This is the case for instance for the words "car" and "bus", which are both vehicles (Sachs et al., 2008). Tse (2009) showed that both types of semantic relationship impact WM recall performance to a similar extent.

The impact of semantic knowledge on verbal WM has also been demonstrated more specifically by the study of brain-damaged patients such as patients with semantic dementia (Patterson et al., 1994). Semantic dementia is characterized by a

progressive loss of grey matter in the left anterior temporal lobe associated with semantic knowledge (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). Several studies have observed that patients with semantic dementia show impaired WM performance for words whose semantic representations are no longer available (the words are not known anymore to the patients) but not for words that are still known. As expected given their preserved phonological processing abilities, these patients also show preserved WM performance for nonword lists (Jefferies, Grogan, Mapelli, & Isella, 2012; Majerus, Norris, & Patterson, 2007; Patterson et al., 1994). Patients with semantic dementia also display a specific pattern of phoneme migrations when recalling words whose semantic content is lost. More specifically, phonemes tend to migrate more often from one item to another (e.g. recalling "mint – pug" instead of "pint – mug"). Interestingly, in healthy adults, this phenomenon is observed more specifically for the recall of nonwords (Jefferies et al., 2006a), and is interpreted as reflecting a stabilization process stemming from semantic knowledge at the phonological level; for Patterson et al. and Jefferies et al., the semantic knowledge of individual items acts as a "semantic glue" which constrains the phonemes that compose the items to stay in a stable and unified representation. Overall, the specific pattern of performance in WM tasks for words and nonwords in these patients clearly demonstrates the importance of the availability of semantic knowledge for accurate performance in WM tasks.

These different studies show the importance of semantic knowledge during WM processing. Semantic knowledge can impact WM performance at the individual item level, in the form of the imageability effect, but also at the inter-item level, in the form of semantic relatedness effects. Furthermore, an impairment to these semantic representations leads to a considerable drop in performance for short-term recall of word lists.

The intervention of linguistic neural substrates in verbal WM as revealed by neuroimaging studies

In line with the behavioral and neuropsychological studies presented in the previous sections, several neuroimaging studies in healthy participants support the intervention of language-related neural substrates during verbal WM tasks. One of the earliest studies used an event-related potentials approach and showed that word stimuli produced a larger negative slow wave component than nonword stimuli, and this during the encoding, maintenance and retrieval stages of WM processing (Ruchkin, Berndt, Johnson, Grafman, & Canoune, 1999). The exact meaning of this increased negativity in ERP components associated with words is however difficult to
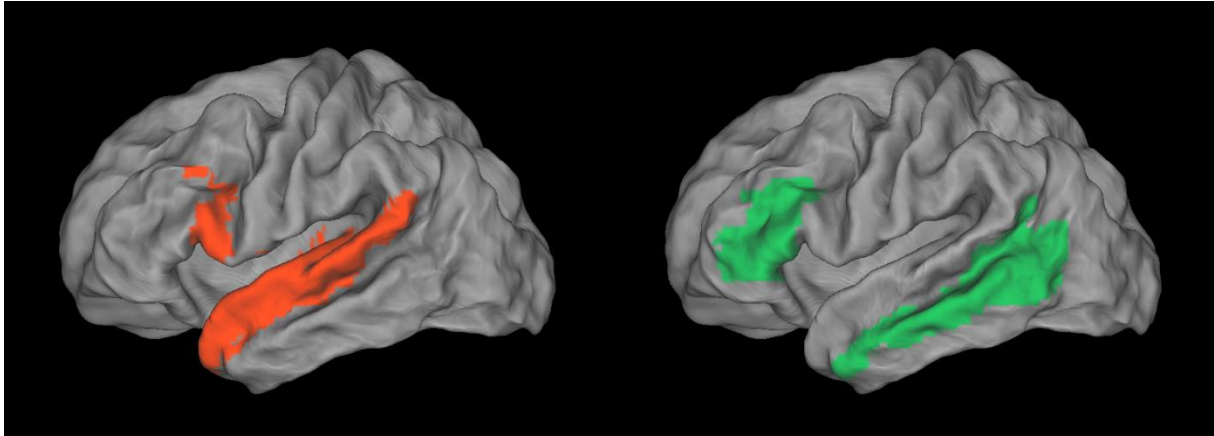
Figure 1.4 - Localisation of the dorsal (left panel) and ventral (right panel) pathways of language processing.

interpret in terms of the linguistic processes that could be involved. Other studies using functional magnetic resonance imaging techniques provide more direct information about the involvement of linguistic cortices in WM tasks, by showing the recruitment of both dorsal and ventral language pathways (see **Figure 1.4**). The dorsal language pathway is composed of the pars opercularis (Broca's area) and the superior temporal gyrus, and is considered to support both input and output phonological processing (Arsenault & Buchsbaum, 2015; Friederici, 2012; Mesgarani, Cheung, Johnson, & Chang, 2014). Parts of this network, such as the superior temporal gyrus, are also recruited during the encoding and the short-term maintenance of phonological information in WM tasks (Buchsbaum, Olsen, Koch, & Berman, 2005; Kalm, Davis, & Norris, 2012; Ravizza, Hazeltine, Ruiz, & Zhu, 2011; Strand, Forssberg, Klingberg, & Norrelgen, 2008). The ventral language pathway is associated with lexico-semantic processing (Friederici, 2012; Friederici & Gierhan, 2013; Hickok & Poeppel, 2007), and involves the pars triangularis (anterior inferior frontal cortex) and the middle temporal gyrus (Lambon Ralph et al., 2017). A study using positron emission tomography showed the involvement of the middle temporal gyrus (Collette et al., 2001) during an immediate serial recall task of word versus nonword lists. Similarly, Fiebach and colleagues showed that parts of the ventral language pathways are activated when the semantic content of memoranda needs to be maintained in a verbal WM task (Fiebach, Friederici, Smith, & Swinney, 2007).

These studies show that linguistic cortices are involved during the short-term maintenance of verbal information, which further highlight the need to integrate language processing and WM processing in models of WM.

Chapter summary

In this chapter, we described in a more detailed manner the different psycholinguistic effects observed in WM. These effects, which are now well-established, appear to strongly impact recall performance, and are observed at all levels of language processing: the phonological, lexical, and semantic levels. In addition, neural regions involved in language processing have also been shown to be recruited when verbal items need to be maintained in WM. These studies demonstrate that the linguistic long-term memory system is a core component to consider when examining the structure and functioning of verbal WM. As we will see in the next chapter, the nature of the interactions between linguistic long-term memory and WM remains a challenge for the theorization of WM. In the next chapter, we will review the different theoretical attempts that have been proposed and identify a set of major research questions aimed at confronting and advancing these different theories.

Chapter 3

What is the nature of linguistic knowledge effects in working memory?

In Chapters 1 and 2, we have reported several lines of evidence suggesting an important role for linguistic knowledge in verbal WM. Even effects considered to be the most specific to verbal WM, such as the word length effect, can be interpreted as reflecting the intervention of linguistic knowledge. At the same time, the nature and locus of the intervention of language knowledge in verbal WM tasks remains a fundamental theoretical question. In Chapter 3, we will review the different theoretical accounts that currently exist regarding the interactions between linguistic knowledge and verbal WM and we will identify several fundamental questions concerning the nature of the linguistic knowledge-verbal WM interactions raised by these accounts.

Linguistic long-term memory effects as a strategic process

A first account that has been proposed is a strategic account (Campoy et al., 2015; H. G. Shulman, 1970). Following this account, participants are considered to implement several strategic processes during the encoding stage of WM processing, thereby enhancing their WM performance. The implementation of such strategies might modulate the magnitude of psycholinguistic effects. The modulation of psycholinguistic effects has already been observed in experiments requiring participants to adopt specific strategies during the encoding stage of WM processing. For instance, under semantic encoding conditions, the phonological similarity effect disappears (Hanley & Bakopoulou, 2003) and the word length effect is less strongly observed (Campoy & Baddeley, 2008), but these phonological effects are exacerbated when participants are encouraged to implement a phonological coding strategy. By extension, it could be argued that the presence of lexico-semantic effects previously observed might be the result of semantic elaborative processes implemented by participants. Indeed, previous studies assessed these psycholinguistic effects in procedures requiring participants to encode relatively short lists of items (i.e. from 5 to 7 items) presented at a comfortable pace (i.e. from 1 to 2 seconds for each item). In these encoding conditions, it has been shown that participants use a wide variety of strategies, such as rehearsal, grouping, mental imagery, or semantic elaboration (Morrison, Rosenbaum, Fair, & Chein, 2016). Whether such strategies are the main

factor responsible for lexico-semantic influences in WM however still remains to be demonstrated.

In an early investigation of the strategic account, Shulman (1970) used a WM probe recognition procedure in which participants had been presented with recognition foils that were either phonologically or semantically related to one of the target items of the memory list. Participants were informed about the type of foil they may encounter before the beginning of each trial. This procedure was supposed to encourage a phonological or a semantic encoding strategy. Furthermore, the words were presented at different presentation rates: 340ms/item, 700ms/item or 1400ms/item. Shulman observed that when participants were tested in the semantic encoding condition, a slower presentation rate increased recognition performance, while this effect was absent in the phonological condition. These results suggest that the encoding of semantic information might be a slow, time-dependent process, contrary to the encoding of phonological information. A potential criticism over this study however, is that Shulman only showed that semantic encoding can benefit from slower presentation rates, but did not directly demonstrated that the nature of semantic coding is purely strategic. Indeed, in the semantic condition, recognition performance in the fast presentation conditions was nevertheless well above what could be expected by chance.

A further study that explored a strategic, time-dependent account of semantic effects in verbal WM is the study by Campoy et al. (2015) that was already briefly mentioned in Chapter 2. In a first experiment, Campoy et al. showed that as presentation rate decreased, recall performance in immediate serial recall tasks increased for high imageability words, while recall of low imageability words remained unaffected. This result might indicate that the semantic features associated with high imageability words need more time to be encoded, or require strategic controlled processes that participants are able to implement during the inter-item stimulus interval. However, the results of the first experiment in the study of Campoy et al. (2015) were contradicted by a second study where the authors showed that an imageability effect was still observed when participants had to perform a concurrent task while encoding the items. More specifically, while hearing the successive memory items, participants were presented with three shapes at the top of the screen, and were then required to judge which one matched a fourth shape presented at the bottom of the screen. This procedure was aimed at preventing participants from implementing strategic controlled processes during the encoding phase, but also at minimizing the opportunity for participants to implement mental imagery, as the concurrent task was a visual task naturally interfering with mental imagery processes. The authors

observed an impact of the concurrent task on recall performance with overall lower performance as compared to a standard encoding condition, and this whatever the imageability of words, low and high. These latter results do not support a purely strategic account of semantic effects in verbal WM. At the same time, it should be noted that the items were presented at a very slow pace in this experiment (2000ms /item) and only one visual judgment had to be carried out within these 2000ms, raising the possibility that participants may still have had sufficient spare time between successive items to implement strategic linguistic encoding processes (the mean time for visual judgment was ~1300ms).

In sum, available evidence does not allow to firmly rule out the intervention of strategic factors as an at least partial explanatory account of psycholinguistic effects in verbal WM.

Linguistic long-term memory effects in verbal WM as a post-encoding reconstructive process

Although the strategic account mentioned in the previous section has remained relatively peripheral in terms of impact on theories of verbal WM, a second account considering that linguistic knowledge effects arise from the reconstruction of degraded phonological traces at the moment of recall has been much more influential. This reconstruction account has been termed redintegration and was initially proposed by Hulme et al. (1991) (Hulme et al., 1991; Lewandowsky, 1999; Schweickert, 1993; Schweickert, Chen, & Poirier, 1999). This account was used by many researchers to account for linguistic long-term effects in the more general framework of the phonological loop model (Baddeley & Hitch, 1974) which, in its original version, does not explicitly take into account interactions with linguistic knowledge. According to the redintegration account, shortly after encoding of memoranda in an exclusively phonological format, phonological traces are subject to degradation due to decay or interference. At the moment of recall, these degraded phonological traces undergo a "clean-up" process, whereby they are reconstructed, by comparing them to stored lexical representations. Thus, the partial phonological features that remain available at the moment of recall can be viewed as retrieval cues. The recall advantage for words over nonwords is explained by assuming that lexical knowledge allows to reconstruct word but not nonword stimuli via the clean-up process. Schweickert (1993) proposed a formal implementation of this process by using a multinomial processing tree, displayed in **Figure 1.5**. Let the parameter "I" be the probability that an item will be intact at the moment of recall. If the item is intact, the system will produce a correct response. The probability that an item will not be intact at the moment of recall

corresponds to "1-I". In this case, there is an "R" probability that the item will be correctly reconstructed (redintegrated) based on partially degraded phonological traces, leading to a correct response. Thus, the probability that an item will be incorrectly recalled corresponds to "1-R". Recall performance for an individual item corresponds then to $I + (1 - I)R$. Nonword recall will exclusively rely on the I parameter (intact phonological traces), because for these items the R parameter will always be equal to 0. This simple mathematical model can also account for other psycholinguistic effects than the lexicality effect. The R parameter can be associated to continuous values to account for the word frequency effect: high frequency words will have a higher R value than low frequency words, because high frequency items are supposed to be more easily accessed in the lexicon. This model can also account for other linguistic effects such as the phonological similarity effect, by assuming that traces for phonologically similar items are more likely to be confused due to trace similarity. The increased rate of serial order recall errors for phonological similar items could be accounted by the intervention of redintegration errors: in the target sequence "CAT – FAT – BAT", at the moment of recall, the remaining trace "_AT" is equally likely to reconstruct "FAT" "CAT" or "BAT" for all serial positions.

An important prediction of the model is that as more a given representation is degraded, as more the redintegration process – and therefore, the R parameter – will have an important weight during WM performance (Schweickert et al., 1999). Put in another way, the more degraded the phonological traces are, the greater the psycholinguistic effects should be observed. Therefore, the model predicts stronger psycholinguistic effects across serial position, because phonological traces are considered to be more degraded for items at the end of the list. This has been observed by manipulating the word frequency effect (Hulme et al., 1997; Quinlan, Roodenrys, & Miller, 2017), but also by manipulating the phonological and semantic similarity effects across different task difficulty conditions (Neale & Tehan, 2007). In Neale & Tehan study, participants were required to recall lists in which the semantic and phonological similarity effects were manipulated. The participants had to perform different interfering tasks of increasing difficulty. As the task became more difficult, the psycholinguistic effects were more strongly observed, as expected if a stronger degradation of phonological traces would lead to a stronger impact of redintegration.

Further evidence supporting the plausibility of a redintegration process stems from the analysis of recall latencies for word and nonword items. Hulme, Newton, Cowan, Stuart, and Brown (1999) analysed the inter-item output time taken by participants to recall word and nonword items. They observed that the pauses between nonword items were longer than those between word items. They argued that this
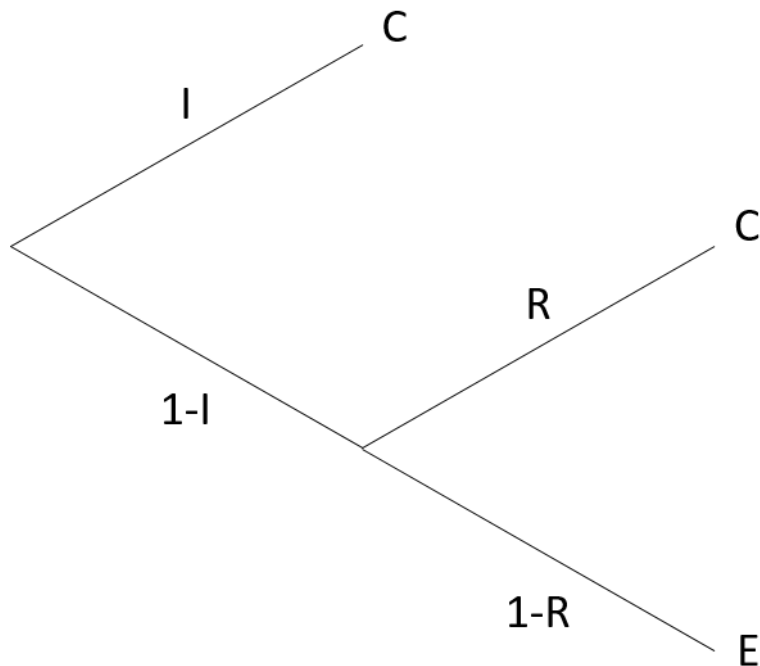
Figure 1.5. Schweickert's multinomial processing tree.

phenomenon stems from the redintegration mechanism acting as a search process, the search taking longer for nonwords as no matching lexical representation can be found in the lexical network. There is however a problem with this interpretation. Indeed, when a nonword is correctly recalled, performance is supposed to rely on the I parameter, i.e. phonological traces are supposed to be complete at the moment of recall. Hence, for correctly recalled nonwords, no search mechanism is supposed to have occurred. Consequently, the search mechanism should logically occur only for words. Therefore, the model would actually predict a *reversed* lexicality effect on response latencies for correctly recalled items. It could however be argued that there may still have some form of lexical redintegration also occurring for nonwords (Gathercole et al., 1999). This is based on the observation that nonwords that are more word-like lead to higher WM recall performance (Ritchie, Tolan, & Tehan, 2015). Further specifications of the redintegration framework have proposed the existence of redintegration processes based on sublexical phonological knowledge (Thorn, Gathercole, & Frankish, 2005), in order to account for the observation of phonotactic frequency effects in nonword (Gathercole et al., 1999).

Overall, the redintegration account faces several issues that need to be considered. First, it is difficult to explain how the redintegration mechanism could account for psycholinguistic effects at the semantic level. The semantic similarity effect

has been explained by assuming that knowledge about the semantic category is supposed to sharpen and facilitate the search within stored long-term memory knowledge (Neale & Tehan, 2007; Saint-Aubin, Ouellette, & Poirier, 2005), which could be considered as the mathematical equivalent of increasing the R parameter. Even though in the original framework, semantic knowledge is not supposed to be accessed at encoding, participants may rapidly become aware of the relevant semantic category after recalling the two first words, leading to more efficient redintegration of the remaining words to be recalled. However, it is even more difficult to imagine how the model could explain the imageability effect without also assuming some form of semantic maintenance, because the redintegration mechanism is supposed to occur at the lexical level. This issue is more related to the lack of extensive specification of the model, which has not been formally impemented in order to account for semantic effects.

Second, the consequence of trace degradation on the amount of redintegration is difficult to predict and test. While some authors proposed that stronger degradation of phonological traces increases redintegration processes and hence psycholinguistic effects in WM, as explained above, other authors considered that massive trace degradation may actually reduce the success of trace redintegration. For instance, given the word "ELEPHANT", the mildly degraded trace "_L_PHANT" would allow for efficient reconstruction, but this would not be the case for the strongly degraded trace "_L____A__", because the cues that remain are too few and uninformative. In line with this assumption, Ritchie et al. observed that when participants were tested in a dual-task WM paradigm where phonological traces were likely to be strongly degraded at recall, weaker lexicality effects were observed (Ritchie et al., 2015). Overall, these predictions are however difficult to test in a systematic manner as irrespective of the outcome of results (increased vs. decreased psycholinguistic effects), the redintegration hypothesis would always be true: if the psycholinguistic effects are weakened, then it can be argued that traces were probably too degraded to allow for efficient redintegration; if the psycholinguistic effects are increased, then it can be argued that traces were only mildly degraded, allowing for efficient redintegration. The fundamental problem here is that the quantity of degraded phonological information cannot be measured in an empirical manner. Another problem for the redintegration hypothesis is related to the neighborhood density effect. Following the model's predictions, words drawn from dense phonological neighborhood should be more poorly recalled under the R criterion. This is because for words from dense neighborhoods, there will be much more candidates to select for recall, and the redintegration process has no means to know which candidate is the

most relevant. However, from work already described in Chapter 2, we know that words drawn from dense phonological neighborhoods actually lead to *higher* recall performance (Roodenrys et al., 2002).

Despite these different problems for the redintegration account, this account is still widely used in the literature for the theoretical interpretation of psycholinguistic effects in WM tasks (see for instance Clarkson, Roodenrys, Miller, & Hulme, 2016; Guitard, Saint-Aubin, Tehan, & Tolan, 2018; Quinlan et al., 2017).

Linguistic long-term memory effects in verbal WM as a direct intervention of the language system

Another family of models takes a more direct stance by considering that the language system directly and automatically supports performance in verbal WM. These models assume that verbal WM is grounded within the language system, with or without the intervention of additional specific WM processes.

A first approach is the one proposed by N. Martin and Saffran (N. Martin, Saffran, & Dell, 1996). They adapted to the WM domain the interactive activation model of language production initially developed by Dell, as illustrated in **Figure 1.6**, (Dell, 1986; Dell et al., 1997) to account for errors produced in naming tasks. The starting point of N. Martin and Saffran's approach was the observation of the strong similarities of errors that language-impaired patients produce in linguistic and verbal WM tasks. Interactive activation models (McClelland & Rumelhart, 1981) are more generally composed of at least two levels of representations. Adjacent levels can reactivate each other simultaneously via bi-directional connection weights. This unique property of interactive activation models allows them to produce very complex behaviour. In Dell's model, the linguistic system is represented by phonological, lexical and semantic representational levels, with each level being composed of unitary representations (i.e., one unit = one phoneme, lexical representation or semantic feature). A decay function is often required in such models, in order to decrease the activation level of each node once activated, otherwise the system would become rapidly saturated. It is precisely this decay function that also allows for a WM function within the same linguistic architecture, without needing to assume the existence of additional short-term buffer systems. For N. Martin and Saffran, WM maintenance is performed via *sustained* activation within the linguistic system, with this activation decaying over time. In this type of framework, WM recall performance depends on the amount of activation available at the moment of recall; the greater or more robust the activation, the greater the probability that an item will be correctly recalled. In order to keep items in their correct serial order, they furthermore proposed to implement an
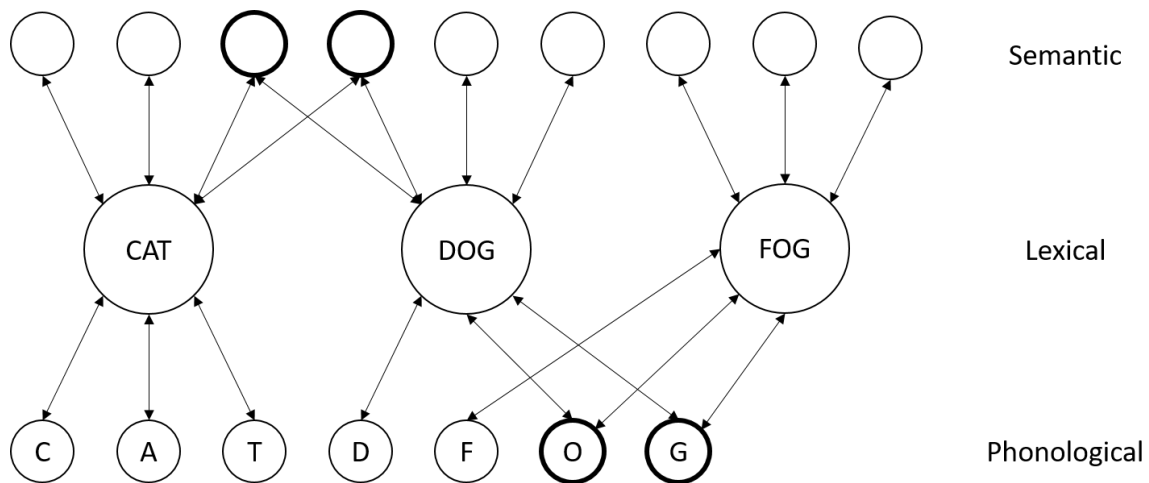
Figure 1.6. A simplified version of Dell's interactive activation model of language processing. Overlapping representations at the semantic and phonological levels are marked in bold.

activation gradient (see Chapter 1) over the activation occurring within the linguistic system: items presented at the beginning of the list would receive stronger activation at the moment of encoding. Hence, at the moment of recall, earlier items will also be selected and recalled first, based on their higher activation level (Page & Norris, 1998) which naturally produces forward recall.

R. C. Martin, Lesch, & Bartha (1999) developed a similar account based on an interactive activation model of language processing. However, R.C Martin and colleagues also included in their model a phonological and a lexico-semantic buffer, which are supposed to actively maintain the representations activated within the linguistic long-term memory knowledge structure. One of the main reason to include separate buffers, contrary to the model proposed by N. Martin & Saffran, was to be able to maintain multiple times the same item. More specifically, a pure activation-based model of WM would be unable to recall sequences composed of repeated items. For instance, given the target sequence "5 9 4 5 3 5", the system would be unable to represent – and therefore recall – more than one time the item "5". Instead, this item would simply be activated more strongly than other items due to its repeated occurrence and resulting re-activation. Nowadays, the need for distinct mechanisms responsible for WM maintenance is supported by neuroimaging studies showing the involvement of specific neural regions in addition to those involved in language processing. For instance, the dorsal attention network appears to be strongly recruited during WM maintenance and has been given the role of a buffer function by some authors (see Majerus et al., 2010, 2016 for a discussion). The plausibility of specialized

buffers is further suggested by a recent study by Yue et al. (2018). The authors trained a classifier to differentiate the neural patterns elicited by speech and nonspeech stimuli during a perceptual task. The classifier was then tested during WM maintenance for the same stimuli. The classifier was able to successfully decode speech and nonspeech stimuli within the inferior parietal cortex (supramarginal gyrus), but not in the superior temporal gyrus during WM maintenance. This result led them to argue that the inferior parietal cortex could act as a WM buffer, and that linguistic perceptual regions (i.e. the superior temporal gyrus) are not responsible for the active maintenance of verbal stimuli. One potential criticism of the buffer system approach proposed by R.C. Martin and colleagues is that the exact properties of these buffers are poorly specified and the buffers are merely conceptual boxes. Serial order information is also assumed to be maintained within the buffers, but no specific mechanism has been proposed to explain how this would be achieved.

Although the embedded-processes model proposed by Cowan is not a purely language-based model, as it assumes modality independent attentional processes, its core assumptions have however a number of similarities with language-based models such as the model proposed by N. Martin and Saffran. The embedded-processes model, like the approach by N. Martin and Saffran, considers that items are directly activated in long-term memory. The main difference is that these activations are further maintained via attentional focalization, as already explained in Chapter 1. This approach is supported by neuroimaging studies showing that sensory-related neural regions are more strongly recruited when items are maintained in the focus of attention (Bettencourt & Xu, 2015; Christophel, Hebart, & Haynes, 2012; Ester, Sprague, & Serences, 2015; S. Lee, Kravitz, & Baker, 2013; Peters, Kaiser, Rahm, & Bledowski, 2015; Yu & Shim, 2017). These studies have also shown that information maintained in WM can be decoded based on neural patterns in the intraparietal sulcus, the intraparietal sulcus being associated with attentional control processes. Note however that other studies did not report reliable decoding of items maintained in WM within the intraparietal sulcus (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Emrich, Riggall, Larocque, & Postle, 2013; LaRocque, Riggall, Emrich, & Postle, 2016; Linden, Oosterhof, Klein, & Downing, 2012). Furthermore, these studies mainly concerned visual information and it remains to be shown to what extent verbal memoranda can be decoded based on neural activity patterns, either in language-processing cortices or the intraparietal sulcus. In any case, the embedded-processes model of WM solves the placeholder problem that many buffer accounts face; instead of simply assuming that items are maintained in a box, whose precise nature is poorly described, the embedded-processes model assumes that items are directly maintained

in the focus of attention, whose properties can be very directly tested and studied in attentional-demanding tasks. The advantage of this explicit assumption makes the model much more prone to falsification, so that accurate predictions can be made. The model, however, also faces the difficulty to solve the problem of serial order maintenance.

Other models integrate linguistic processing and serial order processing within a multi-component architecture while providing a detailed implementation of the representation of serial order information. The model developed by Burgess and Hitch is one example of this type of architecture (Burgess & Hitch, 1999, 2006; Hitch et al., 1996). In this model, items are represented via lexical and phonological nodes representing the language system, and the lexical nodes are further connected to nodes in a contextual system whose activation levels vary continuously over time, meaning that items presented at different moments are associated with different activation states of the contextual system, allowing for the encoding of serial position information. Each position is represented by a distributed set of contextual nodes partially overlapping with those for adjacent positions. Hence, adjacent positions are more likely to be confused than more distant positions, because they also share more similar representations in the contextual system. Recall is performed by sequentially resetting/reactivating the contextual nodes, thereby reactivating the items linked to them. This model is interesting, because, in addition to providing a detailed implementation of the representation of serial order information, it also takes into account interactions with the language system, although in a less developed manner than in the approach of N. Martin and Saffran. This model has shown to successfully simulate important benchmark phenomena of the WM literature, such as the typical shape of serial position curves (i.e. primacy and recency effects), phonological similarity effects, the occurrence of transposition errors and the impact of temporal grouping on recall accuracy and transposition errors. In its actual state, the model is however unable to account for semantic effects, because this level of processing remains unspecified. Other models also explicitly represent the processing of serial order information, such as the Start-End Model (Henson, 1998), the primacy model (Page & Norris, 1998) or OSCAR (Brown, Preece, & Hulme, 2000), but these models do not incorporate linguistic long-term memory systems.

All these aforementioned models have different strength and weaknesses. The N. Martin and Saffran approach provides a very precise description of how activation within the linguistic system propagates and how this propagation of activity (and its decay) explains both linguistic and WM processing. On the other hand, the embedded processes model by Cowan, while assuming that items are activated in long-term

memory, does not make specific assumptions about how this activation occurs. However, this model provides a much more accurate description of how items are maintained in WM tasks via the interaction between attentional processes and long-term memory activation. These two models have in common the fact that they do not focus directly on mechanisms involved in the retention of serial order information. Other models, such as the model proposed by Burgess and Hitch, have been specifically developed for explaining the mechanisms involved in serial order processing, but they do not provide a detailed implementation of the interactions with the language system and they ignore the role of attentional processes. These different aspects have been considered simultaneously in an integrative framework proposed by Majerus (2013, 2018), as illustrated in **Figure 1.7**. This framework considers that verbal memoranda are directly activated within the linguistic system, and that this activation provides the representational basis for WM maintenance, by considering phonological and semantic levels of representations characterized by interactive activation as in the model proposed by N. Martin and Saffran. Serial order information is maintained thanks to specific serial order mechanisms distinct from those involved in the representation of item identity and linking items activated in the linguistic components to specific serial position markers similar to the Burgess and Hitch computational approach. Activity in the serial order and linguistic item components is further modulated by an attentional control mechanism allowing to direct attention to those types of representations that are most task-relevant at a given time and to maintain these representations within the focus of attention as in Cowan's embedded-processes model of WM. A further specificity of the framework proposed by Majerus is that it explicitly relates the different components of the architecture to underlying neural substrates. Representation of item information for verbal memoranda is considered to directly involve linguistic cortices, with the dorsal language pathway (superior temporal and posterior inferior frontal cortices) for the representation and maintenance of phonological information, and the ventral pathway (middle/inferior temporal and anterior inferior frontal cortices) for the representation and maintenance of lexico-semantic information. The processing of serial order information is considered to be supported by several neural substrates depending on the type of codes used for the representation of serial order information, with the involvement of a fronto-parietal network centered on the right intraparietal sulcus when spatial codes are used (Majerus, 2019; Majerus & Attout, 2018). Attentional control and focalization is considered to involve a left-lateralized fronto-parietal network involving the dorsal attention network centred on the left intraparietal sulcus. Recent studies have shown that the left intraparietal sulcus is involved both in quantitative (number of
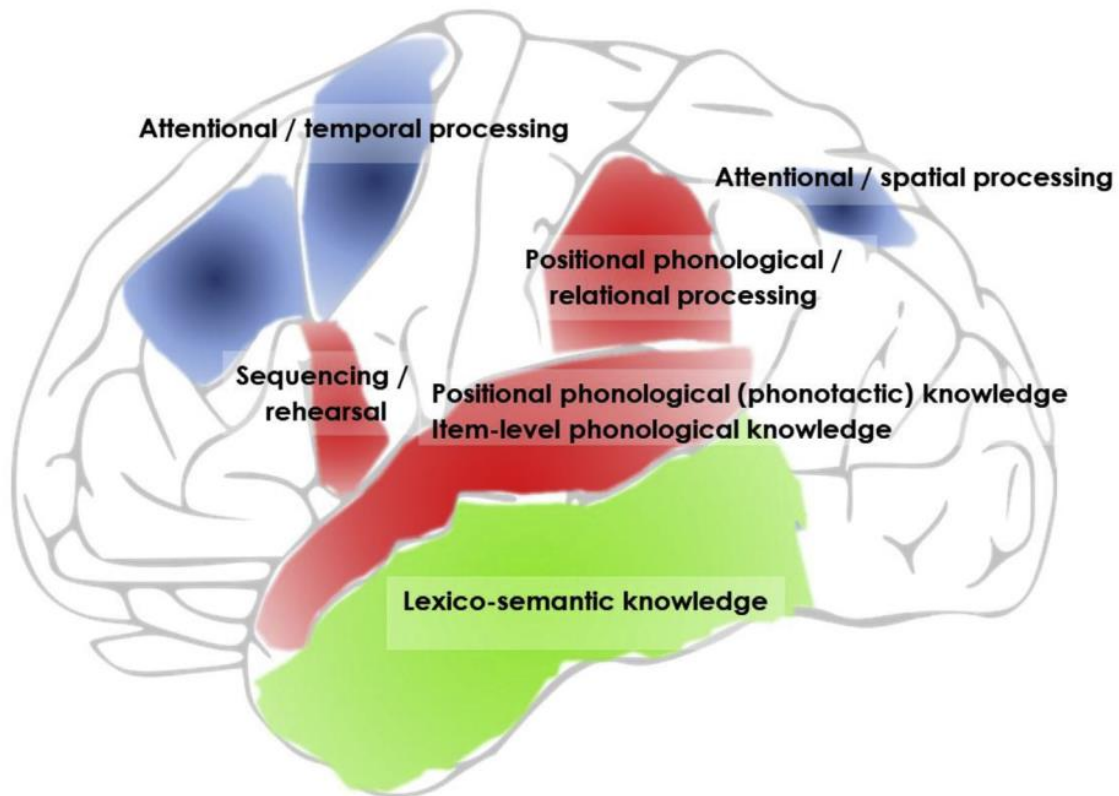
Figure 1.7. The integrative architecture proposed by Majerus (2013, 2018).

memoranda) and qualitative (type of memoranda) aspects of attentional control (Majerus et al., 2016). This integration of cognitive and neural aspects of verbal WM allows the framework to make specific predictions regarding the impact of specific brain lesions on the nature of verbal WM impairment a patient could present. For instance, an impairment to the right fronto-parietal network should result to greater deficits in the ability to maintain serial order information, while a deficit in the ventral language pathway should result in a deficit to maintain the lexico-semantic content (item information) of verbal memoranda.

The language-based models and equivalents presented in this section provide a very straightforward explanation for most of the psycholinguistic effects that have been described in Chapter 2. Any aspect that will facilitate or make more robust the activation of language representations in these models will automatically also translate in an advantage for short-term maintenance, as short-term maintenance (at least of item information) is considered to use the same cognitive and neural substrates as the language system. The lexicality effect in linguistic and WM tasks can be explained by assuming that phonemes activated at the phonological level will receive stronger stabilizing feedback activation from the lexical and semantic levels of language processing. Likewise, the language-based models can also account for the fact that

nonwords that are more word-like are more easily recalled (Ritchie et al., 2015); although nonwords do not have a lexical representation strictly speaking, they will nevertheless activate to some extent partially matching representations at the lexico-semantic level via the phonological features they share with real words, and could therefore also receive stronger feedback stabilizing activation. For instance, the nonword "leofard" is likely to also activate the lexical word forms "leopard" and "leotard". This hypothesis is congruent with neuroimaging studies showing strongly overlapping neural regions when words and nonwords are processed (Davis & Gaskell, 2009; Kotz, Cappa, von Cramon, & Friederici, 2002; Newman & Twieg, 2001; Orfanidou, Marslen-Wilson, & Davis, 2006; Raettig & Kotz, 2008; Rissman, Eliassen, & Blumstein, 2003; Sabri et al., 2008; Xiao et al., 2005). The lexical frequency effect could be explained by assuming that high frequency words will be more strongly activated at the lexical level via stronger connection weights between the lexical and phonological levels of language processing (Besner & Risko, 2016). Alternatively, the lexical frequency effect can also be explained via lateral excitatory connections between lexical nodes which are supposed to be stronger for high frequency words as these words also co-occur more often (Hulme et al., 2003). The neighborhood density effect naturally emerges from interactive activation models: it is a direct consequence of the interactive activations occuring between the phonological and lexical levels of language processing. Words located in dense neighborhoods will be more strongly activated due to increased redundant feedback activation between neighboring items as compared to words surrounded by a small number of neighbors. Interactive activation models can also deal with semantic effects. The imageability effect is supposed to occur because high imageability words have more or richer semantic features (Martin & Saffran, 1992; Pexman, Lupker, & Hino, 2002; Yap, Lim, & Pexman, 2015). When activated, they are supposed to send more activation back to lower levels of processing, thereby producing a recall advantage for high as compared to low imageability words. Likewise, semantically related words are supposed to share common semantic features, and will reactivate each other via redundant feedback activation, thereby more strongly counteracting the deleterious effect that decay has on WM representations.

When framed through interactive activation models, the sometimes complex language and WM profiles of many aphasic patients can also be explained at small costs. One of the most illustrative examples is the case of deep dysphasia. This (rare) aphasic syndrome is characterized by oral language comprehension difficulties with a marked concreteness effect (advantage for concrete over abstract words), by the production of semantic paraphasias during single word repetition (e.g., 'sun' repeated

as 'planet') and by strongly impaired nonword recall, as well as severe verbal WM deficits (Majerus, Lekeu, de Linden, & Salmon, 2001; N. Martin & Saffran, 1992). N. Martin and Saffran showed that this complex profile can be explained in their interactive activation architecture by assuming a deficit of a single component: an abnormally increased decay rate of activated language representations. In this architecture, when words are presented auditorily, phonological representations will be the first to decay (as they are the first to have been activated), thereby reducing the subsequent feedback from phonological to lexical levels of representations. Hence, performance on comprehension tasks will rely to a stronger extent on feedback from semantic to lexical levels of representation, which are supposed to be activated later than phonological representations and will therefore have less strongly decayed at the moment of response selection. This will lead to a strong impact of lexico-semantic variables on both language processing and verbal WM tasks. For example, the semantic paraphasias in single word repetition can be explained by the combined effects of reduced feedback from phonological level and stronger feedback from semantic level, allowing for a semantically related lexical competitor to be selected instead of the target word. Finally, a higher decay rate will also naturally produce a global verbal WM impairment.

One challenging question for language-based models is however the question of serial order and the way it will interact with language knowledge. For example, the models by Majerus (2013, 2018) and Burgess and Hitch (1999, 2006) assume direct links between the linguistic and serial order levels, and hence, at the theoretical level allow for interactions between these different levels. Although we have shown in the previous section that linguistic knowledge leads to a benefit at the level of item recall, this is often considered not to be the case for serial order recall measures. However, there are currently a few studies providing some preliminary evidence for interactions between linguistic activation and serial order processing. For instance, as regards the lexicality effect, it has been shown that, while item recall is increased for words as opposed to nonwords, words are actually more often recalled in a wrong serial position relative to nonwords (after controlling for the fact that overall a larger amount of words than nonwords is recalled) (Fallon, Mak, Tehan, & Daly, 2005; Guérard & Saint-Aubin, 2012; Jefferies et al., 2006a; Saint-Aubin & Poirier, 1999b). Likewise, it has been shown that similarity effects, either at the phonological or semantic level, increase the number of items recalled in WM tasks (Gupta, Lipinski, & Aktunc, 2005; Poirier & Saint-Aubin, 1995), but decrease recall performance at the serial order level. This has been shown at least as regards phonological similarity, evidence for an interaction

between serial order processing and semantic similarity being much more subtle. (e.g., Tse, Li, & Altarriba, 2011).

Finally, language-based models of WM should also be assessed by exploring the full range of psycholinguistic effects that have been identified in the linguistic domain. Although many psycholinguistic effects have been shown to impact verbal WM performance, some effects remain unexplored. One such effect is the lexical cohort effect, or cohort competition effect, which reflects the competition among items within the lexicon. Typically, words sharing their initial cohort (e.g. /al/ in /aligator/, /almond/, /aliv/) with many competitors are identified more slowly in language processing tasks (Tyler, Voice, & Moss, 2000). This phenomenon reflects the fact that the activation of lexical representation is performed in parallel, as illustrated in **Figure 1.8**. In large cohorts, the selection of the target word is more ambiguous, because an important number of items compete for activation in parallel during the lexical selection process. The fact that words drawn from large cohorts are more difficult to activate is supported by the observation of interactions between cohort competition and word meaning in lexical decision tasks, with effects of word imageability only observed for words drawn from larger cohorts (Tyler et al., 2000; Zhuang, Randall, Stamatakis, Marslen-Wilson, & Tyler, 2011). When lexical activation is slow to settle on a specific word form (due to many words competing for selection), the semantic features of high imageability words provide an additional source of information to facilitate selection at the lexical level (Evans, Lambon Ralph, & Woollams, 2012). This effect of cohort competition is also associated with a stronger recruitment of inferior frontal regions associated with selection and retrieval processes (Kocagoncu, Clarke, Devereux, & Tyler, 2017; Zhuang et al., 2011). Very interestingly, contrary to other psycholinguistic effects, the lexical cohort effect is characterized by differences in speed of lexical access/selection, not by the inherent richness of linguistic representations. Language-based models predict that words drawn from large cohorts should also be activated more slowly in verbal WM tasks, and it could therefore be argued that this dimension should also impact WM recall performance.

In sum, language-based models offer a very parsimonious and straightforward explanation of the different psycholinguistic effects that have been observed in verbal WM tasks. Multi-component versions of these models can also account for other fundamental aspects of WM such as the representation of serial order processing. At the same time, the interactions between serial order processing and linguistic components require deeper investigation. Also, the occurrence of all psycholinguistic effects in WM tasks has not yet been examined so far.

```
                                          ╱→ /bark/
                                        ╱  → /baf/
                                      ╱   → /bat/
        /ba…/  ●─────────────────────────→ /bal/
                                      ╲   → /bato/
                                        ╲  → /barak/
                                          ╲→ /barb/
```
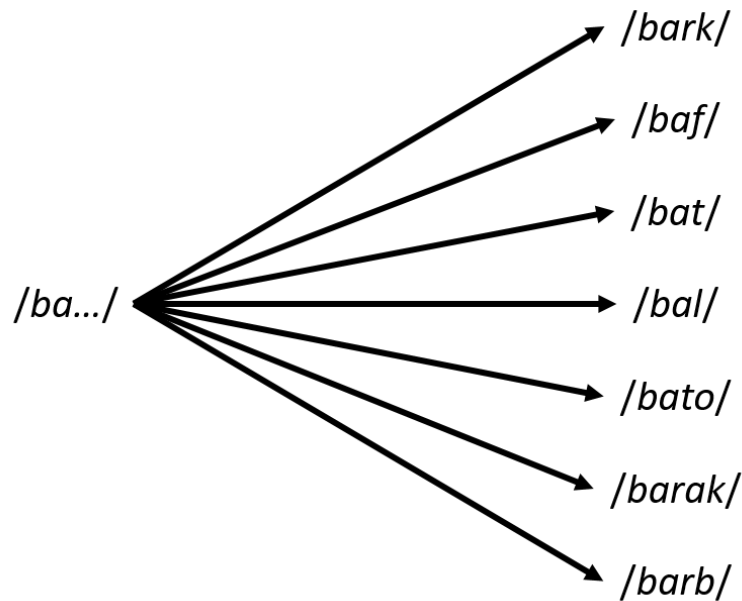
Figure 1.8. Illustration of the cohort competition effect.

How to dissociate the different accounts of linguistic LTM effects in verbal WM?

In the previous sections of this chapter, three different accounts of psycholinguistic effects have been presented. A first account considers that psycholinguistic effects occur because participants implement slow controlled strategic processes during the encoding stage of WM that favour semantic elaboration of verbal memoranda when possible. A second account, the redintegration framework, considers that these influences stem from a post-encoding clean-up process that reconstructs the degraded phonological traces maintained via a phonological buffer system. Finally, the language-based account considers that psycholinguistic effects are the consequence of the interactions occuring directly and in an automatic manner within the linguistic system. These different accounts continue to co-exist in the WM literature and are all supported, to some extent, by empirical evidence. We currently do not have firm evidence that would enable us to favour a specific account over another account, or to determine that psycholinguistic effects could involve multiple mechanisms at the same time, potentially deployed in a context-dependent manner. In this section, we will discuss possible methods to disentangle the different accounts of psycholinguistic effects that have been proposed. Up to now, no systematic investigation has been made in order to deconfound these different hypotheses. This investigation will be the aim of the current PhD thesis.

First, regarding the strategic account, we should note that most psycholinguistic effects have been tested in paradigms in which strategic processes, such as researsal, mental imagery and semantic elaboration are very likely to occur (Morrison et al., 2016), but less consistently in paradigms preventing the implementation of such processes. In addition, early investigations of the strategic account, when aiming at preventing the intervention of linguistic encoding and elaboration strategies, may not have used optimal procedures. For instance, Campoy et al. (2015) assessed the occurrence of the imageability effect under a dual-task paradigm considered to prevent strategic encoding. But dual tasks also lead to interference effects between the memory task and the concurrent non-memory task, making the interpretation of their results (irrespectively of the observation of the imageability effect or not) very difficult. For example, the reduction of performance observed in the study by Campoy et al. in the dual-task paradigm could have reflected the lack of intervention of strategic linguistic encoding processes, but also the intervention of between-task interference effects (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). Due to this situation, dual tasks may not be the optimal paradigm for testing the intervention of strategic processes. A possibility to test the strategic account in a more direct manner while also avoiding interference effects could be the use of the running span procedure (Pollack, Johnson, & Knaff, 1959). In this procedure, participants attend to continuous lists of words of unpredictable length and presented at a very fast rate (i.e. 2 – 3 items every second). Participants are usually invited to recall the items they have in mind when the recall instruction appears, that is the most recent items, as far as they can remember, but in *forward* order. Under these encoding and recall conditions, it has been shown that participants adopt a "passive listening" attitude toward the task instead of trying to constantly rehearse or actively maintain the items (Morrison et al., 2016; Palladino & Jarrold, 2008). Hence, the running span procedure appears to be a promising procedure for preventing the implementation of strategic processes during encoding of memoranda. Indeed, in these paradigms, the digit span estimate is about 4 items (Bunting et al., 2006) instead of the usually observed estimate of ~7 digits (Miller, 1956). This finding has been interpreted by the fact that participants are unable to chunk items or use other elaborative processes during encoding of memoranda in a running span procedure as compared to standard immediate serial recall procedures (Cowan, Johnson, & Saults, 2005). Importantly, the running span procedure has the further advantage of avoiding any task-related interference effects as compared to dual-task paradigms.

When using a running span procedure for assessing the presence of psycholinguistic effects in a WM context, several predictions can be made. Most importantly, for the strategic account, all psycholinguistic effects should disappear when using a running span procedure given that all of these effects are considered to stem from linguistic elaboration strategies, be it at the semantic or phonological level. Indeed, during this procedure, participants will be maximally prevented from implementing strategic processes such as mental imagery and semantic elaboration. On the other hand, for language-based models, the influence of linguistic knowledge is supposed to be direct, automatic and purely non-strategic and hence all psycholinguistic effects should still be observed when using a running span procedure instead of a standard immediate serial recall task. The redintegration account would also predict a persistence of psycholinguistic effects, given that redintegration is considered to occur only at the recall stage, which, in a running span procedure, remains unchanged relative to an immediate serial recall task.

In sum, the use of a running span procedure allows to dissociate the strategic account from the two other accounts, but does not allow to distinguish between the language-based and redintegration accounts. A first manipulation allowing to disentangle the two latter accounts could be the comparison of recall and recognition procedures for probing memory content in WM tasks. When information does not need to be fully recalled, the original redintegration framework predicts that psycholinguistic effects should disappear, because in the absence of overt recall output, the redintegration mechanism is not supposed to occur (Gathercole, Pickering, Hall, & Peaker, 2001; Jefferies, Frankish, & Lambon Ralph, 2006b). For language-based models however, psycholinguistic effects should be reliably observed in any type of verbal WM tasks, irrespective of type of response modality, since stabilizing feedback activation from lexico-semantic knowledge is considered to occur at any stage of WM processing and, most importantly, already during encoding. Previous studies have shown that in paradigms not requiring full recall of memoranda, such as item recognition and matching-span tasks, reliable psycholinguistic effects can actually be observed, which does not support the redintegration account (Jarrold, Cocksey, & Dockerill, 2008; Jefferies et al., 2006b; Turner & Henry, 2004). However, it could be argued that the redintegration mechanism is not specific to overt recall or the recall phase, but can already be performed during encoding itself through rehearsal during inter-item intervals, or even covertly during the recognition procedure if sufficient time is allowed for participants to do so (Jefferies et al., 2006b; Turner & Henry, 2004); note that this assumption is already going slightly beyond the canonical version of the redintegration account which considers that redintegration is limited to the stage of

recall. Previous studies used recognition paradigms in which sufficient time was available for participants to covertly reconstruct the memoranda at both the encoding and recognition stages as items were presented at the relatively slow presentation rate that characterizes most verbal WM tasks (Jarrold et al., 2008; Jefferies et al., 2006b; Turner & Henry, 2004). A strong test of the redintegration hypothesis should therefore prevent participants from rehearsing items or covertly reconstructing phonological traces during all WM stages. In order to provide this test, we could actually combine the running span procedure discussed above with a fast item probe recognition procedure in which participants must decide as fast as possible whether a probe item matches an item in the memory list without any opportunity for redintegration at any WM stage. Only direct redintegration at the perceptual level could still occur at the moment when an item is presented by cleaning up noisy perceptual input based on existing phonological categories in the language system ; however, this kind of redintegration, also referred to as 'predictive coding' in the psycholinguistic literature, is considered to be an integral part of the language processing system itself (Hannemann, Obleser, & Eulitz, 2007; Heald & Nusbaum, 2014; Leonard, Baud, Sjerps, & Chang, 2016; Sohoglu, Peelle, Carlyon, & Davis, 2012). This form of perceptual redintegration allows to identify speech percepts in an efficient and fast manner despite the extreme variability that characterizes the acoustic envelope of speech stimuli.

This type of paradigm also has the potential to distinguish the different accounts of psycholinguistic effects at the neural level. Indeed, the three accounts are not equivalent as regards the neural substrates and their timing when processing words and nonwords during a running span procedure. First, the strategic account predicts that there should be no difference between the neural activity elicited by word and nonword items during the presentation of memoranda (encoding), because in the absence of strategic elaborative processes, lexico-semantic knowledge will have only a limited opportunity to influence verbal memoranda. Second, the redintegration account also predicts that neural activity between word and nonword memoranda should not differ, because they should be encoded and maintained only at the phonological level, and hence word and nonword should be processed in an equivalent manner, at least during encoding and maintenance stages. Third, language-based models predict that a clear difference of neural substrates involved in word and nonword processing should already be observed at the encoding and maintenance stages, due to the direct and obligatory access of lexico-semantic knowledge at any stage of processing of verbal memoranda. It should be noted here that even within the psycholinguistic research field, there is currently a controversy as regards the

involvement of specific neural regions associated to word and nonword processing. Previous studies directly comparing the neural activation between words and nonwords in linguistic tasks have shown the involvement of strongly overlapping neural regions or distinct neural substrates within language cortices (Davis & Gaskell, 2009; Kotz et al., 2002; Newman & Twieg, 2001; Orfanidou et al., 2006; Raettig & Kotz, 2008; Rissman et al., 2003; Sabri et al., 2008; Xiao et al., 2005). These neuroimaging studies are however difficult to interpret as they used tasks in which the processing of word and nonword stimuli may have been difficult to compare. For example, in lexical decision tasks, nonwords may recruit additional neural regions because they are also more difficult to process (Graves, Boukrina, Mattheiss, Alexander, & Baillet, 2016; Mattheiss, Levinson, & Graves, 2018). Studies in which word and nonword items were presented without any specific judgement requirements showed an absence of difference in terms of overall neural activation (Binder et al., 2000). Critically, none of these studies has used more sensitive multivariate neural pattern analysis techniques which do not simply look at locally elevated voxel responses but which consider the distribution of activity patterns over a set of voxels. Words and nonwords may not involve distinct neural substrates: they are all language stimuli after all and will elicit elevated activity in the language system but their representations within the language system should be different, and these differences in representational content can be investigated by comparing multivariate signals of voxel patterns[1] for words and nonwords over the linguistic cortices (Hebart & Baker, 2018).

Chapter summary

To sum up, in this PhD thesis, we will disentangle different hypotheses that have been proposed to account for the presence of psycholinguistic effects during WM tasks. These accounts include a strategic account whereby psycholinguistic effects are the result of controlled strategic processes, a redintegration account assuming that the

---

[1]Contrary to classical univariate approaches in which a signal can be detected only if a difference in the intensity of brain activity is triggered, multivariate approaches assess the *pattern* of voxel activation within brain regions. As such, no elevated neural activity is required in order to extract information stemming from the BOLD signal. These analyses are usually done by training/fitting a classifier to differentiate between two or several conditions, and then by assessing the performance of the classifier by testing it on a new set of data. This can also be performed using alternative procedures. For instance, one can use a leave-one-out cross-validation procedure, which consists in (1) discarding a trial from the whole dataset, (2) training the classifier on the dataset containing N-1 trials, (3) determining whether the discarded trial is correctly classified and (4) repeating the process across all trials. In all cases, this results in a percentage representing classification accuracy for each individual participant.

presence of psycholinguistic effects is the consequence of a post-encoding reconstruction process, and finally language-based models assuming direct interactions with the linguistic system. These different theoretical accounts will be confronted by manipulating different psycholinguistic effects in fast presentation running span procedures, using both behavioural and functional neuroimaging study designs.

Experimental part

Goals and hypothesis

Although a substantial number of studies indicate an effect of linguistic knowledge on verbal WM, the nature of this effect and its locus remains highly controversial. The theoretical accounts that have been proposed consider either fast, automatic effects arising directly within the language network, optional controlled strategic processes or post-memory reconstructive processes. Furthermore, an important debate concerns the question of whether the effects of linguistic knowledge include all aspects of WM (including serial order processing) or whether they are strictly limited to the retention of item information. The aim of this PhD thesis is to confront the different accounts that have been proposed to explain the influence of linguistic knowledge in verbal WM and to better characterize the breadth and nature of linguistic effects on verbal WM.

In **Study 1**, we tested the strategic account of linguistic effects, and this for both lexical and semantic knowledge. In standard WM tasks, participants may intuitively implement several types of strategies in order to support their performance, such as mental imagery or semantic elaboration. In order to assess whether these strategies are necessary for the observation of psycholinguistic effects in verbal WM tasks, we used a running span procedure, involving the rapid presentation of memoranda in lists of unpredictable length and designed for preventing the intervention of encoding strategies. Several psycholinguistic effects were assessed, including lexicality, lexical frequency, semantic similarity and imageability. The lexicality and lexical frequency variables are assumed to assess linguistic influences at the level of lexical processing, while the semantic similarity and imageability effects are supposed to occur at the semantic level. This study allowed us to investigate in a systematic manner whether all psycholinguistic effects appear under non-strategic encoding conditions, and if not, under which specific conditions they may begin to appear. This study will allow us to assess the plausibility of the strategic account as an explanation of psycholinguistic effects in WM.

In **Study 2**, we examined the plausibility of the redintegration hypothesis as the main and only account of psycholinguistic effects in verbal WM, by using a novel method aimed at maximally reducing the likelihood of reconstructive processes potentially occurring during a WM task. In this study, we combined the running span procedure and a fast item probe recognition procedure for word and nonword lists. Participant heard a sequence of words or nonwords of unpredictable length presented at a very fast pace, aimed at preventing inter-item reconstructive processes,

in addition to preventing strategic encoding processes. At the end of the memory list, participants were invited to judge as fast as possible whether a probe item matched an item of the memory list or not, and this within a very limited amount of time (i.e. 1750 ms post-stimulus onset) in order to minimize post-memory list reconstruction processes. We reasoned that if a lexicality effect is still present in this type of paradigm, as characterized by higher recognition performance for word lists than nonword lists, redintegration processes as assumed by Schweickert (1993) and Hulme et al. (1991, 1997) are not a necessary condition for psycholinguistic effects such as the lexicality effect to occur.

In **Study 3**, we used a functional neuroimaging approach in order to tease further appart redintegration and language-based accounts of psycholinguistic effects in verbal WM. The aim of this study was to track the neural representations associated with word and nonword processing over the encoding and maintenance stages for stimuli presented again using a running span procedure. As for Study 1 and Study 2, we used a combined running span – probe recognition paradigm in order to minimize the intervention of redintegration and strategic processes during the encoding stage. Furthermore, contrary to previous neuroimaging studies, we used a multivariate pattern analysis approach in order to directly track the distributed neural representations that may distinguish words and nonwords, rather than simply exploring local univariate activity changes in language cortices. More specifically, while lying in an fMRI scanner, participants were presented with lists of unpredictable length composed of words or nonwords using a fast presentation procedure. At the end of the memory list presentation, participants were instructed to either actively maintain the memoranda, or to just rest, and this in order to have a better assessement of the specific stages at which neural patterns begin to differentiate word and nonword stimuli. According to the strategic and redintegration accounts, no difference between words and nonwords should be observed during the encoding stage. Indeed, the strategic account predicts that lexico-semantic knowledge should not have the opportunity to intervene if strategic and elaborative encoding processes are prevented due to the fast presentation paradigm of the memoranda. Likewise, the intervention of reconstruction mechanisms should also be very limited during the encoding stage of the running span procedure, leading to an absence of difference in neural patterns between word and nonword items given that according to the redintegration account, memoranda are exclusively encoded and maintained via phonological codes. Only the language-based account predicts that a difference between word and nonword items should be detected in the form of distinct multivariate neural patterns already at the encoding stage, and, critically, this differentiation should be observable all over the

maintenance delay but should disappear as soon as participants are instructed to not maintain the information anymore.

In **Study 4**, we assesed a psycholinguistic effects which has never been manipulated in WM tasks so far, namely the lexical cohort or cohort competition effect. This effect refers to the number of lexical representations competing for activation over the timecourse of the lexical selection process. More specifically, words sharing their first phonemes with many other words (e.g. *al*cove, *al*ligator, *al*cohol, …) are usually responded to more slowly than words sharing few phonemes with other words early during language processing, which reflects the ambiguity of lexical access (Marslen-Wilson, 1987). This is a good test of language-based models, because these models assume that variables influencing language processing in purely linguistic tasks should also influence the processing of verbal stimuli in WM tasks. The examination of the lexical cohort effect is particularly interesting as it involves a linguistic dimension that has not yet been explored in terms of its potential impact on WM performance: speed of access. Indeed, contrary to other psycholinguistic effects such as the imageability or the lexicality effect, words stemming from larger lexical cohorts are not characterized in terms of richness and robustness of lexico-semantic representations relative to words stemming from smaller lexical cohorts, but they are accessed by the rapidity of lexical access, with words from *smaller* cohorts leading to faster access. If speed of lexical access also characterizes WM performance, then we should observe an advantage for words from small cohorts in immediate serial recall tasks.

In **Study 5**, in order to assess more fully the breadth and functional importance of the impact of psycholinguistic knowledge on verbal WM performance, we assessed the extent to which the availability of psycholinguistic knowledge may not only facilitate the immediate maintenance of verbal memoranda, but may also have a more active role in protecting verbal memoranda against longer term interference and forgetting. We examined this situation by presenting to participants verbal memoranda differing in terms of semantic content (semantic relatedness, word imageability) and by asking them to recall the stimuli directly after encoding of the word list, or only after a secondary task had been performed (backward countint task). If semantic knowledge protects memoranda against the deleterious effect of task-related interference (i.e., the backward counting task), recall of semantically related and high imageability words should be less impacted by the interfering task, as compared to recall of semantically unrelated and low imageability words.

In **Study 6**, we also assessed the breadth of the impact of psycholinguistic knowledge on verbal WM performance, by addressing the critical question of the

impact of linguistic knowledge on recall of serial order information, as opposed to recall of item information. As we have seen, the impact of language knowledge on recall of item information is well established but its impact on serial order recall is much less understood. The investigation of this question is of fundamental importance for language-based models of WM that explicitly or implicitly acknowledge interactions between language components and the storage of (arbitrary) serial order information. In this final study, the semantic relatedness effect was manipulated and its impact on serial order recall performance was investigated in a detailed manner based on an analysis of the specific patterns of serial order errors for recall of semantically related versus unrelated word lists. We examined whether the semantic associations constrain serial position migration errors: in line with language-based models of WM (Burgess & Hitch, 1999; Majerus, 2018), we predicted that these errors should preferentially involve migrations between words from the same semantic group. This observation would support these integrative, language-based models of verbal WM by showing that semantic knowledge not only supports recall of item information, as shown in previous studies, but has also a direct and specific effect on the encoding (or retrieval) of serial position information.

Study 1

The non-strategic nature of linguistic long-term memory effects in verbal short-term memory

Benjamin Kowialiewski and Steve Majerus

Abstract. The contribution of lexical and semantic knowledge to verbal short-term memory (vSTM) span is explained by language-based models, assuming that vSTM is deeply grounded within the linguistic system with to-be-remembered items being activated in a non-strategic and automatic manner. However, direct evidence for a non-strategic account of lexical and semantic contributions to vSTM span is scarce. In this study, we assessed the influence of several types of long-term linguistic knowledge (lexicality, lexical frequency, semantic similarity and imageability) on vSTM using a fast encoding running span procedure preventing any strategic processes during encoding. We observed reliable effects of lexicality (words vs. nonwords, Experiment 1), lexical frequency (high vs. low frequency words, Experiment 2) and semantic similarity (related vs. unrelated lists, Experiment 3) on running span performance. However, word imageability (high vs. low imageability words, Experiment 4) did not consistently impact running span performance. Experiment 5 showed that the imageability effect only appears in standard immediate serial recall conditions which do not prevent list-strategic encoding. This study provides novel evidence for linguistic accounts of vSTM by demonstrating a robust impact of lexical and surface-level semantic knowledge on vSTM in non-strategic, fast-encoding conditions.

Introduction

Verbal Short-Term Memory (vSTM) and language processing have been shown to strongly interact, with many properties that define the linguistic system also impacting vSTM performance (Guérard & Saint-Aubin, 2012; Majerus et al., 2010; N.

Martin & Saffran, 1997; Patterson et al., 1994; Romani et al., 2008). This situation has been explained by language-based models which assume that vSTM is the by-product of the temporary activation of phonological, lexical and semantic long-term knowledge stored in the linguistic system (Acheson & MacDonald, 2009; Gupta, 2009; Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999). These models implicitly assume that activation of linguistic knowledge in vSTM tasks operates in a non-strategic and automatic manner. The aim of this study was to test this fundamental assumption by determining the extent to which linguistic long-term memory effects appear in non-strategic encoding conditions of vSTM tasks.

VSTM and linguistic knowledge are two closely related processes as observed by strong similarities between language processing and vSTM (Gupta, 2009; Majerus, 2013; Majerus, Martinez, et al., 2012; Majerus, Van der Linden, Mulder, et al., 2004; N. Martin & Saffran, 1990, 1997; N. Martin et al., 1996; R. C. Martin et al., 1999). First, verbal stimuli that activate robust representations in the language system are also associated with better performance in vSTM. Higher vSTM recall accuracy has been observed for nonwords characterized by phonotactic probabilities that are frequent in the native language of the participants (Gathercole et al., 1999; Majerus, Van der Linden, Mulder, et al., 2004), for words over nonwords (Brener, 1940; Hulme et al., 1991; Jefferies et al., 2006a) and for high frequency over low frequency words (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Watkins & Watkins, 1977). Other studies have also reported effects specific to semantic knowledge, showing that lists composed of words drawn from similar *vs.* dissimilar semantic categories lead to better vSTM span (Monnier & Bonthoux, 2011; Poirier & Saint-Aubin, 1995; Tse, 2009; Tse et al., 2011). Likewise, vSTM span is better for high over low imageability words (Acheson, Postle, & MacDonald, 2010; Campoy et al., 2015; L. M. Miller & Roodenrys, 2009; Romani et al., 2008; Walker & Hulme, 1999). Hence, many psycholinguistic effects that are known to affect the language system also appear to affect vSTM (Anderson & Holcomb, 1995; Connine, Mullennix, Shernoff, & Yelen, 1990; Tyler et al., 2000). This is further supported by neuropsychological studies showing a strong association between vSTM and language deficits in language-impaired patients (Majerus et al., 2007; N. Martin & Saffran, 1990, 1997; N. Martin et al., 1996; R. C. Martin et al., 1999; Papagno, Vernice, & Cecchetto, 2013; Patterson et al., 1994). Neuroimaging studies have also observed that neural regions supporting phonological and semantic processing are activated during all stages of vSTM tasks (Fiebach et al., 2007; Majerus et al., 2010).

This sensitivity of vSTM tasks to psycholinguistic variables is accounted for by language-based models of vSTM (Acheson & MacDonald, 2009; Gupta, 2009; Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999). These models consider that vSTM

reflects the direct and automatic activation of underlying phonological, lexical and semantic representations in a manner identical to what is happening in other language tasks. In auditory speech perception tasks, it has been shown that lexical and semantic levels of representations are activated in less than 150 ms after a spoken word's onset (MacGregor, Pulvermuller, van Casteren, & Shtyrov, 2012; Moseley, Pulvermu, & Shtyrov, 2013; Shtyrov & Lenzen, 2017), leading to the assumption that lexico-semantic activation occurs very rapidly and automatically (Dell, 1986; MacGregor et al., 2012; Marslen-Wilson, 1987; McClelland & Elman, 1986; Shtyrov & Lenzen, 2017). This is also in line with studies in the visual STM domain using Rapid Serial Visual Presentation (RSVP) procedures. Potter (1976) showed that conceptual knowledge can be accessed very rapidly, up to 13ms (Potter, Wyble, Hagmann, & McCourt, 2014) in a visual object STM recognition task. However, no equivalent studies currently exist for the auditory-verbal STM domain.

In the vSTM domain we know that the impact of linguistic knowledge is robust as it has been observed in different experimental conditions. Lexicality, word frequency, semantic similarity and imageability/concreteness effects have been shown to occur in various experimental procedures including forward recall paradigms (Besner & Davelaar, 1982; Hulme et al., 1991; Walker & Hulme, 1999), recognition paradigms (Gathercole et al., 2001; Jarrold et al., 2008; Jefferies et al., 2006b; Romani et al., 2008; Tse et al., 2011), mixed word/nonword and high frequency/low frequency lists recall paradigms (Hulme et al., 2003; Jefferies et al., 2006a; L. M. Miller & Roodenrys, 2012), backwards recall paradigms (Guérard & Saint-Aubin, 2012) and serial reconstruction tasks (Quinlan, Roodenrys, & L. M. Miller, 2017; Romani et al., 2008). At the same time, the automatic, non-strategic nature of these interactions has not yet been demonstrated in an unambiguous manner as the different vSTM tasks in which linguistic long-term memory effects have been observed allow for slow, strategic processing during encoding which could have inflated at least some of the psycholinguistic effects observed in these vSTM tasks. Several studies have highlighted the role of encoding strategies in standard vSTM tasks, such as associating a visual image to memoranda (visual imagery), remembering items in groups (grouping) or voluntarily focusing on the meaning of memoranda (semantic elaboration) (Bailey, Dunlosky, & Kane, 2011; Corbin & Marquer, 2009; Dunlosky & Kane, 2007; Logie et al., 1996; Morrison et al., 2016). The use of mental imagery and semantic elaboration strategies could theoretically increase the impact of lexico-semantic variables such as lexicality and word concreteness on vSTM performance. Campoy and Baddeley (2008) observed that participants using a semantic encoding

strategy showed indeed reduced phonological effects in vSTM tasks (Campoy & Baddeley, 2008).

More directly, Shulman (1970) observed that a semantic encoding strategy led to enhanced vSTM performance only when stimuli were presented at a slow rate, leading him to consider that semantic effects in vSTM result from slow, strategic encoding processes. A second, more recent study by Campoy, Castellà, Provencio, Hitch, & Baddeley (2015) led to similar results. The authors manipulated the word concreteness effect under slow (0.5 items/s) versus standard (1 item/s) presentation rates. They observed that concrete words benefited more from slow presentation rates as compared to abstract words, suggesting that semantic activation may at least partially be the result of strategic elaboration. At the same time, the concreteness effect still occurred in a dual task-paradigm reducing strategic encoding processes, indicating that the concreteness effect may not be exclusively the result of strategic processing.

The aim of this study was to examine a core prediction of language-based models, by assessing in a systematic manner whether the main psycholinguistic effects studied so far still appear in vSTM tasks that prevent strategic encoding processes. In order to achieve this goal, we used a running span procedure. In running span procedures (Pollack et al., 1959), subjects attend to continuous sequences of spoken items presented at a very fast rate, up to 3-4 items every second, and of unpredictable length. At the end of each list, subjects are invited to recall the items they still have in mind, which typically are the most recent items of the continuous sequence. Even with comfortable presentation rates (e.g., 1 or 2 seconds per item), it has been shown that participants tend to use a "passive listening" attitude toward the task (Botto, Basso, Ferrari, & Palladino, 2014), or just try to "concentrate" on the items during encoding as compared to other standard recall tasks (Morrison et al., 2016). It has further been shown that the attempt of using encoding strategies such as rehearsal in fast encoding paradigms deteriorates vSTM performance (Hockey, 1973). Fast encoding procedures have also been shown to be particularly effective in preventing the use of list segmentation and inter-item grouping strategies (Cowan, 2001; Cowan, Elliott, et al., 2005). Due to the very fast presentation rate, each attended item is activated only briefly as subjects cannot refresh them during encoding and maintenance. Using this procedure, the usually observed span size of 7 digits (G. A. Miller, 1956) drops to 4 (Bunting, Cowan, & Saults, 2006; Cowan, 2001; Cowan, Elliott, et al., 2005). If linguistic representations are activated in a non-strategic manner in vSTM tasks, as predicted by language-based models, then we should observe preserved psycholinguistic effects even under these encoding conditions. If on the other hand, the impact of psycholinguistic knowledge in vSTM tasks is the result of strategic lexico-semantic

elaboration of memoranda, psycholinguistic effects should not be observed under fast encoding conditions. So far, no study has used a running span procedure to test the impact of linguistic knowledge on vSTM.

We conducted five different experiments in which we manipulated, using a running span recall procedure, the psycholinguistic effects most typically studied in the verbal STM literature. In Experiment 1, we manipulated the lexicality effect (words *vs.* nonwords), which is the most commonly studied and most robust psycholinguistic variable impacting vSTM and reflects the influence of both phonological-lexical and semantic knowledge. Experiment 2 manipulated the word frequency effect (high *vs.* low frequency words). In Experiment 3 through 5, we assessed more directly the impact of semantic knowledge, by manipulating the two most commonly studied semantic variables, semantic relatedness (related *vs.* unrelated semantic categories) and imageability (high *vs.* low imageability words). Note that the running span procedure used here differs from backward recall procedures as participants are instructed to recall all the items they have in mind in *forward* serial order; the fact that in this task participants recall mostly items from the end of the list is the consequence of items from earlier positions being increasingly more difficult to maintain as the number of presented items augments.

Experiment 1

Experiment 1 aimed at assessing the impact of lexico-semantic knowledge on a vSTM task using a fast-presentation, running span procedure. Lexical-semantic knowledge is considered to bind the word's constituent phonemes into coherent units (Jefferies et al., 2006a; Patterson et al., 1994; N. Savill et al., 2016), thereby enhancing the stability of word versus nonword representations at encoding and recall, and leading to a lexicality effect.

We made two predictions: First, if the impact of lexico-semantic knowledge on vSTM performance does not require slow, strategic encoding of items, a difference in recall performance for words versus nonwords should be observed. Second, in line with previous studies using the running span procedure, we expected strong recency but near-to-zero primacy effects, indicating that participants are not able to rehearse and elaborate on earlier presented items during encoding (Bunting et al., 2006; Fiore, Borella, Mammarella, & De Beni, 2012; Palladino & Jarrold, 2008; Ruiz & Elosúa, 2013).

Method

Participants. Thirty-two undergraduate students (24 females and 8 males) aged between 19 and 28 years (*M* = 22.81, *SD* = 2.21), recruited from the university

community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had received approval from the local ethics committee.

Materials. A set of 270 CVC words was selected from a French lexical database (New, 2006). Because the stimuli used in this experiment had to be as short as possible, we restricted our pool of stimuli to monosyllabic words. We chose the most frequent nouns available in the database (M = 100.69 count per million). Phonologically ambiguous words or homonyms were also excluded (e.g., "vers – verre – vert" in French). We then pseudorandomly combined French consonant and vowels using an algorithm programmed under MATLAB® in order to create a larger set (N > 1000) of phonotactically regular CVC nonwords. The algorithm automatically discarded items that matched existing lexical entries in the database, based on their phonological form. We then extracted from this set the nonwords that matched the biphone frequency values of the word stimuli (Tubach & Boë, 1990, M = 742.79, SD = 570.42 and M = 742.95, SD = 571.90 for words and nonwords, respectively). All the stimuli were recorded by a French-native female speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The length of each item was normalized to 375ms without altering the pitch.

Procedure. There were four versions of the experimental task. The word and the nonword conditions were presented in a blocked design, with each of the two blocks being composed of 30 trials, with order of presentation being counterbalanced across the 4 different versions of the task. Lists were identical in versions 1 and 2, except that, in version 1, participants received first the word block condition and, in version 2, they received first the nonword block condition. Lists from versions 3 and 4 were constructed using the same stimuli, but with items presented in a different order as in versions 1 and 2. Participants received the word block condition first in version 3 and the nonword block condition first in version 4. All participants performed 5 practice trials for each condition.

Each trial began with an on-screen countdown starting from 3, announcing the beginning of the trial. This countdown was followed by a blank screen and an auditory list composed of 6, 9 or 12 items. The presentation order of the different list lengths was pseudorandomised, with the constraint that a given list length was not repeated more than twice on two consecutive trials. The presentation rate was fixed to 2.5 items per seconds, with an inter-stimulus interval of 25ms. The end of each list was signaled by a brief (135ms) sinusoidal tone of 1952 Hz, prompting the subject to recall aloud the sequence. Participants were instructed to recall the last items of the list, as far as they

could remember, in the *initial order of presentation*, and to substitute any item that could not be remembered with the word "blanc" ("blank" in French). For instance, given the target sequence "item1 – item2 – item3 – item4 – item5 – item6 – item7 – item8 – item9", a correct output response could be "item6 – item7 – item8 – item9" or "item8 – item9". Participants were informed to recall those items that were immediately available in their mind at the moment of the recall cue. When participants had finished recalling the items they remembered, they were invited to press the SPACEBAR of the keyboard to initiate the next trial. Total experiment duration was approximatively 20 minutes.

Task presentation and timing was ensured using OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) running on a desktop station computer. The auditory stimuli were presented at comfortable hearing loudness via headphones connected to the computer in a soundproof booth. Participant's responses were recorded using a digital recorder for later transcription and scoring.

In order to ensure that the scoring procedure was not affected by perceptual and articulatory planning errors, we allowed distortions covering one articulatory transformation on the whole item, as used in other studies of nonwords repetition (Dollaghan, Biber, & Campbell, 1993; Dollaghan & Campbell, 1998; M. W. Moore, Tompkins, & Dollaghan, 2016; Poncelet & Van der Linden, 2003). The list of allowed distortions is provided in Appendix A.

We performed three different scoring procedures. The first was an item recall scoring procedure, in which an item was scored as correct even when the serial output of the item did not correspond to the initial presentation order. For instance, given the target sequence "… Item 9 – Item 10 – Item 11 – Item 12", and the response sequence "Item 9 – Item 11 – Item 12" or the response sequence "Item 11 – Item 9 – Item 12", items 9, 11 and 12 were scored as correctly recalled. This scoring procedure is particularly sensitive to item recall. Although the running span procedure is not an ideal procedure for scoring serial recall errors as there is no fixed starting point from which the participant should start to recall the items, and as there will be a substantial amount of omission errors, we nevertheless also computed recall accuracy as a function of serial position. We used a strict serial recall criterion, in which an item was considered correct only if it was recalled at the correct serial position and by anchoring the target sequence on the last item recalled which in general was also the final item of the running span sequence (see Results). For instance, given the target sequence "… – Item 9 – Item 10 – Item 11 – Item 12" and the output sequence "Item 10 – Item 9 – Item 11 – blank", only item 11 was scored as correct. Finally, we computed the proportion of order errors, which is the number of items recalled in wrong serial position divided by the overall amount of items recalled (Murdock, 1976).

Statistical analysis. We used a Bayesian analysis approach instead of the frequentist statistical analysis approach which has been shown to overstate the evidence for an effect (Rouder & Morey, 2011), especially as sample size increases (Kruschke, 2011). In Bayesian model comparison, the null hypothesis is directly compared to the alternative hypothesis (variable of interest), and evidence for both the null effect and the effect of interest can be simultaneously tested (Dienes, 2014), while frequentist statistics only test the effect of interest by rejecting the null hypothesis. Statistical results are interpreted using the Bayesian Factor (BF), which reflects the likelihood ratio of a given model (effect of interest). The model with the highest BF value must be favoured over others. Here, the BF_{10} is used to determine the likelihood ratio of the alternative model (H_1) relative to the null model (H_0), and the BF_{01} to determine the likelihood ratio of H_0 relative to H_1. We use the classification proposed by previous studies (Jeffreys, 1998; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011): A BF of 1 provides no evidence, $1 < BF < 3$ provides anecdotal evidence, $3 < BF < 10$ provides moderate evidence, $10 < BF < 30$ provides strong evidence, $30 < BF < 100$ provides very strong evidence and $100 < BF$ provides extreme/decisive evidence. All the analyses were performed using JASP (JASP Team, 2017).

Results and discussion

A first analysis assessed recall performance as a function of serial position and stimulus condition (see **Figure 1.1**). We report in **Table 1.1** specific effect values ($BF_{inclusion}$), which reflect the likelihood of all the models including a given effect as compared to any other model not including the effect. The full statistical results are reported in supplementary material **S1** and **S2**. Since there were 3 different list lengths (6, 9 and 12), separate analyses were conducted for each list length. As can be seen in **Table 1.1**, we found decisive evidence for lexicality and serial position effects across all list lengths, and this both using an item recall and strict serial recall criterion. Similar finding were found for the interaction term, except for list length 6 where anecdotal evidence and strong evidence were found using an item recall criterion and a strict serial recall criterion, respectively.
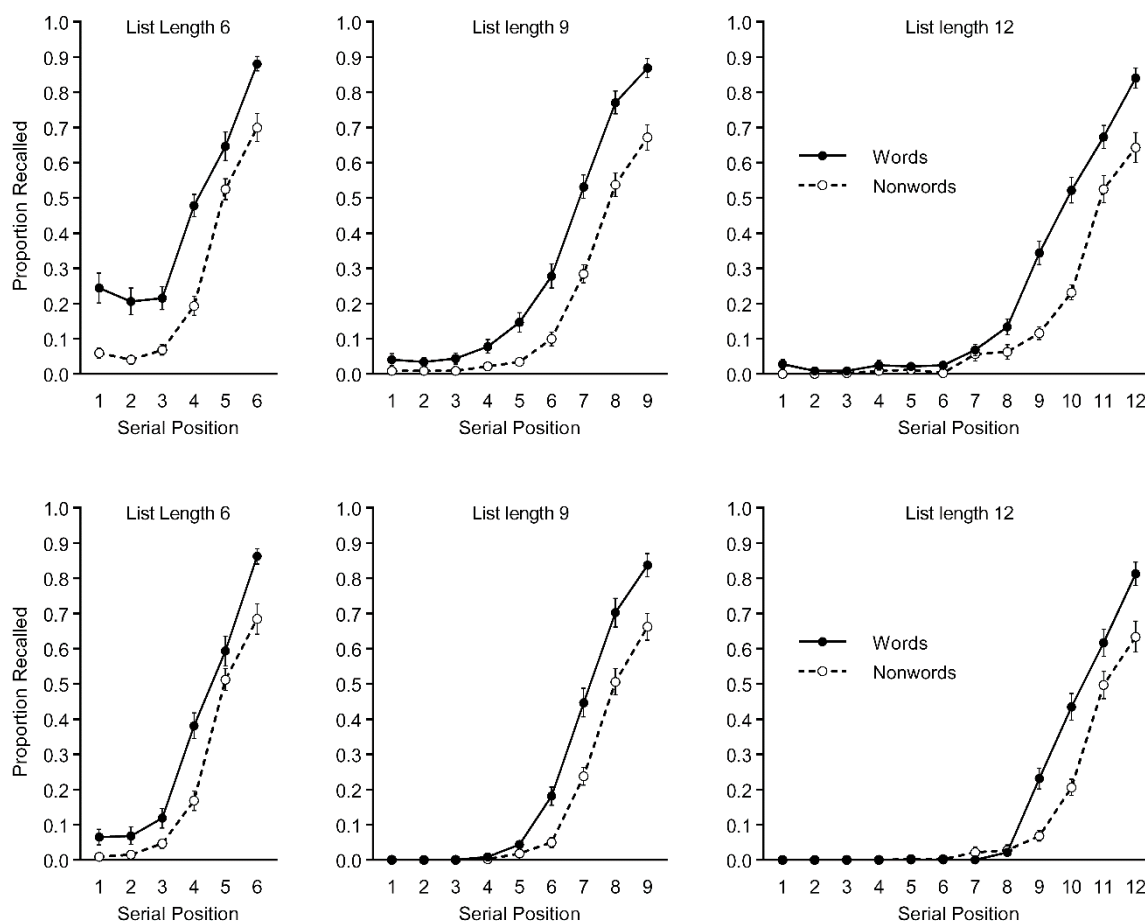
Figure 1.1. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 1 (Lexicality effect) using an item recall (top panel) and strict recall (bottom panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

Serial position effects were analysed in a more detailed manner, but by grouping the positions in pre-recency and recency in order to reduce the number of statistical contrasts to be conducted while increasing their reliability given the expected poor recall performance for initial positions (see **Figure 1.1**). Recency positions were always the three last positions for all list lengths, while pre-recency where defined as the three positions just before the recency portion, and hence correspond to positions 1-3 in list length 6, 4-6 in list length 9 and 7-9 in list length 12. As shown in **Table 1.1**, we found decisive evidence supporting the lexicality effect over the pre-recency and recency portions for all list lengths using an item recall criterion. Using a strict serial recall criterion, we found decisive evidence over the recency portions of all list lengths. While decisive evidence were found for list length 9 in the pre-recency portion, moderate evidence were found for list length 6 and 12. Descriptive statistics of mean items recalled across list length is provided in **Table 1.2**. As expected, strong recency

Table 1.1. *Bayesian Factor values as a function of serial position and stimulus condition – Experiment 1.*

	List Length					
	Item recall criterion			**Strict serial recall criterion**		
	6	**9**	**12**	**6**	**9**	**12**
Effects	**Bayesian Repeated Measure ANOVA**					
Lexicality	$\rightarrow +\infty$	$\rightarrow +\infty$	$\rightarrow +\infty$	6.45e+9	$\rightarrow +\infty$	$\rightarrow +\infty$
Serial position	$\rightarrow +\infty$	$\rightarrow +\infty$	$\rightarrow +\infty$	6.01e+15	$\rightarrow +\infty$	$\rightarrow +\infty$
Lexicality * Serial position	.95	2.50e+8	$\rightarrow +\infty$	24.26	2.37e+10	$\rightarrow +\infty$
Effects	**Bayesian Paired T-Test**					
Pre-recency	9617	218897.06	1218.263	4.40	44085	7.461
Recency	2.90e+6	1.26e+9	1.96e+7	20241.11	184736	256504.46

Note. For main effects and the interaction term (Bayesian ANOVA), the values represent $BF_{inclusion}$. For pre-recency and recency effects (Bayesian T-Test), values represent BF_{10}.

and weak primacy effects were observed, and this for all list lengths (see **Figure 1.1**), showing that participants recalled the most recently presented items, i.e. those that were still active in their mind when instructed to recall. In sum, over the different list lengths, a robust recall advantage was observed for words over nonwords, and this advantage was also observed in the recency portion of the serial position curves.

Although the majority of items were recalled in the recency portion of the serial position curve, we performed an additional analysis by removing the trials for which participants recalled items in the very beginning of the lists, that is, from position 1 through 3 for list length 9, and from position 1 through 6 for list length 12. This analysis allowed us to ensure that recall performance reflected immediate output of information from pre-recency and recency positions and that it was not delayed by recall of earlier items potentially associated with strategic retrieval processes. Serial position curves and BF associated with this re-analysis of the data are reported in supplementary material **S3**. The results still showed very reliable evidence for the presence of lexicality effects when ensuring that the *only* items that were output were items from the pre-recency and recency portions.

We further assessed whether participants directly recalled information available "in their mind", as instructed, or whether they used effortful retrieval strategies when attempting to recall items. We computed response latencies by measuring the duration between the beginning of the recall cue and the onset of the first item recalled, for each

Table 1.2. *Descriptive statistics for Experiment 1.*

| | Lexicality | Length | | |
		6	9	12
Item recall	Word	2.672 (.582)	2.794 (.451)	2.702 (.398)
	Nonword	1.588 (.331)	1.678 (.327)	1.663 (.387)
Strict serial recall	Word	2.091 (.607)	2.216 (.580)	2.113 (.559)
	Nonword	1.438 (.387)	1.478 (.392)	1.463 (.474)

Note. The values represent the mean absolute number of items recalled. Standard deviations are marked in parenthesis.

participant and each trial. Hesitations such as "hum" at the beginning of trials were considered part of the response latency. A controlled retrieval strategy would be reflected by longer response latencies. Median response latencies averaged across participants were indeed rather short, $M = 1475$, $SD = 948$, and are only slightly longer than the response latencies observed in single word speeded repetition paradigms (800 to 1100 ms; Vitevitch & Luce, 1998, 1999). In order to determine whether participants also kept responding very fast once the first response had been produced, we further computed the average median response duration times, corresponding to the duration between the onset of the first and final response. Response durations were also very short: $M = 1494$, $SD = 977$, for an average of 2.72 words and 1.64 nonwords recalled.

Finally, we computed the proportion of order errors, and observed decisive evidence in favour of the Lexicality effect ($BF_{10} = 559.5$) using a Bayesian T-Test, but with the effect going in the opposite direction: nonwords, when recalled, were less often recalled in a wrong serial position ($M = .12$, $SD = .141$) than words ($M = .21$, $SD = .179$).

The results of this experiment showed decisive evidence for the existence of a lexicality effect in a running span recall procedure. As in standard immediate serial recall tasks (Brener, 1940; Jefferies et al., 2006a; Saint-Aubin & Poirier, 1999b, 2000), the lexicality effect was expressed through better recall performance for words as compared to nonwords. The pattern of recall performance as a function of the serial position curve is also compelling; as predicted, the vast majority of items were recalled over the 3 last positions, and strong lexicality effects were observed over these positions. Finally, we also observed a lexicality effect as regards to order errors: words led to higher proportion of order errors as compared to nonwords, in line with previous results (Fallon et al., 2005; Guérard & Saint-Aubin, 2012; Jefferies et al., 2006a; Saint-Aubin & Poirier, 2000). This has been taken as evidence for a stronger reliance on sequential encoding and retrieval mechanisms for phonological codes as compared to lexico-semantic codes (Romani et al., 2008).

Experiment 2

The second experiment explored the impact of the lexical frequency effect on a running span recall procedure. Lists composed of frequently used words typically lead to higher vSTM span than lists composed of infrequent words in standard immediate serial recall tasks (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Watkins & Watkins, 1977). This lexical frequency effect has been robustly demonstrated through many experimental conditions and is, as the lexicality effect, one of the most studied psycholinguistic variables in vSTM.

Models of language processing consider that high frequency words have a lower activation threshold, leading to faster activation in linguistic tasks (McClelland & Elman, 1986). However, if high and low frequency words only stem from item-level activation characteristics, a sawtooth pattern of performance should be expected in lists with alternating high and low frequency items. Instead, overall list recall performance has been shown to be intermediate as compared to pure lists (Hulme et al., 1997, 2003). Hence the word frequency effect may also stem from the higher inter-item associations of high frequency words as compared to low frequency words (Hulme et al., 2003). This is notwithstanding the fact that frequency effects are observed even when inter-item association frequency is controlled (Poirier & Saint-Aubin, 2005; Tse & Altarriba, 2007), suggesting that frequency effects are also determined by intrinsic lexical properties at an item-level. Note that this experiment did not aim at assessing which specific aspect of the frequency effect contributes to vSTM span performance, but rather to test the non-strategic nature of the lexical frequency effect in vSTM tasks.

Method

Participants. Thirty-nine undergraduate students (25 females and 14 males) aged between 18 and 29 years (M = 22.82, SD = 2.81), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had been approved by the local ethics committee.

Materials. All the stimuli were selected from a French lexical database (New, 2006). As in Experiment 1, we used monosyllabic words. We first chose the 135 most frequent nouns available in the database. From this first set, we then selected 135 low-frequency words matched for phonological length (M = 3.48, SD = 0.57 for both high and low frequency words) and imageability (M = 4.93, SD = 1.44 and M = 4.79, SD = 1.23 for high and low frequency words, respectively) as well as for neighborhood

density (M = 14.17, SD = 7.88 and M = 13.02, SD = 8.41 for high and low frequency words, respectively). The neighborhood density was defined here as the number of real words that it was possible to create by substituting one phoneme from a target. There was no overlap between the two lists in terms of lexical frequency (M_{log} = 2.11, *min* = 1.64, *max* = 3.12 for high frequency words and M_{log} = -.24, *min* = -2, *max* = .23 for low frequency words). Each word was heard twice across the whole experiment. All the stimuli were recorded by a French-native male speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The length of each item was normalized to 375ms without altering the pitch.

Procedure. Contrary to Experiment 1, the lists from the two linguistic conditions were presented in an unblocked manner in Experiment 2 and participants were not informed of the existence of two different conditions, in order to further reduce the intervention of any strategic expectation effects. The experiment was divided into 2 blocks composed of 30 trials each to allow participants to take a very short break and refocus on the task if they needed to. The high and low frequency stimulus conditions were presented in pseudorandom succession within each of the two blocks, so that each condition was not repeated more than twice on two consecutive trials. All participants performed 10 practice trials before the beginning of the main task. All other aspects of the procedure, including scoring and statistical analysis, were the same as in Experiment 1.
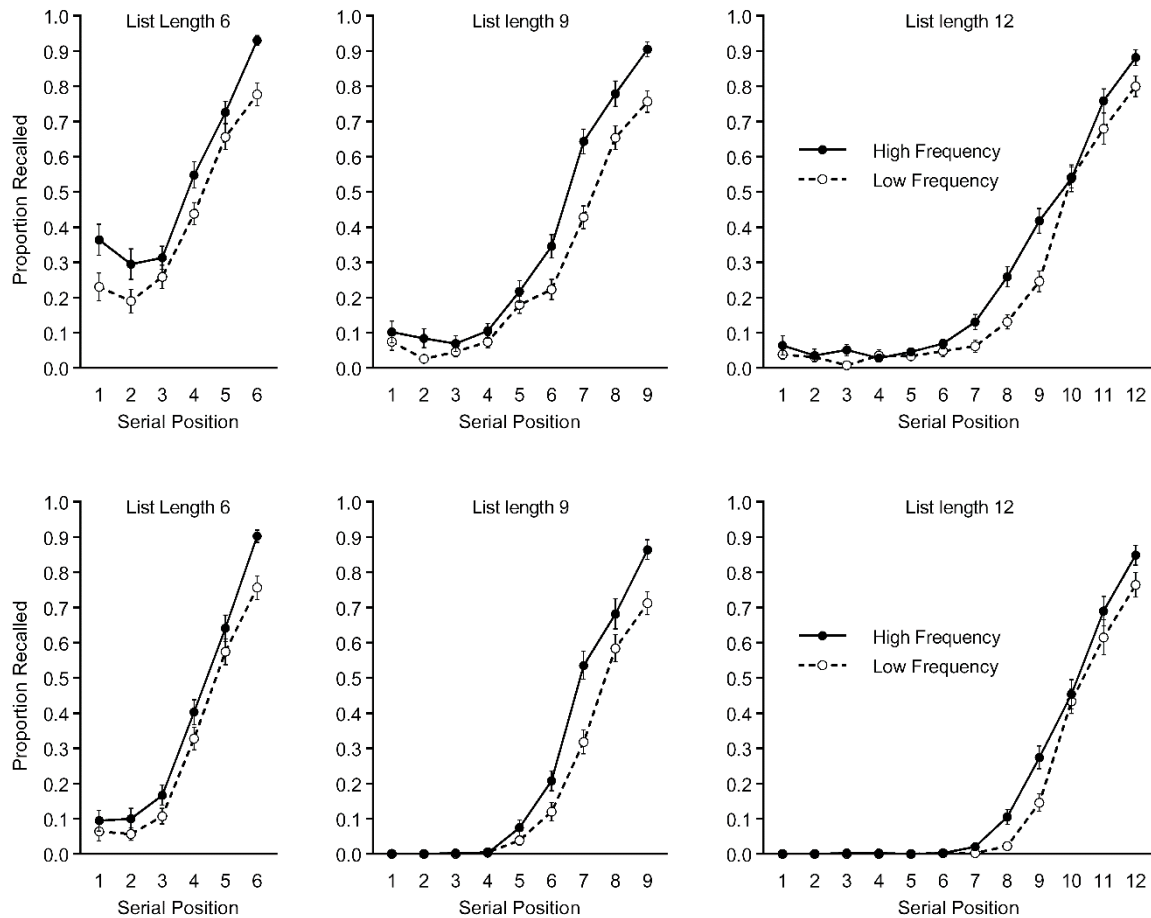
Figure 1.2. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 2 (Lexical frequency effect) using an item recall (top panel) and strict recall (bottom panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

Results and discussion

We first assessed recall performance as a function of serial position and lexical frequency (see **Figure 1.2**). As can be seen in **Table 1.3**, effects of lexical frequency and serial position provided decisive evidence for all list lengths regardless of whether an item or strict serial recall criterion was used. Using an item recall criterion, the interaction term provided very strong evidence for list length 9 and 12, and moderate evidence for the null hypothesis for list length 6. Using a strict serial recall criterion, the interaction term provided decisive and strong evidence for list length 9 and 12, respectively. For list length 6, we found moderate evidence for the null hypothesis. Relevant descriptive statistics are provided in **Table 1.4.** The full statistical results are reported in supplementary material **S4** and **S5**.

Table 1.3. *Bayesian Factor values as a function of serial position and stimulus condition – Experiment 2.*

	List Length					
	Item recall criterion			**Strict serial recall criterion**		
	6	**9**	**12**	**6**	**9**	**12**
Effects	**Bayesian Repeated Measure ANOVA**					
Lexical frequency	6.08e+5	9.8e+9	1.47e+6	2734.23	6.82e+13	1243.97
Serial position	$\rightarrow +\infty$	6.01e+15	6.01e+15	3e+15	$\rightarrow +\infty$	$\rightarrow +\infty$
Lexical frequency * Serial position	.121	60.33	50.87	.291	1.32e+7	13.27
Effects	**Bayesian Paired T-Test**					
Pre-recency	23885	95.208	1.283e+6	3.911	57.594	203817.262
Recency	45655	4.565e+8	6.916	516.19	4.969e+6	6.816

Note. For main effects and the interaction term (Bayesian ANOVA), the values represent $BF_{inclusion}$. For pre-recency and recency effects (Bayesian T-Test), values represent BF_{10}.

We next assessed lexical frequency effects across pre-recency and recency portions of the serial position curves (**Table 1.3**). Using an item recall criterion, decisive evidence supported the lexicality effect over the pre-recency portion for list length 6 and 12, while list length 9 was associated with very strong evidence. The same was observed over the recency portion, with effect of lexical frequency associated with decisive evidence for list length 6 and 9. List length 12 was associated with moderate evidence. Using a strict serial recall criterion over the pre-recency portion, list length 6, 9 and 12 were associated with moderate, very strong and decisive evidence, respectively. The lexical frequency effect over the recency portion was associated with decisive evidence for list length 6 and 9, and moderate evidence for list length 12.

As for experiment 1, in a separate analysis, we removed the trials for which participants recalled items in the primacy portion in order to insure that recall performance over the pre-recency and recency portions were not biased by output delay. As can be observed in supplementary material **S6,** the results remained the same, and indicated very robust evidence in favour of a lexical frequency effect over the recency portions for all list lengths. Next, we computed the proportion of order errors, and observed moderate evidence in favour of the null hypothesis ($BF_{10} = 0.179$; $BF_{01} = 5.58$). Participants produced as many order errors for high frequency ($M = .26$, $SD = .19$) as for low frequency ($M = .26$, $SD = .19$) words.

Table 1.4. *Descriptive statistics for Experiment 2.*

	Frequency	Length		
		6	**9**	**12**
Item recall	High	3.185 (.738)	3.246 (.582)	3.294 (.662)
	Low	2.548 (.671)	2.474 (.471)	2.646 (.496)
Strict serial recall	High	2.308 (.728)	2.369 (.793)	2.4 (.822)
	Low	1.887 (.601)	1.779 (.653)	1.987 (.725)

Note. The values represent the mean absolute number of items recalled. Standard deviations are marked in parenthesis.

As for Experiment 1, we also examined response latencies and response durations. Response latencies and response durations, although slightly longer than in Experiment 1, were still very short: $M = 1627$ ($SD = 698$) for response latencies and $M = 2033$ ($SD = 1340$) for response durations. The slightly longer response durations reflect the fact that a larger number of items were recalled as compared to Experiment 1 (3.24 and 2.56 items for the high and low frequency conditions, respectively).

Finally we assessed the extent to which the lexical frequency effect also reflected differences in between-item lexical co-occurrence values which partly define the lexical frequency effect. We computed lexical co-occurrence values using Latent Semantic Analysis (LSA) (http://lsa.colorado.edu/, using the semantic space "Francais-Monde-Extended") for each trial. LSA measures the extent to which two words co-occur within similar contexts using large corpora (Landauer & Dumais, 1997). This analysis was restricted to the last 6 positions of the serial positions curves since the vast majority of items were recalled in this portion. As expected, we found a large difference between the two stimulus list conditions in terms of lexical co-occurrence values with almost no overlap between the two conditions ($M = .3$, $SD = .07$ for high frequency lists and $M = .03$, $SD = .04$ for low frequency lists, $BF_{10} = 8.71e+46$), in agreement with previous studies (Hulme et al., 2003).

In line with previous studies using standard immediate serial recall tasks (Hulme et al., 1997, 2003; Watkins & Watkins, 1977; Tse & Altarriba, 2007; Parmentier, Comesaña, & Soares, 2007), we observed a lexical frequency effect. This result was found in a running span recall task, and as in Experiment 1, we observed very strong recency effects and poor primacy effects, indicating that participant recalled the most recently presented items. Critically, this lexical frequency effect was observed for the last positions of the serial position curve, which are positions that are directly recalled. No effect was found as regards to order errors, consistent with previous studies showing that lexical frequency mostly influences item recall, rather than order recall in immediate serial recall tasks (Hulme et al., 2003; L. M. Miller & Roodenrys, 2012).

Experiment 3

In Experiment 3, we manipulated the impact of semantic knowledge on running span recall performance, by comparing words with close semantic relations as opposed to words with no direct semantic relations. These semantic relations could be thematic or taxonomic (e.g., "leaf – three – green" or "corn – grain – rice"). It has been shown that thematic relationships produce very similar effects as those observed for taxonomic categories in vSTM tasks (Tse, 2009). The related lists contained series of triplets, with items within each triplet being related at the semantic level.

Language-based models of vSTM consider that semantic inter-item associations should stabilize semantic activations for items within each thematic or taxonomic group via spreading activation of the shared semantic features within the same semantic space (N. Martin et al., 1996), leading to higher vSTM span as compared to words sharing no semantic features. This in agreement with more recent evidence showing the existence of spreading activation in vSTM leading to false memories (Atkins & Reuter-lorenz, 2008; Flegal & Reuter-Lorenz, 2014; Melnik, Mapelli, & Özkurt, 2017). At the same time, effects of semantic relatedness could also stem from more strategic semantic grouping processes based on the shared semantic features of the words (Schleepen, Markus, & Jonkman, 2014). Using a running span recall procedure, we determined the extent to which semantic relatedness effects still occur in encoding conditions minimizing the intervention of strategic encoding processes.

Method

Participants. Thirty-nine undergraduate students (22 females and 17 males) aged between 19 and 31 years (M = 23.03, SD = 3.02), recruited from the university community took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had been approved by the local ethics committee.

Materials. Ninety triplets of semantically similar short, monosyllabic words were selected. The semantic associations could be taxonomic (e.g., "gold – iron – lead") or thematic (e.g., "tree – branch – trunk"). The semantic triplets were then randomly distributed over the 30 lists of the semantically similar list conditions. There were three list lengths (6, 9 and 12) as in the previous experiments, with respectively 2, 3 and 4 triplets. Thirty non-similar lists were created by pseudorandomly sampling items from the semantically related lists, with the constraint that no word from the same semantic category could directly follow another word from the same semantic category, and that the number of words within a list stemming from the same semantic category was

minimized. The a priori defined semantic associations between the words were further validated and corrected when necessary via an online survey. In this survey, an independent group of participants had to judge on a scale ranging from 0 to 5 the semantic similarity of pairs of words occurring in the similar and non-similar lists. The total number of pairs to be judged was 540. Because of this large number of stimuli to judge, participants were free to stop the survey at any moment, with no restriction regarding the number of pairs to judge. One-hundred and thirty-six participants took part in this online survey. Final results showed that similar pairs were judged as being strongly similar (M = 4.197, SD = 0.482) and non-similar pairs were judged as being very dissimilar (M = 0.45, SD = 0.415), and decisive evidence supported this difference ($BF_{10} \rightarrow +\infty$). Final data acquisition showed that each pair had been judged 75.25 times on average. All the stimuli were recorded by a French-native male speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The length of each item was normalized to 375ms without altering the pitch.

Procedure. Experimental procedure as well as statistical analysis were identical to Experiment 2.

Results and discussion

We first assessed recall performance as a function of serial position and stimulus condition (see **Figure 1.3** and **Table 1.5**). Decisive evidence supported the semantic similarity and serial position effects across all list lengths and using both item recall and strict serial recall criterion. The interaction term provided very strong evidence across list length 6 and 12 and moderate evidence for list length 9 using an item recall criterion. Using a strict serial recall criterion, the interaction term provided strong evidence for list length 9 and 12, but substantial evidence for list length 6. See **Table 1.6** for relevant descriptive statistics. The full statistical results are reported in supplementary material **S7** and **S8**.

We next assessed semantic similarity effects across pre-recency and recency portions of the serial position curves (**Table 1.5**). Using an item recall criterion, decisive evidence supported the semantic similarity effect over pre-recency and recency positions for all list lengths. Using a strict serial recall criterion, decisive evidence supported the semantic similarity effect over pre-recency and recency portions for list length 6. For list length 9, strong and moderate evidence supported the semantic similarity effect over the pre-recency and recency portions, respectively. Finally, strong evidence supported the semantic similarity effect over pre-recency and recency portion for list length 12. Consistent with Experiments 1 and 2, strong recency and
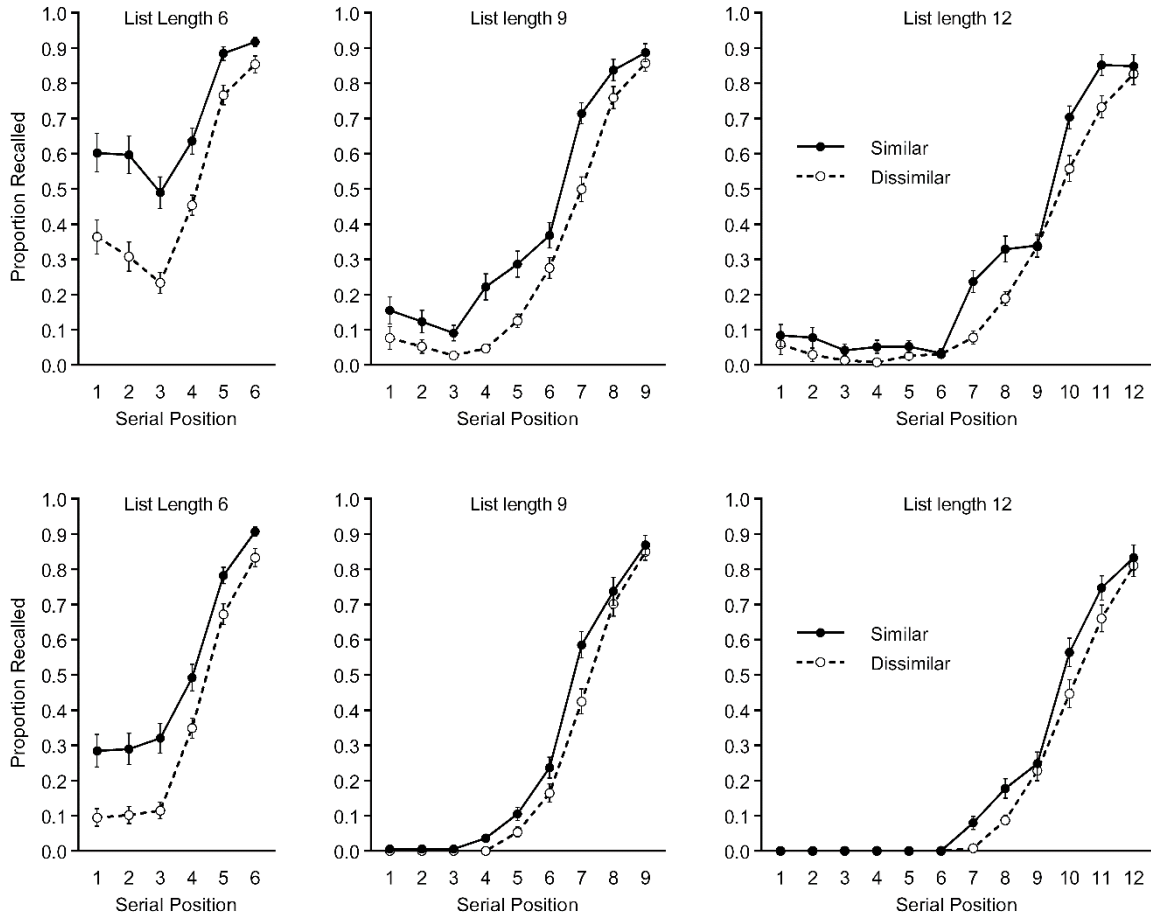
Figure 1.3. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 3 (Semantic similarity effect) using an item recall (top panel) and strict recall (bottom panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

weak primacy effects were observed, and this for all list lengths; we found reliable semantic similarity effects over the recency portions of the serial position curves reflecting the portion where items are considered to be most directly accessible.

Similarly to Experiment 1 and 2, a separate analysis investigated the reliability of these results after removing trials for which items were recalled in the primacy portion. An impact of semantic similarity was still observed over the recency portion of the serial position curves, as shown in supplementary material **S9**. Next, we computed the proportion of order errors, and observed moderate evidence supporting the semantic similarity effect (BF_{10} = 3.186). Participants produced more errors for semantically similar lists (*M* = .26, *SD* = .18) as compared to non-similar lists (*M* = .22, *SD* = .17).

As in Experiments 1 and 2, response latencies and response durations were determined. The values were very similar to those observed in Experiments 1 and 2 (response latencies: *M* = 1472, *SD* = 665; response durations: *M* = 1877, *SD* = 1062),

Table 1.5. *Bayesian Factor values as a function of serial position and stimulus condition – Experiment 3.*

	List Length					
	Item recall criterion			Strict serial recall criterion		
	6	9	12	6	9	12
Effects	**Bayesian Repeated Measure ANOVA**					
Semantic similarity	→ +∞	6.69e+12	8.71e+7	→ +∞	4625.78	872.36
Serial position	→ +∞	→ +∞	6.01e+15	→ +∞	→ +∞	→ +∞
Semantic similarity * Serial position	27.57	5.45	43.64	1.11	19.45	10.42
Effects	**Bayesian Paired T-Test**					
Pre-recency	6.990e+8	11355.76	291.457	127038	12.637	18.87
Recency	221400	5316.89	1634.108	11059	3.748	22.49

Note. For main effects and the interaction term (Bayesian ANOVA), the values represent $BF_{inclusion}$. For pre-recency and recency effects (Bayesian T-Test), values represent BF_{10}.

suggesting that participants followed the task instructions and recalled information that was directly available in their mind when cued for recall.

Finally, we computed lexical co-occurrence values to rule out the possibility that the effect of semantic similarity could have been inflated by differences in lexical co-occurrence values between the semantically related and unrelated lists. A first analysis showed that lexical co-occurrence values were slightly higher for related lists ($M = .211$, $SD = .114$) as compared to unrelated lists ($M = .09$, $SD = .05$), with $BF_{10} = 5.14e+8$. Therefore, in a second analysis, we resampled our lists by keeping only those lists (six lists per length) that showed overlapping lexical co-occurrence ($M = .13$, $SD = .055$ for related lists and $M = .121$, $SD = .037$ for unrelated lists, $BF_{01} = 3.156$) while still differing at the level of semantic relatedness. When restricting our analyses of semantic relatedness effects to these lists matched for lexical co-occurrence, robust semantic relatedness effects across all list lengths were still observed using an item recall ($BF_{10} = 1.06e+13$ for list length 6, $BF_{10} = 5.98e+7$ for list length 9 and $BF_{10} = 41376.14$ for list length 12) and a strict serial recall criterion ($BF_{10} = 2.14e+13$ for list length 6, $BF_{10} = 90.01$ for list length 9 and $BF_{10} = 131.86$ for list length 12). Hence, the semantic relatedness effect observed in this experiment is not likely to be biased by differences in lexical co-occurrence values.

We observed a semantic similarity effect using a running span recall procedure, and as in Experiment 1 and 2, the impact of linguistic knowledge was observed for the

Table 1.6. *Descriptive statistics for Experiment 3.*

		Length		
	Similarity	**6**	**9**	**12**
Item recall	Similar	4.136 (1.009)	3.686 (.805)	3.649 (.703)
	Dissimilar	2.982 (.690)	2.721 (.489)	2.885 (.489)
Strict serial recall	Similar	3.033 (1.009)	2.56 (.723)	2.633 (.846)
	Dissimilar	2.133 (.686)	2.158 (.617)	2.215(.73)

Note. The values represent the mean absolute number of items recalled. Standard deviations are marked in parenthesis.

recency portions of the serial position curves. Moreover, the pattern of the serial position curve suggests that semantically similar triplets tended to be processed as a broader chunk unlike semantically non-similar triplets (see **Figure 1.3**); items within semantically similar triplets were associated with more similar recall performance levels than items from equivalent, semantically dissimilar triplets for which recall performance gradually increased from the first to the third item of the triplet. This was particularly the case for the two last triplets of list length 9 and 12. Finally, we found a higher proportion of order errors for semantically similar lists, in agreement with previous studies that have assessed the impact of semantic similarity on serial order processing (Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011). This may be caused by a similar level of activation for items sharing common semantic features, leading to more errors when items are associated to their serial positions (Poirier, Saint-Aubin, Mair, Tehan, & Tolan, 2015).

Experiment 4A

In Experiment 4A, we assessed the effect of word imageability on running span performance, allowing us to further define the nature of semantic effects revealed in Experiment 3. This experiment assessed the impact of the nature of concrete versus abstract semantic features associated with individual words on running span performance. If we assume that semantic features are activated in a non-strategic manner in vSTM wathever their nature, an imageability effect should be observed over the recency portion of the serial position curve. On the other hand, even if semantic access is considered to be performed very quickly in language processing tasks (Evans et al., 2012; Evans, Lambon Ralph, & Woollams, 2017; Pexman et al., 2002; Yap et al., 2015), the deeper, context-dependent visual and other semantic features associated with concrete, high imageability words may take more time to get co-activated and to stabilize the phonological and lexical representation than the redundant and overlapping semantic associations characterizing semantic relatedness as tested in

Experiment 3. Semantic features associated with concrete words are indeed known to be distributed over distinct semantic subsystems and neural networks (Binder, 2016; Lambon Ralph et al., 2017). Hence, in conditions of very fast presentation of memoranda, these semantic features may not reach a stable level of activation during encoding. Moreover, if the imageability effect in vSTM tasks depends on elaborative strategies such as mental imagery, no word imageability effects should be observed in a running span procedure.

Method

Participants. Fourty undergraduate students (30 females and 10 males) aged between 18 and 30 years (M = 21.75, SD = 2.93), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had been approved by the local ethics committee.

Materials. Ninety high imageability (HI) words and 90 low imageability (LI) words were selected from the Lexique database (New, 2006). All the stimuli were monosyllabic words matched for lexical frequency (M_{log} = 1.38, SD_{log} = .57 for high imageability words and M_{log} = 1.4, SD_{log} =.54 for low imageability words), subjective frequency (M = 4.46, SD = .85 for high imageability words and M = 4.62, SD = .91 for low imageability words) phonological length (M = 3.31, SD = .68 for high and low imageability words) and neighborhood density (M = 14.62, SD = 8.46 for high imageability words and M = 14.03, SD = 8.94 for low imageability words). There was no overlap between each list in terms of imageability (M = 6.69, min = 6, max = 6.9 for high imageability words and M = 2.71, min = 1.7, max = 3.2 for low imageability words). Given that imageability ratings were available for only a subset of the words listed in the Lexique database (Content, Mousty, & Radeau, 1990) we were able to select only 180 words. This results in each word being repeated 3 times through the whole experiment. After lists were created, we used the same type of online survey as in Experiment 3 to control for semantic similarity between adjacent stimuli; in this survey, 43 participants had to judge on a scale ranging from 0 to 5 whether two words were semantically related. They could stop the survey at any moment and they were not obliged to judge the entire set of 913 pairs. The survey showed that high and low imageability pairs did not differ for semantic similarity and semantic relatedness rating scores were on average very low (<1) (M = .93, SD = .91 for high imageability lists, and M = .93, SD = .76 for low imageability lists, BF_{01} = 13.47). Each pair had been judged 7.04 times on average. All the stimuli were recorded by a French-native male
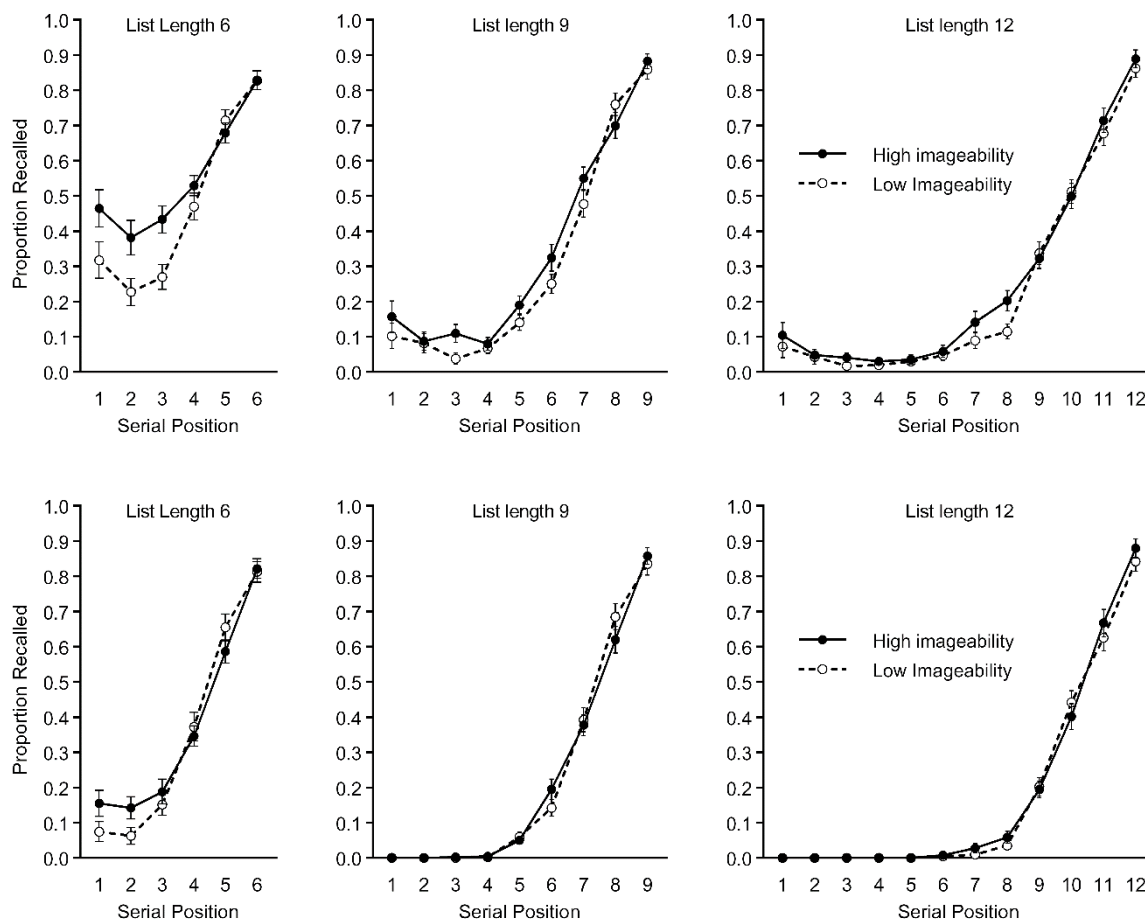
Figure 1.4. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 4A (Imageability effect) using an item recall (top panel) and strict recall (bottom panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The length of each item was normalized to 375ms without altering the pitch.

Procedure. Experimental procedure as well as statistical analysis and scoring were identical to Experiments 2 and 3.

Results and discussion

We assessed recall performance as a function of serial position and word imageability condition (see **Figure 1.4** and **Table 1.7**). Using an item recall criterion, the imageability effect provided anecdotal evidence, except for list length 6 where decisive evidence *supported* the effect. When using a strict serial recall criterion, the imageability effect was associated with moderate evidence *against* the effect for list length 6 and strong evidence *against* the effect for list length 9 and 12. The serial

Table 1.7. *Bayesian Factor values as a function of serial position and stimulus condition – Experiment 4A.*

	List Length					
	Item recall criterion			Strict serial recall criterion		
	6	9	12	6	9	12
Effects	Bayesian Repeated Measure ANOVA					
Imageability	862.65	1.47	.61	.217	.061	.086
Serial position	6.01e+15	6.01e+15	→ +∞	→ +∞	6.01e+15	→ +∞
Imageability * Serial position	11.89	.103	.006	.44	.016	.003

Note. For main effects and the interaction term (Bayesian ANOVA), the values represent $BF_{inclusion}$.

position effect provided decisive evidence across all list lengths both using item recall and strict serial recall criterions. Using an item recall criterion, we found strong evidence supporting the interaction for list length 6, moderate evidence against the interaction for list length 9, and decisive evidence against the interaction for list length 12. Using a strict serial recall criterion, we found anecdotal evidence for list length 6, very strong evidence against the interaction for list length 9, and decisive evidence against the interaction for list length 12. Overall, no reliable evidence supported the imageability effect, except for list length 6 using an item recall criterion. This imageability effect for list length 6 as a function of serial position was further investigated. We found decisive evidence supporting the imageability effect only over positions 1-3 (BF_{10} = 3.33e+6); a Bayesian T-Test over positions 4-6 provided moderate evidence supporting the *absence* of the imageability effect (BF_{01} = 5.39), showing that this absence was reliable over the recency portion. Relevant descriptive statistics are provided in **Table 1.8**. The full statistical results are reported in supplementary material **S10** and **S11.**

In sum, the results of this experiment are in striking contrast to the results of the three previous experiments. No effect of word imageability was observed for longer list lengths, and even for the shortest list lengths, the effect was observed for initial list positions but not for recency positions using an item recall criterion. When using a strict serial recall criterion, the effect disappeared completely across all list lengths. We should note that the stimuli were drawn from a more restricted pool as compared to Experiments 1 – 3 and that for 41% of the stimuli in the low imageability word condition, the words, although all being nouns, could belong to an additional grammatical class. Particularly the latter could potentially have led to a decrease of the

Table 1.8. *Descriptive statistics for Experiment 4A.*

	Imageability	Length		
		6	9	12
Item recall	High	3.318 (.844)	3.085 (.719)	3.094 (.686)
	Low	2.828 (.75)	2.778 (.63)	2.825 (.693)
Strict serial recall	High	2.235 (.821)	2.108 (.565)	2.225 (.654)
	Low	2.13 (.775)	2.118 (.665)	2.163 (.626)

Note. The values represent the mean absolute number of items recalled. Standard deviations are marked in parenthesis.

imageability effect, by allowing, in low imageability word lists, for sentence-level word associations (e.g., "fair" followed by "health"). In Experiment 4B, we attempted to replicate the results of Experiment 4A by controlling for the potential issues described here.

Experiment 4B

Method

Participants. Thirty-eight undergraduate students (26 females and 12 males) aged between 18 and 31 years (M = 21.55, SD = 2.76), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had been approved by the local ethics committee.

Materials. One hundred and thirty-five high imageability words and 135 low imageability words were selected from Lexique database (New, 2006). All the stimuli were monosyllabic words matched for lexical frequency (M_{log} = .99, SD = .57 for high imageability words and M_{log} = .99, SD = .87 for low imageability words), subjective frequency (M = 3.97, SD = .8 for high imageability words and M = 3.95, SD = .89 for low imageability words) phonological length (M = 3.54, SD = .7 for high imageability and M = 3.53, SD = .83 for low imageability words) and neighborhood density (M = 11.02, SD = 7.27 for high imageability words and M = 10.86, SD = 7.54 for low imageability words). There was no overlap between each list in terms of imageability (M = 6.55, *min* = 6.2, *max* = 6.9 for high imageability words and M = 3.17, *min* = 1.9, *max* = 4.2 for low imageability words). However, as compared to Experiment 4A, Experiment 4B included low imageability words with slightly higher imageability scores as we wanted to avoid repetition effects of the stimuli and thus, a larger set of stimuli was selected for each condition; in Experiment 4B, each word was repeated twice.

Moreover, we also insured that all words were nouns in their most frequent grammatical form (New, 2006). After the stimuli were selected and arranged in lists, we used the same type of online survey as in Experiment 3 and 4A to control for semantic relatedness between adjacent stimuli, by having 121 participants judge the semantic relatedness for the 1821 word pairs; as before, participants could stop the survey at any moment. We observed that high and low pairs differed mildly for semantic similarity ($M = .65$, $SD = .62$ for high imageability lists, and $M = .74$, $SD = .55$ for low imageability lists, $BF_{10} = 7.16$), but this difference had no impact at the semantic level as for both conditions the ratings were very low (<1). Each pair had been judged 11.57 times on average. All the stimuli were recorded by a French-native female speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The length of each item was normalized to 375ms without altering the pitch.

Procedure. Experimental procedure as well as statistical analysis and scoring procedures were identical to Experiments 2, 3 and 4A.
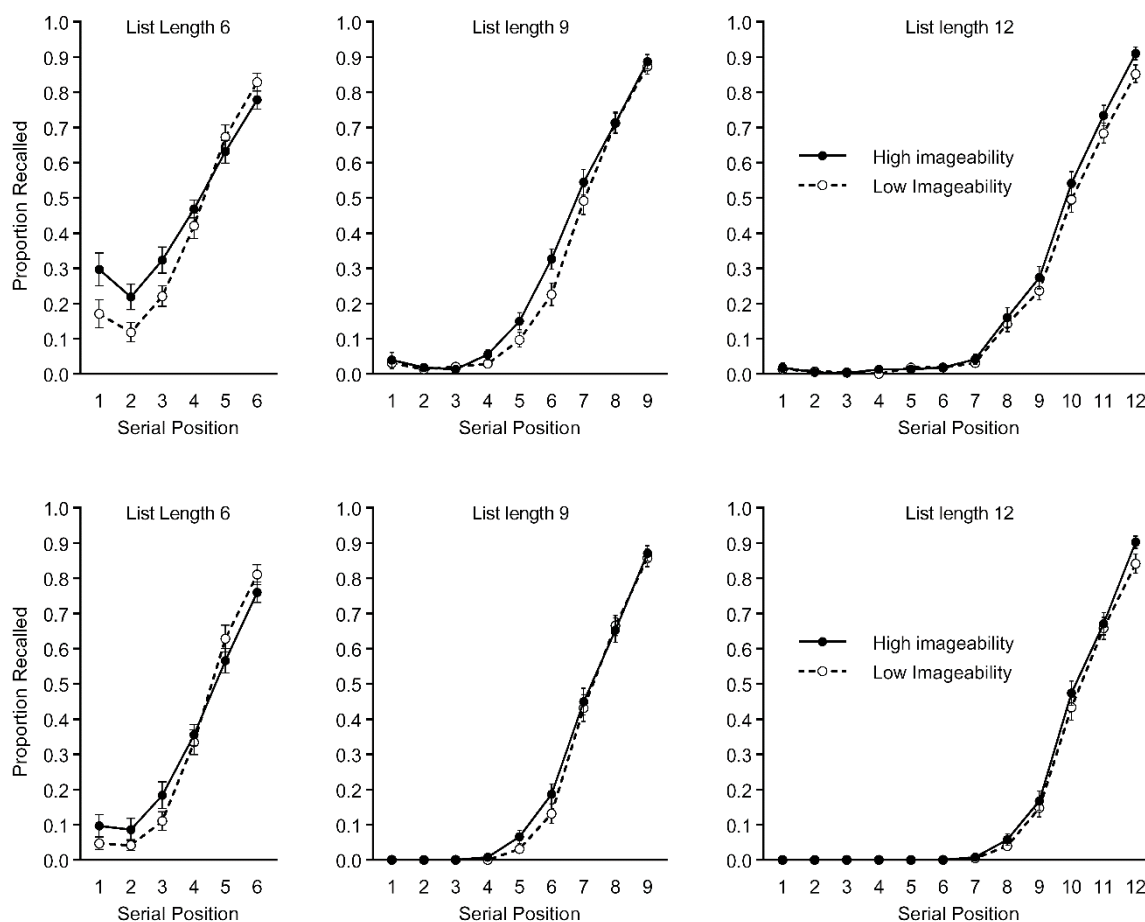
Figure 1.5. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 4B (Imageability effect) using an item recall (top panel) and strict recall (bottom panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

Results and discussion

An analysis of recall performance separately for each list length replicated the results of Experiment 4A (see **Figure 1.5** and **Table 1.9**): When using an item recall criterion, the imageability effect was associated with anecdotal evidence for list length 9 and 12, and moderate evidence for list length 6. When using a strict serial recall criterion, moderate evidence *against* the effect was found across all list lengths. Decisive evidence supported the serial position effect across all list lengths and both using an item recall and strict serial recall criterion. We found moderate evidence for the interaction term for list length 6, moderate evidence against the interaction for list length 9 and decisive evidence against the interaction term for list length 12. When using a strict serial recall criterion, we found moderate evidence against the interaction

Table 1.9. *Bayesian Factor values as a function of serial position and stimulus condition – Experiment 4B.*

	List Length					
	Item recall criterion			**Strict serial recall criterion**		
	6	**9**	**12**	**6**	**9**	**12**
Effects	**Bayesian Repeated Measure ANOVA**					
Imageability	6.26	1.91	1.11	.145	.154	.312
Serial position	$\rightarrow +\infty$	$\rightarrow +\infty$	$\rightarrow +\infty$	$\rightarrow +\infty$	6.01e+15	$\rightarrow +\infty$
Imageability * Serial position	7.17	.105	.01	.268	.005	.004

Note. For main effects and the interaction term (Bayesian ANOVA), the values represent $BF_{inclusion}$.

term for list length 6 and decisive evidence against the interaction term for list length 9 and 12. As in Experiment 4A, the imageability effect found for list length 6 was further assessed with specific Bayesian T-Test. We found decisive evidence supporting the imageability effect over positions 1-3 (BF_{10} = 975.87) when using the item recall criterion. Again, a Bayesian T-Test over positions 4-6 provided moderate evidence supporting the *absence* of imageability effect (BF_{01} = 4.2), showing that this absence was reliable over the recency portion. Relevant descriptive statistics are reported in **Table 1.10**. The full statistical results are reported in supplementary material **S12** and **S13**.

Finally, the values of response latencies and response durations were comparable to those observed in the preceding experiments (response latencies: *M* = 1833, *SD* = 620; response durations: *M* = 1761, *SD* = 1253).

To sum up, Experiment 4B replicated Experiment 4A by showing a strongly reduced imageability effect relative to the other linguistic effects observed through Experiments 1 – 3. The initial serial positions for which an imageability effect was observed correspond to the portion where items can be retrieved from long-term memory (Nee & Jonides, 2011). At first glance, the results of Experiment 4A and 4B appear to be inconsistent with a purely non-strategic activation hypothesis as regards the intervention of lexico-semantic knowledge associated with the abstract/concrete dimensions. However, the absence of an imageability effect could also have been the result of the the very fast presentation rate used in the running span procedure, not leaving sufficient time for the distributed and distant semantic features of high imageability to reach a stable level of co-activation. Experiment 5A directly addressed this question.

Table 1.10. *Descriptive statistics for Experiment 4B.*

	Imageability	Length 6	Length 9	Length 12
Item recall	High	2.708 (.668)	2.745 (.490)	2.724 (.491)
	Low	2.421 (.529)	2.497 (.511)	2.5 (.476)
Strict serial recall	High	2.05 (.739)	2.234 (.608)	2.282 (.547)
	Low	1.974 (.582)	2.118 (.565)	2.126 (.580)

Note. The values represent the mean absolute number of items recalled. Standard deviations are marked in parenthesis.

Experiment 5A

In experiment 5A, the occurrence of an imageability effect was tested for a running span procedure using a slow stimulus presentation rate (1 item every 1.5 seconds). If absent imageability effect in Experiment 4A and 4B is the result of the very fast presentation rate of the word stimuli, then the effect should re-appear when presenting stimuli at a slower speed. It is important to note that in running span procedures with slow presentation rates, participants still adopt a passive listening attitude (Botto et al., 2014; Ruiz & Elosúa, 2013; Ruiz, Elosúa, & Lechuga, 2005) provided that memory load is kept sufficiently high (i.e. recalling as much items as possible).

Method

Participants. Thirty-five participants (22 females and 13 males) aged between 18 and 22 years ($M = 20$, $SD = 1.19$), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had been approved by the local ethics committee.

Materials. The same stimuli as in Experiment 4B were used. Hence, all the stimuli were matched for relevant linguistic properties. Four different versions of the task were randomly created, with the constraint that adjacent items within a list belonged to the pairs participants judged in the online survey of Experiment 4B. This allowed us to control for semantic similarity between high and low imageability lists for the 6 finals items, which are the positions in which the vast majority of items are recalled. For the final pairings across the four different versions we observed that both high and low imageability lists had a similar degree of semantic relatedness ($M = .67$, $SD = .37$ for high imageability words, $M = .68$, $SD = .247$ for low imageability words, $BF_{01} = 5.82$).

Procedure. Items were presented at a rate of 1 item every 1.5 seconds. There were 4 list lengths: 5, 9, 10 and 11. Only list lengths 9, 10 and 11 were used for stimulus by serial position analysis and were analyzed as one length by conducting an analysis on the 9 last serial positions. Each list length was presented 7 times, resulting in 21 trials in each stimulus condition (high and low imageability). Lists length 5 served as a filler condition making list length unpredictable and ensured that participants encoded the entire list and did not discard early items during encoding. Lists length 5 occurred 4 times in each stimulus condition. These modifications of the experimental procedure ensured that task duration was kept at a reasonable length (participants performed the task in approximatively 20-25 minutes). This also had the advantage of increasing the number of trials per linguistic condition entered in the statistical analyses as the linguistic condition effect was not analysed separately as a function of list length. The experiment was divided into 2 blocks composed of 25 trials each to allow participants to take a very short break and refocus on the task if they needed to. The high and low imageability word conditions were presented in pseudorandom succession within each of the two blocks, so that each condition was not repeated more than three times on three consecutive trials. All participants performed 5 practice trials before the beginning of the main task. The same recall instructions as in Experiment 1 through 4 were given to participants, except that participants were informed that stimuli were presented at a comfortable pace. All other aspects of experimental procedure as well as scoring procedures were identical to Experiments 2 through 4.

Results and discussion

We performed an analysis of recall performance as a function of stimulus condition and serial position (see **Figure 1.6**). Regardless of recall criterion, there was no evidence in favour of an effect of imageability, with the BF instead supporting the null hypothesis although at an anecdotal level ($BF_{01} = 2.25$ and $BF_{01} = 1.92$ using an item recall and strict serial recall criterion, respectively). The serial position effect was supported by decisive evidence ($BF_{inclusion} \rightarrow +\infty$ using an item recall and strict serial recall criterion) and very strong evidence against the interaction term was observed ($BF_{01} = 58.82$ and $BF_{01} = 31.25$ using an item recall and strict serial recall criterion, respectively).

When analysing response latencies and durations, slightly higher values were observed as compared to the preceding experiments, and this particularly for response durations (response latencies: $M = 1753$, $SD = 621$; response durations: $M = 2521$, $SD = 1144$). This longer response duration does not stem from a higher amount of items recalled, the average number of words recalled (2.46) being similar to that observed in
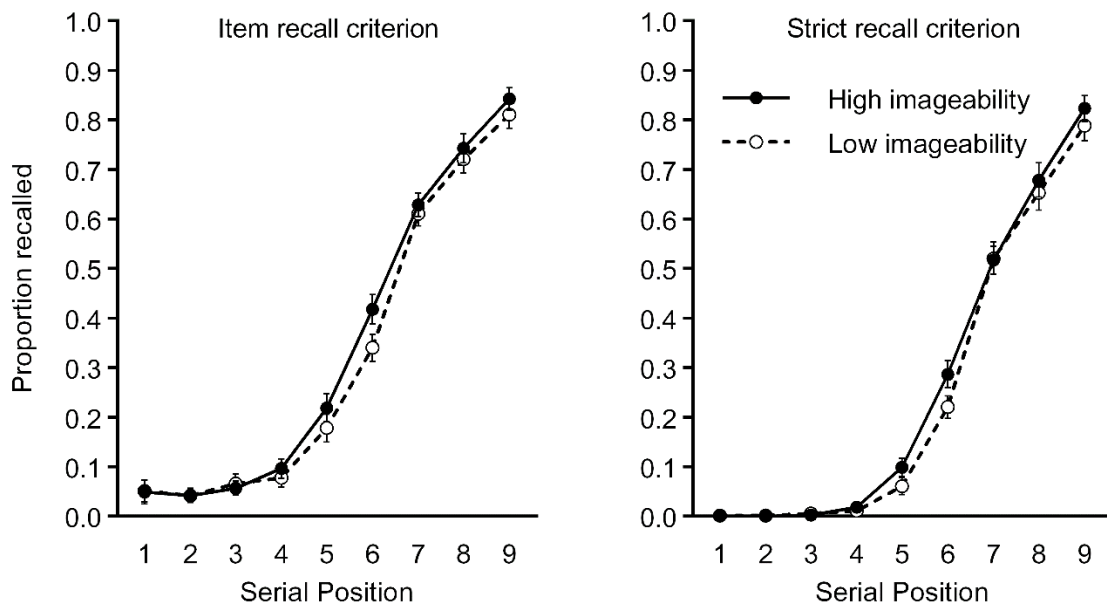
Figure 1.6. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 5A (Imageability effect) using an item recall (left panel) and strict recall (right panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

the other experiments. The larger response durations in this experiment may reflect the mimicking of the slower presentation times of items during encoding; a temporal mimicking effect at recall has also been observed for temporal grouping of items during encoding (Hurlstone et al., 2014). Despite the longer response durations, no imageability effect was found.

Overall, Experiment 5A reproduced the results observed for Experiments 4A and 4B. No reliable evidence supporting an imageability effect was observed despite using a slow presentation rate. These results indicate that a slow presentation rate is not a determining factor for the expression of an imageability effect in a vSTM task. Furthermore, studies having analysed strategic processing in slow presentation running span tasks indicate that participants still adopt a "passive listening" strategy like for fast-presentation running span procedures (Botto et al., 2014; Ruiz & Elosúa, 2013; Ruiz et al., 2005). Although no formal assessment of strategy use had been conducted in this experiment, our participants reported to mainly rely on phonological aspects of memoranda when encoding the items, without trying to rehearse them.

Experiment 5B

In this final experiment, we determined whether an imageability effect could be observed in standard immediate serial recall conditions using the same stimulus material and the same presentation rate (1 item every 1.5 seconds) as in Experiment 5A. If the absence of an imageability effect in previous experiments is due to reduced opportunity for implementing item-level and list-level elaborative encoding strategies, then a word imageability effect may re-appear in an immediate serial recall task in which the use of these strategies is not prevented (Morrison et al., 2016). Furthermore, in order to allow for a direct comparison with Experiment 5A, the same sample size was used.

Participants. Thirty-five participants (21 females and 14 males) aged between 18 and 24 years ($M = 20.71$, $SD = 1.81$), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had been approved by the local ethics committee.

Method

Materials. We applied the same procedure for list generation as in Experiment 5A, with high and low imageability lists being paired for semantic similarity using judgement rating of Experiment 4B, producing very similar levels of semantic similarity for both conditions across the 6 different versions created ($M = .67$, $SD = .32$ for high imageability words, $M = .7$, $SD = .25$ for low imageability words, $BF_{01} = 5.58$). Given that we used shorter list lengths and that the number of trials between Experiment 5A and 5B were relatively equivalent (see procedure), we only used a subset of stimuli (132 instead of 135 in each stimulus condition). This also had the consequence of each stimulus being presented once instead of twice, which should not be problematic when comparing results with Experiment 5A as lexico-semantic effects are known to be diminished only when stimuli are repeatedly sampled from a very restricted pool of items (e.g. 6 items) (Quinlan et al., 2017; Romani et al., 2008; Roodenrys & Quinlan, 2000).

Procedure. The six-item lists were presented at a rate of 1 item every 1.5 seconds, and participants were invited to immediately recall any item they could remember in their exact serial position at the end of the list. As in standard immediate serial recall procedures, participants were also instructed to substitute any item they could not remember with the word "blank". There were 22 trials for each stimulus condition (high and low imageability). The experiment was divided into 2 blocks composed of

22 trials each to allow participants to take a very short break and refocus on the task if they needed to. The high and low imageability stimulus conditions were presented in pseudorandom succession within each of the two blocks, so that each condition was not repeated more than three times on three consecutive trials. All participants performed 4 practice trials before the beginning of the main task.

Each item was scored according to an item recall and strict serial criterion. In the item recall criterion, an item was scored as correct regardless of its serial position at recall. For instance, given the target sequence "Item1 – Item2 – Item3 – Item4 – Item5 – Item6" and the output sequence "Item2 – Item3 – Item6 – Item1 – Item4 – Item5", all item were scored as correct. In the strict serial recall criterion, an item was scored as correct only if it was recalled at its strict serial position. Given the target sequence "Item1 – Item2 – Item3 – Item4 – Item5 – Item6" and the output sequence "Item1 – Item2 – Item4 – Item5 – blank – Item6", only items 1, 2 and 6 were scored as correct.

Results and discussion

An analysis of recall performance as a function of stimulus condition and serial position was first performed (see **Figure 1.7**). The imageability effect was supported by decisive evidence, and this regardless of whether items were scored using an item recall ($BF_{inclusion}$ = 1.06e+12) or a strict serial recall ($BF_{inclusion}$ = 8.63e+10) criterion. Similarly, the position effect was also associated with decisive evidence both using an item recall ($BF_{inclusion}$ = 3.00e+12) and strict serial recall ($BF_{inclusion}$ = 6.01e+15) criterion. The interaction term provided anecdotal evidence using an item criterion ($BF_{inclusion}$ = .706) and substantial evidence using a strict serial recall criterion ($BF_{inclusion}$ = 4.08). Contrary to all previous experiments, recall performance was characterized by marked primacy effects, reflecting the fact that participants recalled information by starting at the beginning of the list.

Response latencies and response durations were also determined like in the other experiments. Results show that the participants recalled the first item very quickly (*M* = 1182, *SD* = 415), even slightly quicker than in the preceding experiments, mirroring the marked primacy effects of recall performance. At the same time, recall duration was much longer as in the other experiments (*M* = 6837, *SD* = 2111). These longer response durations could be expected if participants encode information in a more strategic manner and use the same strategies to retrieve item information at recall. The large response durations are however also the consequence of the fact that, as in any standard immediate serial recall paradigms, participants had to attempt recall of all 6 items and, when not being able to retrieve an item for a given serial position, had to say "blank".
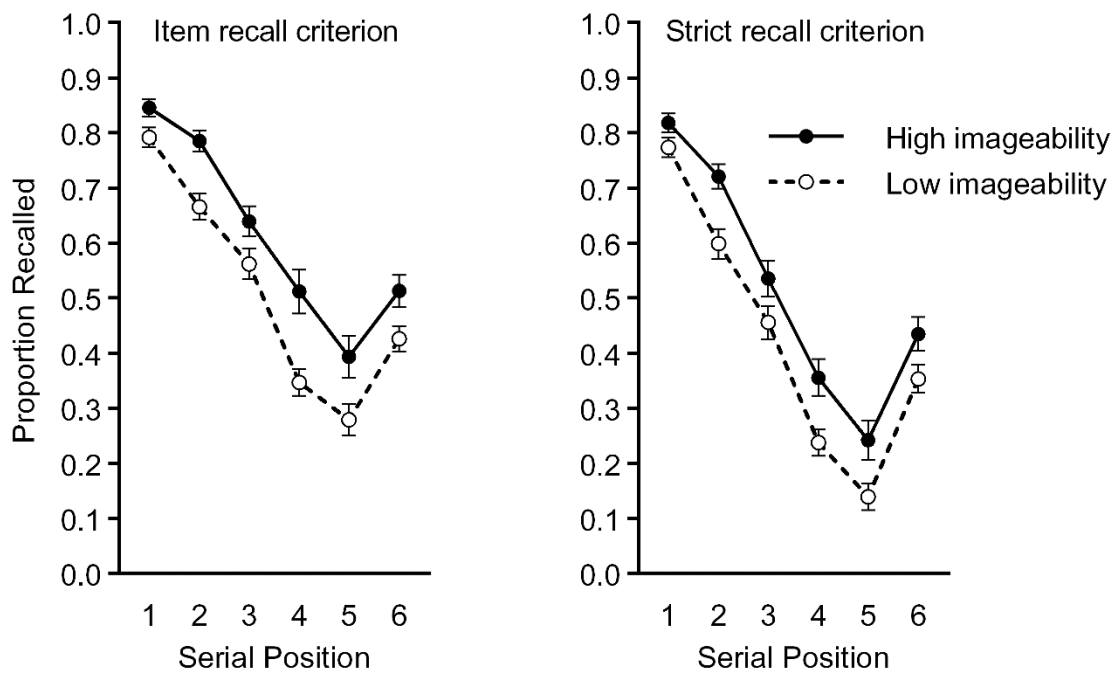
Figure 1.7. Proportion of items correctly recalled (y axis) across serial position (x axis) for Experiment 5B (Imageability effect) using an item recall (left panel) and strict recall (right panel) criterion. The x axis indicates the serial positions for target items. Error bars represent Standard Errors.

Finally, we assessed recall performance as regards to order errors. Proportion of order errors were very similar across stimulus condition, with moderate evidence supporting the null hypothesis (M = .16, SD = .09 for high imageability words and M = .17, SD = .10 for low imageability words, BF_{01} = 3.8).

In this experiment, a clear imageability effect was observed. This shows that the absence of an imageability effect in Experiments 4A, 4B and 5A cannot be attributed to the specific stimuli used in this study. The fact that an imageability effect was observed in Experiment 5B but not in Experiment 5A also indicates that slow encoding is not a sufficient condition for making an imageability effect appear in a vSTM task. A major difference between the tasks used in Experiments 5A and 5B is that list length was predictable in Experiment 5B, favouring list-level elaboration and rehearsal of memoranda (Botto et al., 2014; Morrison et al., 2016; Palladino & Jarrold, 2008; Ruiz et al., 2005). An informal assessment of strategy use indicated that participants tried to associate the words according to their meaning and rehearsed items sequentially (1; 1,2; 1,2,3; 1,2,3,4…) in Experiment 5B. We need however to remain cautious about these observations as they need to be confirmed by studies specifically designed to assess strategy use. A further important difference between immediate serial recall

paradigms, as used in Experiment 5B, and the running span procedures used in the other experiments, is that, in immediate serial recall paradigms, participants are instructed to attempt recall for all serial positions, even when information is not directly accessible anymore in their mind. This may further favor the intervention of strategic processes such as controlled retrieval processes (see below for further discussion). Finally, a further difference between Experiments 5A and 5B is the number of stimuli to be maintained. Longer list lengths were used in Experiment 5A (9-11 items) than in Experiment 5B (6 items). It could be argued that the reduced opportunity for inter-item interference in the shorter list lengths favored the expression of the word imageability effect. However, in that case robust imageability effects should also have been observed for the shortest running span lists in Experiments 4A and 4B which were also comprised of 6-word lists. This was not the case as only the item recall criterion in Experiments 4A and 4B supported an imageability effect for 6-word running span lists; evidence against an imageability effect was actually observed when using a strict serial recall criterion for the 6-word running span lists.

General discussion

Using fast-encoding, running span recall tasks, we tested the non-strategic nature of linguistic long-term memory effects in vSTM tasks, as predicted by language-based models of vSTM. In Experiment 1, we manipulated the lexicality effect and observed that words led to higher running span performance as compared to nonwords. In Experiment 2, the word frequency effect also modulated running span performance with higher recall performance for high frequency items over low frequency items. In Experiment 3, semantic relatedness led to higher running span recall performance; furthermore, semantic grouping effects were observed. Critically, in these three experiments, the linguistic effects were observed over the recency portions of the serial position curves for which items are considered to be directly available at the moment of recall. By contrast, in Experiment 4A, no reliable word imageability effect was observed. This result was replicated in Experiment 4B controlling for potential confounds of the stimulus material. When presenting items at a slowed pace in Experiment 5A, word imageability effects were still absent. However, a word imageability effect was observed in Experiment 5B using a standard immediate serial recall task. These results show that linguistic long-term memory effects can be observed in vSTM tasks minimizing strategic processes during encoding. A notable exception appears to be the word imageability effect whose expression appears to be task-dependent.

Are LTM effects reduced under non-strategic conditions?

Although we observed effects of lexicality, lexical frequency and semantic similarity by using a running span procedure, we nevertheless need to consider the possibility that these effects could be reduced relative to immediate serial recall procedures that do not minimize strategic encoding processes. If that was the case, this would indicate that these effects are determined partly by strategic and partly by non-strategic processes. We tentatively compared the effect sizes observed in this study with those of comparable studies, by computing partial eta-squared values (Lakens, 2013) and their corresponding 90% confidence intervals from F statistics and their degree of freedom using the MBESS package software under R (Kelley, 2007, 2017). We included psycholinguistic effects from studies that (1) used an open set of stimuli like in our study, (2) used a standard immediate serial recall procedure, (3) included young healthy adult participants and (4) used a within-subject design. This led to only one study for the lexicality effect (as most studies exploring lexicality effects used closed lists) and to 3 to 5 studies for the other effects (depending on the type of recall measure retained). As shown in **Tables 1.11** and 1.**12**, the effect size of the lexicality effect was comparable for the running span procedure used in the present study (all list lengths) and the immediate serial recall procedure used by Guérard and Saint-Aubin (2012). The same was true for the word frequency effect, with most studies reporting similar effect sizes as those observed in the present study, except for one study (Guérard & Saint-Aubin, 2012, Exp. 1B); the frequency effect observed in that study was particularly strong as compared to other studies. Finally, similar results were observed for the semantic similarity effect, with comparable effect sizes across experiments. This tentative comparison with studies using standard immediate serial recall procedures does not support the hypothesis of a substantive reduction of lexicality, lexical frequency and semantic relatedness effects in the running span experiments described in this study. However, this hypothesis could be explored further in future studies by directly contrasting running span and standard immediate serial recall procedures in the same experiment.

Table 1.11. *Synthesis of effect sizes observed for lexicality, frequency and semantic similarity effects in previous studies using a standard immediate serial recall procedure (item recall criterion).*

	Study	F	df$_{(effect, error)}$	η_p^2	90% CI
Lexicality	List length 6	91.834	1,31	.748	[.594, .816]
	List length 9	257.27	1,31	.892	[.82, .922]
	List length 12	147.19	1,31	.826	[.713, .873]
	Guérard & Saint-Aubin (2012) (Exp. 2)	160.99	1,19	.89	[.789, .927]
Frequency	List length 6	101.872	1,38	.728	[.587, .798]
	List length 9	105.387	1,38	.735	[.596, .803]
	List length 12	72.214	1,38	.655	[.487, .743]
	Guérard & Saint-Aubin (2012) (Exp. 1B)	168.23	1,27	.86	[.76, .901]
	Poirier & Saint-Aubin (1996)	49.23	2,34	.743	[.586, .806]
	Tse & Altarriba (2007) (Exp 1)	98.01	1,27	.784	[.634, .845]
Similarity	List length 6	124.579	1,38	.766	[.641, .826]
	List length 9	115.919	1,38	.753	[.622, .816]
	List length 12	90.165	1,38	.704	[.552, .779]
	Poirier & Saint-Aubin (1995)	82.82	1,23	.783	[.613, .847]
	Guérard & Saint-Aubin (2012) (Exp. 3)	183.97	1,19	.906	[.812, .935]
	Saint-Aubin & Poirier (1999) (Exp. 1)	49.29	1,23	.682	[.459, .777]
	Saint-Aubin & Poirier (1999) (Exp. 2)	7.64	1,23	.249	[.035, .449]

Note. The 90% confidence intervals were computed using the MBESS package software from R (Kelley, 2007, 2017).

Table 1.12. *Synthesis of effect sizes observed for lexicality, frequency and semantic similarity effects in previous studies using a standard immediate serial recall procedure (strict serial recall criterion).*

	Study	F	df(effect, error)	η_p^2	90% CI
Lexicality	List length 6	44.358	1,31	.589	[.377, .7]
	List length 9	78.31	1,31	.716	[.549, .794]
	List length 12	100.98	1,31	.765	[.62, .829]
	Guérard & Saint-Aubin (2012) (Exp. 2)	58.89	1,19	.756	[.543, .833]
Frequency	List length 6	27.575	1,38	.421	[.214, .56]
	List length 9	65.78	1,38	.634	[.459, .727]
	List length 12	29.983	1,38	.441	[.235, .577]
	Guérard & Saint-Aubin (2012) (Exp. 1B)	82.333	1,26	.76	[.593, .829]
	Roodenrys & Quinlan (2000) (Exp. 1)	74.04	1,43	.633	[.471, .722]
	Poirier & Saint-Aubin (1996)	20.95	2,34	.552	[.328, .659]
	Tse & Altarriba (2007) (Exp. 1)	67.24	1,27	.713	[.528, .795]
	Hulme, Stuart, Brown & Morin (2003) (Exp. 2)	23.2	1,15	.607	[.281, .739]
Similarity	List length 6	58.54	1,38	.606	[.424, .706]
	List length 9	16.995	1,38	.309	[.115, .467]
	List length 12	26.254	1,38	.409	[.203, .551]
	Guérard & Saint-Aubin (2012) (Exp. 3)	59.6	1,19	.758	[.546, .834]
	Poirier & Saint-Aubin (1995)	49.37	1,23	.682	[.46, .777]
	Tse, Li & Altarriba (2011)	126.29	1,150	.457	[.361, .533]
	Tse (2009)	31.81	1,24	.570	[.318, .696]

Note. The 90% confidence intervals were computed using the MBESS package software from R (Kelley, 2007, 2017).

How does the linguistic system contribute to vSTM?

The very robust lexicality, lexical frequency and semantic grouping effects observed in this study are incompatible with the idea that linguistic knowledge influences vSTM exclusively via strategic elaborative processes. As concerns the lexicality effect, it is assumed that lexical-semantic knowledge provides stability to a word's constituent phonemes, in such a way that a given word will be represented as one unitary chunk, unlike nonwords that will be represented via multiple subsyllabic units (Jefferies et al., 2006a; Patterson et al., 1994; N. Savill et al., 2016). This lexico-semantic stabilisation appears to be achieved very quickly and in the absence of strategic control in a running span paradigm. With regard to the semantic grouping effect, words sharing close semantic features are considered to be represented within the same semantic space (Lambon Ralph et al., 2017). Once a word is activated, it will

rapidly trigger the activation of all words sharing common semantic features within this semantic space, leading to the activation of a supra-item semantic context representation which associates the different semantically related words (Atkins & Reuter-lorenz, 2008; N. Martin et al., 1996). As the present data show, this semantic activation in vSTM does not require controlled encoding processes. The lexical frequency effect is likely to reflect two different levels of lexical linguistic knowledge. On the one hand, high frequency words are considered to have a lower activation threshold (McClelland & Elman, 1986) leading to facilitated activation during encoding. On the other hand, high frequency words are not only more frequent, but they also tend to co-occur more frequently (as observed by Hulme et al., 2003 and in this study), leading to stronger inter-item associations and inter-item stabilizing and cueing effects. Although the exact locus of the lexical frequency effect remains controversial (e.g. Hulme et al., 2003; L. M. Miller & Roodenrys, 2012; Parmentier, Comesaña, & Soares, 2017; Poirier & Saint-Aubin, 2005), it is very likely that both inter-item associative knowledge and item-based properties contribute to the lexical frequency effect (Poirier & Saint-Aubin, 2005; Tse & Altarriba, 2007). The critical finding of the present study is that these item and inter-item activations occur in a non-strategic manner in vSTM.

A further important finding of this study is the fact that the word imageability dimension did not lead to detectable effects on recall performance when using a fast-encoding running span procedure. In the psycholinguistic literature, the advantage in reaction times commonly observed for high over low imageability words in tasks such as lexical decision is generally attributed to rapid and automatic feedbacks between lexical and semantic levels of language processing contributing to accelerated lexical access for high imageability words (Evans et al., 2012, 2017; Pexman et al., 2002; Yap et al., 2015). High imageability, concrete words are indeed considered to be associated with richer and more stable semantic information as compared to low imageability, abstract words (Binder, Desai, Graves, & Conant, 2009; Hill, Korhonen, & Bentz, 2014; Roxbury, McMahon, & Copland, 2014; Sabsevitz, Medler, Seidenberg, & Binder, 2005). Critically, this faster lexical access for high over low imageability words does not appear to lead to higher performance in a running span recall task as shown by the results of this study. It could be argued that semantic features for a given word are distributed over distant neural networks (Binder, 2016; Lambon Ralph et al., 2017) and may require a longer time-scale to reach stabilized levels of co-activation. Experiment 5A showed that the lack of an imageability effect in Experiments 4A and 4B was not due to the fast presentation of these experiments. Other studies have shown that slower presentation rates can enhance imageability effects (Campoy et al., 2015),

meaning that a slower presentation rate is beneficial for the imageability effect, but only in paradigms in which it can be observed at first hand. As shown by Experiment 5B, the imageability effect only appeared in standard immediate serial recall tasks which have been shown to allow for the implementation of list-level and item-level elaboration strategies (Bailey et al., 2011; Morrison et al., 2016). List-level elaboration of memoranda, including repeated reactivation of the same stimuli, may favour the activation of the deep semantic features that distinguish high imageability from low imageability words; these features may not necessarily be immediately activated each time a given word is encountered but their activation may depend on the context in which the word appears. This is in agreement with data suggesting that full semantic activation does not necessarily occur in vSTM tasks under standard conditions (N. Savill et al., 2015). Moreover, access to semantic features in the concrete/abstract dimension has been shown to be task-dependent in lexicality judgements (Evans et al., 2012). On the other hand, surface-level associative semantic features defining semantic relatedness will easily and automatically get co-activated each time a list of words sharing semantic features is presented, via the redundant semantic features that characterize and associate these words (Acheson & MacDonald, 2009; Dell, 1986; N. Martin et al., 1996). Although the present study shows that the expression of word imageability effects in vSTM is task-dependent and seems to appear in tasks in which strategy use is more likely to occur, future studies need to define the exact strategies and processes that lead to the expression of the word imageability effect in immediate serial recall tasks.

At the same time, Campoy et al. (2015) observed concreteness effects when an immediate serial recall task was performed simultaneously with an attention-demanding concurrent task, contradicting the observation of a lack of imageability effect in the present study for the non-strategic, running span encoding conditions. This raises the question whether strategic processing was prevented to the same extent in both studies. In the present study, an analysis of response latencies indicated that participants recalled information very quickly, in line with task instructions asking participants to recall items directly as available in their "mind". Also during encoding, time for implementing strategic processes other than careful listening to memoranda was highly reduced (25msec) in the running span procedure used in Experiment 4B while in the study by Campoy et al. there was an idle time of on average 700ms between items, after taking into account the time needed to perform the concurrent task. The use of short (i.e. 5 items) lists of predictable length in Campoy et al. could have further facilitated the implementation of encoding strategies (Bunting et al., 2006; Palladino & Jarrold, 2008). Hence, despite the use of a concurrent task, there could

have more time available for implementing strategies in the paradigm used by Campoy et al. than with the running span paradigm used in the present study.

Implications for vSTM models

The present study was framed by linguistic models of vSTM which consider that vSTM reflects the activated part of the linguistic system (Acheson & MacDonald, 2009; Gupta, 2009; Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999). Our results provide critical evidence for these models by showing that the non-strategic activation properties of knowledge stored in the linguistic system also seem to characterize performance in vSTM tasks, except for deep semantic knowledge associated with the word imageability effect. Our results also support attention-based models (Barrouillet et al., 2004; Barrouillet & Camos, 2007; Cowan, 1995, 1999, 2001; Oberauer, 2002) for which short-term maintenance is the product of the interactions between an attentional system (such as the focus of attention) holding a limited quantity of information and the part of long-term memory which is currently activated. For instance, in Cowan's embedded processes model, vSTM span is constrained by the number of chunks (distinct units) that can be consciously maintained in the focus of attention. Direct interactions between the properties of the language system and the amount of information held in the focus of attention are a critical prediction of this model, as language knowledge will allow memoranda to be regrouped in larger chunks when available. As regards the lexicality effect, semantic knowledge is assumed to chunk phonemes into coherent word units (Jefferies et al., 2006a; Patterson et al., 1994). Word frequency can also be considered to lead to more stable chunks in the focus of attention due to the stronger and easier-to-activate lexical representations for high frequency words (Jefferies et al., 2006a). The same principle can be applied to semantic similarity, with semantically related words being chunked in triplets at the semantic level. These predictions are supported by the results of the present study given that the running span task, as used in this study, has been considered to directly tap the amount of information held in the focus of attention (Cowan, Elliott, et al., 2005).

Finally, it is also important to examine our results in the light of the redintegration hypothesis, which considers that psycholinguistic effects in vSTM tasks stem from item reconstructive processes occurring during the recall stage of vSTM, and allow for completion of degraded phonological traces in vSTM by selecting, during recall, the linguistic representations in the language network that match best the degraded vSTM traces (Hulme et al., 1997, 1991; Schweickert, 1993). It could be argued that the different psycholinguistic effects that have been observed in this study are merely the result of item reconstructive processes occurring during recall. Indeed,

provided that an item is degraded, it has been proposed that redintegration is performed based on phonological and/or semantic STM traces and enhances vSTM span more strongly for easier-to-reconstruct items such as items associated with richer or more distinctive lexico-semantic representations (Walker & Hulme, 1999). Hulme, Newton, Cowan, Stuart and Brown (1999) further proposed that redintegration processes are likely to be reflected by slowed processes during recall such as longer inter-item pauses. The fast response durations in the running span tasks (Experiment 1 through 5A) suggest that participants recalled information whose traces were well activated in their mind, and trace redintegration requirements may have been minimal for these experiments. However, in Experiment 5B, where a word imageability effect was observed, participants were instructed to attempt item recall for all serial positions, as required in standard immediate serial recall tasks. This may have led to a stronger need for strategic retrieval and redintegration processes as participants had to recall items even when the items were strongly degraded. Experiment 5B was also associated with much longer response output times as compared to the other experiments (note however that these longer response output times were at least partially due to the need to say "blank" when a given item could not be recalled at all). In other words, particularly the word imageability effect observed in Experiment 5B could have been due to redintegration processes at recall, in addition to list-level elaborative encoding strategies that characterize standard immediate serial recall paradigms.

Conclusions

This study shows that lexical and semantic levels of language processing influence vSTM performance in conditions minimizing strategy use during encoding, confirming and refining the predictions of current language-based models of vSTM. Our study further suggests that the non-strategic expression of linguistic long-term memory effects in vSTM characterizes lexical and surface-level associative semantic knowledge, while deep semantic knowledge appears to be activated in a task-dependent manner in vSTM tasks. Future studies need to determine the exact strategies and processes that lead to the expression of word imageability effects in standard immediate serial recall tasks.

Appendix A. Allowed distortions covering one articulatory transformation

Consonants :	Vowels :
b : p – m – d – v	**i** : e – y
d : t – b – n – g – r – l	**e** : i – ɛ – ɛ̃ – œ
f : v – p – s – ʃ	**ɛ** : ɛ̃ – e – a – ə – œ̃
g : k – d – ɲ – j	**a** : ɑ̃ – ɛ – ɛ̃ – ə – œ̃ – ɔ – o
ɲ : g – n	**y** : œ – i – u
j : g – l	**œ** : y – ə – œ̃ – e – o
k : g – t	**ə** : œ̃ – œ – a – ɑ̃ – ɛ – ɛ̃ – ɔ – ɔ̃
l : r – d – j	**u** : y – o
m : n – b	**ɔ** : ɔ̃ – o – a – ɑ̃ – ə – œ̃
n : m – d – ɲ	**o** : u – ɔ – ɔ̃ – œ
p : b – t – f	**ɛ̃** : ɛ – e – a – ə – œ̃
r : l – d	**œ̃** : ə – œ – a – ɑ̃ – ɛ – ɛ̃ – ɔ – ɔ̃
s : z – t – f – ʃ	**ɔ̃** : ɔ – o – a – ɑ̃ – ə – œ̃
t : d – p – k – s	**ɑ̃** : a – ɔ – ɔ̃ – ə – œ̃
v : f – b – z – ʒ	
ʒ : v – s – ʃ	
ʃ : ʒ – f – s	
z : s – d – v – ʒ	

Study 2

Testing the redintegration hypothesis by a single probe recognition paradigm

Benjamin Kowialiewski and Steve Majerus

Abstract. The lexicality effect in verbal short-term memory (STM), in which word lists are better recalled than nonword lists, is considered to reflect the influence of linguistic long-term memory (LTM) knowledge on verbal STM performance. The locus of this effect remains however a matter of debate. The redintegrative account considers that degrading phonological traces of memoranda are reconstructed at recall by selecting lexical LTM representations that match the phonological traces. According to a strong version of this account, redintegrative processes should be strongly reduced in recognition paradigms, leading to reduced LTM effects. We tested this prediction by contrasting word and nonword memoranda in a fast encoding probe recognition paradigm. We observed a very strong lexicality effect, with better and faster recognition performance for words as compared to nonwords. These results do not support a strong version of the redintegrative account of LTM effects in STM which considers that these LTM effects would be the exclusive product of reconstruction mechanisms. If redintegration processes intervene in STM recognition tasks, they must be very fast, which at the same time provides support for models considering direct activation of lexico-semantic knowledge during verbal STM tasks.

Introduction

Verbal short-term memory (STM) performance is influenced by several lexico-semantic effects (Besner & Davelaar, 1982; Hulme et al., 1991) such as the lexicality effect in which word stimuli are better recalled than nonword stimuli. The lexicality effect was first observed in standard immediate serial recall tasks, requiring participants to recall in forward order short lists of items (usually 5 or 6) (Besner & Davelaar, 1982; Hulme et al., 1991). This effect has subsequently been observed in other experimental paradigms, including recognition paradigms (Gathercole et al., 2001;

Jarrold et al., 2008; Jefferies et al., 2006b), mixed word/nonword list recall paradigms (Jefferies et al., 2006a) and backwards recall paradigms (Guérard & Saint-Aubin, 2012). Despite the robustness of this effect, its nature remains a matter of theoretical debate.

Although different theoretical approaches agree that this effect represents the influence of linguistic long-term memory (LTM) representations on verbal STM performance, they disagree on the mechanisms that are supposed to give rise to these LTM-STM interactions. One theoretical approach is the redintegrative account considering that verbal STM primarily relies on phonological codes for storing memoranda in a dedicated STM buffer; these phonological traces are subject to degradation via decay or interference unless rehearsal is provided via an articulatory mechanism (Hulme et al., 1997, 1991; Hulme, Roodenrys, Brown, & Mercer, 1995; Schweickert, 1993; Schweickert et al., 1999). At the moment of recall, pattern completion is performed, in which degraded phonological traces are reconstructed, either through a comparison mechanism with lexical knowledge (Schweickert, 1993) or through an automatic reconstruction process via the language production system (Hulme et al., 1995). More specifically, when comparing immediate serial recall for word versus nonword lists, the superiority of word recall is supposed to stem from the lexical reconstruction of the degraded phonological representations at the moment of recall. As lexical reconstruction is not possible for nonword stimuli, pattern completion cannot be performed and recall performance will be reduced. Critically, early accounts of the redintegration framework consider that redintegration processes operate exclusively during restrieval: "…*any item may benefit if a long-term phonological representation of it is available during the retrieval process*" (Hulme et al., 1991, p. 699). Other theoretical approaches consider that VSTM is the product of the direct activation of linguistic knowledge stored in LTM (Acheson & MacDonald, 2009; Burgess & Hitch, 1999, 2006; Cowan, 1995; Gupta, 2009; Jones & Macken, 2017; Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999; Patterson et al., 1994). These language-based models strongly differ from the redintegration hypothesis as the intervention of linguistic knowledge is not supposed to be the exclusive result of post-encoding retrieval processes, but is supposed to occur already during encoding and maintenance by providing stabilizing feedback activation to memory items.

Evidence proposed to support the redintegration hypothesis comes from studies showing strongly reduced LTM effects in recognition paradigms when no overt output is required (Gathercole et al., 2001; Nimmo & Roodenrys, 2005; Thorn, Gathercole, & Frankish, 2002). Since the pattern completion mechanism is supposed to occur when retrieving an item for output, effects stemming from reconstructive processes such as the lexicality effect should disappear in the absence of recall output. Gathercole et al.

(2001) were the first to observe an important reduction of the lexicality effect in both children and adults when using a list probe recognition paradigm. Likewise, Thorn, Gathercole, and Frankish (2002) observed, in bilingual speakers, that the superior STM performance for words presented in first versus second language disappeared when using a probe recognition paradigm instead of a recall paradigm.

At the same time, the results of these studies are difficult to interpret as the recognition paradigms that were used tested serial order recognition rather than item recognition, by presenting negative probe lists in which the serial position of two adjacent items had been exchanged relative to the target list. This is problematic as linguistic LTM knowledge has been shown to facilitate STM for item information but much less for serial order information (Allen & Hulme, 2006; Campoy et al., 2015; Hulme et al., 1997, 2003; Poirier & Saint-Aubin, 2005; Romani et al., 2008; Roodenrys et al., 2002; Walker & Hulme, 1999). Moreover, when impacting serial order information, LTM effects have been shown to go in an opposite direction, with serial order recall errors being less frequent for nonwords than words, after controlling for the overall amount of words and nonwords recalled (Fallon et al., 2005; Guérard & Saint-Aubin, 2012; Jefferies et al., 2006a; Saint-Aubin et al., 2005; Saint-Aubin & Poirier, 2000; Tse et al., 2011). More generally, serial order recognition tasks are not optimal for assessing the impact of linguistic LTM on STM as these tasks tap item processing only minimally (Majerus, 2009, 2013).

Indeed, studies using recognition tasks specifically designed to assess memory for item information have been shown to lead to robust lexicality effects (Jarrold et al., 2008; Jefferies et al., 2006b; Turner & Henry, 2004). These results appear to contradict the original account of the redintegration hypothesis, which predicts the absence of lexicality effects in recognition paradigms, irrespective of the type of information – item or serial order – that is being probed. However, it could still be argued that these LTM effects stem from covert redintegration processes occurring during list presentation in recognition paradigms, and particularly during the inter-item interval. Rehearsal of memoranda during list encoding and during probe list presentation could be considered as a covert recall process, and hence also provides the opportunity for trace reconstruction (see Jefferies, Frankish, & Ralph, 2006; Turner & Henry, 2004). Furthermore, in matching-span tasks requiring the comparison of two lists, rehearsal and reconstruction of the first encoding list may also occur as the second list is being presented.

The aim of this study was to provide a critical test of a strong version of the redintegration account, considering that the lexicality effect is exclusively the product of reconstruction mechanisms during recall, by using an item probe recognition

paradigm that minimizes any opportunities for overt or covert redintegration, both during encoding and recognition. In order to achieve this, we used a recognition variant of the running span procedure (Pollack et al., 1959) in which participants are presented with continuous lists presented at a very fast rate (>2items/s), and their memory for the lists is tested at unpredictable moments. The fast presentation rate and the unpredictability of sequence length forces participants to attend to each item as it appears while preventing rehearsal and the use of other types of controlled strategies (Bhatarah, Ward, Smith, & Hayes, 2009; Bunting et al., 2006; Cowan, Elliott, et al., 2005; Hockey, 1973). Serial positions curves for list recall in this type of paradigm are typically associated with strong recency effects and poor primacy effects, reflecting the fact that participants passively attend each item until the end of the list (Bunting et al., 2006; Hockey, 1973; Palladino & Jarrold, 2008; Ruiz et al., 2005). In the present study we presented items at a very fast rate (2 items/s), with items being presented in lists of varying length, strongly reducing the likelihood of any overt or covert reconstruction processes that could occur between the presentation of successive items during encoding. In addition, to diminish the likelihood of reconstruction processes during retrieval, participants were instructed to determine as quickly as possible (within 1750 ms post-stimulus onset) whether a given probe item matched one of the items in the memory list. Hence, if the lexicality effect is the product of a redintegration process occurring during retrieval or encoding phases of the task, this effect should be strongly reduced or disappear in the present experiment. Alternatively, language-based model predict preserved lexicality effects in these experimental conditions as LTM knowledge is supposed to directly determine the stability of memoranda, and this already during the encoding phases of the task, when no overt or covert reconstruction is required.

Experiment

Method

Participants. A total of forty-nine participants (39 females, 10 males) aged between 18 and 29 years (M = 22.59, SD = 2.4) were recruited from the university community after giving their informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials. We selected 75 words stimuli with a mean log frequency of .43 from the French lexical database Lexique 3 (New, 2006). From this initial set, we constructed 75 nonwords stimuli that were furthermore matched to the words as regards the number of phonemes (all stimuli were 5 phonemes long), phonological structure

(consonant, vowel and semivowel; a strict matching was applied between words and nonwords), lexical competitors (i.e. the number of words sharing their first two phonemes with the target stimulus, then log-transformed) (M_{log} = 1.71, SD_{log} = .51 for words and M_{log} = 1.64, SD_{log} = .47 for nonwords, BF_{01} = 3.96, supporting the null over the alternative hypothesis; see the description of Bayes factor values in the Methods section), competitor cumulative frequency (the summed and then log-transformed frequency of all competitors for each stimulus) (M_{log} = 2.54, SD_{log} = .88 for words and M_{log} = 2.53, SD_{log} = .95 for nonwords, BF_{01} = 5.67), biphone frequency (M = 785.62, SD = 426.02 for words and M = 687.93, SD = 447.63 for nonwords, BF_{01} = 2.41) and neighborhood density (i.e. number of words that can be created from the target stimulus after substitution of one phoneme) (M = .11, SD = .39 for words and M = .04, SD = .26 for nonwords, BF_{01} = 2.81). The word and nonword stimuli had no other competitors except themselves (i.e. for words) from the third phonological point. In variance with previous studies, we controlled for lexical competition and uniqueness point since it has been shown that these variables modulate effort required to access linguistic information (Gaskell & Marslen-Wilson, 2002; Marslen-Wilson, 1987; Zhuang et al., 2011; Zhuang, Tyler, Randall, Stamatakis, & Marslen-Wilson, 2014). We opted for using a very strictly matched word and nonword stimulus set in order to ensure that any differences in performance and response time between the two stimulus categories was due to differences in lexical status and not to other phonological variables; this procedure however also implied that our stimulus set was closed and that items were sampled repeatedly (10 times) for creating the STM lists. The list of stimuli used in their phonetic transcription is available in appendix A.

The stimuli were recorded by two different native French speakers, one female and one male speaker, using a neutral voice. This allowed us to present target lists and probe stimuli in two different voices, and to prevent participants from using low-level perceptual recognition strategies. Each item was recorded as a separate .wav file. Background noise was removed using Audacity® software and each stimulus was normalized to a duration of 470 ms without altering the pitch.

Procedure. The same word and nonword stimuli were used for each participant but they were pseudorandomly sampled for each list with the following constraints: a same stimulus could not appear twice in the same sequence and a given item only appeared once in the same serial position throughout the entire task. For half of the participants, the target lists were presented with a male voice and the probe stimulus was presented with a female voice, and the reverse voice pairing was used for the second half of participants. The word and nonword conditions were presented in pseudorandomized order, with the constraint that a given condition did not occur on

more than three consecutive trials. There were four different list lengths varying between 11 and 14 stimuli. A given list length could not be repeated more than twice on two consecutive trials. There were 15 trials per list length and per condition, with a total of 60 trials in each stimulus condition. No individual stimulus was probed twice through the whole experiment. We furthermore probed memory only for the last eight positions of each sequence in order to probe information held in STM and to avoid biasing the results by episodic long-term memory retrieval processes or other strategic processes, in line with the original requirements of running span paradigms (Cowan, Elliott, et al., 2005). The non-matching probe stimuli were randomly sampled from the stimulus set, with the constraint that a given stimulus could be used only once as a non-matching probe stimulus. Two thirds (40) of trials were matching trials. This unbalance between matching and non-matching trials ensured that each serial position was probed an equal number of times (5 times for matching trials) while keeping task length at a reasonable level. To ensure that recognition performance was not biased by differences in phonological similarity of the negative probe stimuli relative to the word versus nonword stimuli of the memory lists, we computed the average Levenshtein distance (Levenshtein, 1966) between negative probes and the items of the respective memory lists, on a trial by trial basis. Values across trials were then averaged for each participant and as a function of word versus nonword conditions. Although paired T-Tests indicated that negative probes in the nonword condition had a slightly larger Levenshtein distance than negative probes in the word condition (M = 4.64, SD = .04 for words, M = 4.70, SD = .04 for nonwords, BF_{10} = 1.641e+7), this difference was very small in absolute values (M_{diff} = .06, SD = .05) and not likely to be meaningful at the cognitive level; a value of 1 would indicate a one-phoneme change between the source and target stimulus. Furthermore, if this difference is to convey any processing advantage, it should favour the nonword condition and hence go against the observation of a lexicality effect.

At the beginning of each trial, a timer counting down from 3 was presented, followed by a blank screen and the presentation of the target list. The items within the lists were separated by a very brief inter-stimulus interval of 10ms, leading to a presentation rate of 2.08 items/s. The probe stimulus was presented in average 300ms (plus or minus a value randomly selected from a continuous uniform distribution with min = 0 and max = 75) after the final item of each list. Participants had to decide whether the probe stimulus matched one of the items in the target list or not, by pressing the "S" key for "yes" and the "L" key for "no". Speeded responses were required by limiting response time to 1750ms post stimulus-onset. When participants did not respond within the allocated time period, they received an on-screen message

reminding them to respond faster. Participants performed 6 practice trials before starting the main STM task and received a feedback of their performance directly after each practice trials.

Task presentation was controlled via OpenSesame software (Mathôt et al., 2012) running on a desktop computer. The auditory stimuli were presented at an individually adjusted, comfortable listening level, via headphones directly connected to the computer.

Statistical analysis. A Bayesian statistical framework was used instead of a classical frequentist statistical approach as the latter approach has been shown to overstate evidence for an effect (Dienes, 2011; Rouder & Morey, 2011; Wagenmakers, 2007), and this particularly as sample size increases (Kruschke, 2011). Also, Bayesian statistics allow to directly test for the presence of the null effects (Dienes, 2014); this is a critical test for the present study as the redintegration account is predicting a null effect for the influence of lexicality on STM performance in a recognition paradigm. Bayesian statistics can be interpreted using Bayesian model comparison, which quantifies the strength of evidence associated to a given model as compared to other models, and correspond to the ratio of the Bayes Factor. By default, equal a priori probabilities were assumed for all models. Hence, Bayesian statistics use continuous values to update one's beliefs in favour of a given model rather than an arbitrary threshold. The BF_{10} value indicates the likelihood ratio of the alternative model (H_1) relative to the null model (H_0), and the BF_{01} values indicates the likelihood ratio of H_0 relative to H_1. To interpret the Bayes Factor (BF), the following classification was used (Jeffreys, 1998; Wagenmakers et al., 2011): no evidence for BF < 1, anecdotal evidence for 1 < BF < 3, moderate evidence for 3 < BF < 10, strong evidence for 10 < BF < 30, very strong evidence for 30 < BF < 100 and extreme/decisive evidence for 100 < BF . All the analyses were performed using JASP (JASP Team, 2017) and all parameters were set to the default Cauchy prior distribution as implemented in JASP (Version 0.8.5.1).

Table 2.1. *Proportions for hit and false alarms.*

Lexicality	Hit	False alarms
Word	.873 (.073)	.268 (.125)
Nonword	.802 (.116)	.318 (.166)

Note. Standard deviations are marked in parenthesis.

Results

First, we assessed the presence of a lexicality effect by comparing recognition accuracy, using d′ scores, for word versus nonword list conditions (Stanislaw & Todorov, 1999). Overall, d′ recognition scores were higher for words as compared to nonwords, as shown in **Figure 2.1**, left panel. Decisive evidence supported this
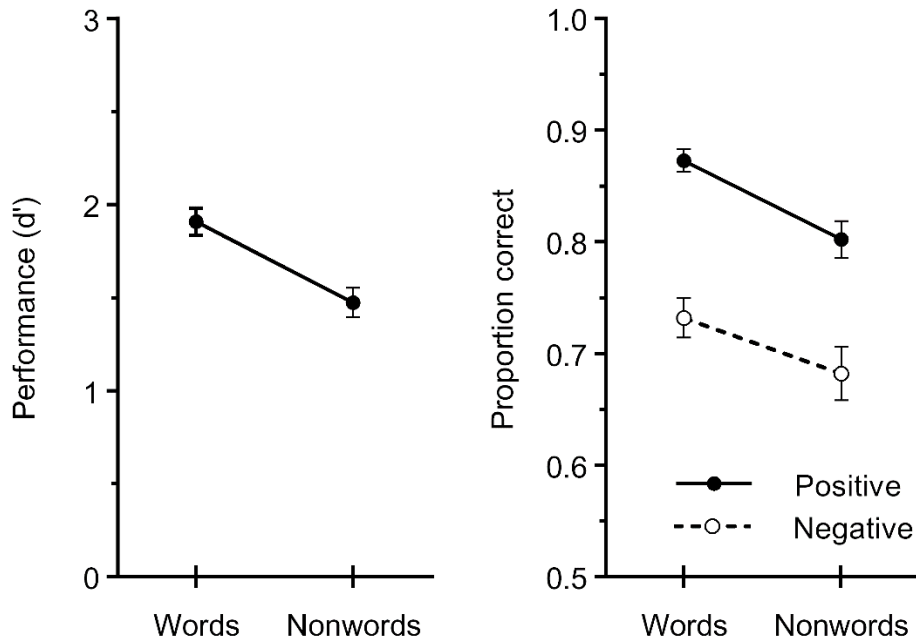
Figure 2.1. Discrimination scores (d') as a function of word/nonword stimulus conditions (left panel), and recognition accuracy as a function of stimulus condition (lexicality) and probe type (positive/negative) (right panel). Error bars represent Standard Errors.

lexicality effect, as shown by a Bayesian paired T-Test (BF_{10} = 121.9). These results clearly show that the lexicality effect was preserved in the item probe recognition paradigm used in this study. Proportions for hits and false alarms are presented in **Table 2.1**.

We also assessed the impact of lexicality for positive and negative trials separately. It could be argued that redintegration processes may be needed for rejecting negative probes as the target needs to be retrieved, but not for accepting positive probes which directly cue the target item. In that case, if lexicality effects are the consequence of redintegration processes, a lexicality effect may be predicted for negative but not positive probes. Using a Bayesian repeated measures ANOVA with probe type (positive, negative) and stimulus material (words, nonwords) factors, the model providing the highest BF value was the model including both effects of probe type and lexicality (BF_{10} = 5.76e+10) and was 3.97 times more likely as compared to the second best model including both effects and the interaction term (BF_{10} = 1.45e+10). As shown in **Figure 2.1**, right panel, recognition accuracy was higher for positive probes as compared to negative probes, with a similar impact of lexicality for the two types of probes. Full Bayesian statistical results are available in supplementary material **S1**. These results indicate that positive and negative trials both led to robust lexicality effects.
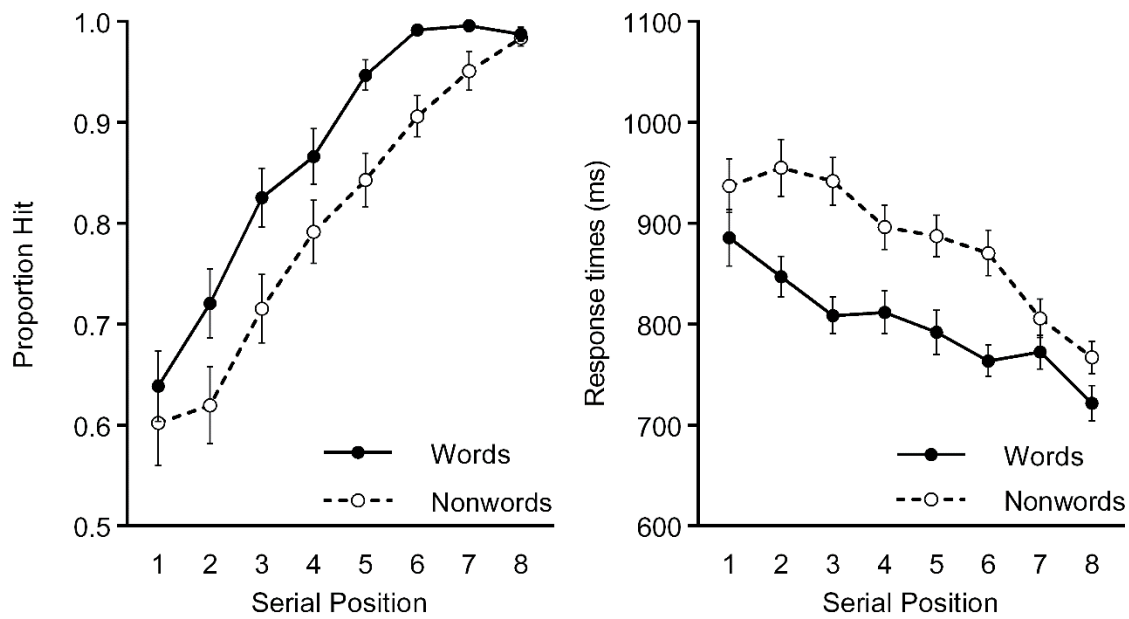
Figure 2.2. Recognition accuracy (left panel) and response times (right panel) for hits as a function of serial position and word/nonword stimulus condition. Error bars represent Standard Errors.

Next, we assessed the proportion of hit responses as a function of serial position and stimulus condition using a Bayesian Repeated Measure ANOVA. The model providing the strongest evidence was the model including both effects of lexicality and serial position ($BF_{10} = 1.211e+73$), and was 34.29 more likely as compared to the second best model including the effects of lexicality, serial position and the interaction term ($BF_{10} = 3.532e+71$). These results show that the hit rate was higher for words as compared to nonwords, and the lexicality effect did not interact with serial position. Moreover, strong recency effects were observed, with hit rates being progressively higher across serial position (see **Figure 2.2**, left panel). Full statistical results are available in supplementary material **S2**.

A similar analysis was conducted for responses times for hit trials. A Bayesian Repeated Measure ANOVA with stimulus condition (word, nonword) and serial position (1 through 8) factors showed that the model with the highest BF was the model including all effects of lexicality, serial position and the interaction term ($BF_{10} = 1.216e+61$) and was 41.70 times more likely as compared to the second best model not including the interaction ($BF_{10} = 2.916e+59$). The interaction was explored using Bayesian T-Tests, showing decisive evidence for the lexicality effect across serial positions 2 through 6 ($BF_{10} > 100$), moderate evidence at position 7 ($BF_{10} = 5.53$) and strong evidence at position 8 ($BF_{10} = 22.99$); however, no reliable evidence for a

lexicality effect was observed for serial position 1 (BF_{10} = 2.09). As show by this analysis, the interaction stemmed from the reduced effect size observed in positions 1, 7 and 8, as compared to the other positions. Word trials were associated with faster response times as compared to nonword trials, with response times progressively decreasing across serial position (**Figure 2.2**, right panel). Full statistical results are available in supplementary material **S2**.

Discussion

In this study, we assessed the impact of lexicality on a running span memory task which severely restricts the possibility of overt or covert redintegration processes during all stages of STM. Using an item probe recognition procedure, we observed a preserved lexicality effect, with better and faster recognition performance for words as compared to nonwords. These results do not support the redintegration account of LTM effects on verbal STM, and which would have predicted the absence of a lexicality effect under the given testing conditions.

For the original redintegration account (Hulme et al., 1997, 1991; Schweickert, 1993), phonological information is stored in a dedicated buffer during the encoding stage of STM. These phonological item representations degrade very quickly (as a result of temporal decay and/or interference) unless they can be reactivated via rehearsal (Schweickert et al., 1999). LTM effects are supposed to enhance STM span at the post-encoding retrieval stage, through a reconstruction mechanism during which degraded phonological traces and LTM representations are compared: "*When attempting to recall items […], knowledge of the spoken form of words may help the person perform pattern completion on decayed traces and so successfully recall them*" (Hulme et al., 1997, p. 1218). Since recognition paradigms do not require recall output, there is no need for pattern completion and hence LTM effects should disappear or be strongly reduced (Gathercole et al., 2001; Thorn et al., 2002; Walker & Hulme, 1999). Note, however, that the observation of a preserved lexicality effect in recognition paradigms does not rule out the existence of redintegration processes in other type of tasks such as immediate serial recall tasks. In these tasks, it could be that STM trace completion is also taking place at the moment of recall in addition to direct LTM activation during encoding. Our results show that, if any redintegration mechanism exists, it is not the sole process contributing to the lexicality effect.

Although our experimental procedure was designed to limit opportunities for redintegration via overt or covert retrieval attempts as much as possible, one could still argue that the response time limit of 1750 ms during the response stage allowed for rapid rehearsal of at least one or two items during the recognition phase. However,

when looking at the average reaction times for hit responses (see **Figure 2.1**, right panel), it appears that participants initiated their response at maximum ~950ms post-stimulus onset for nonword stimuli and at maximum ~900ms for word stimuli. Since probe stimulus length was normalized to 470ms, this amounts to a mean post-stimulus decision time of 480ms, suggesting that participants responded very rapidly, leaving only a very small amount of time for rehearsal of the memory sequence. Furthermore, one could argue that in recognition paradigms, no LTM redintegration processes are needed for positive trials as the positive probe directly cues and completes the representation of the target stimulus, but negative trials may need the re-activation of the memory items providing opportunity for LTM redintegration. However, this argument was also contradicted by the results: both negative and positive trials led to robust lexicality effects.

Our results are consistent with linguistic accounts of STM-LTM interactions, considering that verbal STM is the product of the activation of phonological, lexical and semantic knowledge stored in the linguistic system, and that this activation provides the representational basis for STM retention (Acheson & MacDonald, 2009; Burgess & Hitch, 1999, 2006; Cowan, 1995; Gupta, 2009; Jones & Macken, 2017; Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999; Patterson et al., 1994). In these models, encoding of information in verbal STM implies the temporary activation of corresponding representations directly in the linguistic knowledge base after sensory input. Since activation of lexical information within the linguistic system is implicitly supposed to operate very rapidly (Egorova, Pulvermüller, & Shtyrov, 2014; MacGregor et al., 2012; N. Martin et al., 1996), lexicality effects in fast running span paradigm with very short response time can easily be accounted by these models, by assuming that lexico-semantic knowledge stabilizes phonological information already during the encoding stage, leading to a general memory advantage for words over nonwords irrespective of the type of STM task. At the same time, linguistic accounts do not exclude the existence of reconstruction mechanisms. These mechanisms are actually considered to be an integral part of linguistic accounts but, contrary to the redintegration accounts, these mechanisms are considered to operate at any stages of verbal stimulus processing, from perception to maintenance and recall stages. Receptive speech processing is characterized by constant interactions between pre-existing knowledge (at phonetic, phonological, lexical and semantic levels) and the incoming speech signal, the latter being predicted and completed by the former as well as by contextual information, a process which has been termed predictive coding (Hannemann et al., 2007; Heald & Nusbaum, 2014; Leonard et al., 2016; Sohoglu et al.,

2012). These linguistic reconstruction processes are considered to be very fast and are part of both receptive and productive language processing stages.

Other studies have also reported results that are difficult to reconcile with the original redintegration account (Jefferies et al., 2006a; Lambon Ralph et al., 2017). Jefferies, Frankish, and Lambon Ralph (2006a) observed in an immediate serial recall task that words were associated with poorer recall performance when presented in mixed word/nonword lists than when presented in pure word lists. Since the pattern completion mechanism in the redintegration hypothesis is supposed to operate on an item-to-item basis, the lexical status of neighboring items should not affect recall success for individual words. They further observed a higher rate of inter-item phoneme migrations errors (e.g. recalling "rug – bite" instead of "bug – rite") for pure nonword lists than for mixed nonword lists. Redintegration is supposed to affect *whole-item identity* and does not make any predictions regarding *phoneme order* in nonwords, except that recall should operate in a linear manner. However, linguistic accounts of STM are compatible with these findings by considering that contrary to nonword items, word items are stored as lexical chunks already during the encoding stage (Patterson et al., 1994); this strongly reduces the opportunity for inter-item phoneme migrations for word stimuli. Poorer recall for words in mixed word/nonword lists can also be explained by assuming that lexico-semantic activation during encoding is noisy and unstable due to the presence of the nonwords which draw linguistic processing resources to phonological levels of processing (McDermott, Petersen, Watson, & Ojemann, 2003). Furthermore, for pure word lists, items can be maintained using associative semantic representations that bind several items or even all items of a list, thereby regrouping items into a smaller number of conceptual units (Bailey et al., 2011; Morrison et al., 2016).

Conclusions

The robust lexicality effect observed in this study using a paradigm minimizing covert or overt output redintegration processes is inconsistent with the original redintegration framework stating that reconstruction does not operate in the absence of output requirements. Instead, direct stabilizing activation of lexical representations and/or very fast pattern completion processes during encoding and STM maintenance are more likely to explain the lexicality effect observed in STM recognition experiments.

Appendix A

Phonetic transcription of the words used in the experiment.

/ãbjãs/	/enɔʁm/	/igwan/	/pãflɛ/	/sytyʁ/
/adɛpt/	/etwal/	/imœbl/	/pɛ̃bɛʃ/	/tɵ̃syʁ/
/aməsɵ̃/	/ɛ̃bɛʁb/	/iɲɔbl/	/pjœvʁ/	/teflɵ̃/
/amjãt/	/ɛmabl/	/jauʁt/	/pœzœl/	/tɛmpo/
/anɥɛl/	/ɛnəmi/	/kesje/	/pokɛʁ/	/tigʁe/
/aʃwaʁ/	/fəmɛl/	/kjɔsk/	/pubɛl/	/tiɲas/
/bãljø/	/fuʒɛʁ/	/kojɔt/	/ʁejyʁ/	/vajãs/
/baɲɔl/	/fyʃja/	/kugaʁ/	/ʁɛ̃gaʁ/	/vodka/
/bɵ̃ʒuʁ/	/gepaʁ/	/lɛ̃pid/	/ʁyʁal/	/vədɛt/
/bemɔl/	/gimov/	/ljɵ̃so/	/seizm/	/vipɛʁ/
/ɔsəlɛ/	/gitaʁ/	/losjɵ̃/	/ʃekje/	/ylsɛʁ/
/ɔskaʁ/	/glwaʁ/	/lubaʁ/	/ʃetif/	/zenit/
/dəgʁe/	/goʁij/	/lydik/	/sɛgmã/	/ʒãvje/
/dopaʒ/	/gozje/	/məʁɛ̃g/	/sfɛ̃ks/	/ʒufly/
/dulœʁ/	/ibʁid/	/məsjø/	/svɛlt/	/ʒuvãs/

Phonetic transcription of the nonwords used in the experiment.

/aʁɛ̃li/	/fɛflã/	/nɔlkɛ/	/ʁɛ̃vyt/	/vedwɛ̃/
/ãswɛ̃p/	/fydɵ̃ʁ/	/nogap/	/sɵ̃ʁas/	/vɛ̃pil/
/aʒɛʁs/	/gɵ̃laʁ/	/nulob/	/ʃigwi/	/vufãk/
/bãpaw/	/gelpu/	/okəzɵ̃/	/smyʁd/	/vyfaf/
/bãpjɛ/	/gokpa/	/oʁøks/	/suwav/	/ymɔms/
/bɔmɛf/	/ikjɵ̃t/	/pabup/	/ʃwefɔ/	/ynɛkʁ/
/bɛ̃ʁãl/	/iʁdod/	/pãlad/	/tɵ̃pjã/	/ysɛʁt/
/bœwɛ̃k/	/juɛʁg/	/pɔbwe/	/tɛʃag/	/ytʁet/
/eduʁɛ̃/	/kevop/	/pɔvym/	/tiɛ̃pl/	/zilɛʁ/
/ɛdalf/	/kœwɛn/	/pɛ̃lav/	/tozɛ̃k/	/zomɛn/
/ɛljof/	/kyzlœ/	/psɛfk/	/tozuv/	/zopʁœ/
/ɛlkiʁ/	/lɛbwɛ̃/	/pswit/	/tyzwi/	/ʒekta/
/ɛpjɵ̃g/	/movaʒ/	/pwokl/	/uswaʒ/	/ʒjeps/
/ɛ̃swat/	/mɛ̃pli/	/ʁɵ̃mup/	/vɔpaʁ/	/ʒopja/
/fapɔp/	/mɛvig/	/ʁɛbãʃ/	/vedɵ̃t/	/ʒoʁuf/

Study 3

Neural patterns in linguistic cortices discriminate the content of verbal working memory

Benjamin Kowialiewski, Laurens Van Calster, Lucie Attout, Christophe Phillips, and Steve Majerus

Abstract. An influential theoretical account of Working Memory (WM) considers that WM is based on representational substrates in sensory cortices. While there is empirical support for this position in the visual WM domain, direct evidence is scarce in the verbal WM domain. In this study, we examined the extent to which the short-term maintenance of words versus nonwords can be differentiated by neural patterns in linguistic cortices by using a running span task minimizing strategic encoding mechanisms. Multivariate analyses showed specific neural patterns for word versus nonword WM conditions. These patterns were not detectable anymore when participants were instructed to stop maintaining the memoranda. The patterns involved the superior temporal sulcus and the pars opercularis of the dorsal language pathway underlying phonological processing. Specific neural patterns were also identified in the middle temporal gyrus and pars triangularis of the ventral pathway supporting semantic processing. This study provides evidence for a role of linguistic cortices in the representation of verbal WM content.

Introduction

Working Memory (WM), the ability to temporarily hold information in mind, is considered to rely on direct and obligatory activation of corresponding representations in long-term memory (LTM) by a number of theoretical accounts (Cowan, 1995, 1999, 2001, Majerus, 2013, 2019; N. Martin et al., 1996; Nee & Jonides, 2011, 2013; Oberauer, 2002). In the visual domain, this position is supported by studies showing that WM content can be decoded by neural patterns in visual sensory cortices (Harrison & Tong, 2009).

In the verbal domain, interactions between WM and corresponding representations in linguistic LTM are supported by the fact that verbal items associated with richer lexico-semantic representations, such as words, lead to strongly increased WM recall performance as compared to stimuli with minimal lexico-semantic content, such as nonwords (Brener, 1940; Jefferies et al., 2006a). At a neuroimaging level, direct evidence for an involvement of linguistic LTM representations in the short-term maintenance of verbal memoranda is still lacking. So far, previous studies have shown that frontal and temporal cortices supporting semantic knowledge also demonstrate sustained BOLD responses during the maintenance of verbal information (Fiebach et al., 2007). Neural patterns in linguistic cortices have also been shown to distinguish between word and nonword maintenance (Lewis-Peacock & Postle, 2012). However, these studies involved the recruitment of explicit phonological versus semantic task processing strategies and therefore they do not provide direct support for an obligatory recruitment of linguistic cortices in WM encoding and maintenance. Finally, neural patterns in the superior temporal gyrus associated with phonological processing have been shown to differentiate different types of nonwords during the encoding and recall phases of an immediate serial recall task (Kalm & Norris, 2014). Linguistic and WM processing stages are however confounded in this type of task and do not provide clear evidence for a role of linguistic cortices in WM maintenance.

This study tested the involvement of linguistic cortices in the representation and short-term maintenance of words and nonwords by using a fast encoding, running span procedure (Pollack et al., 1959) minimizing linguistic encoding strategies (Botto et al., 2014; Hockey, 1973; Morrison et al., 2016). We examined the extent to which multivariate neural patterns in linguistic cortices are able to differentiate the short-term maintenance of word versus nonword stimuli. We focused on temporal and frontal language regions, and more specifically on the dorsal and ventral language pathways whose role in the representation of phonological and lexico-semantic aspects of verbal information, respectively, has been firmly established (Friederici, 2012; Friederici & Gierhan, 2013; Hickok & Poeppel, 2007). Many studies have shown the involvement of the temporal section of the dorsal pathway in the processing and representation of phonological information, while its frontal section is critical for sensori-motor integration of phonological information (Arsenault & Buchsbaum, 2015; Mesgarani et al., 2014; Murakami, Kell, Restle, Ugawa, & Ziemann, 2015; Restle, Murakami, & Ziemann, 2012). The temporal section of the ventral pathway supports the content of semantic knowledge, while its frontal section is involved when semantic control over to-be-processed information is required (Lambon Ralph et al., 2017). In addition, we examined whether the linguistic nature of memoranda could also be

decoded in posterior intraparietal sulci (IPS). IPS involvement has been associated with attentional and task control processes in WM (Emrich et al., 2013; Majerus, Péters, Bouffier, Cowan, & Phillips, 2017; Todd, Fougnie, & Marois, 2005) and there is conflictual evidence concerning its role in the representation of WM content as opposed to WM task control (Bettencourt & Xu, 2015; Emrich et al., 2013; LaRocque et al., 2016; Yu & Shim, 2017).

Experiment

Method

   Participants. Data were obtained for 31 right-handed native French-speaking young adults (14 males; mean age = 21.42 years; age range 18-29) recruited from the university community, with no history of psychiatric or neurological disorders. The data from 2 participants had to be discarded due to excessive peaks in head movement (volume-to-volume displacement exceeding 4 mm and/or 4°). The data of one additional participant had to be excluded due to difficulties with task compliance, as indicated by response omissions for a significant amount (26.67%) of trials. Functional data acquisition were incomplete for one participant due to premature stopping of the scanner, but resulted in only 5% loss of the whole data set; we decided to retain this participant for data analysis. The final sample was composed of 28 valid data sets. The study was approved by the ethics committee of the Faculty of Medicine of the University of Liège and was performed in accordance with the ethical standards described in the Declaration of Helsinki (1964). All participants gave their written informed consent before their inclusion in the study.

   Task material. The stimuli consisted of 200 words and 200 nonwords. The words were selected from the Lexique 3 database and had an average lexical frequency of 18.99 counts per million (SD = 71.1) (New, 2006). The nonwords were created by generating under MATLAB® a very large ($N > 10^5$) number of stimuli that did not match any entry within the Lexique 3.0 database. The words and nonwords were matched for several critical phonological dimensions: number of phonemes (M = 4.59, SD = .65 and M = 4.6, SD = .67 for words and nonwords, respectively, BF_{01} = 8.93), biphone frequency (M = 726.37, SD = 418.51 and M = 691.96 and SD = 485.05 for words and nonwords, respectively, BF_{01} = 8.93) (Tubach & Boë, 1990), number of competitors from the same lexical cohort (e.g. *al*cove, *al*ligator, *al*cohol… Marslen-Wilson, 1987; Tyler, Voice, & Moss, 2000) (M_{log} = 1.92, SD_{log} = .58 and M_{log} = 1.97, SD_{log} = .55 for words and nonwords, respectively, BF_{01} = 6.02), uniqueness point (M = 4.17, SD = .91 and M = 4.09, SD = .81 for words and nonwords, respectively, BF_{01} = 5.65) and phonological

structure (the syllabic structure was matched at a pairwise basis for 98.5% of words and nonwords).

The stimuli were recorded by a French-native female speaker using a neutral voice. Each item was recorded as a separate stereo .wav sound file (44,100 Hz sampling frequency). Background noise was removed using Audacity® software using a Fourier transform analysis. Each stimulus was normalized to a duration of 475ms without altering the pitch, using the SBSMS algorithm implemented in Audacity®. Task presentation was controlled via the Cogent toolbox implemented under MATLAB®.

Task procedure. Participants underwent a one-hour MRI session during which they performed a passive listening task, followed by a running span task. The passive listening task was administered in order to identify the activation patterns that discriminate between word and nonword stimuli when no WM maintenance is required and when no explicit lexico-semantic judgments have to be made (Graves et al., 2016; Mattheiss et al., 2018). For each participant, 35 stimuli within each stimulus conditions were randomly sampled from the whole set of data and were used for the passive listening task, while the remaining stimuli were used for the running span task. This procedure of sampling ensured that any difference between the passive listening and the running span tasks could not be imputed to the specific set of stimuli used in either task, while also avoiding between-task learning and lexicalization processes. Both tasks were separated by a ~5minute break during which the structural scan was acquired (see MRI acquisition). Each participant received a different version of both tasks, with each version being constructed with the constraints mentioned below.

Passive listening task. In the passive listening task, each trial began within a fixation cross lasting for 1000ms, followed by a black screen and an auditory sequence composed of 4 items pacing at 500ms/item. The extra 25ms inter-stimulus interval insured that all stimuli were correctly perceived. Participants were not informed about the stimulus condition before the beginning of a sequence. Participants were instructed to carefully listen the sequences, without trying to memorize or rehearse them. Each sequence was followed by an inter-trial interval lasting for 9000ms (plus or minus a random duration sampled from a normal (Gaussian) continuous distribution with SD = 750ms). During this inter-trial interval, the participants were instructed to rest and do nothing, without trying to rehearse or remember the items from the auditory sequence. Each sequence was constructed such that adjacent items could not share their first or last two phonemes to avoid phonological overlap and potentially over-activation at the lexical level (Acheson & MacDonald, 2009; Gupta et al., 2005). Each stimulus was repeated between 3 and 4 times throughout the task. In addition, we

insured that each stimulus was not repeated twice in a given serial position or within the same sequence. Each sequence was randomly presented with the constraint that a given stimulus condition could not be repeated more than three consecutive trials. There were 30 trials within each stimulus condition (word, nonword), resulting in 60 experimental trials and 15 minutes total task duration.

Running span – Encoding phase. Each trial started with the presentation of the auditory sequence, pacing at 500ms/item, where participants were instructed to carefully memorize the items as much as they could. At the start of a trial, the participants were not informed about the stimulus (word, nonword) or the delay ('hold' or 'release' condition, see below) conditions. Auditory sequences were composed of 9, 10, 11, 12 or 13 items, ensuring that participants could not predict the length of a given sequence in advance, furthermore reducing the use of encoding strategies (Botto et al., 2014; Palladino & Jarrold, 2008). Each auditory sequence was constructed such that adjacent items could not share their two first or two last phonemes to avoid phonological overlap. Each stimulus was repeated 4 times throughout the task, and did not appear twice in the same serial position.

Running span – Delay phase. The auditory sequence was followed by an on-screen colored 40x40 pixels square lasting for 1000ms, directly followed by a white cross at the center of the screen lasting for 5500ms. If the square color was green, participants were instructed to maintain the items they had heard (hold condition), while if the square color was red, participants were instructed to just rest and do nothing (release condition).

Running span – Recognition phase. The delay phase was followed by a black screen and an auditory probe stimulus requiring participants, in the hold condition, to judge whether the probe was presented in the list or, in the release condition, to just press any key when they heard the auditory stimulus. Consequently, in the hold condition, memory for the auditory sequence was always assessed, while this was never done in the release condition. In the hold condition, participants were invited to use their index finger of the right hand for "yes" (the probe appeared in the auditory sequence) and their major finger for "no" (the probe did not appear in the auditory sequence). During this recognition phase, 60% (36 out of 60) of the active trials were composed of matching probes. This unequal distribution of matching and non-matching probes ensured that each serial position was sufficiently probed (between 3 and 4 times) to analyze serial position effects (see results). Non-matching probes were randomly sampled from the pool of stimuli while also ensuring that they never appeared in the current memory sequence. The stimuli were never used twice as a probe. Participants had a 3000ms upper limit to respond after probe onset. After the participant's response

or after the 3000ms time limit in case of no-responses, the next trial was initiated, separated by an inter-trial interval of 9000ms (plus or minus a random duration sampled from a normal (Gaussian) continuous distribution with SD = 750ms).

Sequences were constructed such that any given condition (word/nonword, hold/release), could not be repeated on more than three consecutive trials. There were 30 trials per experimental conditions (word – hold; word – release; nonword – hold; nonword – release), with a total of 120 experimental trials. Participants took approximatively 45 minutes to perform the running span task. To ensure that participants complied with task requirements, they performed a training version of both tasks (with stimuli not used in the experimental task) outside the scanner during a one-hour information session preceding the session in the scanner by at least 1 day and a maximum of 7 days. During the MRI session, the instructions were repeated before the beginning of the task. The passive listening and running span tasks were presented in separate EPI sequences, and the passive listening task was always presented first to avoid spontaneous use of maintenance processes that could have occurred if the running span task had been presented before the passive listening task.

MRI acquisition

The experiments were carried out on a 3-T whole-body scanner (Prisma, Siemens Medical Solutions, Erlangen, Germany) operated with a standard transmit–receive quadrature head coil. Multislice T2*-weighted functional images were acquired with a gradient-echo echo-planar (EPI) imaging sequence using axial slice orientation and covering the whole brain/most of the brain (30 slices, FoV = 192x192 mm², voxel size 3x3x3 mm³, TR = 1830ms, TE = 30ms). The five initial volumes were discarded to avoid T1 saturation effects. After each functional acquisition, a gradient-recalled sequence was applied to acquire two complex images with different echo times (TE = 10ms and 12.46ms respectively) and generate field maps for distortion correction of the functional images. For anatomical reference, a high-resolution T1-weighted image was acquired for each subject (T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence, TR = 1900ms, TE = 2.19ms, inversion time (TI) = 900ms, FoV = 256x240 mm², matrix size = 256x240x224, voxel size = 1x1x1 mm³). Around 1450 functional images per participants were acquired for the running span task. Head movement was minimized by restraining the participant's head using a vacuum cushion. Stimuli were displayed on a screen positioned at the rear of the scanner, which the participant could comfortably see through a mirror mounted on the head coil.

fMRI analysis

Image preprocessing. Data were preprocessed and analyzed using SPM12 software (version 12.0; Wellcome Department of Imaging Neuroscience, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB® for univariate analyses. The default parameters as defined in SPM12 were used. EPI time series were corrected for motion and distortion with "Realign and Unwarp" (J. L. R. Andersson, Hutton, Ashburner, Turner, & Friston, 2001) using the generated field map together with the FieldMap toolbox (Hutton et al., 2002) provided in SPM12. A mean realigned functional image was then calculated by averaging all the realigned and unwarped functional scans and the structural T1-image was coregistered to this mean functional image (rigid body transformation optimized to maximize the normalized mutual information between the two images). The mapping from subject to Montreal Neurological Institute space was estimated from the structural image with the "unified segmentation" approach (Ashburner & Friston, 2005). The warping parameters were then separately applied to the functional and structural images to produce normalized images of resolution $2 \times 2 \times 2$ mm^3 and $1 \times 1 \times 1$ mm^3, respectively. Finally the warped functional images were spatially smoothed with a Gaussian kernel of 4 mm FWHM to improve signal-to-noise ratio while preserving the underlying spatial distribution (Schrouff, Kussé, Wehenkel, Maquet, & Phillips, 2012); this smoothing also diminishes the impact that residual head motion can have on MVPA performance, even after head motion correction (Gardumi et al., 2016).

Univariate analysis. Univariate analyses first assessed brain activity levels associated with stimulus condition (word vs. nonword) in the passive listening task, and with stimulus condition (word vs. nonword) within each task condition (hold and release) in the running span task. For each participant, brain responses were estimated at each voxel, using a general linear model with event-related regressors. In the passive listening task, the design matrix included two regressors modelling the two stimulus conditions (word and nonword), and this for the entire duration of the list. In the running span task, the design matrix contained one regressor for the encoding and maintenance phase for each condition resulting from the crossing of stimulus and task conditions (word – hold; word – release; nonword – hold; nonword – release) and a single regressor for the recognition phase (all conditions were confounded for this regressor as this study focused on condition effects for the encoding and maintenance stages). For both models, passive listening and running span tasks, the time course of the events was convolved with the canonical hemodynamic response function (HRF) to account for the shape of the BOLD response. Each model also included the realignment parameters to account for any residual movement-related effect. A high-

pass filter was implemented using a cutoff period of 128 sec to remove the low-frequency drifts from the time series. Serial autocorrelations were estimated with a restricted maximum likelihood algorithm with an autoregressive model of order 1 (plus white noise).

In the passive listening task, linear contrasts were defined for the two stimulus conditions (word vs. nonword). In the running span task, linear contrasts were defined for each stimulus conditions (word vs. nonword), and this within each task condition (hold and release). The resulting contrast images, after additional smoothing by 6 mm FHWM, were entered in a second-level, random effect ANOVA analysis to assess the effect of stimulus conditions responsive brain areas at the group level. The additional smoothing was implemented to reduce noise due to inter-subject differences in anatomical variability and to reach a more conventional filter level for group-based univariate analyses ($\sqrt{(4^2 + 6^2)}$ = 7.21mm; Mikl et al., 2008). All the univariate analyses were performed using a cluster-level Family-Wise Error Rate (FWER) corrected threshold at p<.05, with a voxel-level cluster forming threshold of p<.001. For regions of interest (ROI) analyses, a small volume correction was applied to the contrasts of interests.

Multivariate analysis. Multivariate analyses of the 4mm smoothed functional time series were conducted using PRoNTo, a pattern recognition toolbox for neuroimaging (www.mlnl.cs.ucl.ac.uk/pronto; J. Schrouff et al., 2013). It was used to determine the voxel patterns discriminating between the different stimulus and/or task condition trials at an individual subject level. Binary support vector machines were used to classify whole-brain voxel activation patterns associated with word versus nonword stimuli in the passive listening task (Burges, 1998). This was also performed in the running span task for the word versus nonword stimulus conditions. We additionally conducted classifications over the time-course of the running span task, by assessing classifier accuracies second by second. A leave-one-block-out cross-validation procedure was used. At the subject level, classifier performance was assessed by running permutation tests on individual balanced classification accuracies ($N_{permutation}$ = 1000, p < .05). At the group level, classifier performance was tested by comparing the group-level distribution of classification accuracies to a chance-level distribution using a Bayesian One Sample T-Test. Bayesian statistics were used given their robustness in case of small-to-moderate sample sizes and non-normal distributions (T. M. Moore, Reise, Depaoli, & Haviland, 2015) and because, with these analyses, the bias toward accepting or rejecting the null hypothesis does not change with sample size. Furthermore, Bayesian statistics assess evidence for a model under investigation in the light of the data, whereas group-level classical T-Tests make

population-level inferences, which have been shown to be problematic when comparing classification accuracies against chance-level (Allefeld, Görgen, & Haynes, 2016). The BF_{10} is used to determine the likelihood ratio of the alternative model (H_1) relative to the null model (H_0), and the BF_{01} to determine the likelihood ratio of H_0 relative to H_1. We use the classification proposed by previous studies (Jeffreys, 1998; Wagenmakers et al., 2011): A BF of 1 provides no evidence, 3 > BF > 1 provides anecdotal evidence, 10 > BF > 3 provides moderate evidence, 30 > BF > 10 provides strong evidence, 100 > BF > 30 provides very strong evidence and BF > 100 provides extreme/decisive evidence. All the analyses were performed using the BayesFactor package (Morey & Rouder, 2014) implemented in R (R Development Core Team, 2008) and using JASP (JASP Team, 2017) with all parameters being set to the default Cauchy prior distribution as implemented in JASP (Version 0.9.0.1). We also report $BF_{inclusion}$ values which compare the evidence of all models including a given factor relative to all other models not including this factor. A standard mask removing voxels outside the brain was applied to all images, and all models included timing parameters for HRF delay (5 sec) and HRF overlap (5 sec), ensuring that stimuli from different categories falling within the same 5 sec were excluded (Schrouff et al., 2013). The whole-brain multivariate analyses were followed up by ROI analyses to determine the role of specific brain regions in the discrimination pattern of the different stimulus and task conditions.

Regions of Interest. We selected ROIs associated with the dorsal and ventral language pathways (Friederici, 2012; Friederici & Gierhan, 2013; Hickok & Poeppel, 2007; Majerus, 2013) using the IBASPM 71 and IBASPM 116 atlases (http://www.thomaskoenig.ch/Lester/ibaspm.htm).

The dorsal language pathway was further subdivided according to its temporal and frontal sections given their distinct roles in perceptual versus sensori-motor aspects of phonological processing (Arsenault & Buchsbaum, 2015; Mesgarani et al., 2014; Murakami et al., 2015; Restle et al., 2012). The temporal region covered the left superior temporal gyrus, encompassing the anterior temporal sulcus up to the planum temporale region. The frontal region covered the pars opercularis, which is located in the posterior part of the inferior frontal gyrus (BA44).

The ventral language pathway was also subdivided as a function of its temporal and frontal components, given their distinct roles in semantic representation and semantic control, respectively (Fiebach et al., 2007; Gagnepain et al., 2008; Gold et al., 2006; Lambon Ralph et al., 2017; Rissman et al., 2003; Sabri et al., 2008; Snijders et al., 2009; Visser, Jefferies, Embleton, & Lambon Ralph, 2012; Whitney, Jefferies, & Kircher, 2011). The temporal component covered the middle temporal gyrus, encompassing the
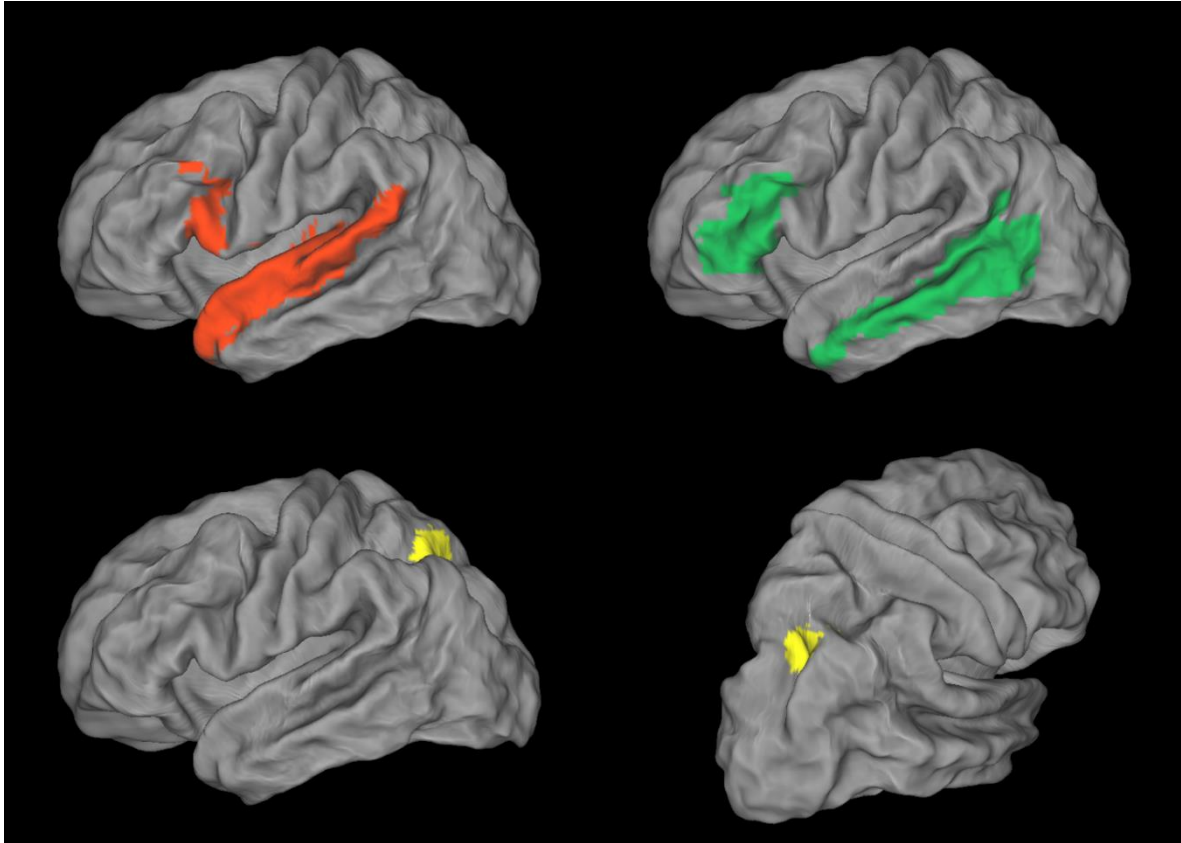
Figure 3.1. Regions of interest (ROI) used in the univariate and multivariate analysis. The upper part of the image shows the ROI within the dorsal (left, orange) and ventral language (right, green) pathways, with their respective frontal and temporal components. The lower part of the image shows the left and right IPS ROI.

anterior temporal lobe up to the posterior middle temporal gyrus, including the middle temporal-occipital junction. The frontal component covered the entire pars triangularis, located in the anterior part of the inferior frontal gyrus (BA45).

ROIs in the intraparietal cortex were defined as 10mm radius spheres, on mean coordinate values for the left posterior IPS (x = -25, y = -64, z = 43), and the right posterior IPS (x = 27, y = -62, z = 38), taken from previous studies that have focused on interactions between attentional and WM processing (Asplund, Todd, Snyder, & Marois, 2010; Majerus et al., 2016; Majerus, Attout, et al., 2012; Todd et al., 2005; Todd & Marois, 2004). An overview of these ROIs is given in **Figure 3.1**.

Results

Behavioural analysis – running span task. A first analysis assessed response accuracy as a function of lexical condition (word vs. nonword) and as a function of the serial positon being probed (1 through 9) using a Bayesian Repeated Measures ANOVA. We found moderate evidence against the presence of a lexical condition effect ($BF_{01} = 6.76$), and decisive evidence supporting the serial position effect ($BF_{inclusion}$

= 1.012e+10). We also found decisive evidence against the interaction term (BF_{01} = 100). As shown in **Figure 3.2**, upper panel, there were strong recency and poor primacy effects. The same pattern of results was observed when running the same analysis on response times. We found strong evidence against the effect of lexical condition (BF_{01} = 13.33), but decisive evidence supporting a serial position effect ($BF_{inclusion}$ = 1164.871). We found decisive evidence against the interaction term (BF_{01} = 125). The serial position effect was characterized by faster responses for the last presented items as compared to items presented in the beginning of the list (**Figure 3.2**, lower panel). In line with our predictions, the strong recency and poor primacy effects for both accuracy and response times show that participants passively encoded each subsequent item without being able to use refreshing or rehearsal strategies for earlier presented items during the running span task (Botto et al., 2014; Ruiz & Elosúa, 2013; Ruiz et al., 2005). A lexicality effect on recognition performance was not strongly expected as these effects have been shown to occur most reliably at the behavioural level when using full recall rather than recognition paradigms, or when using recognition paradigms with a particularly large number of trials and no additional maintenance delay (Gathercole et al., 2001; Jefferies et al., 2006b; Kowialiewski & Majerus, 2018a).
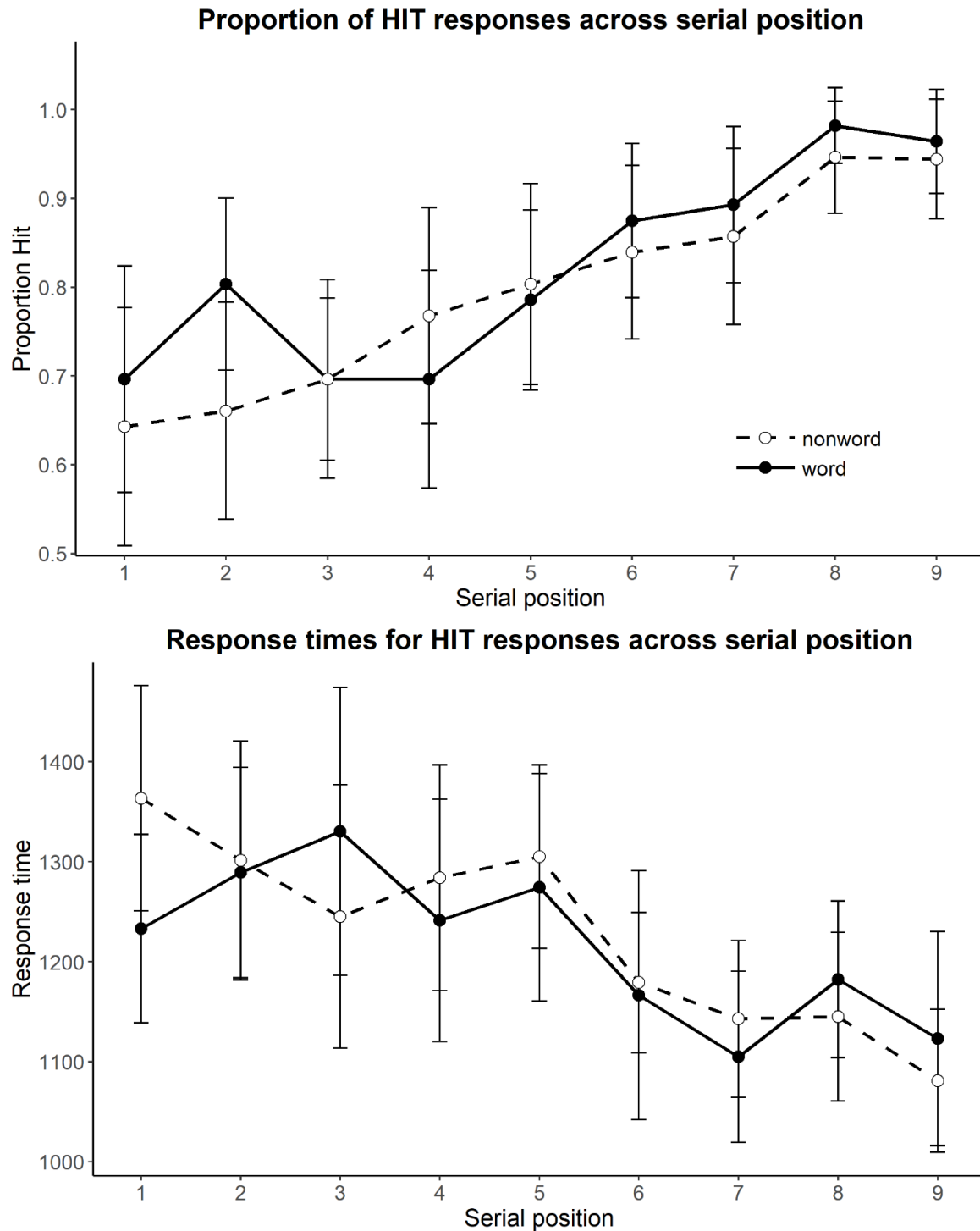
Proportion of HIT responses across serial position

Response times for HIT responses across serial position

Figure 3.2. Proportion of hit responses (upper panel) and response times for hit responses (lower panel) across serial position, averaged across participants. Error bars represent 95% confidence intervals, after controlling for between-subject variability (Baguley, 2012; Cousineau, 2005; Morey, 2008).

Neuroimaging – Univariate analyses. For the passive listening task, no linguistic condition effect was observed when directly contrasting the two (word vs. nonword) stimulus conditions (cluster-level FWE corrected threshold p<.05, with p<.001
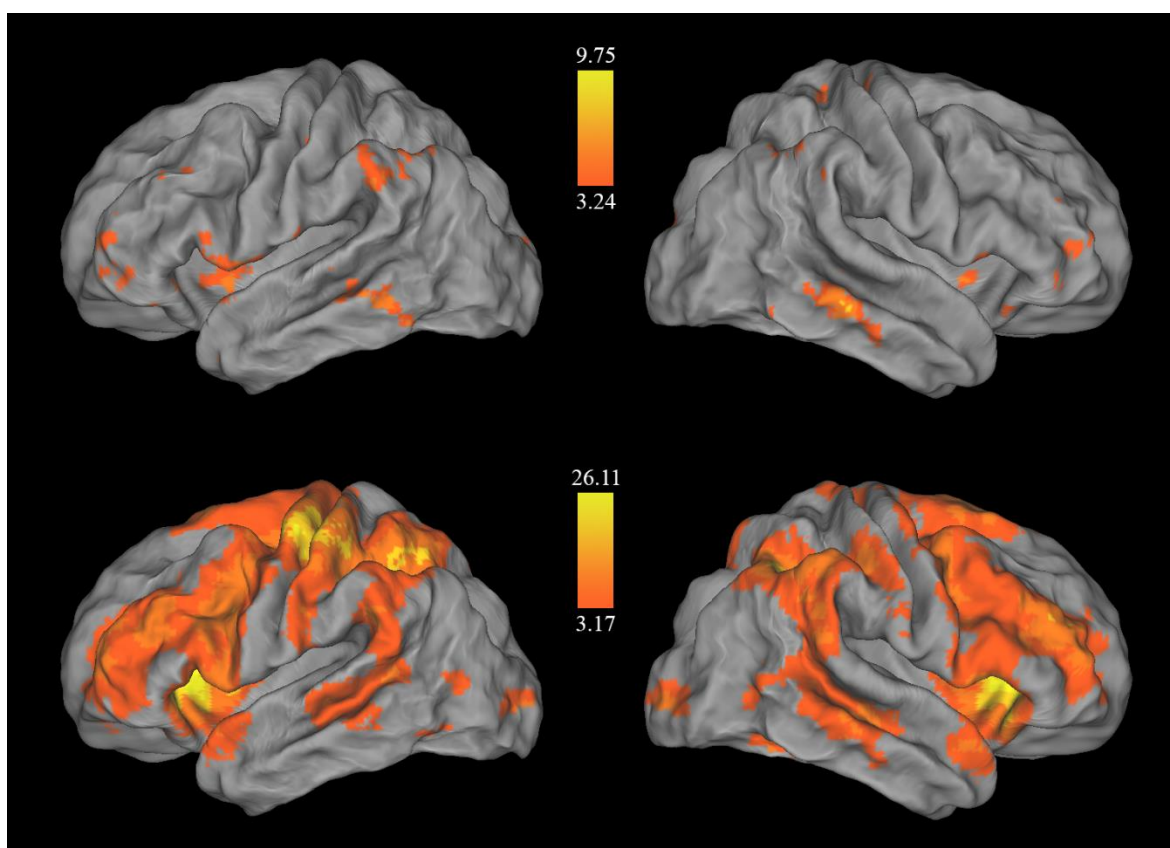
Figure 3.3. Univariate results when word and nonword stimuli are contrasted against baseline in the passive listening task (top panel) and in the running span task (bottom panel). The regions are displayed at an uncorrected voxel-level threshold of p<.001. The scale colour indicates minimum and maximum t values, with (1, 56) and (1, 108) degrees of freedom in the passive listening and in the running span task, respectively.

uncorrected voxel-level cluster-forming threshold). The same result was observed when assessing univariate responses associated with the running span task, regardless of the task condition (hold, release). When using the ROIs as defined in the Methods section, no linguistic condition effect was observed for both tasks. This finding is in line with the majority of studies showing overlapping univariate neural responses for word and nonword conditions (Davis & Gaskell, 2009; Kotz et al., 2002; Newman & Twieg, 2001; Orfanidou et al., 2006; Raettig & Kotz, 2008; Rissman et al., 2003; Sabri et al., 2008; Xiao et al., 2005). As shown in **Figure 3.3**, word and nonword conditions activated temporal and frontal cortices of the ventral and dorsal language pathways when looking at the effect of both conditions, with these activity foci being much more extended in the running span task.

Neuroimaging – Multivariate analyses: Passive listening task. As can been seen in **Figure 3.4**, permutation tests and Bayesian One Sample T-Tests showed that the accuracy of the linguistic condition classifier did not differ from chance-level performance. This was the case for all ROIs as well as for a whole-brain level analysis.

Neuroimaging – Multivariate analyses: Running span task. By contrast, when running the same analyses for the running span task, robust discrimination of word versus nonword conditions was observed. As shown in **Figure 3.5**, above chance-level discrimination was observed in the temporal and frontal components of both the dorsal and ventral language pathways, and this particularly in the hold condition where decisive evidence for above-chance discrimination was observed for all language regions. Interestingly, the IPS regions presumably involved in attentional and task control also showed decisive evidence for above-chance discrimination in the hold condition, and strong evidence was still observed in the release condition.
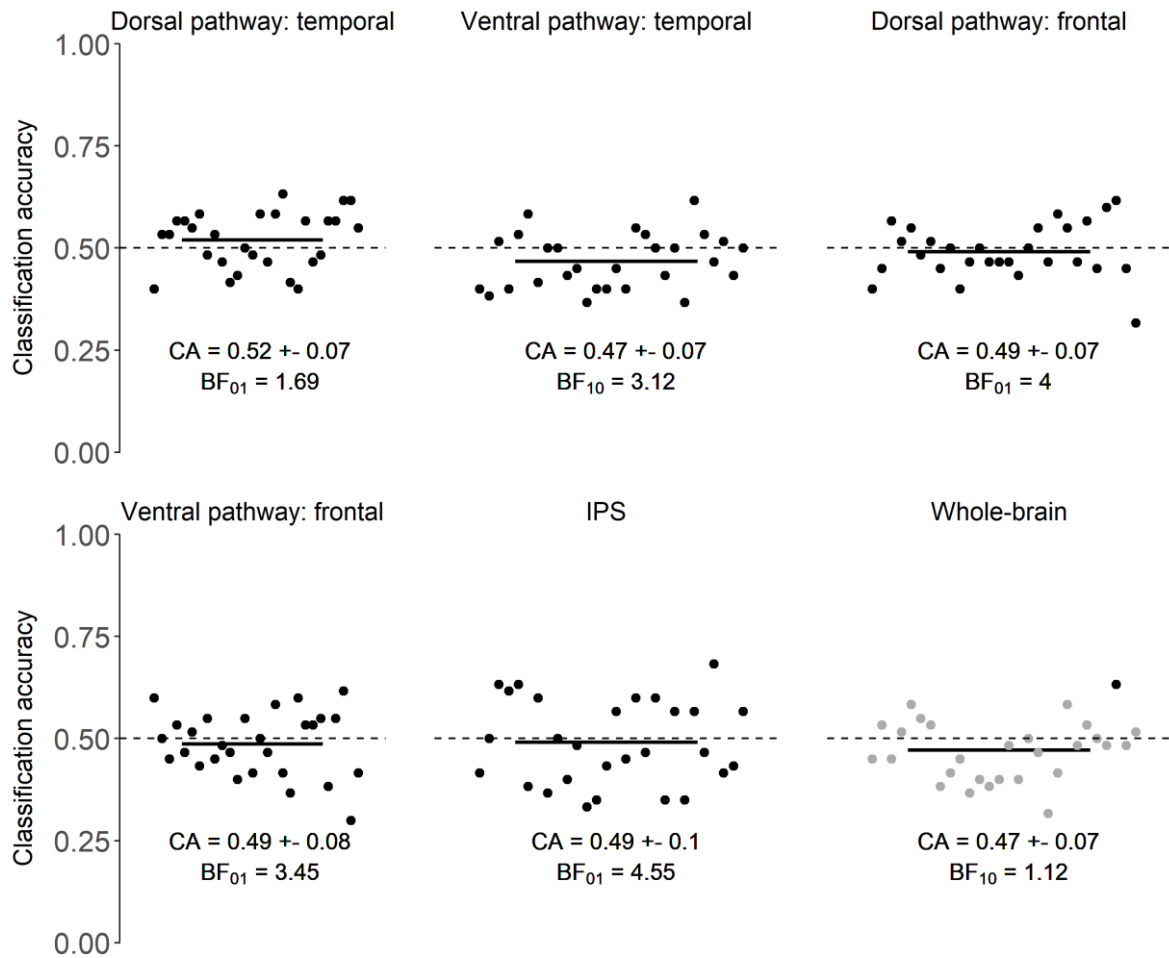
Figure 3.4. Classification accuracies in the passive listening task, for ROI and whole-brain analyses. Each point represents classification accuracy for one participant. For the whole-brain analysis, individual classification accuracies significant at p<.05 (permutation tests) are marked in black. Chance level classification accuracy is indicated by the horizontal black line.
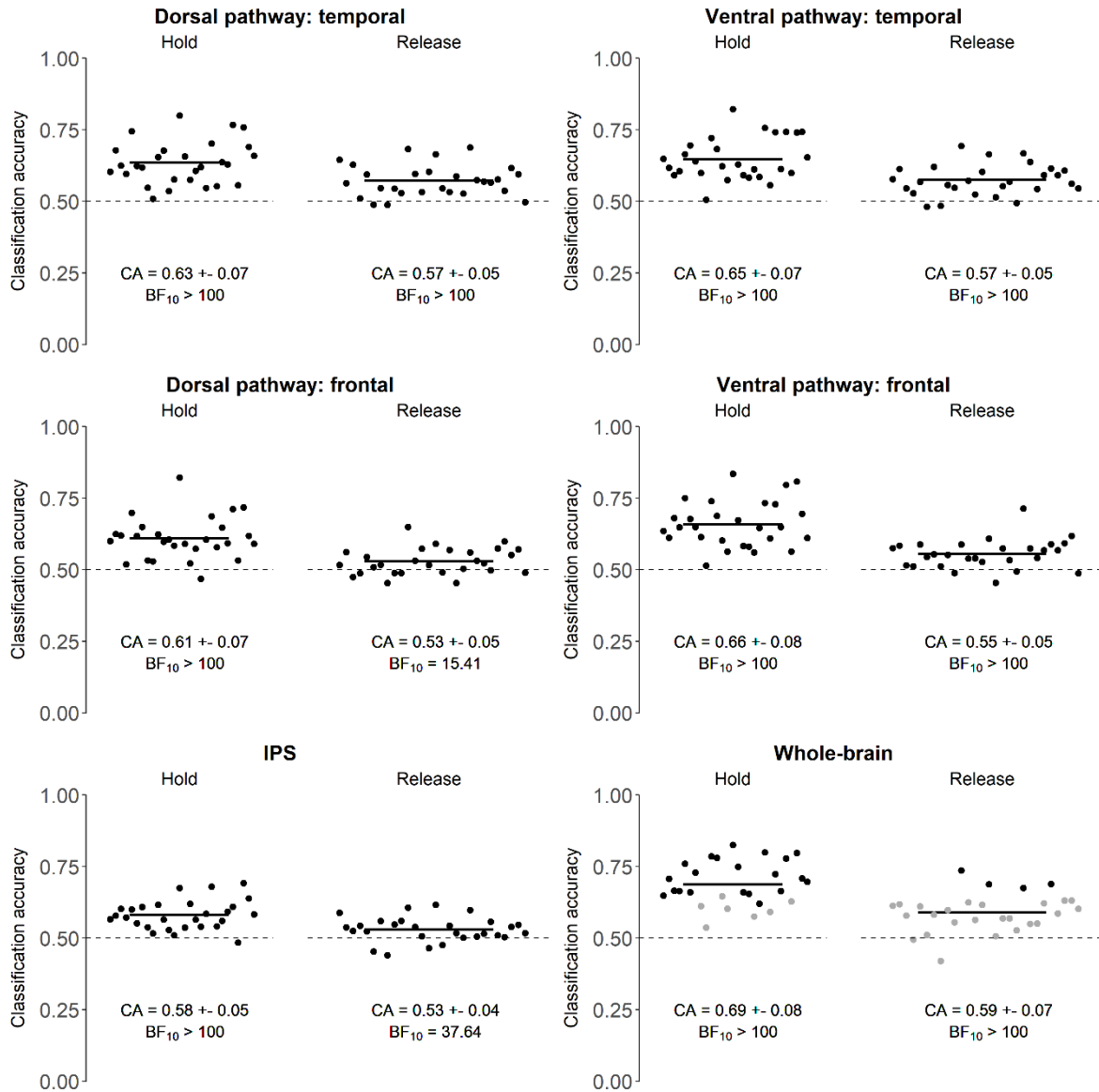
Figure 3.5. Classification accuracies in the running span task, for ROI and whole-brain analyses. Each point represents classification accuracy for one participant. For the whole-brain analysis, individual classification accuracies significant at p<.05 (permutation tests) are marked in black. Chance level classification accuracy is indicated by the horizontal black line

The same conclusions were drawn when assessing classification accuracy in the whole-brain analysis; decisive evidence for above-chance discrimination was found in the hold and release conditions. In addition, using permutation tests at an individual level, significant multivariate discrimination between word and nonword stimuli was observed for 75% of the subjects in the hold condition, when participants were required to maintain the items during the delay. In the release condition, when participants were instructed to stop holding the items, significant discrimination was observed for only 14.9% of the participants.
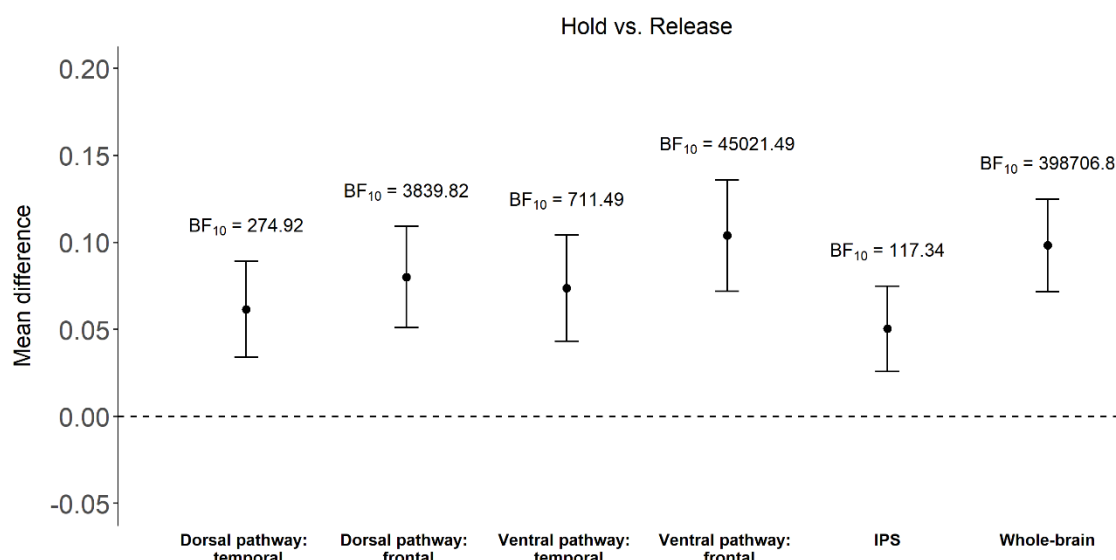
Figure 3.6. Difference of classification accuracies between the hold and release task conditions, for ROI and whole-brain approaches. Error bars represent 95% confidence intervals.

In a next analysis, we directly compared word versus nonword classification accuracy between the hold and release task conditions. As can be seen in **Figure 3.6**, classification accuracy dropped in the release condition as compared to the hold condition, and this was consistently observed across all ROIs and in the whole-brain analysis, as assessed by Bayesian One Sample T-Tests. This drop of classification accuracy in the release condition suggests that participants had stopped maintaining the memoranda, leading to a disappearance of stimulus condition specific neural activity patterns in linguistic cortices.

In order to obtain a more precise understanding of the moment at which linguistic condition informative neural patterns disappeared, we explored the time-course of classification accuracy over the entire duration of the running span task trials (see **Figure 3.7**). Classification accuracy was assessed for time points ranging from 15 seconds before and 2 seconds after the presentation of the probe, to ensure a window of 17 seconds. This procedure ensured that the same time points were examined for each trial despite the variable duration of the encoding event. Note that an HRF delay of 5 seconds was used in all multivariate analyses, thereby shifting classification events by 5 seconds relative to trial time.

Dorsal pathway. For the temporal component of the dorsal pathway, above-chance level discrimination in the hold condition was found over time points 3 through 9, and over time points 11 through 13, corresponding to the encoding and delay phases. In the release condition, this was observed only for time points corresponding to the encoding phases, and more specifically over time-points 2, 4 and 5. Direct comparisons

between task conditions showed significant differences in classification accuracy at time points 8, 9 and 13, corresponding to the delay phase. Over the frontal component of the dorsal pathway, above-chance level discrimination was observed in the hold condition over time points 5 through 13, all located in the encoding and delay phases. In the release condition, this was the case only for time points 4 and 5 (encoding phase). Direct comparisons showed that the two task conditions differed at time points 6 to 13, corresponding mostly to the delay phase. For simplicity, these results are reported in the **Appendix**.

Ventral pathway. In the temporal component of the ventral pathway, in the hold condition, above-chance level discrimination was found over time points 2 through 11, and 13 (encoding and delay phases). In the release condition, this was observed for time points 2 through 6 (encoding phase only). The two delay conditions differed over time points 8 through 11, and 13 through 14 (mainly in the delay phase). The same results were observed when focusing on the frontal component of the ventral pathway: above-chance level discrimination was found over time points 2 through 14 in the hold condition (encoding and delay phases), and 2 through 5 in the release conditions (encoding phase only). The difference between the two delay conditions was observed over time-points 6 through 14, corresponding again mainly to the delay phase.

IPS. In the hold condition, above-chance level discrimination was found at time point 2, as well as over time points 4 through 11 (encoding and delay phases). This was observed over time-points 2 and 4 only in the release condition (encoding phase only). The two task conditions differed over time-points 7 through 10, which corresponds to late encoding/early delay phases.
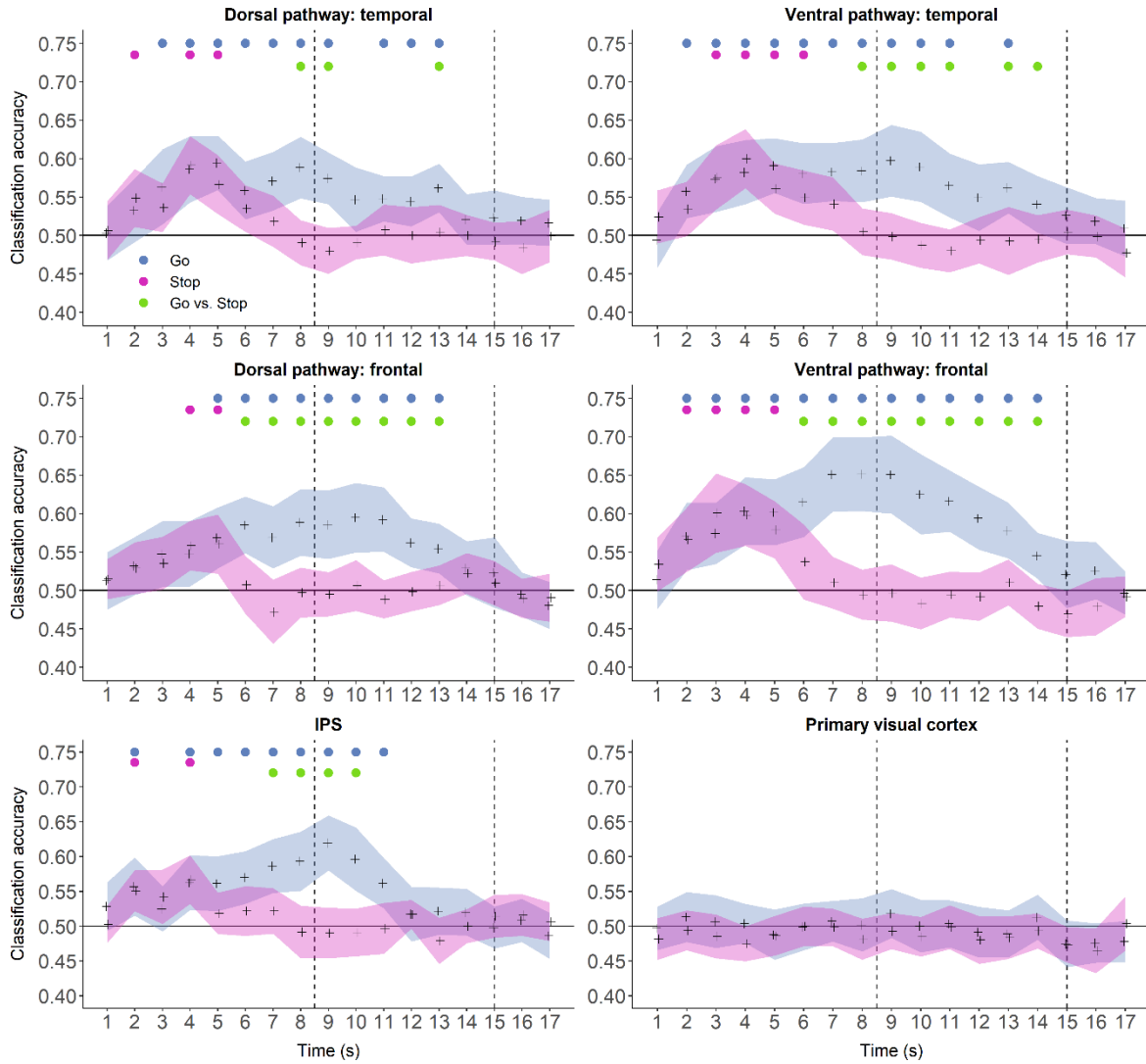
Figure 3.7. Classification accuracies (word vs. nonword) as a function of trial time in the running span task (in seconds), for ROI and whole-brain approaches. The dashed vertical lines represent the beginning and the end of the maintenance phase. The recognition probe directly appeared after the end of the maintenance phase. BF_{10} values for word-nonword classification accuracies above 3 are indicated via the coloured lines on top of the figure. Blue: $BF_{10} > 3$ for hold condition (word-nonword classification accuracy relative to theoretical random classification distribution). Purple: $BF_{10} > 3$ for release condition (word-nonword classification accuracy relative to theoretical random classification distribution). Green: $BF_{10} > 3$ for direct comparison of word-nonword classification accuracies in the hold versus release task conditions. The ribbon represents 95% confidence intervals of the mean.

In order to show that the observed discrimination patterns for word and nonword conditions are specific to language and WM-related cortices, we performed a final ROI analysis over the primary visual cortex V1. Given the purely auditory-verbal nature of the stimuli used in this experiment, no discrimination of word versus nonword conditions was expected in this region associated with processing of visual

sensory information (Kamitani & Tong, 2005; Tootell et al., 1998). The region was defined using the probabilistic atlas Anatomical toolbox (Eickhoff et al., 2005) implemented in SPM. The region covered the whole primary visual cortex region (Amunts, Malikovic, Mohlberg, Schormann, & Zilles, 2000), and has shown to successfully decode different visual features such as colour, line orientation or movement direction in tasks requiring WM maintenance (Emrich et al., 2013; Ester, Anderson, Serences, & Awh, 2013; Harrison & Tong, 2009; Weber, Peters, Hahn, Bledowski, & Fiebach, 2016). As can be seen in **Figure 3.7**, classification accuracy was at chance-level in both task conditions throughout the entire trial duration.

In sum, we observed decisive evidence for a differentiation of neural patterns involved in encoding and short-term maintenance of words versus nonwords. These patterns could be found both in dorsal and ventral pathways of the language network, but also in IPS areas presumably associated with WM attentional control processes.

Discussion

This study provides evidence for a non-strategic, bottom-up involvement of linguistic cortices in the encoding and maintenance of verbal information. Some studies have hinted to a role of linguistic cortices during WM tasks by showing sustained activity in these cortices or by identifying neural patterns that discriminate between different types of verbal memoranda in explicit linguistic judgment tasks or between speech and non-speech stimuli (Buchsbaum et al., 2005; Kalm et al., 2012; Lewis-Peacock & Postle, 2012; Ravizza et al., 2011; Strand et al., 2008; Yue et al., 2018). The present study shows that linguistic cortices represent the type of WM content as soon as verbal information enters WM, in the absence of any explicit linguistic encoding strategies, and this representation continues over the maintenance stage.

These results provide novel and critical support for theoretical models considering that WM involves the temporary activation of LTM knowledge (Cowan, 1995, 1999, 2001, Majerus, 2013, 2019; N. Martin et al., 1996; Nee & Jonides, 2011, 2013; Oberauer, 2002) and more precisely, the temporary activation of representations in the linguistic system (Acheson & MacDonald, 2009; Gupta, 2003; M. C., 1999; Majerus, 2013, 2018; Martin et al., 1996). The present study shows that WM content is represented in dorsal and ventral streams of the linguistic system as soon as information is encoded and continues to be represented there all over maintenance. The present neuroimaging results are also consistent with neuropsychological data showing that patients can show selective impairment for word or nonword stimuli in WM tasks, depending on whether their lesions involve parts of the ventral or the dorsal language pathway, respectively (Hoffman, Jefferies, Ehsan, Jones, & Lambon

Ralph, 2009, 2012; Leff et al., 2009; Majerus et al., 2007; N. Martin & Saffran, 1997; Patterson et al., 1994). Our data also provide support to repetitive transcranial magnetic stimulation studies that suggested a reduction of WM performance for word or nonword stimuli, depending on stimulation of areas within the ventral or the dorsal language stream (Acheson, Hamidi, Binder, & Postle, 2011; N. J. Savill, Cornelissen, Pahor, & Jefferies, 2018).

At the same time, it is important to note that we were not able to discriminate the neural patterns associated with the word versus nonword stimuli in a simple passive listening task. This finding could be explained by several factors. One possibility is that participants may not have fully attended the stimuli in the linguistic task. Given the repetitive aspect of the linguistic task, participants may have wandered in their minds, leading to continuously changing neural patterns in language cortices independently of the appearance of the word versus nonword stimuli of the passive listening task (Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011). In addition, it has been shown that under dual-task paradigms, consistent lexicality effect are observed only under fully attended speech conditions (Sabri et al., 2008). More generally, passive perceptual tasks have been shown to lead to strongly impoverished neural activity in regions involved in sensory perception as compared to tasks requiring items to be fully attended (L. Andersson, Sandberg, Olofsson, & Nordin, 2018; Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005). Studies reporting successful decoding of different types of verbal stimuli have used tasks requiring at the least the active monitoring of the different stimuli (Correia et al., 2014; Murphy et al., 2017; Stevens, Kravitz, Peng, Tessler, & Martin, 2017; Yue et al., 2018).

A final important finding of the present study is that maintenance of word versus nonword stimuli could also be decoded based on neural patterns in IPS regions that have been associated in the past with attentional processes in WM tasks. In the visual WM domain, there is a controversy with respect to the role of the posterior intraparietal cortex in the representation of WM content (Albers et al., 2013; Emrich et al., 2013; LaRocque et al., 2016; Linden et al., 2012). While a number of studies suggest that the posterior intraparietal cortex is involved in attentional control processes, and more precisely the representation of task set (Cowan et al., 2011; Majerus et al., 2010; Todd & Marois, 2004), other studies have shown that visual features of memoranda such as line orientations (Bettencourt & Xu, 2015; Ester et al., 2015), objects features (S. Lee et al., 2013), abstract visual patterns (Christophel et al., 2012), spatial locations (Peters et al., 2015), but also colours (Yu & Shim, 2017), can be decoded from patterns in posterior IPS. This raises the important question of the role of the intraparietal cortex in WM : is it merely involved in attentional and task control (Emrich et al., 2013;

Majerus, 2019), or does it also have a WM buffering function in which traces of memoranda are stored? On the one hand, the fact that patterns in IPS regions were able to decode word versus nonword memoranda suggests that parietal regions also encode at least some characteristics of memoranda in verbal WM. This possibility is suggested by further results from the study by Yue et al. (2018): they showed that their classifier trained on a perceptual task successfully decoded between speech and nonspeech stimuli in the supramarginal gyrus but not in the superior temporal gyrus during WM maintenance, leading them to argue that this neural region acts as a WM buffer. On the other hand, the fact that identical IPS regions and neural patterns are consistently involved in both verbal and visual WM tasks (Majerus, 2019; Majerus et al., 2010, 2016) reduces the likelihood of a modality specific buffering function of the parietal cortex (this could however be different for the more lateral and inferior region explored by Yue et al.). Furthermore, encoding of word versus nonword lists is likely to be associated with different levels of perceptual load, encoding of word stimuli being faster and less effortful at the perceptual level as compared to nonwords due to their familiarity. A recent study has shown that the posterior IPS is sensitive to perceptual load during WM encoding (Majerus et al., 2017).

Conclusions

This study shows a non-strategic involvement of linguistic cortices in lexico-semantic processing during verbal WM processing, providing critical support for theoretical statements assuming that verbal WM relies on direct activation of the linguistic LTM system and that this activation supports the representation of WM content.

Appendix

Table 3.1. *Time-point specific Bayesian Factor (BF$_{10}$) values of classification accuracies for word versus nonword conditions in the temporal portion of the dorsal language pathway ROI, separately for the hold and release conditions (Bayesian One Sample T-Test), or resulting from a direct comparison between the two conditions (Bayesian Paired Samples T-Test).*

Time-point	Bayesian One Sample T-Tests		Bayesian Paired Samples T-Tests
	Hold	Release	Hold vs. Release
1	.204	.212	.203
2	.636	3.653	.239
3	3.615	2.035	.408
4	86.813	698.991	.207
5	2507.772	23.628	.384
6	11.597	2.148	.347
7	43.347	.362	2.173
8	239.366	.245	107.258
9	242.666	.469	128.207
10	1.867	.283	2.328
11	15.301	.219	1.400
12	4.765	.200	1.524
13	73.023	.206	3.971
14	.437	.200	.388
15	.454	.245	.519
16	.421	.298	.752
17	.354	.201	.275

Table 3.2. *Time-point specific Bayesian Factor (BF₁₀) values of classification accuracies for word versus nonword conditions in the frontal portion of the dorsal language pathway ROI, separately for the hold and release conditions (Bayesian One Sample T-Test), or resulting from a direct comparison between the two conditions (Bayesian Paired Samples T-Test).*

Time-point	Bayesian One Sample T-Tests		Bayesian Paired Samples T-Tests
	Hold	Release	Hold vs. Release
1	.250	.366	.202
2	.749	.788	.203
3	1.769	1.357	.245
4	1.879	35.124	.222
5	23.925	11.792	.218
6	396.394	.216	14.022
7	21.750	.464	25.449
8	11.315	.203	24.520
9	58.273	.213	22.579
10	139.775	.216	27.440
11	226.951	.302	65.689
12	72.958	.202	22.892
13	19.445	.229	5.490
14	.699	.737	.210
15	.329	.248	.230
16	.215	.275	.207
17	.433	.239	.222

Table 3.3. *Time-point specific Bayesian Factor (BF$_{10}$) values of classification accuracies for word versus nonword conditions in the temporal portion of the ventral language pathway ROI, separately for the hold and release conditions (Bayesian One Sample T-Test), or resulting from a direct comparison between the two conditions (Bayesian Paired Samples T-Test).*

Time-point	Bayesian One Sample T-Tests		Bayesian Paired Samples T-Tests
	Hold	Release	Hold vs. Release
1	.211	.509	.545
2	16.829	1.108	.407
3	23.338	34.547	.201
4	75.521	1721.624	.288
5	1384.396	45.246	.490
6	107.277	5.562	.612
7	227.373	2.351	.772
8	118.983	.210	52.345
9	133.046	.201	32.997
10	7.220	.291	341.100
11	11.091	.520	104.656
12	2.020	.219	2.225
13	44.977	.211	51.904
14	1.726	.209	3.405
15	.505	.210	.361
16	.417	.201	.315
17	.227	.525	.473

Table 3.4. *Time-point specific Bayesian Factor (BF₁₀) values of classification accuracies for word versus nonword conditions in the frontal portion of the ventral language pathway ROI, separately for the hold and release conditions (Bayesian One Sample T-Test), or resulting from a direct comparison between the two conditions (Bayesian Paired Samples T-Test).*

Time-point	Bayesian One Sample T-Tests		Bayesian Paired Samples T-Tests
	Hold	Release	Hold vs. Release
1	.261	1.133	.313
2	13.653	12.405	.205
3	44.702	69.602	.324
4	502.707	761.167	.204
5	536.792	161.234	.328
6	1232.405	.592	6.447
7	23929.963	.238	1347.057
8	28161.262	.212	556.857
9	10917.482	.204	118.525
10	632.203	.323	325.001
11	8364.911	.214	12326.180
12	384.704	.228	85.223
13	196.445	.252	16.792
14	9.056	.506	4.929
15	.309	1.237	1.513
16	.493	.368	.704
17	.207	.240	.205

Table 3.5. *Time-point specific Bayesian Factor (BF₁₀) values of classification accuracies for word versus nonword conditions in the intraparietal sulcus ROI, separately for the hold and release conditions (Bayesian One Sample T-Test), or resulting from a direct comparison between the two conditions (Bayesian Paired Samples T-Test).*

| Time-point | Bayesian One Sample T-Tests | | Bayesian Paired Samples T-Tests |
	Hold	Release	Hold vs. Release
1	.669	.204	.441
2	4.834	24.099	.206
3	.593	1.587	.263
4	12.749	59.321	.203
5	11.271	.411	.973
6	4.566	.408	1.355
7	277.189	.459	8.831
8	239.323	.221	38.544
9	15719.783	.231	285.936
10	157.872	.230	3.207
11	21.443	.204	2.282
12	.292	.690	.201
13	.430	.437	.768
14	.395	.200	.323
15	.202	.303	.284
16	.230	.349	.214
17	.281	.222	.317

Study 4

Verbal working memory and linguistic long-term memory: Exploring the lexical cohort effect

Benjamin Kowialiewski and Steve Majerus

Abstract. Numerous studies have shown that verbal working memory (vWM) performance is strongly influenced by linguistic knowledge, with items more familiar at sublexical, lexical and/or semantic levels leading to higher vWM recall performance. Among the many different psycholinguistic variables whose impact on vWM has been studied, the lexical cohort effect is one of the few effects that has not yet been explored. The lexical cohort effect reflects the fact that words sharing their first phonemes with many other words (e.g. alcove, alligator, alcohol…) are typically responded to more slowly as compared to words sharing their first phonemes with a smaller number of words. In a pilot experiment (Experiment 1), we manipulated the lexical cohort effect in an immediate serial recall task and found no effect. Experiment 2 showed that, in a lexical decision task, participants responded more quickly to items stemming from small cohorts, showing that the material used in Experiment 1 allowed for a valid manipulation of the cohort effect. Experiment 3, using stimuli from Experiment 2 associated with maximal cohort effects during lexical decision, failed again to reveal a cohort effect in an immediate serial recall task. We argue that linguistic knowledge impacts vWM performance via continuous interactive activation within the linguistic system, which is not the case for the lexical cohort variable which may influence language processing only at the initial stages of stimulus activation.

Introduction

Language-based models of verbal working memory (vWM) assume that temporary storage of verbal information relies on direct activation of corresponding representations within the linguistic system (Acheson & MacDonald, 2009; Gupta,

2009; Majerus, 2013; N. Martin, Saffran, & Dell, 1996; R. C. Martin, Lesch, & Bartha, 1999). This is supported by the fact that many psycholinguistic variables affecting language processing also affect vWM (Brener, 1940; Guérard & Saint-Aubin, 2012; Hulme et al., 1997; Kowialiewski & Majerus, 2018b; Majerus, Van der Linden, Mulder, et al., 2004; Poirier & Saint-Aubin, 1995, 1996; Romani et al., 2008; Watkins & Watkins, 1977). A specific psycholinguistic variable has, however, never been investigated in vWM: the lexical cohort competition effect. This effect is characterized by the fact that words drawn from large lexical cohorts (e.g. *al*cove, *al*ligator, *al*cohol…) are usually responded to slower than words drawn from small cohorts in lexical decision tasks, as a result of many lexical competitors getting co-activated for words from large lexical cohorts (Marslen-Wilson, 1987). The purpose of the present study was to investigate whether the lexical cohort variable can also impact vWM, as predicted by language-based models of vWM.

VWM closely interacts with phonological, lexical and semantic linguistic variables. At the sublexical/phonological level, this is illustrated by studies showing that nonwords containing structures of high phonotactic probability (i.e., high biphone frequencies) are associated with higher vWM performance than nonwords containing structures of low phonotactic probabilities (Gathercole et al., 1999; Majerus, Van der Linden, Mulder, et al., 2004). Likewise, the lexical levels of representation have also been shown to impact vWM performance, with higher recall performance for words than nonwords (Brener, 1940; Hulme et al., 1991; Jefferies et al., 2006a), and high frequency words also leading to higher recall performance as compared to low frequency words (Hulme et al., 1997, 2003; Poirier & Saint-Aubin, 1996; Watkins & Watkins, 1977). Furthermore, lists composed of words having many versus few lexical neighbors also lead to differential recall performance in vWM tasks (Clarkson et al., 2016; Roodenrys et al., 2002; Vitevitch, Chan, & Roodenrys, 2012). It is important to distinguish here the lexical neighborhood and the lexical cohort effects: while the lexical cohort effect characterizes words sharing the same onset, the lexical neighborhood effect characterizes words differing from each other by a single phoneme substitution, deletion or addition independently of phoneme position. Finally, semantic variables also affect vWM performance, with higher recall performance for lists composed of high versus low imageability words, and for semantically related versus unrelated words lists (Campoy et al., 2015; Poirier & Saint-Aubin, 1996; Romani et al., 2008).

These psycholinguistic effects in vWM tasks can be explained by language-based models of vWM processing, assuming fast and direct interactions between vWM and the linguistic system (Acheson & MacDonald, 2009; Gupta, 2009; Majerus, 2013; N.

Martin et al., 1996; R. C. Martin et al., 1999), where to-be-remembered items are activated within the linguistic system as soon as they are presented in a vWM task. Interactive activation models of language processing are particularly suited for explaining these results (Dell et al., 1997; McClelland & Rumelhart, 1981). Contrary to purely feedforward activation models, in interactive activation models, each level of language processing (phonological, lexical and semantic) is allowed to directly activate adjacent levels via bi-directional connexion weights. In the case of single word repetition, initial activation at the phonological level spreads toward the lexical level. At the same time, activation at the lexical level spreads to the semantic level and reactivates the phonological level. Finally, the semantic layer reactivates the lexical level. These interactions are supposed to occur iteratively over the time-course of a language processing task. The impact of lexical and semantic knowledge on vWM can be explained in the same manner: verbal memoranda associated with richer or more stable lexico-semantic representations will receive stronger feedback activations from adjacent layers, and hence will be less prone to decay over time. This approach has been modelled in a computational model of single word repetition (N. Martin, Dell, Saffran, & Schwartz, 1994; N. Martin et al., 1996) by extending Dell's spreading activation model of picture naming. Although this computational model was constructed to explain single word repetition performance, a conceptual attempt has been made to extend it to whole list repetition (N. Martin et al., 1996). This conceptual approach is also consistent with attention-based models assuming direct interactions between the attentional and long-term memory systems (Barrouillet & Camos, 2007; Cowan, 2001; Oberauer et al., 2012), and with other language-based models assuming strong interactions between vWM and language activation (Acheson & MacDonald, 2009; Gupta, 2009).

As we have shown, many psycholinguistic variables have been assessed as regards their impact on vWM performance. However, one specific effect has not yet been investigated: the effect of lexical cohort competition. As already noted, this effect is related to the number of lexical competitors sharing their first phonemes with a target stimulus (Marslen-Wilson, 1987; Tyler et al., 2000): "alligator" shares the onset syllable /æl/ with the words "alcohol", or "alcove". Words sharing their first phonemes with many other words (i.e., from large cohorts) are usually associated with slower response times in lexical decision tasks as compared to words drawn from small cohorts (Gaskell & Marslen-Wilson, 2002; Kocagoncu et al., 2017; Marslen-Wilson, 1987; Tyler et al., 2000; Zhuang et al., 2011). Words from a given cohort are co-activated and compete for selection when a given speech input is analysed, leading to larger competition effects for words stemming from large cohorts. Furthermore, these lexical

competition effects interact with semantic access: in lexical decision tasks, concreteness/imageability effects are most pronounced for words stemming from larger cohorts (Tyler et al., 2002; Zhuang et al., 2011). This situation has been explained by considering that, when direct and fast mapping between phonological and lexical levels is difficult, semantic levels of processing intervene by providing additional information that will help to disambiguate the target stimulus (Evans et al., 2012).

As regards vWM, like for language processing tasks, language-based models of vWM predict that words drawn from larger cohorts will be more difficult to activate during the encoding stage due to their larger ambiguity, causing item interference effects. This should result in poorer recall performance for words drawn from large vs. small cohorts, either via an increased number of omissions or an increased number of intrusions, or both. At the same time, it should be noted that the lexical cohort effect involves the early stages of lexical access and lexical selection. Once a lexical representation has been activated, the lexical cohort variable is no longer considered to exert any effect. This strongly contrasts with other psycholinguistic effects such as lexicality and word imageability effects which are considered to have a more continuous impact on vWM maintenance as the underlying lexical and semantic variables provide stabilizing feedback all over the vWM maintenance phase, at least according to interactive activation models of language processing. Therefore, we could also expect a reduced or even an absent effect of the lexical cohort variable in vWM performance. In this study, we explored, via two experiments (Experiment 1 and 3), the impact of cohort competition on a word list immediate serial recall task, with all words (for a given list) stemming from large or small lexical cohorts. Experiment 2 was a control experiment checking the validity of the cohort competition manipulations using a lexical decision task.

Experiment 1

Method

Participants. A total of sixteen (12 females, 4 males) participants aged between 18 and 33 years (M = 22.27, SD = 3.83) were recruited from the university community after giving their informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials. The list of stimuli consisted of 210 words associated with large cohorts and 210 words associated with small cohorts. The cohort competition variable was computed using the following procedure. First, we selected all existing French words, including nouns, verbs and adjectives using "Lexique381" from the Lexique 3.0

database (Lexique 3; New, 2006). From this pool, only the canonical (lemma) form of the stimuli was retained. Hence, different words sharing the same lemma (e.g. *sister – sisters*) were considered as the same word in the cohort. When several words shared a common phonological form (i.e., homonym), only the most frequent word form was considered, since its lexical form is supposed to win the competition over the others, less frequent lexical forms. From this final pool, the number of words sharing their first phonemes with a target stimulus was computed, and this for each individual word and for increasing numbers of onset phonemes until a given word's phonological uniqueness point was reached (i.e., when the word can be identified in an unambiguous manner) (Marslen-Wilson, 1987; Zhuang et al., 2014). We derived a first measure quantifying the number of competitors in the initial cohort, which we refer to here as the *number of competitors*, as a function of the number of shared onset phonemes. We also computed a *cohort competition* variable, which for a given target word, corresponds to its lexical frequency, divided by the summed frequency of all its competitors. For instance, given the cohort composed of "cat", "cab" and "car", with lexical frequencies of 5, 3 and 2, respectively, the cohort competition value for "cat" will be equal to $\frac{5}{(3+2)} = 1$. Likewise, the cohort competition value for "cab" will be equal to $\frac{3}{(5+2)} = .43$. As can be seen in this example, *smaller* values represent *higher* competition because the target has less weight (in terms of lexical frequency) in the cohort. Both measures of competition have been shown to impact lexical access in linguistic tasks, but cohort competition seems to be the most reliable variable (Tyler et al., 2000; Zhuang et al., 2011), since it also takes into account the weight of each competitor within the cohort. As expected, the two measures were highly correlated for our pool of stimuli (r = -.84, r^2 = .71 after log transformation, BF_{10} > 100, see statistical analysis below for the interpretation of the Bayes Factor). Critically, we also controlled for lexical neighborhood density by measuring the number of real words in the Lexique381 pool that could be created by adding, deleting or substituting one phoneme in the target word. To do that, we used the Levenshtein distance (Levenshtein, 1966) which computes the minimal distance between two character arrays. These changes include deletion, suppression and substitution. The Levenshtein distance was computed between the target word and all other words of the Lexique381 database (after the word selection process detailed above) allowing us to compute the number of lexical neighbors: a word associated with a Levenshtein distance of 1 was considered to be a neighbor. From this pool, the words containing 5 or 6 phonemes and having a high lexical frequency ($freq_{log}$ > -1.52) were selected and then divided in small and large cohort stimuli using a median split. This set of stimuli was then further

reduced by selecting by hand only the small and large cohort words that were matched according to the different psycholinguistic variables mentioned below.

Both lists of stimuli differed according to the number of competitors (M_{log} = 2.65, SD_{log} = .26 and M_{log} = 1.71, SD_{log} = .45 for high and low competition words, respectively, BF_{10} > 100) and cohort competition values (M_{log} = -.75, SD_{log} = .31 and M_{log} = .76, SD_{log} = .57 for high and low competition words, respectively, BF_{10} > 100) by considering the two first phonemes. Cohort competition values also differed between the two lists when considering the four initial phonemes. The two sets were matched for several other psycholinguistic variables: biphone frequency (M = 887.64, SD = 293.19 and M = 883.81, SD = 347.49 for high and low competition words, respectively, BF_{01} = 9.18, Tubach & Boë, 1990), lexical frequency (M_{log} = .84, SD_{log} = .27 and M_{log} = .84, SD_{log} = .53 for high and low competition words, respectively, BF_{01} = 9.24 taken from the "freqlemfilm2" variable in the Lexique database) and number of phonemes (M = 5.47, SD = .50 and M = 5.47, SD = .50 for high and low competition words, respectively, BF_{01} = 9.25). The two set of stimuli did not differed according to the lexical neighborhood density variable (M = 3.68, SD = 3.02 and M = 3.03, SD = 2.75 for high and low competition words, respectively, BF_{10} = 1.38). Finally, since imageability ratings are available for only a restricted set of stimuli in French, we conducted an online survey in which we invited the participants to judge the degree of imageability of our stimuli on a scale ranging from 1 to 7. Because of this very large number of stimuli to judge, the participants were free to stop the survey at any moment. Sixty-seven participants took part in the survey, and each stimuli was judged 16.15 times on average. The two sets of stimuli were equivalent in terms of imageability ratings (M = 4.60, SD = 1.42 and M = 4.64, SD = 1.42 for high and low competition words, respectively, BF_{01} = 8.93). All the linguistic properties for this set of stimuli are summarized in **Table 4.1**. Note that an additional analysis in which homonyms were included for computing the number of competitors and cohort competition values led to a similar pattern of results, with low and high cohort stimuli still reliably differing on these values.

Table 4.1. *Values for linguistic matching variables between high and low cohort word stimuli used in Experiments 1 & 2.*

Linguistic variables	Cohort competition		BF
	High	**Low**	
Number of competitors (M_{log})	2.65 (.26)	1.71 (.45)	$BF_{10} = 5.32e+86$
Cohort competition (M_{log})	-.75 (.31)	.76 (.57)	$BF_{10} = 1.99e+118$
Biphone frequency	887.64 (293.19)	883.81 (347.49)	$BF_{01} = 9.18$
Lexical frequency (M_{log})	.84 (.27)	.84 (.53)	$BF_{01} = 9.24$
Number of phonemes	5.47 (.5)	5.47 (.5)	$BF_{01} = 9.25$
Neighborhood density	3.68 (3.02)	3.03 (2.75)	$BF_{10} = 1.38$
Imageability	4.60 (1.42)	4.64 (1.42)	$BF_{01} = 8.93$
Acoustic length	751 (100)	751 (88)	$BF_{01} = 9.25$
LSA values	.05 (.04)	.06 (.04)	$BF_{01} = 12.21$

Note. Log transformation of mean values is signalled by "(M_{log})". Values in parenthesis represent standard deviations. BF values are based on Bayesian Independent Samples T-Tests. LSA = Latent Semantic Analysis.

The items were recorded by a French-native female speaker in a neutral voice. Each item was then isolated in a separate file whose length corresponded to its acoustic duration. A Bayesian Independent Samples T-Test showed that high and low competition stimuli did not differ according to their length (M = 751 ms, SD = 100 ms and M = 751 ms, SD = 88 ms for high and low competition words, respectively, BF_{01} = 9.25). We removed the residual background noise via Audacity® which uses a Fourier analysis (see http://wiki.audacityteam.org/wiki/How_Audacity_Noise_Reduction_Works for more information).

Procedure. Each participant received a different version of the vWM task. We manipulated the cohort competition effect by presenting 6-word lists composed of words drawn from either high or low cohort competition, such that a given list was composed of words drawn exclusively from one stimulus condition. The lists were pseudorandomly presented with the constraint that a given condition could not appear on more than three consecutive trials. In order to avoid phonological overlap, two adjacent words could not share the same two first or two last phonemes within each list, because phonological similarity has been shown to strongly influence vWM processing (Baddeley, 1966), and this both at the item and serial order levels of processing (Gupta et al., 2005). In addition, for each trial, we computed Latent Semantic Analysis (LSA) values (http://lsa.colorado.edu/, using the semantic space "Francais-Monde-Extended"). LSA measures reflect the extent to which two words co-

occur in the same linguistic corpora. The higher the co-occurrence of two words, the higher their (theoretical) semantic association values. This lexical variable is important to control for because it has been shown to impact vWM performance and has previously been shown to drive, at least partially, the lexical frequency effect (Hulme et al., 1997, 2003, but see Poirier & Saint-Aubin, 2005; Tse & Altarriba, 2007). We computed LSA values for adjacent items within a given word list (Saint-Aubin, Guérard, Chamberland, & Malenfant, 2014). For each adjacent pair of stimuli, the LSA values were then averaged for each trial. We observed that the two stimulus conditions did not differ (M = .05, SD = .04 and M = .06 and SD = .04 for high and low cohort trials, respectively), and this was supported by strong evidence (BF_{01} = 12.21). There was 35 experimental trials in each stimulus conditions. Participants could take a short break after 35 trials if they needed to. The whole experiment took approximately 35 minutes to be performed.

Participants performed 3 unrecorded practice trials before the beginning of the main vWM task. At the beginning of each trial, an on-screen countdown display starting from 3 was first presented, followed by a blank screen and the auditory items presented at a rate of 1 item every 1200 ms. Each list was directly followed by a sinusoidal tone of 440 Hz lasting for 150 ms, signaling the start of the recall phase. After participants had recalled the items, they were invited to initiate the next trial using the SPACEBAR of the keyboard. Participants were told that they had to recall aloud any item they could remember and in the serial order in which the items had been presented. In order to ensure for accurate scoring of serial recall performance, participants were invited to use a sheet when recalling each item. The sheet was placed directly in front of them on the desk in landscape orientation, and was composed of 6 squares placed along the horizontal axis (see **Appendix A**). The participants were invited to move their finger to the right by one square each time they recalled an item. Pilot tests had shown that participants often failed to recall all 6 items because they struggled to count how many items they had already recalled. The pointing procedure helped participants to keep track of the number of recalled items and allowed the experimenters to accurately score serial recall performance. When participants could not remember a given item in the list, they were invited to say the word "blanc" (i.e., "blank" in French). Task presentation was controlled via OpenSesame software running on a desktop station computer (Mathôt et al., 2012). The auditory stimuli were presented via headphones directly connected to the computer. The loudness was adjusted to comfortable listening levels for each participant during the practice trials. The participants' responses were recorded with a digital recorder and stored on computer disk for later transcription and scoring.

As regards the scoring procedure, we performed different analyses. First, we used an item recall scoring procedure in which an item was scored as correct if it was recalled regardless of its recall position. For instance, given the target sequence "Item1 – Item2 – Item3 – Item4 – Item5 – Item6" and the output sequence "Item1 – Item2 – Item4 – Item3 – blank – Item6", items 1, 2, 3, 4 and 6 were scored as correct. This scoring procedure is particularly sensitive to item recall. In addition, we also performed a strict scoring procedure in which an item was scored as correct only if it was recalled at the correct serial position. Using this scoring procedure, only items 1, 2 and 6 would be scored as correct in the previous example.

Statistical analysis. We performed a Bayesian analysis instead of the traditional frequentist analyses in order to substantially reduce Type-1 false error probabilities (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). The Bayesian approach has the advantage of computing continuous values against or in favour of a given model, rather than deciding for the presence of an effect based on an arbitrary statistical threshold. Evidence in favour of a model is given by the Bayesian Factor (BF), reflecting the likelihood ratio of a given model relative to other models, including the null model. Both the null model and the effect of interest can be simultaneously tested, by directly comparing the alternative hypothesis against the null hypothesis, and vice versa. The BF_{10} is used to determine the likelihood ratio for the alternative model (H_1) relative to the null model (H_0), and the BF_{01} to determine the likelihood ratio for H_0 relative to H_1. We use the classification of strength of evidence proposed in previous studies (Jeffreys, 1998; Wagenmakers et al., 2011): A BF of 1 provides no evidence, $1 < BF < 3$ provides anecdotal evidence, $3 < BF < 10$ provides moderate evidence, $10 < BF < 30$ provides strong evidence, $30 < BF < 100$ provides very strong evidence and $100 > BF$ provides extreme/decisive evidence. We also report $BF_{inclusion}$ values which compare the evidence of all models including a given factor relative to all other models not including this factor. All the analyses were performed using JASP (JASP Team, 2017) and we used default Cauchy prior distribution parameters as implemented in JASP (Version 0.8.5.1).

Results and discussion

We first assessed the effect of cohort competition (high, low) as a function of serial position (1-6) using a Bayesian Repeated Measures ANOVA. For the item recall measure, we found moderate evidence *against* the effect of cohort competition (BF_{01} = 7.93), decisive evidence supporting the serial position effect ($BF_{inclusion} \rightarrow +\infty$) and strong evidence against the interaction term (BF_{01} = 24). Similar results were observed when using a strict recall criterion, with moderate evidence against the cohort competition
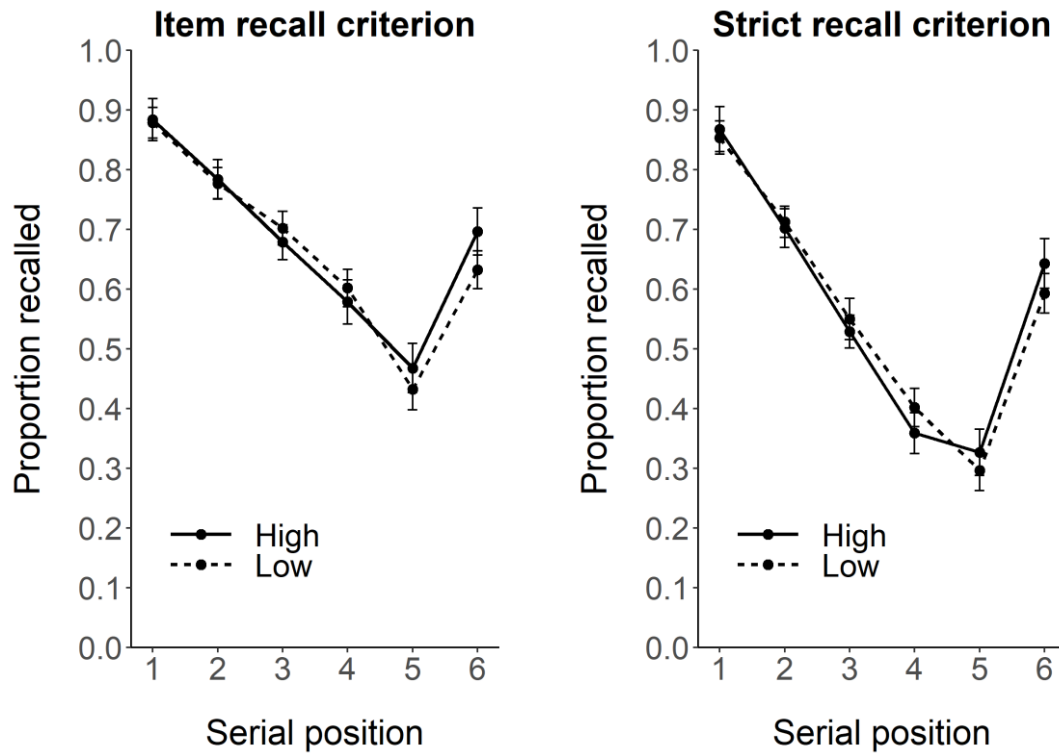
Figure 4.1. Experiment 1 - Proportion of items correctly recalled (y axis) across serial positions (x axis) as a function of the lexical cohort variable, for item recall (left panel) and strict recall (right panel) scoring procedures. The solid and dashed lines represent high and low cohort stimuli, respectively. Error bars represent standard errors, after correction for between-subject variability (Cousineau, 2005).

effect (BF_{01} = 9.09), decisive evidence in favour of the serial position effect ($BF_{inclusion}$ = 3.002e+15) and very strong evidence against the interaction term (BF_{01} = 30.30). These results are displayed in **Figure 4.1**. Hence, we observed no impact of cohort competition on recall accuracy, and this absence of difference was reliably supported, as shown by the BF_{01}.

This pilot experiment provides evidence for the *absence* of a cohort competition effect on vWM recall performance. At the same time, we cannot rule out the possibility that this absence reflects an insufficient contrasted lexical cohorts, or because this effect does not characterize the French language (although the latter possibility is rather unlikely, English and French sharing many lexical properties; but see Sadat, Martin, Costa, & Alario, 2014). So far, no study has investigated the cohort competition effect in French, and therefore we need to check whether our stimulus material is appropriate for eliciting this effect in French, by using a lexical decision task which has been most frequently used to evidence this effect in other languages (mostly English).

Experiment 2 – Cohort competition in lexical decision

Experiment 2 assessed the effect of cohort competition in a linguistic, lexical decision task where participants are invited to judge the lexical status of word and nonword stimuli. The primary goal of Experiment 2 was to assess whether the absence of a cohort competition effect observed in Experiment 1 was due to the specific set of stimuli we had created. Second, examining the occurrence of a cohort competition effect in French is important to demonstrate its generalizability across languages. Third, since we used word and nonword stimuli for lexicality judgement, the impact of cohort competition was factorially manipulated and hence was also assessed on nonword stimuli. Even though one study investigated effects of cohort competition on nonwords using correlational methods (Zhuang et al., 2014), cohort competition in nonwords has never been directly manipulated and evidence supporting its existence is scarce.

Method

Participants. A total of twenty-nine (28 females, 1 male) participants aged between 19 and 25 years (M = 21.14, SD = 1.43) were recruited from the university community after giving their informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials. The same set of words as in Experiment 1 was used. We additionally created nonword stimuli for the lexical decision task used in Experiment 2. We first constructed a large (N > 10e+5) set of nonwords using an algorithm programmed under MATLAB® (the script and the modified Lexique381 pool was made available in the Open science framework using the following link: https://osf.io/3rkh5/) with the constraint that a given item could not match any entry within Lexique381. In addition, all the nonwords were 5 to 6 phonemes long, and were created by randomly assembling phonemes from the French language, by constraining the program to use the syllabic structures as those characterizing the stimuli of the word pool. Two sets of 210 nonwords were selected, with the constraint that they had to strongly differ in their number of lexical competitors. Like for the word stimuli, this was made by computing the number of words in the Lexique381 database (after discarding homophones and non-lemma forms, see Experiment 1 for the details of the cleaning process) sharing their onset with the target nonword. For instance, given the nonword "caz", the words "cat", "cab" and "car" were considered as competitors. As for the word stimuli, we also computed the summed frequency of all the competitor words as an equivalent to the cohort competition variable (Zhuang et al., 2014).

The high and low competition nonwords differed according to their number of competitors (M_{log} = 2.61, SD_{log} = .34 and M_{log} = 1.462, SD_{log} = .43 for high and low competition nonwords, respectively, BF_{10} > 100) and variable of competitors summed frequency (M_{log} = 3.29, SD_{log} = .55 and M_{log} = 1.74, SD_{log} = .95 for high and low competition nonwords, respectively, BF_{10} = 1.67e+61). The high and low competitors nonwords were matched for biphone frequency (M = 646.76, SD = 335.29 and M = 646.206, SD = 378.22 for high and low competition nonwords, respectively, BF_{01} = 9.5), neighborhood density (M = .3, SD = .7 and M = .18, SD = .61 for high and low competition nonwords, respectively, BF_{01} = 1.75) and number of phonemes (M = 5.47, SD = .5 and M = 5.47, SD = .5 for high and low competition nonwords, respectively, BF_{01} = 9.25). The stimuli did not differed according to their acoustic duration (M = 688 ms, SD = 95 ms and M = 704 ms, SD = 89 ms for high and low competition words, respectively, BF_{01} = 2.01).

Both word and nonword stimuli were matched according to their number of phonemes (M = 5.47, SD = .5 and M = 5.47, SD = .5 for word and nonword stimuli, respectively, BF_{01} = 12.961) and syllabic structures (98.81% had a similar consonant/vowel/semivowel structure). The word and nonword stimuli could however not be perfectly matched for all psycholinguistic variables. This is however not problematic for the purpose of the present experiment as we were not interested in directly comparing the word and nonword stimuli. More specifically, the word and nonword stimuli differed strongly according to their phonotactic frequency (M = 885.73, SD = 321.11 and M = 646.48, SD = 356.972 for word and nonword stimuli, respectively, BF_{10} = 9.14e+19), acoustic duration (M = 751.038 ms, SD = 94.06 ms and M = 696.215 ms, SD = 92.423 ms for word and nonword stimuli, respectively, BF_{10} = 6.32e+13), neighborhood density (M = 3.58, SD = 2.9 and M = .24, SD = .66 for word and nonword stimuli, respectively, BF_{10} = 1.64e+78) and to a lesser extent, their number of lexical competitors (M = 2.18, SD = .6 and M = 2.04, SD = .7 for word and nonword stimuli, respectively, BF_{10} = 8.38). All these values are summarized in **Table 4.2**.

Table 4.2. *Values for linguistic matching variables between high and low cohort nonword stimuli used in Experiment 2.*

Linguistic variables	Cohort competition		BF
	High	Low	
Number of competitors (M_{log})	2.61 (.34)	1.462 (.43)	$BF_{10} = 3.31e+103$
Competitors summed freq. (M_{log})	3.29 (.55)	1.74 (.95)	$BF_{10} = 1.67e+61$
Biphone frequency	646.76 (335.29)	646.206 (378.22)	$BF_{01} = 9.5$
Number of phonemes	5.47 (.5)	5.47 (.5)	$BF_{01} = 9.25$
Neighborhood density	.3 (.7)	.18 (.61)	$BF_{01} = 1.75$
Acoustic length	688 (95)	704 (89)	$BF_{01} = 2.01$

Note. Log transformation of mean values is signalled by "(M_{log})". Values in parenthesis represent standard deviations. BF values are based on Bayesian Independent Samples T-Tests. LSA = Latent Semantic Analysis.

All the nonword items were recorded by a French-native female speaker in a neutral voice, using the same voice as for the words. Each item was then isolated in a separate file whose length reliably corresponded to its acoustic duration. We removed the background noise via Audacity®. The word and nonword stimuli were recorded at different moments. In order to ensure that word and nonword stimuli did not differ at the level of general acoustic parameters, we checked fundamental frequency and intensity values. An analysis of the fundamental frequency (F0) using the "freqz" function implemented under MATLAB® showed evidence for a very small difference between the word and nonword stimuli, with significant overlap of values (M = 446.5, SD = 262.06 and M = 500.3, SD = 326.3 for word and nonword stimuli, respectively, BF_{10} = 2.31). Intensity values were associated with positive evidence for an absence of difference (M = 208.8, SD = 76.31 and M = 200.9, SD = 90.62 for word and nonword stimuli, respectively, BF_{01} = 5.11).

Procedure. For each participant, stimuli were presented in pseudorandom order, with the constraint that a given stimulus condition (word/nonword, high/low cohort) could not be repeated on more than three consecutive trials. There were 210 trials for each of the 4 stimulus conditions, and it took approximatively 45 minutes to perform the whole experiment. Participants were allowed to take a maximum of three short breaks if they needed to. Participants performed 14 practice trials before administration of the main task. If participants made a mistake during the practice trials, they received corrective feedback on their performance.

Each trial began with an on-screen fixation cross lasting on average 1000ms, plus/minus a random duration sampled from a continuous uniform distribution

ranging from 0 to 250ms. The fixation cross was directly followed by a blank screen and the target item. Participants were told that they had to judge the lexical status of the item (i.e., "You will have to judge whether the presented item is a word or a nonword."), and had to press the "S" (for word) or "L" (for nonword) key on the keyboard to indicate their response. The next trial directly began after each keypress. Participants were told that they had to respond as fast as possible, without sacrificing accuracy. In order to stress rapidity, throughout the entire experiment, an on-screen message instructed participants to respond faster when they failed to respond within 2500 ms after a stimulus' onset (scored as a no-response). Stimulus presentation was controlled via OpenSesame software running on a desktop station computer (Mathôt et al., 2012). The auditory stimuli were presented via headphones directly connected to the computer. The loudness was adjusted to comfortable listening levels during the practice trials. The experiment was separated in four blocks, allowing participants to take a very short break between blocks if they needed to. Both response accuracy and time were recorded. Hits and false alarms were combined via d' prime scores (Stanislaw & Todorov, 1999).

Results and discussion

Participants were on average very accurate, and this both for words (M = .95, SD = .04) and nonwords (M = .97, SD = .03). A Bayesian Repeated Measures ANOVA on d' scores with the factors lexicality (word, nonword) and cohort competition (high, low) showed that discrimination scores did not differ as a function of lexicality (BF_{01} = 5.71) or cohort competition (BF_{01} = 2.14), and strong evidence supported the absence of an interaction (BF_{01} = 15.63).

After removing incorrect trials, response times smaller or larger than 2.5 absolute deviations from the median on an individual basis (Leys, Ley, Klein, Bernard, & Licata, 2013) were discarded from the analysis, leading to discarding a total of 1161 observations (4.77%) from the entire set of data. The vast majority (97.42%) of these extreme values were located in the upper part of the distribution, and comprised response times between min = 970 and max = 2493. In the lower part of the distribution, these data comprised response times between min = 164 and max = 684. For each participant, average response times were computed across all four stimulus conditions (word-high competition; word-low competition; nonword-high competition; nonword-low competition). The presence of a cohort competition effect (high, low) as a function of lexicality (word, nonword) was assessed using a Bayesian Repeated Measures ANOVA. We found decisive evidence supporting both effects of lexicality ($BF_{inclusion}$ = 458447.174) and cohort competition ($BF_{inclusion}$ = 5342.209). The interaction
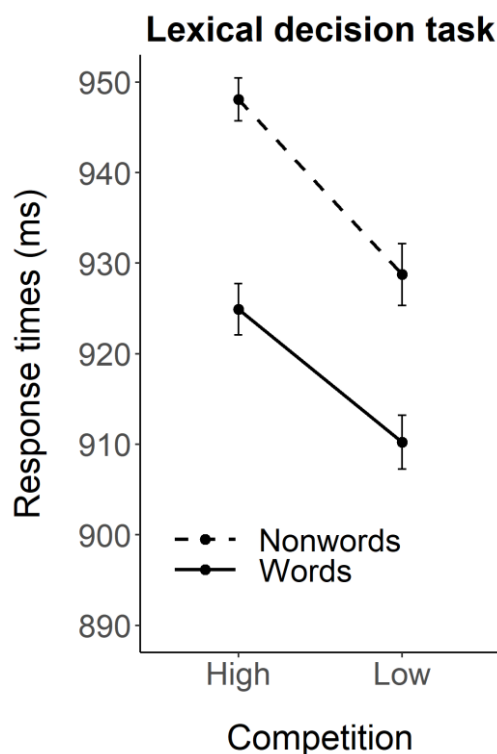
Lexical decision task

Figure 4.2. Experiment 2 – Mean response times averaged across participants (y axis) for high and low cohort stimuli (x axis) separately for words (solid line) and nonwords (dashed line). Error bars represent standard errors, after correction for between-subject variability (Cousineau, 2005).

term was ambiguous ($BF_{inclusion}$ = 1.108). Hence, the cohort competition effect was reliably observed across the word and nonword conditions. The same analysis using median response times instead of average response times led to similar results. These results are displayed in **Figure 4.2**. Note that a mixed model analysis using frequentist statistics was also conducted, and led to very similar conclusions. This latter analysis is available in **Appendix B**.

The observation of slower response times for words from large lexical cohorts is consistent with previous psycholinguistic studies. Interestingly, the present study also revealed a cohort competition effect for nonwords. This finding suggests that lexical activation also occurs for these stimuli, possibly in the form of lexical search processes (e.g. Vitevitch & Luce, 1999). Critically, the presence of a lexical cohort effect in the lexical decision task of Experiment 2 shows that the absence of a cohort competition effect for the vWM task in Experiment 1 cannot be imputed to the specific characteristics of our stimulus material. It should be noted that cohort competition effects, when occurring, are typically very small, with a mean difference of ~15ms for word stimuli in the present study (see also Tyler et al., 2000 and Zhuang et al., 2011 for similar values). It could therefore be argued that the impact of the lexical cohort variable may have been too subtle to influence performance in vWM paradigms. To

assess this possibility, Experiment 3 used the same vWM task setup as Experiment 1 but by selecting only those word items that had led to the largest cohort competition effects in Experiment 2.

Experiment 3

This third experiment assessed the cohort competition effect in a vWM task by using only the word stimuli that had been shown to be the most responsive to the cohort manipulation variable in Experiment 2. We retained, from Experiment 2, those word stimuli associated with the slowest response times (for the large cohort category), and with the fastest response times (for the small cohort category).

Method

Participants. A total of thirty (23 females, 7 males) participants aged between 18 and 28 years (M = 21.37, SD = 2.30) were recruited from the university community after giving their informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials. The word stimuli were identical to those used in Experiment 1, except that we selected the items with the largest response time differences in the lexical decision task of Experiment 2. We were able to select 150 word items for each stimulus condition. More specifically, we first considered the median response times obtained for each word in the lexical decision task in Experiment 2. Next, we removed one by one items from the large and small cohort word sets such that for the remaining words the difference in terms of response times was maximized, while also ensuring that the words were still matched at the level of psycholinguistic variables between both sets. Although there was still a slight overlap between the two sets in terms of response times (M = 930.8, SD = 67.76 and M = 885.65, SD = 58.82 for high and low cohort stimuli, respectively, BF_{10} = 4.02e+6), the gap was now larger as compared to the initial set: ~45ms. The two stimulus sets were matched for biphone frequency (M = 907.84, SD = 283.30 and M = 905.35, SD = 353.23 for high and low cohort stimuli, respectively, BF_{01} = 7.85), imageability (M = 4.76, SD = 1.34 and M = 4.76, SD = 1.45 for high and low cohort stimuli, respectively, BF_{01} = 7.87), lexical frequency (M_{log} = .84, SD_{log} = .25 and M_{log} = .85, SD_{log} = .42 for high and low cohort stimuli, respectively, BF_{01} = 7.7), number of phonemes (M = 5.47, SD = .50 and M = 5.46, SD = 50 for high and low cohort stimuli, respectively, BF_{01} = 7.67), neighborhood density (M = 3.71, SD = 2.97 and M = 3.2, SD = 2.75 for high and low cohort stimuli, respectively, BF_{01} = 2.5) and acoustic length (M = 750, SD = 97 and M = 749, SD = 83 for high and low cohort stimuli, respectively, BF_{01} =

Table 4.3. *Values for linguistic matching variables between high and low cohort word stimuli used in Experiments 3.*

Linguistic variables	Cohort competition		BF
	High	Low	
Number of competitors (M_{log})	2.64 (.26)	1.69 (.42)	$BF_{10} = 9.08e+66$
Cohort competition (M_{log})	-.74 (.29)	.76 (.56)	$BF_{10} = 5.54e+85$
Biphone frequency	907.84 (283.3)	905.35 (353.23)	$BF_{01} = 7.85$
Lexical frequency (M_{log})	.84 (.25)	.85 (.42)	$BF_{01} = 7.7$
Number of phonemes	5.47 (.5)	5.46 (.5)	$BF_{01} = 7.67$
Neighborhood density	3.71 (2.97)	3.2 (2.75)	$BF_{01} = 2.5$
Imageability	4.76 (1.34)	4.76 (1.45)	$BF_{01} = 7.87$
Acoustic length	750 (97)	749 (83)	$BF_{01} = 7.81$
LSA values	.06 (.04)	.05 (.04)	$BF_{01} = 5.21$

Note. Log transformation of mean values is signalled by "(M_{log})". Values in parenthesis represent standard deviations. BF values are based on Bayesian Independent Samples T-Tests. LSA = Latent Semantic Analysis.

7.81). The high and low cohort stimuli still differed at the level of number of competitors (M_{log} = 2.64, SD_{log} = .26 and M_{log} = 1.69, SD_{log} = .42 for high and low cohort stimuli, respectively, BF_{10} = 9.08e+66) and cohort competition values (M_{log} = -.74, SD_{log} = .29 and M_{log} = .76, SD_{log} = .56 for high and low cohort stimuli, respectively, BF_{10} = 5.54e+85). A summary of matching variables is provided in **Table 4.3**.

Procedure. Ten different experimental lists were created, counterbalanced across subjects. Like for Experiment 1, we computed LSA values for each trials. The two stimulus conditions did not differ at the level of LSA values (M = .06, SD = 0.04 and M = .05, SD = .04 for high and low cohort stimuli, respectively), and this absence of difference was supported by moderate Bayesian evidence (BF_{01} = 5.21). Due to a smaller amount of items available for each word condition relative to the previous experiments (150 instead of 210), we decided to reduce list length by creating lists of 5 items (instead of 6 in Experiment 1). This enabled us to create 30 trials (instead of 35 in Experiment 1) per experimental conditions and to present each word only once (as in previous experiments). In addition, presentation rate was set to 1000ms per items to further shorten task duration. The response sheet for immediate serial recall was the same as in Experiment 1 (see **Appendix A**), except that it included 5 squares. All other aspects of the experimental procedure, including statistical analysis and scoring procedures were identical to Experiment 1.
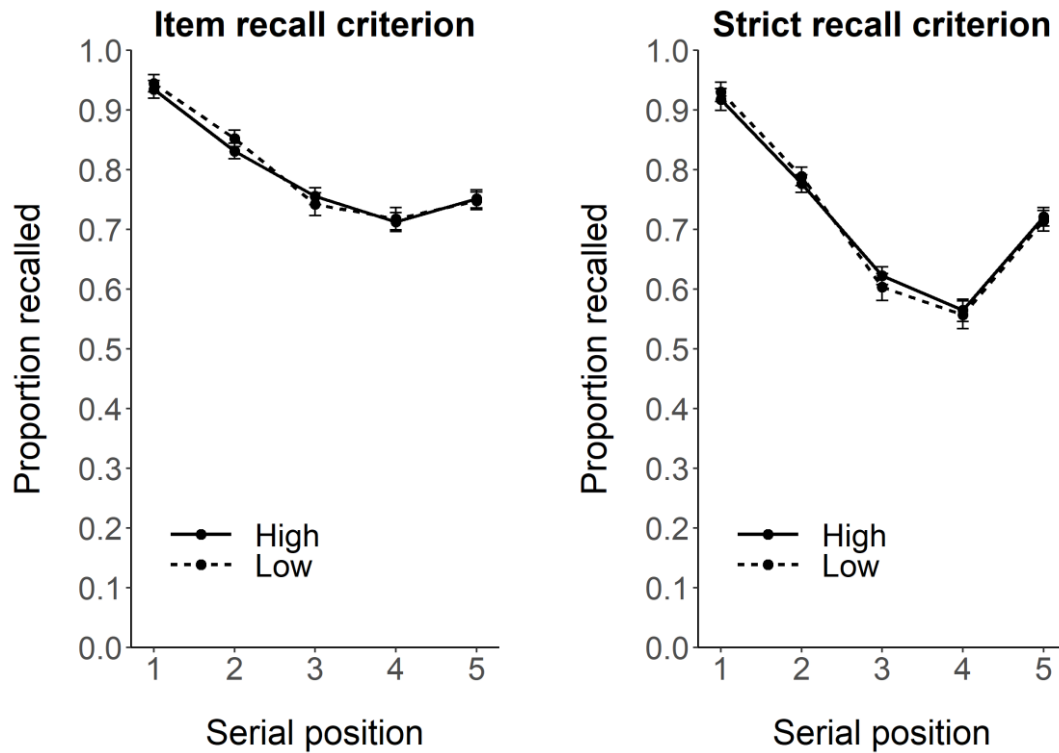
Figure 4.3. Experiment 3 - Proportion of items correctly recalled (y axis) across serial positions (x axis) as a function of the lexical cohort variable, for item recall (left panel) and strict recall (right panel) scoring procedures. The solid and dashed lines represent high and low cohort stimuli, respectively. Error bars represent standard errors, after correction for between-subject variability (Cousineau, 2005).

Results and discussion

We observed very similar results to those reported in Experiment 1. Using a Bayesian Repeated Measure ANOVA on item recall performance, we found moderate evidence against the cohort competition effect (BF_{01} = 9.52), decisive evidence supporting the effect of serial position ($BF_{inclusion} \rightarrow +\infty$) and very strong evidence supporting the absence of interaction (BF_{01} = 55.56) (see **Figure 4.3**). When conducting the same analyses using the strict recall criterion, there was again strong evidence supporting the absence of a cohort competition effect (BF_{01} = 10.87), decisive evidence supporting the serial position effect ($BF_{inclusion} \rightarrow +\infty$) and very strong evidence supporting the absence of interaction (BF_{01} = 62.5).

Experiment 3 confirms the absence of a cohort competition effect in vWM, even when using stimuli that had been shown to lead to maximal cohort effects in a lexical decision task. In addition, this absence appears to be reliable given that we included a higher number of participants as compared to Experiment 1.

General discussion

This study explored the impact of the cohort competition effect on vWM. We observed in two experiments that this linguistic variable did not influence immediate serial recall performance for lists composed of words stemming from large versus small lexical cohorts. This result cannot be attributed to a problem at the level of stimulus material as, with the same stimulus set, a reliable cohort effect was observed in a lexical decision task, the task most typically used in previous studies for studying cohort competition effects.

Why is the lexical cohort variable associated with a null effect in vWM tasks?

For cohort models of language processing, stimuli drawn from large cohorts are considered to be more ambiguous during lexical selection because a greater amount of lexical competitors are activated simultaneously during the initial stages of speech perception, and have to be inhibited during the lexical selection process (Gaskell & Marslen-Wilson, 2002; Kocagoncu et al., 2017; Marslen-Wilson, 1987; Tyler et al., 2000; Zhuang et al., 2011). Computational implementations of cohort effects consider that words drawn from large cohorts receive initially less activation and hence need more time to reach their activity threshold (Chen & Mirman, 2012). A possible explanation for the observed lack of a cohort competition effect in an immediate serial recall task is that the rapidity of lexical activation (selection) at the encoding stage is not a strong contributor to vWM performance. Contrary to other psycholinguistic effects, the cohort competition variable influences language processing only during the initial stages of lexical selection process, which may not be sufficient to produce measurable differences in terms of recall performance in vWM tasks. For other psycholinguistic effects, such as the imageability effect for example, items associated with richer and stable semantic features are considered to be more highly activated due to continuous interactive activations between lexical and semantic representations (Pexman et al., 2002; Yap et al., 2015) and this during all stages (encoding, maintenance and recall) of vWM processing, leading to higher vWM recall performance. Similarly, the semantic similarity effect has been explained by assuming that semantically related words will continuously reactivate each other through interactive activations via their shared semantic features (Dell et al., 1997), leading to overall higher activation levels.

It could be argued that other psycholinguistic effects can also be explained in terms of speed of lexical activation while still producing measurable effects in vWM tasks such as the lexical frequency effect. The lexical frequency effect has indeed been explained by assuming that high frequency words have a higher resting activation level and hence can be activated more easily and rapidly (McClelland & Elman, 1986;

McClelland & Rumelhart, 1981). At the same time, the lexical frequency effect can also be explained in terms of connection strengh between phonological and lexical levels of representations (Besner & Risko, 2016), with higher connection strength for high frequency words. It results that for a same quantity of activation at the phonological level, high frequency words will receive more activation and will be more strongly activated as compared to low frequency words. Finally, it must be noted that the frequency effect is also partly driven by inter-item associations in immediate serial recall tasks, with high frequency words co-occuring more frequently than low frequency words (Hulme et al., 2003; Stuart & Hulme, 2000; Tse & Altarriba, 2007). Due to these higher inter-item associations, high frequency words may also activate and support each other during WM encoding and maintenance, similarly to the semantic similarity effect described above.

More generally, items associated with faster response times in linguistic tasks do not necessarily have a positive effect on WM maintenance and recall performance, as further illustrated by the lexical neighborhood density effect. Indeed, while slower response times have been observed in auditory comprehension tasks for dense neighborhood density stimuli, a reverse effect is observed in vWM tasks, with words from dense neighborhoods facilitating recall performance. If vWM performance was to be explained exclusively by the rapidity of lexical activation, reduced performance for dense neighborhood stimuli should be expected. Note that for now, the null effect observed here for the lexical cohort variable in vWM only holds for the type of task that was used in the different experiments. In immediate serial recall tasks, after encoding, memoranda are maintained via internal mechanisms such as refreshing and rehearsal and hence are no longer externally driven, contrary to lexical decision tasks. The null effect observed in this study could be re-examined using running span tasks relying on very fast presentation of memoranda and unexpected, immediate output diminishing the role of internally generated representations. It should however be noted that, for those psycholinguistic effects that have been examined with this type of task, the effects are very similar to those observed in immediate serial recall tasks (see Kowialiewski & Majerus, 2018b).

Consequences for theoretical frameworks

The absence of a cohort competition effect on vWM performance suggests that language-based models of vWM need to distinguish between the speed of lexical activation and the stability of lexical activation. To our knowledge, this is the first study investigating the effect of speed of lexical activation on vWM performance as reflected by the cohort competition effect. For language-based, and more broadly,

activation-based models of vWM (Acheson & MacDonald, 2009; Cowan, 1995, 2001; Majerus, 2013; N. Martin et al., 1996), verbal items are supposed to be activated in vWM using the same mechanisms as those used in language processing more generally. Hence, mechanisms related to the speed of lexical activation should also operate in these models, although they do not (yet) explicitly include them. In these models, items need to be constantly refreshed using the focus of attention and/or rehearsal, otherwise they will be rapidly forgotten due to decay/interference. It logically follows that items that are more strongly activated are less likely to decay up to the point of being forgotten, leading to higher vWM span. In contrast, we may consider that the speed of initial activation is supposed to have a more negligible impact, because it has only a limited influence on the overall activation level and/or decay.
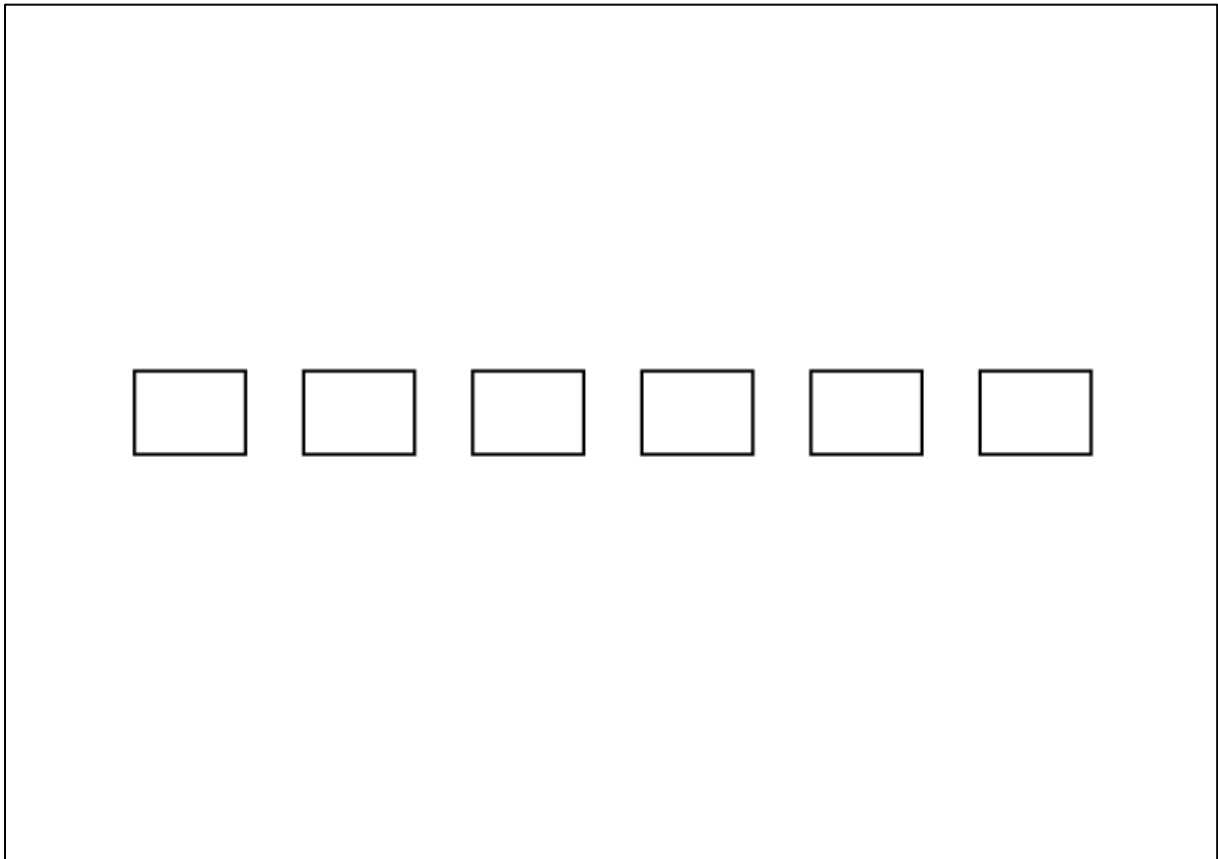
It could also be argued that the results of the present experiments support the redintegration framework which assumes that, during the recall phase of vWM processing, a reconstruction process occurs to "redintegrate" the degraded traces that have been maintained in a phonological buffer. In a strong version of this theoretical framework (Hulme et al., 1991; Lewandowsky, 1999; Schweickert, 1993), lexical and/or semantic knowledge affect vWM processing only during the recall phase, while the encoding stage is characterized by the maintenance of only phonological codes. The recall advantage for words over nonwords, for instance, is explained by assuming that the redintegration process will use the stored lexical representations to clean-up the degraded phonological traces of word stimuli. This model predicts an absence of cohort competition effect because this variable affects only the speed of activation of items during the encoding stage and during the redintegration process but not the quality of their activation in terms of availability, strength or robustness. At the same time, it should be noted that other evidence is not in favour of a redintegration mechanism as the exclusive account of psycholinguistic effects in vWM. For instance, strong lexicality effects have been observed in vWM tasks that do not require overt recall and redintegration (Jefferies et al., 2006b; Kowialiewski & Majerus, 2018a; N. Savill et al., 2016). The neighborhood density effect is also of interest here. As explained above, words from dense neighborhood structures are better recalled in vWM tasks, while the redintegration hypothesis would predict the reverse: when reconstructing degraded phonological traces for words from dense neighborhoods, recall performance should decrease due to the many competing words (neighbors) that can potentially be selected for reconstructing the target word. In contrast, interactive activation models predict that high neighborhood items should be better recalled, because they reactive each other via their shared phonological features (Gordon &

Dell, 2001), resulting in a greater amount of activation and will consequently be less affected by decay/interference. In sum, language-based models of vWM assuming interactive activation within the linguistic system during encoding, maintenance and recall provide a theoretical framework that is able to deal with a wider range of empirical data, including those observed in the present study.

Conclusions

The absence of a cohort competition effect observed in vWM suggests that the speed of lexical activation is not a critical factor for vWM performance. Instead, the psycholinguistic effects that have a robust impact on vWM performance are rather driven by the strength and robustness of lexical activation.

Appendix A. Recall sheet used in Experiment 1.

Appendix B. Mixed model analysis for the effect of cohort competition in Experiment 2.

The mixed model analysis was launched using the lme4 and lmerTest (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017) packages under R (R Development Core Team, 2008). We ran the model on response times as dependent variable, with lexicality (word, nonword), cohort competition (high, low) and the interaction term as fixed effects. The participants and items intercepts were set as random effects. Because the full model failed to converge with maximum random parameters, by-item and by-participant were set as random slopes for the effect of cohort competition, while only by-participant was set as random slope for the effect of lexicality. We found an effect of lexicality (t = -3.067, p = .00274), cohort competition (t = -3.3, p = .00103) but no interaction (t = .664, p = .50672), suggesting that the effect of cohort competition was equally observed for both words and nonwords. The analysis was launched using the following R code:

lexical_decision.model = lmer(RT ~ competition + lexicality + (competition:lexicality) + (competition | items) + (competition + lexicality | participants), data = lexical_decision)

summary(lexical_decision.model)

Study 5

The protective effect of semantic knowledge on working memory

Benjamin Kowialiewski and Steve Majerus
Submitted

Abstract. Several studies have demonstrated an influence of semantic knowledge on verbal working memory (WM) performance, such as shown by the observation of semantic relatedness (related vs. unrelated words) and word imageability (high vs. low imageability words) effects in working memory. The present study extends these observations by examining in four experiments the extent to which semantic knowledge can protect WM representations against interference. We assessed immediate serial recall performance for semantically related vs. unrelated word lists and for high vs. low imageability word lists, with memory lists being followed by an interfering task after encoding or not. Results show that semantic relatedness leads to a stronger protective effect against interference than word imageability. Overall, the semantic relatedness had a stronger impact on WM performance than word imageability; this was further confirmed by a meta-analysis of all relevant studies in the field. These results suggest that inter-item associative semantic knowledge can protect WM content against interference, but less so item-level semantic knowledge. This protective effect may result from between-item recurrent reactivation or from reduced cognitive load via the compression of memoranda into conceptual chunks.

Introduction

Semantic knowledge has been shown to influence WM performance: semantically related words are better recalled as compared to semantically unrelated words (semantic relatedness effect) (Poirier & Saint-Aubin, 1995), and concrete/highly imageable words are better recalled as compared to abstract or low imageability words (word imageability effect) (Walker & Hulme, 1999). The aim of this study is to examine the impact of semantic knowledge on WM performance in a more specific manner, by examining the extent to which semantic knowledge can protect memoranda against interference. In order to do so, we assessed the protective effect of semantic knowledge

associated with semantic relatedness and word imageability when participants have to conduct an interfering task between encoding and recall of memoranda.

In the semantic relatedness effect, words sharing similar semantic features in semantic long-term memory (e.g. leaf – three – branch) lead to higher recall performance as compared to lists composed of unrelated words (e.g. neck – brush – floor) (Kowialiewski & Majerus, 2018b; Monnier & Bonthoux, 2011; Poirier & Saint-Aubin, 1995; Tse, 2009; Tse et al., 2011). In the word imageability effect (or word concreteness effect), words with more concrete and less context-dependent semantic representations (e.g. leaf – hand – bear) lead to better recall performance than abstract words (e.g. phase – doubt – magic) (Acheson et al., 2010; Campoy et al., 2015; Castellà & Campoy, 2018; Chubala et al., 2018; Kowialiewski & Majerus, 2018b; L. M. Miller & Roodenrys, 2009; Romani et al., 2008; Walker & Hulme, 1999). These two effects reflect the intervention of two different types of semantic knowledge on verbal WM. While the semantic relatedness effect reflects the intervention of inter-item associative knowledge (e.g. "cat" and "dog" share more semantic features than do "cat" and "rope", see Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), the imageability effect reflects the richness and consistency of semantic features associated to each individual item (i.e. high imageability words are associated with more, context-independent semantic features than low imageability words, see Pexman, Lupker, & Hino, 2002; Yap, Lim, & Pexman, 2015).

The intervention of semantic levels of representation on WM performance is also supported by other lines of evidence. Several studies have shown that so-called phonological WM effects, such as the phonological similarity and the word length effects, appear to be reduced when participants use a semantic coding strategy instead of a phonological one (Campoy & Baddeley, 2008; Logie et al., 1996). Furthermore, when participants are explicitly instructed to use these semantic coding strategies (Hanley & Bakopoulou, 2003), or when required to perform a semantic judgement over to-be-remembered items at the moment of encoding (N. Savill et al., 2015), WM performance is increased. Similarly, newly acquired phonological word forms are better recalled in a subsequent WM task if these forms are also associated with semantic knowledge (N. Savill et al., 2016). At a theoretical level, semantic effects are explained by language-based models of WM assuming that items associated with richer semantic representations receive stronger feedback stabilizing activation during encoding and maintenance, thereby increasing subsequent recall performance (Acheson & MacDonald, 2009; Jefferies et al., 2006a; N. Martin et al., 1996; Patterson et al., 1994). The aim of this study was to go one step further and to examine to what extent semantic knowledge not only supports WM performance, but can also have a

more active effect, by protecting memoranda against interference when items have to be maintained while a second interfering task has to be carried out.

A few of studies have examined whether semantic effects can be observed in dual-task conditions, but without making a direct comparison between interfering and non-interfering conditions. For example, Campoy et al. (2015) showed word imageability effects in dual-task WM paradigms. Other studies showed that semantic effects such as the semantic relatedness and word imageability can still be observed under articulatory suppression, by considering that articulatory suppression acts like a secondary, interfering task during a WM paradigm (Acheson et al., 2010; Poirier & Saint-Aubin, 1995; Romani et al., 2008; Saint-Aubin et al., 2005; Saint-Aubin & Poirier, 1999a).

In the present study, we determined the extent to which semantic knowledge can counteract the deleterious effect of interference on WM content. We used an immediate serial recall paradigm in which the potential impact of semantic knowledge was varied at the inter-item or individual item level (semantic relatedness in Experiments 1a & 1b; imageability in Experiments 2a & 2b). In the no interference condition, participants were instructed to directly recall the words in the correct serial order after the encoding phase. In the interference condition, participants were required to perform a backward counting task after the encoding phase, until being probed for serial recall of the encoding list. This task was used as the continuous calculation and updating of numerical information fully captures attentional resources and the verbal production of the numbers prevents rehearsal of memoranda (Barrouillet & Camos, 2007), leading to interfering effects at the level of attentional control. Furthermore, the verbal labels of the numbers will interfere with the verbal representations of the memoranda. We predicted that if semantic knowledge provide a protective effect against interference, lists composed of semantically related/high imageability words should be less affected by the interfering task, as compared to lists composed of unrelated/low imageability words.

Experiment 1a

In Experiment 1a, we assessed the impact of semantic relatedness on immediate serial recall performance with and without interference. More specifically, participants were required to attend to lists of words that were semantically related (e.g. leaf – tree – branch – cloud – sky – rain) or unrelated (e.g. knife – rice – dog – chair – wall – lake). After the encoding stage, participants were invited to either directly recall the words in correct serial order (immediate recall condition) or to perform a backward counting task (interference condition) before being invited to recall the items. In the backward

counting task, a number was displayed on the screen (e.g. 35) and the participants were invited to count backwards by steps of 3 (i.e., 35, 32, 29, 26…) until being probed for recall of the memory list.

Method

Participants. Twenty-nine undergraduate students (20 females and 9 males) aged between 18 and 31 years (M = 21.14, SD = 2.45), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had received approval from the local ethics committee.

Materials. The material consisted of a set of 180 word stimuli, drawn from a pool used in a previous study (see Kowialiewski & Majerus, 2018). All stimuli were monosyllable words with a mean frequency of M_{log} = 1.536 and SD_{log} = .695 (Lexique 3 database; New, 2006). They formed 60 different triplets of semantically related words. The semantic relationships included taxonomic (e.g., "gold – iron – lead") or thematic (e.g., "tree – branch – trunk") categories. Semantic relatedness had been determined via an online survey in which participants were invited to judge the semantic similarity of words by resampling them in pairs of items, on a scale ranging from 0 (not related at all) to 5 (strongly related) (for more details, see Kowialiewski & Majerus, 2018). In average, each pair had been judged 75.25 times. The word triplets used in this study all had an average semantic relatedness rating of at least 3.5. All the stimuli were recorded by a French native male speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The length of each item was normalized to 375 ms without altering the pitch.

To create semantically related memory lists, two triplets were pseudorandomly sampled from the whole set of triplets for each list, by ensuring that two triplets from the same list could not belong to the same semantic category (e.g. leaf – tree – branch – cloud – sky – rain). The unrelated lists were created by randomly combining the words taken from the same stimulus set. Each sequence was checked to ensure that each word within a list could not have a specific relationship with another word in the list. Each word appeared twice throughout the entire experiment; once in a related list, and once in an unrelated list, and each time in a different serial position.

There were sixty trials in total, fifteen in each experimental condition. We generated 16 different versions of memory lists in order to neutralize any possible stimulus list effects. We first generated 4 different versions by using the rules described above. For each version, a new version was then created by re-ordering the

items within each list in a random order, resulting in 8 different versions. For related lists, this was done by re-ordering the items within each semantic triplet (e.g. leaf – tree – branch; branch – leaf – tree). Order of presentation of lists was pseudorandomized with no more than three successive presentations of the same stimulus condition. Furthermore, the succession of interference and no-interference trials was pseudorandomized following the same constraints. Lists presented in interference versus no-interference conditions were in addition randomized between participants.

Procedure. Each trial began with an on-screen countdown starting from 3, announcing the beginning of the trial. This countdown was followed by a blank screen and the presentation of the memory list. Each of the six items was presented at a pace of 750 milliseconds per item, with each item lasting for 375 milliseconds and an inter-stimulus interval of 375 milliseconds. In the no-interference condition, the end of each list was signaled by a brief (100 milliseconds) sinusoidal tone of 440 Hz, prompting the subject to recall aloud the sequence. Participants were instructed to recall the sequence in the order in which each word appeared, and to substitute any item they could not remember with the word "blanc" ("blank" in French). In the interference condition, the end of the list was followed by a number displayed on the screen. This number was a two digit number randomly chosen on each trial from a uniform distribution, ranging from 19 to 99. Participants were required to subtract three to this number, and then to subtract three to the result of this subtraction, and so on during 5 seconds until a tone prompted the participants to recall aloud the encoding sequence. Participants were required to say aloud the numbers which were recorded for subsequent scoring and analysis. When participants had finished recalling all items they could remember, they were invited to press the spacebar of the keyboard to initiate the next trial. Participants performed 3 practice trials of each interference condition before the beginning of the main experiment. None of the stimuli used in the practice phase were used in the main experiment. The experiment was divided into 2 blocks composed of 30 trials each to allow participants to take a very short break and refocus on the task if they needed to.

Task presentation and timing was controlled using OpenSesame (Mathôt et al., 2012) running on a desktop computer. The auditory stimuli were presented at comfortable listening level via headphones connected to the computer in a soundproof booth. Participant's responses were recorded using a digital recorder for later transcription and scoring.

Scoring procedure. Two scores were computed. The item recall score included all items recalled, regardless of serial position. The order recall score reflected the

number of items recalled in correct serial position divided by the number of items recalled, irrespective of their serial position (Murdock, 1976). Finally, we also assessed performance on the backward counting tasks by counting the number of digit output by participants before recalling the memoranda, regardless whether the response was correct or not. Note that considering only correct responses did not changed the results across all experiments.

Statistical analysis. We performed a Bayesian analysis as it reduces Type-1 false error probabilities relative to frequentist statistics (Schönbrodt et al., 2017). The Bayesian approach has the further advantage of computing continuous values against or in favour of a given model, rather than deciding for the presence of an effect based on an arbitrary statistical threshold. Evidence in favour of a model is given by the Bayesian Factor (BF), reflecting the likelihood ratio of a given model relative to other models, including the null model. Both the null model and the effect of interest can be simultaneously tested, by directly comparing the alternative hypothesis against the null hypothesis, and vice versa. The BF_{10} is used to determine the likelihood ratio for the alternative model (H_1) relative to the null model (H_0), and the BF_{01} to determine the likelihood ratio for H_0 relative to H_1. We use the classification of strength of evidence proposed in previous studies (Jeffreys, 1998; Wagenmakers et al., 2011): A BF of 1 provides no evidence, $1 < BF < 3$ provides anecdotal evidence, $3 < BF < 10$ provides moderate evidence, $10 < BF < 30$ provides strong evidence, $30 < BF < 100$ provides very strong evidence and $100 < BF$ provides extreme/decisive evidence. We also report $BF_{inclusion}$ values which compare the evidence of all models including a given factor relative to all other models not including this factor. All the analyses were performed using JASP (JASP Team, 2017) and we used default Cauchy prior distribution parameters as implemented in JASP (Version 0.8.5.1). Finally, credible intervals were computed using the 95% highest density interval (https://cran.r-project.org/web/packages/HDInterval/index.html) on the sampled posterior distribution ($N_{iterations} = 100,000$) using the BayesFactor package (Morey & Rouder, 2014) implemented in R (R Development Core Team, 2008).

Results and discussion

We first assessed recall performance as a function of interference condition (no interference, interference), semantic relatedness (related, unrelated) and serial position (1 through 6). Using a Bayesian Repeated Measures ANOVA, the main effects of interference, semantic relatedness and serial position were associated with decisive evidence. As can be clearly seen in **Figure 5.1**, items were more poorly recalled in the interference condition, as compared to the no-interference condition. In addition, recall
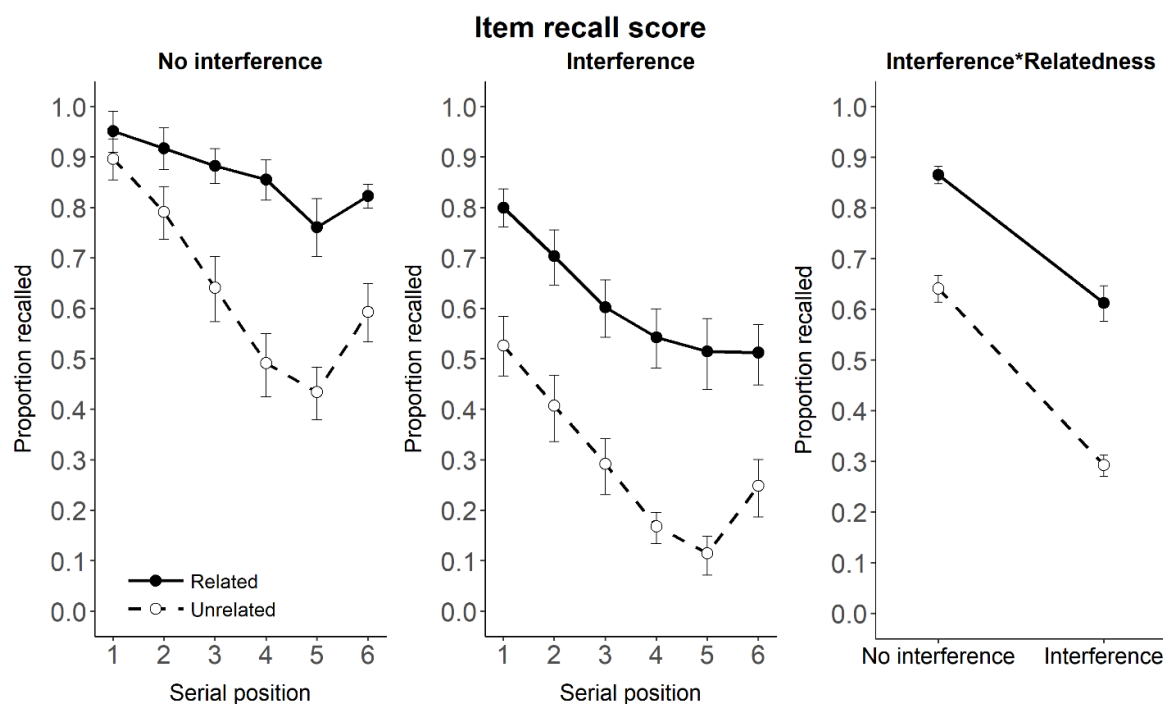
Figure 5.1. Proportion of items correctly recalled (y axis) as a function of serial position (x axis) for Experiment 1a (semantic relatedness effect, two semantic groups) using an item recall score. Error bars represent 95% credible intervals, after correction for between-subject variability (Cousineau, 2005; Morey, 2008).

performance was higher for related lists, as compared to unrelated lists. Importantly, the interaction between interference and semantic relatedness was supported by decisive evidence. This interaction was characterized by the fact that related lists were less strongly affected by the interfering task as compared to unrelated lists (see **Figure 5.1**, right panels). See **Table 5.1** for a full report of BF values.

Next, we assessed order recall performance as a function of interference and semantic relatedness using a Bayesian Repeated Measures ANOVA. We found decisive evidence supporting the main effect of interference and strong evidence supporting the main effect of semantic relatedness. The interaction provided ambiguous evidence. Order recall performance decreased when participants had to perform the interfering task. Likewise, related lists led to better order recall performance as compared to unrelated lists. Descriptive statistics of this analysis are available in **Table 5.2**.

Table 5.1. *Statistical values for Bayesian analyses in Experiment 1a.*

Effects	BF$_{inclusion}$
Item recall performance	
Interference	$\rightarrow +\infty$
Semantic relatedness	$\rightarrow +\infty$
Serial position	$\rightarrow +\infty$
Interference * Semantic relatedness	6579.012
Interference * Serial position	.128
Semantic relatedness * Serial position	1.527e+7
Interference * Semantic relatedness * Serial position	.579
Order recall performance	
Interference	1.585e+11
Semantic relatedness	12.472
Interference * Semantic relatedness	1.354

Table 5.2. *Order recall performance for Experiment 1a.*

Semantic relatedness	Interference condition	
	No interference	Interference
Related	.85 [.819; .882]	.696 [.669; .724]
Unrelated	.807 [.783; .831]	.62 [.571; .666]

Note. The values represent scores averaged across participants. The lower and upper 95% credible intervals are marked in square brackets.

We also assessed whether the semantic condition had an impact on the interfering task. We compared participant's performance on the backward counting task as a function of the type of words to be maintained (related, unrelated) using a Bayesian Paired Samples T-Test. We observed no impact of the semantic condition on the participants ability to perform the backward counting task (M_{diff} = -.011, CI$_{95\%}$ = [-.141; .121], BF$_{01}$ = 5.002).

In sum, the interfering task decreased recall performance, and this less strongly for semantically related lists than semantically unrelated lists. It could however be argued that the semantic relatedness effect was enhanced due to the fact that words were in addition presented in two three-item groups of semantically related words; grouping effects have been shown to improve recall performance overall (Henson, 1996; Hitch et al., 1996). Hence, the protective effect of semantic relatedness may at least partially reflect the intervention of grouping effects. In Experiment 1b, we aimed at replicating the results of Experiment 1a while avoiding grouping effects.

Experiment 1b

In Experiment 1b, we manipulated semantic relatedness in immediate serial recall tasks with or without an interfering task by presenting lists of six words that were either all related at the semantic level (e.g. foot, leg, arm, hand, elbow, thigh) or completely unrelated. This manipulation allowed us to avoid sublist semantic grouping effects while varying the impact of associative semantic knowledge.

Method

Participants. Twenty-nine undergraduate students (21 females and 8 males) aged between 18 and 27 years (M = 21.9, SD = 2.91), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had received approval from the local ethics committee.

Materials. We used a set of 192 words (one to three syllables long) with a mean frequency of M_{log} = 1.164 and SD_{log} = .732 (Lexique 3 database; New, 2006). The stimuli were drawn from thirty-two different semantic categories, with six words per category. All the stimuli were recorded by a French native male speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The stimulus set had a duration of M = .454 and SD = .063.

Unrelated lists were created by pseudorandomly combining the words from the semantically related lists, such that each word within a list could not share an obvious semantic relationship with another word in the list. Like in Experiment 1a, each word appeared twice in different serial positions: once in a related list, and once in an unrelated list.

The entire experiment was comprised of 64 trials, with 16 trials per experimental conditions. To avoid stimulus list effects, we generated 6 different versions of the memory lists, by randomly recombining the words for unrelated lists as described above, and by re-ordering the serial position of words for related lists. The lists within each version were then assigned to recall conditions with or without an interfering task. These pairings were reversed for half of participants, resulting in 12 different list versions. Finally, for a third of participants, the order of items within a list was reversed (i.e. the first item of each lists was shifted to position 6). For another third of participants, the order of the trials was reversed (the first presented list was now the last one), leading to a total of 36 different list versions.

The experiment was divided into 2 blocks composed of 32 trials each to allow participants to take a very short break and refocus on the task if they needed to. All
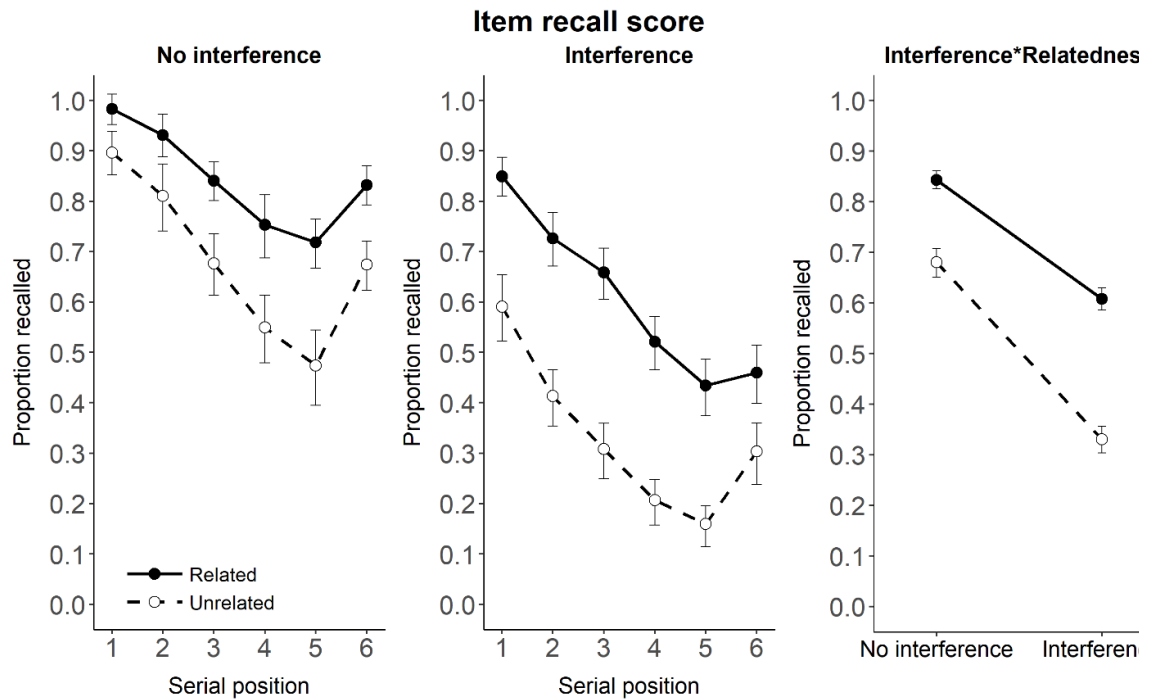
Item recall score

Figure 5.2. Proportion of items correctly recalled (y axis) as a function of serial position (x axis) for Experiment 1b (semantic relatedness effect, one semantic group) using an item recall score. Error bars represent 95% credible intervals, after correction for between-subject variability (Cousineau, 2005; Morey, 2008).

other aspects of the method, including the task procedure (presentation rate, interfering task…), scoring procedure and statistical analysis were identical to Experiment 1a.

Results and discussion

Recall performance as a function of interference (no interference, interference), semantic relatedness (related, unrelated) and serial position (1 through 6) was first assessed using a Bayesian Repeated Measures ANOVA. As can be clearly seen in **Figure 5.2** and **Table 5.3**, all main effects of interference, semantic relatedness and serial position were supported by decisive evidence. Again, the interfering task reduced recall performance, and related lists were better recalled as compared to unrelated lists. Importantly, the interaction between interference and semantic relatedness variables was associated with decisive evidence: the impact of the interfering task was stronger in the unrelated as compared to the related condition.

When running the same analyses on the order score, there was decisive to strong evidence for the interference and semantic relatedness effects. Critically, contrary to Experiment 1a, the semantic relatedness variable had a negative impact on order recall performance, with more serial position confusions for semantically related word lists as can be seen in **Table 5.4**. The interference by semantic relatedness interaction term

Table 5.3. *Statistical values for Bayesian analyses in Experiment 1b.* was

Effects	$BF_{inclusion}$
Item recall performance	
Interference	$\rightarrow +\infty$
Semantic relatedness	$\rightarrow +\infty$
Serial position	$\rightarrow +\infty$
Interference * Semantic relatedness	311051.165
Interference * Serial position	15.126
Semantic relatedness * Serial position	7.339
Interference * Semantic relatedness * Serial position	11.576
Order recall performance	
Interference	3.002e+14
Semantic relatedness	49.8
Interference * Semantic relatedness	6.174

Table 5.4. *Order recall performance for Experiment 1b.*

Semantic relatedness	Interference condition	
	No interference	Interference
Related	.777 [.746; .808]	.556 [.527; .584]
Unrelated	.801 [.778; .825]	.652 [.611; .689]

Note. The values represent scores averaged across participants. The lower and upper 95% credible intervals are marked in square brackets.

supported by moderate evidence, with ambiguous evidence for semantic relatedness effect in the no interference condition ($BF_{01} = 1.754$), and very strong evidence in the interference condition ($BF_{10} = 50.852$).

Finally, we assessed the participants performance on the backward counting task as a function of semantic relatedness using a Bayesian Paired Samples T-Test, and found again evidence for the absence of difference ($M_{diff} = -.011$, $CI_{95\%} = [-.11; .09]$, $BF_{01} = 4.969$).

Experiment 1b replicated the main results of Experiment 1a: semantic knowledge, as assessed by inter-item relatedness effects, had a protective impact on item WM recall performance. Interestingly, this pattern of result was reversed as regards order recall performance: related lists were more strongly affected by the interfering task than were unrelated lists. One possible explanation is that maintenance of serial order processing might be more difficult for related words due to the closeness of their semantic representations. The resulting similarity of items will also increase the similarity of item-serial order mappings, serial order codes for the different serial

positions pointing to partially overlapping loci of the semantic space (by assuming that serial order is coded by codes that are distinct from item representations). The temporary maintenance of overlapping item-serial order mappings will require increased attentional control in order to keep the different item-serial order mappings for a given list separable. Once attentional control is monopolized by by the interfering task, representations of item-serial order mappings might get blurred very quickly. Guérard and Saint-Aubin (2012) observed a similar detrimental impact of semantic relatedness on order recall performance in an attentional control demanding backward recall paradigm, but not when using a less demanding standard immediate serial recall paradigm.

Experiment 2a

Experiments 1a and 1b have shown that semantic knowledge, in the form of inter-item associative knowledge, can protect memoranda against task-based interference, and this particularly for item recall. The aim of Experiment 2a was to determine whether the same protective effects can be observed when semantic knowledge operates on individual item characteristics rather than on inter-item associative knowledge. This question was examined by varying the imageability dimension of memoranda, based on the assumption that high imageability words are associated with richer and more context-independent semantic features than low imageability words (Evans et al., 2012; Pexman et al., 2002; Yap et al., 2015). All other aspects of the experiment were identical to Experiment 1b.

Method

Participants. Thirty undergraduate students (20 females and 10 males) aged between 18 and 25 years ($M = 21.34$, $SD = 2$), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had received approval from the local ethics committee.

Materials. The stimuli were taken from a previous study (Kowialiewski & Majerus, 2018b) in which they produced reliable imageability effects in standard immediate serial recall conditions. The stimuli were two sets composed of ninety unisyllabic words, selected from the Lexique 3 database (New, 2006). One of the sets scored high on the imageability dimension (M = 6.661, SD = .122, min = 6.5, max = 6.9) and the other one scored low on the imageability dimension (M = 2.908, SD = .379, min = 1.9, max = 3.4) (Content et al., 1990). This difference was reliable, as shown by a

Bayesian Independent Samples T-Test ($BF_{10} = 2.907e+145$). Both sets were matched for different psycholinguistic dimensions: lexical frequency ($M_{log} = 1.04$, $SD_{log} = .449$ and $M_{log} = 1.001$, $SD_{log} = .82$ for high and low imageability words, respectively, $BF_{01} = 5.745$), subjective frequency ($M = 4.044$, $SD = .837$ and $M = 3.943$, $SD = .92$ for high and low imageability words, respectively, $BF_{01} = 4.694$) (Content et al., 1990), number of phonemes ($M = 3.522$, $SD = .691$ and $M = 3.478$, $SD = .722$ for high and low imageability words, respectively, $BF_{01} = 5.697$), neighborhood density ($M = 16.444$, $SD = 9.7$ and $M = 16.922$, $SD = 10.802$ for high and low imageability words, respectively, $BF_{01} = 5.914$) and neighborhood frequency ($M_{log} = .115$, $SD = .4$ and $M = .033$, $SD = .388$ for high and low imageability words, respectively, $BF01 = 2.472$). The neighborhood density values were obtained by computing the Levenshtein distance (Levenshtein, 1966) between the phonological form of each word relative to all other entries of the Lexique 3 database. The neighborhood frequency values were obtained by computing the ratio of the log10-transformed frequency of each word against the summed frequency of all neighbors including the target word. All the stimuli were recorded by a French native female speaker in a neutral voice. Background noise was removed via the noise reduction tool implemented in Audacity®. The original stimuli were normalized to a duration of 375 ms.

The entire experiment was comprised of 60 trials, with 15 trials in each experimental conditions. Six different list versions were created using an algorithm programmed under MATLAB®. The algorithm ensured that a word could not be repeated within the same list, or within the same serial position throughout the experiment. Furthermore, two stimulus conditions (high, low imageability words) or interference conditions (no interference, interference) could not be repeated on more than three consecutive trials. Six further list versions were created by inverting the pairings of the stimulus lists and interference conditions, resulting in 12 different versions. The lists of each of these versions were furthermore reversed as regards the serial position of items in each list (i.e. the first item of each lists was shifted to position 6) and then flipped up to down (the first presented list was now the last one), resulting in a total of 24 versions. Throughout the experiment, each word was repeated twice in different serial positions: once in the no interference condition, and once in the interference condition. Finally, we checked that high and low imageability lists were also equivalent in terms of lexical co-occurrence values, which were computed using the Latent Semantic Analysis (LSA) (http://lsa.colorado.edu/). When averaged across all versions, high and low imageability lists showed very similar LSA values ($M = .118$, $SD = .047$ and $M = .123$, $SD = .059$ for high and low imageability words, respectively).
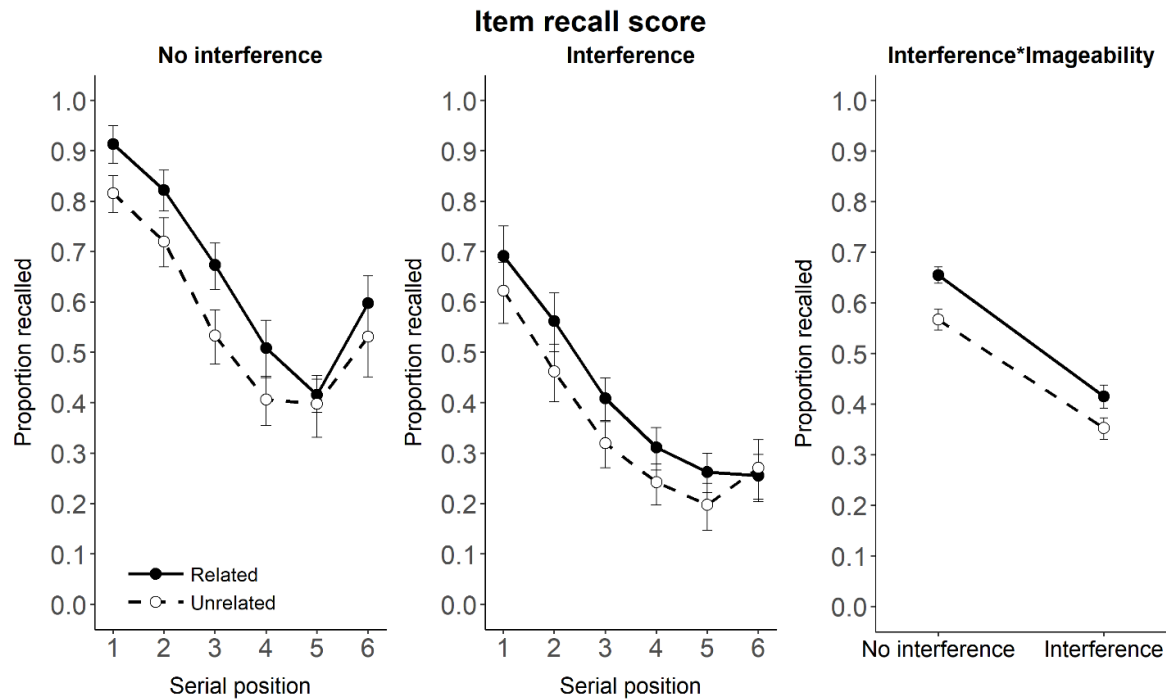
Figure 5.3. Proportion of items correctly recalled (y axis) as a function of serial position (x axis) for Experiment 2a (imageability effect, random interference) using an item recall score. Error bars represent 95% credible intervals, after correction for between-subject variability (Cousineau, 2005; Morey, 2008).

Procedure. Contrary to Experiments 1a and 1b, items within lists were presented at a rate of 1.5 seconds/items. This slower presentation rate was used because the imageability effect is maximized for slow stimulus presentation rates (Campoy et al., 2015). The experiment was divided into 2 blocks composed of 30 trials each to allow participants to take a very short break and refocus on the task if they needed to.

Results and discussion

We first assessed recall performance as a function of interference (no interference, interference), imageability (high, low imageability) and serial position (1 through 6) using a Bayesian Repeated Measures ANOVA (see **Figure 5.3** and **Table 5.5**). The main effect of interference condition was supported by decisive evidence. Recall performance strongly decreased in the interference condition. The imageability effect was also associated with decisive evidence, with high imageability words leading to higher recall performance than low imageability words. Likewise, the serial position effect was also supported by decisive evidence. Critically, there was some (but ambiguous) evidence supporting the absence of interaction between the interference condition and the imageability effect using an item recall score; high and low imageability words were similarly affected by the interfering task.

Table 5.5. *Statistical values for Bayesian analyses in Experiment 2a.*

Effects	BF$_{inclusion}$
Item recall performance	
Interference	→ +∞
Imageability	5.781e+9
Serial position	→ +∞
Interference * Imageability	.499
Interference * Serial position	13.896
Imageability * Serial position	.363
Interference * Imageability * Serial position	.019
Order recall performance	
Interference	802839.748
Imageability	.519
Interference * Imageability	1.058

Table 5.6. *Order recall performance for Experiment 2a.*

	Interference condition	
Imageability	No interference	Interference
High	.824 [.796; .851]	.709 [.677; .739]
Low	.793 [.763; .82]	.696 [.66; .728]

Note. The values represent scores averaged across participants. The lower and upper 95% credible intervals are marked in square brackets.

Next, we assessed order recall performance as a function of interference and imageability. We found decisive evidence supporting an effect of the interference condition, but ambiguous evidence as regards the imageability effect and the interaction term. As in Experiments 1a and 1b, order recall performance decreased in the interference conditions. However, contrary to Experiments 1a and 1b, order recall performance did not differ between high and low imageability words (see **Table 5.6**), in line with previous studies (Campoy et al., 2015; Castellà & Campoy, 2018; Guérard & Saint-Aubin, 2012; Tse & Altarriba, 2007).

In line with the previous experiments, we found evidence slightly favouring the absence of a stimulus condition effect on performance for the backward counting task (M_{diff} = -.064, $CI_{95\%}$ = [-.17; .054], BF_{01} = 2.945).

In this experiment, we observed that the imageability dimension had no obvious protective effect against interference; high and low imageability words were similarly affected by interference. This is in contrast with Experiments 1a and 1b where the impact of the semantic relatedness variable did interact with the interference

condition. This might suggest that different semantic dimensions do not protect against interference in the same way.

Experiment 2b

In this final experiment, we aimed at replicating the results of Experiment 2a. Furthermore, we presented the interference – no-interference conditions in blocks in order to rule out the possibility that the unpredictability of the interference conditions may have led to the negative results observed in Experiment 2a.

Method

Participants. Thirty undergraduate students (21 females and 9 males) aged between 19 and 27 years (M = 22.03, SD = 1.81), recruited from the university community, took part in the experiment. All participants were native French speakers with no history of neurological disorders or learning difficulties. All participants gave their written informed consent prior to their inclusion in the study. The study had received approval from the local ethics committee.

Materials & procedure. The material, testing and scoring procedure were identical as in Experiment 2a, except that the interference and no interference lists were presented in a fully blocked manner. The block conditions were counterbalanced across subjects, such that odd subjects received the interference condition first, while even subjects received the no interference condition first. Participants received 3 practice trials before each block.

Results and discussion

Recall performance as a function of interference, imageability and serial position were assessed using a Bayesian Repeated Measures ANOVA. As can be seen in **Figure 5.4** and **Table 5.7**, all main effects were supported by decisive evidence: recall performance dropped in the presence of task interference and high imageability words were better recalled than low imageability words. We also reproduced the main result of Experiment 2a by observing moderate evidence for the absence of an interference-by imageability interaction. Hence, both high and low imageability words were similarly affected by the interfering task.
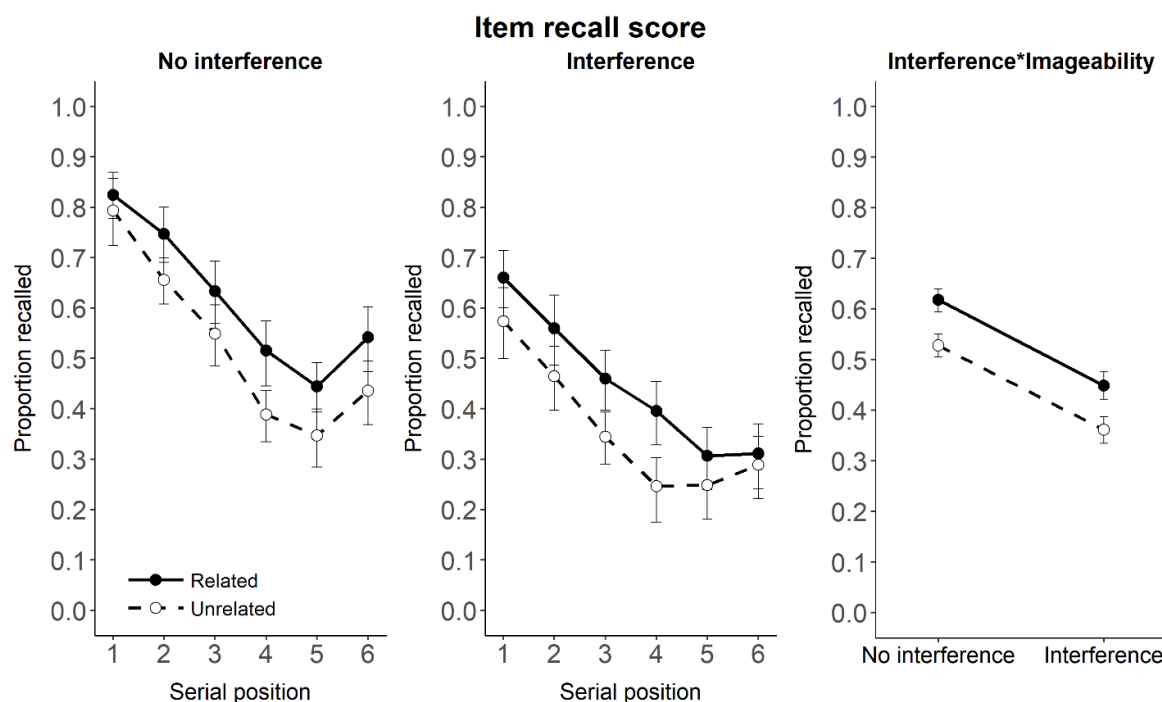
Item recall score

Figure 5.4. Proportion of items correctly recalled (y axis) as a function of serial position (x axis) for Experiment 2b (imageability effect, blocked interference) using an item recall score. Error bars represent 95% credible intervals, after correction for between-subject variability (Cousineau, 2005; Morey, 2008).

Next, we assessed order recall performance as a function of interference and imageability (see **Table 5.8** for descriptive results). We again observed decisive evidence supporting the impact of interference. The imageability effect and the interaction term were associated with ambiguous evidence.

Again, we observed evidence *against* an impact of the stimulus condition on performance during the backward counting task ($M_{diff} = -.013$, $CI_{95\%} = [-.104; .083]$, $BF_{01} = 4.97$).

This final experiment shows that the absence of an imageability-by-interference interaction is a robust finding and was not caused by the blocked versus random order of presentation of task conditions. One possibility is that the word imageability dimension protects less against interference because its impact on WM performance is smaller than the impact of the semantic relatedness dimension. We explored this possibility by directly comparing the mean difference between the two stimulus conditions of Experiments 1b and 2a (related and unrelated in the semantic relatedness study; high and low imageability in the imageability study) using a Bayesian Independent Samples T-Test, and this in the no interference condition only. As can be shown in **Figure 5.5**, the impact of the semantic relatedness manipulation was nearly twice as bigger ($M_{diff} = .163$) as compared to the imageability manipulation ($M_{diff} = .088$),

Table 5.7. *Statistical values for Bayesian analyses in Experiment 2b.*

Effects	BF_{inclusion}
Item recall performance	
Interference	$\rightarrow +\infty$
Imageability	5.436e+9
Serial position	$\rightarrow +\infty$
Interference * Imageability	.259
Interference * Serial position	.107
Imageability * Serial position	.055
Interference * Imageability * Serial position	9.191e-5
Order recall performance	
Interference	6250.14
Imageability	1.054
Interference * Imageability	1.726

Table 5.8. *Order recall performance for Experiment 2b.*

	Interference condition	
Imageability	No interference	Interference
High	.808 [.78; .835]	.75 [.718; .78]
Low	.807 [.777; .835]	.688 [.651; 723]

Note. The values represent scores averaged across participants. The lower and upper 95% credible intervals are marked in square brackets.

and this difference was supported by decisive evidence (BF_{10} = 162.978). We took one step further by assessing the impact of the semantic variables across multiple studies reported in the literature, by performing a random effect meta-analysis, and this using the "metafor" package (Viechtbauer, 2010) implemented in the statistical software R (R Development Core Team, 2008). The detailed statistical procedure and studies inclusion criteria are reported in **Appendix A** and **B**, respectively. The results from **Figures 5.6** and **5.7** largely confirm the observations drawn from the present study: the semantic relatedness effect was twice as larger (Cohen's d = 2.08) as compared to the imageability effect (Cohen's d = .93), without overlap between confidence intervals. Hence, overall, the semantic relatedness dimension seems to provide a stronger boost for WM performance than the imageability/concreteness dimension does. In the next section, we discuss the implications of these results.
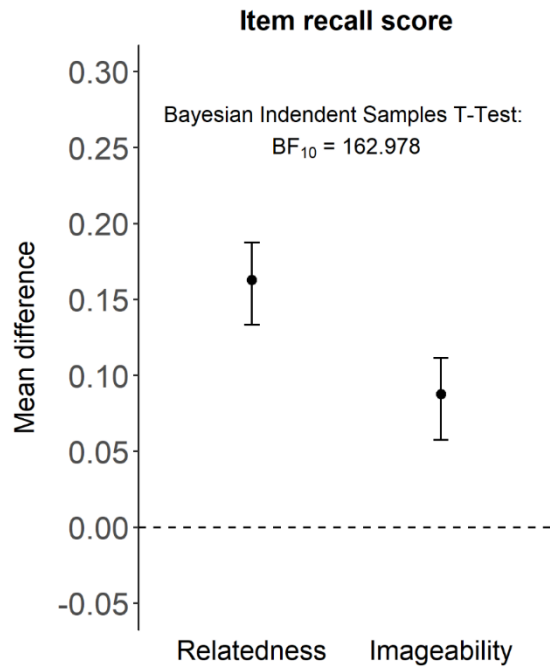
Figure 5.5. Mean difference (y axis) between stimulus conditions for the semantic relatedness effect (Experiment 1b) and the imageability effect (Experiment 1a). Error bars represent 95% credible intervals of the mean.
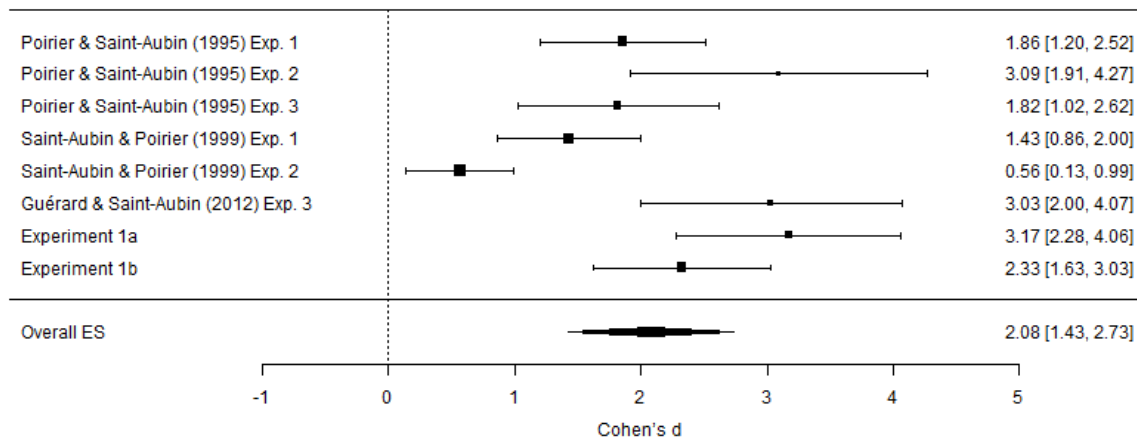

Figure 5.6. Forest plot of the meta-analysis performed on studies assessing the semantic relatedness effect using an item recall score. Each individual point and error bars represent Cohen's d and the corresponding 95% confidence interval.
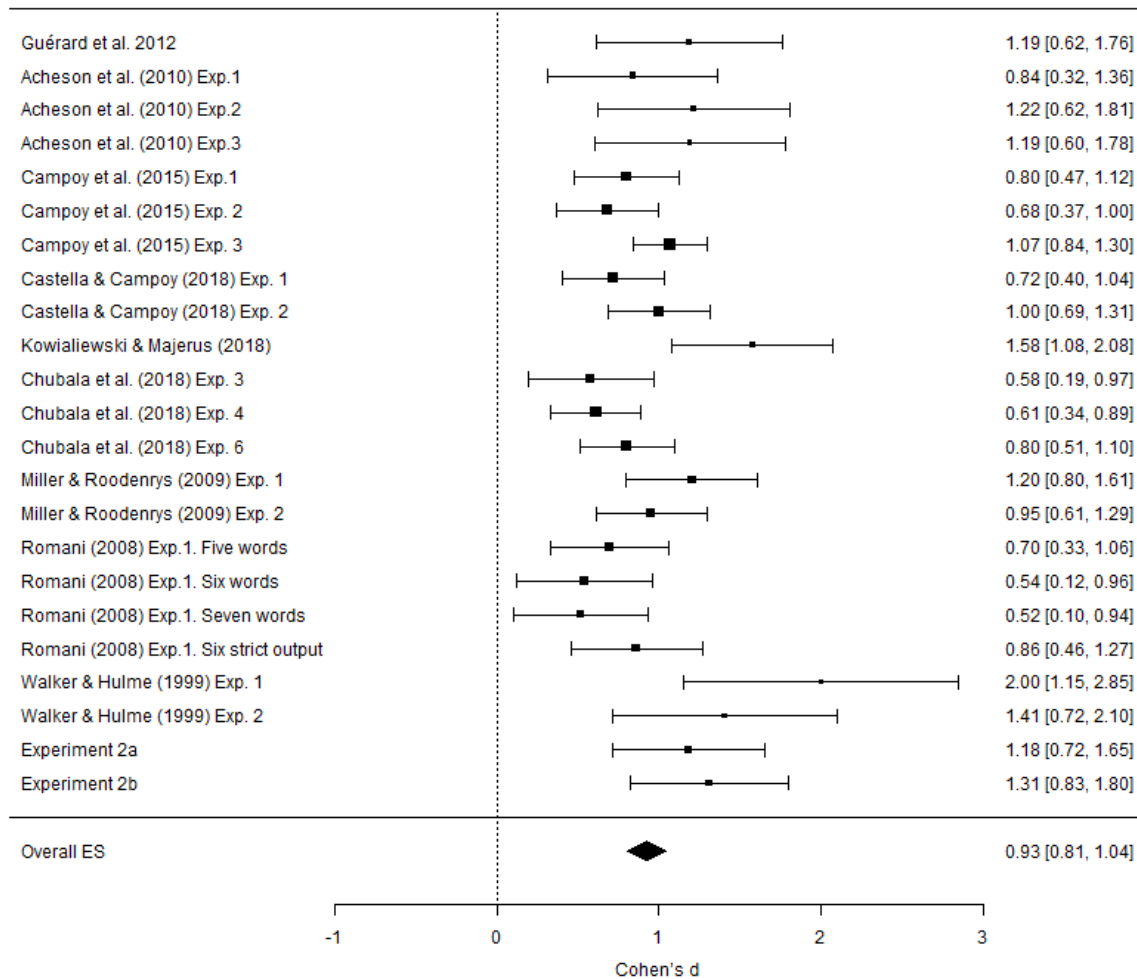
Guérard et al. 2012	1.19 [0.62, 1.76]
Acheson et al. (2010) Exp.1	0.84 [0.32, 1.36]
Acheson et al. (2010) Exp.2	1.22 [0.62, 1.81]
Acheson et al. (2010) Exp.3	1.19 [0.60, 1.78]
Campoy et al. (2015) Exp.1	0.80 [0.47, 1.12]
Campoy et al. (2015) Exp. 2	0.68 [0.37, 1.00]
Campoy et al. (2015) Exp. 3	1.07 [0.84, 1.30]
Castella & Campoy (2018) Exp. 1	0.72 [0.40, 1.04]
Castella & Campoy (2018) Exp. 2	1.00 [0.69, 1.31]
Kowialiewski & Majerus (2018)	1.58 [1.08, 2.08]
Chubala et al. (2018) Exp. 3	0.58 [0.19, 0.97]
Chubala et al. (2018) Exp. 4	0.61 [0.34, 0.89]
Chubala et al. (2018) Exp. 6	0.80 [0.51, 1.10]
Miller & Roodenrys (2009) Exp. 1	1.20 [0.80, 1.61]
Miller & Roodenrys (2009) Exp. 2	0.95 [0.61, 1.29]
Romani (2008) Exp.1. Five words	0.70 [0.33, 1.06]
Romani (2008) Exp.1. Six words	0.54 [0.12, 0.96]
Romani (2008) Exp.1. Seven words	0.52 [0.10, 0.94]
Romani (2008) Exp.1. Six strict output	0.86 [0.46, 1.27]
Walker & Hulme (1999) Exp. 1	2.00 [1.15, 2.85]
Walker & Hulme (1999) Exp. 2	1.41 [0.72, 2.10]
Experiment 2a	1.18 [0.72, 1.65]
Experiment 2b	1.31 [0.83, 1.80]
Overall ES	0.93 [0.81, 1.04]

Figure 5.7. Forest plot of the meta-analysis performed on studies assessing the imageability effect using an item recall score. Each individual point and error bars represent Cohen's d and the corresponding 95% confidence interval.

General discussion

In this study, we investigated the extent to which semantic knowledge can protect memoranda against task interference in WM tasks. We observed in two experiments a protective effect of semantic relatedness against interference caused by a secondary task presented between the WM encoding and recall phases: task-based interference had a lesser detrimental effect on semantically related versus unrelated word lists. However, in two further experiments assessing the same question by manipulating word imageability rather than semantic relatedness, we observed no differential impact of task interference: the secondary task reduced recall performance for high and low imageability word lists to the same extent.

First, we will discuss the potential mechanisms that explain the strong impact of semantic relatedness on WM recall performance, and the resulting protective effect

against task-based interference. At the psycholinguistic level, the semantic relatedness effect can be explained by assuming that similar items share more features at the level of semantic representations in the language network (Dell et al., 1997). When multiple words sharing similar semantic features are presented, they will reactivate each other, leading to overall high and robust activation levels for all items during encoding in a WM task, relative to semantically unrelated words. These redundant semantic representations will also protect the items against degradation by decay (Barrouillet et al., 2004; Barrouillet & Camos, 2007; Camos & Barrouillet, 2014) or interference (Oberauer et al., 2016, 2012, Oberauer & Lewandowsky, 2011, 2013). Furthermore, at recall after an interfering secondary task, causing massive degradation of memoranda, due to the redundant semantic features, the availability of a small number of semantic features of the original memoranda is sufficient to re-activate or reconstruct (see also Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999; Schweickert, 1993) most of the items, which, in turn, will re-activate each other; this process will be much less efficient for semantically unrelated word lists as in these lists the few semantic features still available for a given item will provide no information about the semantic features associated with the other items. Furthermore, given that items in the semantically related lists share an important number of semantic features, they can be grouped in a smaller number of semantic or conceptual categories. This will allow compression of information in WM (Chekaf, Cowan, & Mathy, 2016; Mathy, Chekaf, & Cowan, 2018; Thalmann, Souza, & Oberauer, 2018) by activating a small number of supra-item contextual representations (Atkins & Reuter-lorenz, 2008) reducing overall WM load. For instance, given the target sequence "leaf – tree – branch – cloud – sky – rain", participants could activate a broader semantic and/or conceptual representation, such as "nature" and "weather", and rely on these supra-item representations as a cue to retrieve the items at the moment of recall. This phenomenon may also explain why the amount of serial order errors actually increased when recalling semantically related lists after an interfering secondary task: due to interference, most of the detailed item-level information, including serial position information, will be strongly degraded at the moment of recall and items are recalled mostly by decompressing, directly within the semantic knowledge space, the few supra-list chunks that could be maintained until the moment of recall.

The situation will be very different for the word imageability effect. Concrete or high imageability words are supposed to be better recalled as compared to abstract words because they are represented by a larger number of semantic features at the individual item level, leading to more robust item-level representations for highly imageable words (Pexman et al., 2002; Yap et al., 2015). However, the richer individual

item information will not be informative about the other items (when high imageability items are not also semantically related) and hence no between-item reactivation nor any list compression processes can occur. In other words, the semantic relatedness and the word imageability share the situation of more robust item level representations at encoding, but, unlike word imageability, semantic relatedness is in addition characterized by the opportunity for inter-item reactivation and list compression effects. The latter effects may be particularly powerful when individual item-level information is strongly degraded, explaining why semantic relatedness leads to a protective effect of WM performance when a secondary interfering task has to be carried out. These additional sources of long-term memory support for semantically related words as opposed to high imageability words is also likely to explain the overall more robust effect of semantic relatedness on WM performance.

Conclusions

In this study, we investigated the extent to which semantic knowledge not only supports the maintenance of memoranda in a WM task, but provides additional protection against interference when a secondary task is introduced after the WM encoding phase. We found that only semantic relatedness provided this protective effect, presumably due to inter-item reactivation and conceptual list compression effects that allow heavily degraded information to be reinstantiated at the moment of recall. Richer item-level semantic knowledge only, as involved in the word imageability effect, does not have a protective effect of WM recall after interference. This study highlights the complexity and multiplicity of linguistic long-term memory variables that determine WM performance.

Appendix A

To compute Cohen's d for repeated measures in each study, we transformed F-statistics into T-values. and then divided the T-value by the square root of the sample size (Lakens, 2013). The variance of Cohen's d was computed using this formula:

$$Var[d] = \frac{1}{n} + \frac{d^2}{2n}$$

where n is the sample size and d^2 is Cohen's d raised to the power of 2 (see also Hedges, 1981).

Appendix B

We included in the meta-analysis only experiments that used a repeated-measures design and an immediate serial recall or similar procedure in healthy adult participants. In cases where the authors combined the psycholinguistic effect manipulation with another variable (e.g., forward and backward recall, see Guérard & Saint-Aubin, 2012; or phonological overlap in Acheson et al., 2010), results of the standard condition (i.e. forward recall) were included when available. In addition, we report here only results based on the item recall score. We also included the no-interference conditions of the experiments of the present study. This led to a total sample of 8 studies for the semantic relatedness effect and to 23 studies for the imageability effect.

Study 6

Semantic knowledge constrains the processing of serial order information in working memory

Benjamin Kowialiewski, Simon Gorin, and Steve Majerus

In preparation

Abstract. Long-term memory knowledge is generally considered to impact short-term maintenance of item information in Working Memory (WM) tasks, as opposed to short-term maintenance of serial order information. In the present study, we assess the extent to which long-term memory knowledge can also influence recall of serial order information Across two experiments, we manipulated the semantic relatedness effect by using semantic categories presented in subgroups of items (e.g. leaf – tree – branch – cloud – sky – rain, Experiment 1), or in an interleaved fashion (leaf – cloud – tree – sky – branch – rain, Experiment 2). Our results showed that semantic knowledge constrained the occurrence of serial order errors: when migrating to a non-target serial position, items tended to do so most of the time toward a semantically related item. However, this was observed only when semantically related items were presented in subgroups. A direct comparison with temporal grouping effects suggests that these results may be partly independent from general grouping effects on serial order coding. These results may provide support WM models that consider direct interactions between serial order and linguistic components of WM.

Introduction

Verbal Working Memory (WM), the ability to store verbal information over a short period of time, is thought to emerge from the interaction between different sub-systems involved in the representation of item identity, on the one hand, and the representation of serial order information, on the other hand (Brown et al., 2000; Henson et al., 2003; Hitch et al., 1996; Majerus, 2009, 2013, 2019). This distinction is supported by behavioural (Gorin et al., 2016; Henson et al., 2003), neuropsychological and neuroimaging evidence (Kalm & Norris, 2014; Majerus et al., 2010, 2015; Papagno

et al., 2017). At the same time, there is also some emerging evidence for an interaction between components involved in the representation of item and serial order information (Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011). In this study, we examined the extent to which semantic knowledge, generally thought to support the representation of item information in WM, may also interact with the representation of serial order information.

The mechanisms responsible for the maintenance of serial order information (the temporal order in which memoranda are presented) are generally considered to be independent from those involved in the maintenance of item information (the orthographic, phonological and lexico-semantic characteristics of the memoranda). Behavioural studies have shown that linguistic long-term memory knowledge impacts the recall of item information, while minimally affecting recall of serial order information (Campoy, Castellà, Provencio, Hitch, & Baddeley, 2014; Hulme et al., 1997; Hulme, Stuart, Brown, & Morin, 2003; Poirier & Saint-Aubin, 2005; Romani, McAlpine, & Martin, 2008; Walker & Hulme, 1999; but see Allen & Hulme, 2006; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Tse & Altarriba, 2007). In addition, the maintenance of serial order information appears to be more sensitive to rhythmic and articulatory interfering tasks than is the maintenance of item information (Gorin et al., 2016; Henson et al., 2003). Neuropsychological studies have also revealed double dissociations between item and serial order retention capacities in brain-injured and neurodevelopmental populations (Attout et al., 2014; Brock & Jarrold, 2005; Majerus et al., 2015; Martinez Perez et al., 2015). Finally, neuroimaging and neurostimulation studies have shown the involvement of neural regions specifically sensitive to the maintenance of serial order versus item information (Attout et al., 2014; Kalm & Norris, 2014; Majerus et al., 2010; Papagno et al., 2017). The majority of current models of WM consider the existence of specific codes responsible for the maintenance of serial order, independent from those involved in the maintenance of item identity, with the nature of these codes ranging from relational, episodic, temporal to spatial codes (Abrahamse, Dijck, Majerus, & Fias, 2014; Brown et al., 2000; Burgess & Hitch, 1999, 2006; Hartley et al., 2016; Henson, 1996; Lewandowsky & Murdock, 1989).

Maintenance of verbal item information, on the other side, is considered to depend on interactions with linguistic knowledge stored in long-term memory (Acheson & MacDonald, 2009; Hulme et al., 1991; Majerus, 2019; N. Martin et al., 1996; Schweickert, 1993). This is shown by the presence of a large number of psycholinguistic effects impacting the recall of item information, involving the different levels (phonological, lexical and semantic) of language processing. At the sublexical, phonological level, it has been shown that verbal WM performance

increases for nonwords containing high versus low phonotactic probability structures (Gathercole et al., 1999; Majerus, Martinez, et al., 2012; Majerus, Van der Linden, Mulder, et al., 2004). At the lexical level, WM performance is higher for words as compared to nonwords (Brener, 1940; Jefferies et al., 2006a), as well for high versus low frequency words (Hulme et al., 1991). The same effects are also observed at the semantic level, with concrete words being better recalled as compared to abstract words (Bourassa & Besner, 1994; Campoy et al., 2015; Kowialiewski & Majerus, 2018b; Walker & Hulme, 1999), and with words related at the semantic level (e.g. leaf – tree – branch) being better recalled as compared to unrelated words (e.g. hand – table – goat). Language-based models of WM (Acheson & MacDonald, 2009; Majerus, 2013, 2019; N. Martin et al., 1996) interpret these results as reflecting the direct activation occurring within the linguistic system; items associated with richer or more robust representations will receive stronger feedback stabilizing activation, leading to increased recall performance.

However, recent evidence suggests that the representation of serial order information in WM may not be completely independent from linguistic knowledge involved in the representation of item information. While lexical knowledge has been shown to support item recall, with a robust recall advantage for word lists as compared to nonword lists, a few studies have shown that the availability of lexical knowledge may actually lead to impaired serial order recall with a higher rate of serial order recall errors for word lists than nonword lists (Fallon et al., 2005; Guérard & Saint-Aubin, 2012; Kowialiewski & Majerus, 2018b; Saint-Aubin & Poirier, 1999b). Similarly, lists composed of words sharing semantic features (e.g. leaf – tree – branch) show superior overall recall performance at the item level but decreased performance at the serial order level in WM tasks as compared to semantically unrelated words (Poirier & Saint-Aubin, 1995; Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011).

These studies suggest that serial order representations, even if independent from linguistic knowledge considered to support maintenance of item information, nevertheless interacts with linguistic knowledge. These interactions are implicitly or explicitly assumed by WM models such as the models presented by Burgess and Hitch (1999, 2006) and Majerus (2013; 2018) but they have not yet been studied in a systematic and direct manner. The model of Burgess and Hitch considers a direct connection between contextual nodes representing serial order information and lexical item nodes while the framework of Majerus (2013, 2018) considers a bidirectional connection between the serial order processing component and phonological and lexico-semantic knowledge bases. The present study will focus on the nature of interactions that may

occur between semantic knowledge and serial order recall in immediate serial recall tasks.

According to the trace confusion hypothesis (Saint-Aubin & Poirier, 1999a), words sharing the same semantic knowledge may be more subject to serial recall errors as compared to semantically unrelated items because their shared semantic features provide cues that will facilitate the recall of item information but the similarity of these cues will at the same time lead to confusions at the serial order level. The trace confusion hypothesis considers that at the moment of recall, degraded items will be reconstructed/reactivated based on the few, shared semantic features that have resisted to WM trace degradation; this will lead to a higher proportion of serial order errors as more semantically related items than semantically unrelated items can be reconstructed via this process. As a consequence, more semantically related than unrelated items are available for recall, but without any possibility for the reconstruction of initial serial position information. Second, according to a semantic chunking hypothesis, at the moment of encoding, semantically similar items will be compressed (Chekaf et al., 2016; Mathy et al., 2018) into a single semantic conceptual chunk. This will negatively affect the encoding of serial order information as the codes corresponding to the different serial positions will point to the same chunk, reducing encoding strength of individual item – serial position mappings. At recall this chunk will be decompressed making memoranda available for recall at the item level but without precise information about serial order codes. Finally, at the linguistic level, without considering the intervention of higher-level conceptual chunks, semantically related items are considered to share overlapping semantic features (Dell et al., 1997). Due to these shared semantic features, related items are supposed to reactivate each other via redundant interactive feedback activation, leading to more robust item representations (McClelland & Rumelhart, 1981). At the same time, due to the similarity of representations activated at the linguistic level, serial order codes will point to similar loci of the space of semantic representations, increasing the likelihood of confusions at the level of mappings between item and serial order representations. Evidence for these predictions remains currently scarce. As noted, a few studies appeared to show an increase of serial order errors for recall of semantically related versus unrelated word lists but the effect was weak. Using a very large sample (N = 152) of participants, Tse et al. (2011) found effect sizes around $\eta_p^2 = .07$. This observation suggests that the impact of semantic knowledge on serial order processing may be anecdotal.

The aim of this study was to examine in a systematic manner the interactions that may exist between semantic knowledge and the representation of serial order

information in WM, by focusing on the semantic relatedness effect. In addition to assessing overall order recall accuracy as done in previous studies, we used more fine-grained measures of the pattern of serial order errors that occurred during the recall of semantically related versus unrelated word lists. In line with the theoretical accounts developed here, we predicted that for semantically related word lists, serial order errors should involve more frequently displacements between two related words than between two unrelated words, by presenting words from two different semantic categories in the same list.

Experiment 1

We manipulated the semantic relatedness effect by creating lists composed of semantically unrelated items or lists containing subgroups of semantically related items (e.g. leaf – tree – branch – cloud – sky – rain). The lists were presented for standard immediate serial recall.

In addition to semantic relatedness effects, we also manipulated temporal grouping effects by presenting items in a temporally grouped or ungrouped manner via the insertion of a short temporal pause between items (e.g. leaf – tree – branch – *pause* – cloud – sky – rain). This was done in order to determine whether the possible effect of semantic relatedness on serial order recall performance as assessed via the semantic grouping manipulation is independent from the more general effect of temporal grouping. Temporal grouping effects are characterized by changes in the shape of the serial position curve, with the occurrence of micro-primacy and micro-recency effects within each temporal group, and items within each group being recalled at similar performance levels (Hartley et al., 2016; Henson, 1996; Hitch et al., 1996). In addition, temporal grouping also leads to an increase of so-called interposition errors, in which items, when migrating between groups, keep the same within-group position (e.g. A **B** C – D **E** F transposed to A **E** C – D **B** F) (Henson, 1996). If the effects observed for the semantic grouping manipulation simply reflect the effect of more general grouping mechanisms, then we should observe similar patterns of performance as the temporal grouping manipulation, including micro-primacy and micro-recency effects within each semantic group, similar performance level of items within the same group, but also an increase of between-group interposition errors.

Method

Participants. Thirty-nine undergraduate students (31 females) aged between 18 and 26 years (*M* = 21.31, *SD* = 2.31) were recruited from the university community of the University of Liège. All the participants were French-native speakers, reported no

history of neurological disorder or learning difficulties, and gave their written informed consent before starting the experiment. The experiment had been approved by the Ethic committee of the Faculty of Psychology, Speech and Language Therapy, and Education of the University of Liège.

Material. In the present experiment, we used a set of 180 auditory verbal stimuli consisting in monosyllabic words with a mean log-frequency of M_{log} = 3.452 and SD_{log} = 1.629. Triplets of words sharing similar semantic characteristics, such as "leaf – tree – branch", had been selected based on the results of an online survey conducted on 136 participants (see Kowialiewski & Majerus, 2018); they had been asked to rate the semantic relatedness characterizing pairs of words by using a scale ranging from 0 (not semantically related at all) to 5 (strongly semantically related). The semantic pairs were defined in an a priori manner and were drawn from a mix of taxonomic (e.g. figue [fig], fraise [strawberry], pêche [peach]) and thematic (e.g. toast [toast], pain [bread], beurre [butter]) categories. A previous study showed that taxonomic and thematic categories produce similar effects on item and serial order processing (Tse, 2009). On average, each pair was judged 75.25 times. Based on these results, we first constructed semantic groups of three words for which adjacent words consisted in pairs of words that received a score strictly superior to 3.5 of semantic relatedness in the online survey. In other words, we ensured that in each semantic group the first word was judged as highly related to the second word, and the second word as highly related to the third word. The stimuli were recorded by a French-speaking Belgian native male speaker and normalized to a duration of 375 milliseconds.

Design. The task consisted in the presentation of 60 lists of six words following a 2 × 2 experimental design with a 2-level semantic factor (related vs. unrelated) and a 2-level temporal grouping factor (grouped vs. ungrouped). According to the semantic factor, one half of the sequences was composed of semantically unrelated words (e.g. soir [evening], plante [plant], tête [head], force [strength], nombre [number], crabe [crab]) and the other half consisted in two groups of three semantically related words (e.g. nuage [cloud], ciel [sky], pluie [rain], tigre [tiger], hyène [hyena], lion [lion]). Concerning the temporal grouping factor, one half of the sequences were presented at a regular pace while the other half formed two groups of words separated by an increased inter-stimulus duration (see details below).

The creation of related and unrelated lists was performed by combining two semantic groups of three words while ensuring that the two semantic groups did not cover similar semantic categories. Using the same procedure, we also ensured that each word in a semantic group could not have a specific semantic relation with a word from the other semantic group in the same list. For the construction of the unrelated

lists, we selected six words from our stimuli set by carefully ensuring that the words were not semantically related. Finally, each of the 180 stimuli occurred two times over the 60 trials, once in a related list and once in an unrelated list, but in different serial positions; the two occurrences of the same word had to be separated by at least two trials.

In order to rule out the presence of list effects on task performance, we generated eight different versions of the 60 lists. We first created four different versions of the 60 lists: For each version, the different semantic groups were combined in different pairings within a list. Next, we generated four additional versions by reordering the order of the words within the sequences of the four first versions; for semantically related sequences, this meant that only the order of the words within a semantic group was changed. No more than three lists of the same type (i.e., related, unrelated) could be presented successively.

Procedure. The lists were presented to participants through headphones connected to a portable workstation and at a comfortable sound level. The presentation of each list was announced via a numerical countdown starting at 3 at a pace of 500 milliseconds per digit and presented on the center of the screen. Participants were then required to listen carefully to the 6-word sequences and they had to recall the words in the correct serial order immediately after their presentation. They started recalling the list after a 440-Hertz sine wave tone of 100 millisecond duration presented after the last word of the target list. If participants did not remember the word for a given position, they were asked to say "blanc" (blank in French). In order to help participants in recalling all items and in their correct serial position, they were provided a 6-square horizontal grid and they indicated for each word they recalled its serial position on the grid by pointing to the corresponding square. They pressed the spacebar to move from one trial to the next. The participants' responses were recorded via a digital audio recorder for later transcription.

The temporal grouping factor was implemented in the following manner. For the ungrouped lists, words were presented with an interstimulus interval (ISI) of 250 milliseconds. For grouped lists, the intra-group ISIs were of 100 milliseconds while the inter-group ISI (between the third and fourth item of the list) was of 750 50 milliseconds leading to a temporal separation of the first and second group. This procedure ensured that overall list presentation durations were identical for the two conditions (3500ms). . Participants were presented three practice before starting the task. The practice trials contained semantically unrelated words presented at a regular presentation speed. Stimulus presentation was controlled via the OpenSesame software running on a desktop station computer (Mathôt et al., 2012).

The temporal grouping factor was blocked, with the 30 first trials being temporally ungrouped while the 30 last trials were temporally grouped, the two blocks being separated by a short pause. This was done in order to minimize the possibility that participants deploy grouping strategies already on the temporally ungrouped sequences (Farrell & Lewandowsky, 2004; Henson, 1999). The semantic grouping factor was manipulated within each temporal factor; with each temporal grouping condition containing the same number of semantically related and unrelated sequences.

Statistical analysis. We used Bayesian statistical analysis techniques as they have the advantage of not being influenced by the intention underlying data collection or by the use of optional stopping rules (Berger & Berry, 1988; Rouder, 2014). Furthermore, they test evidence for both against and in favour of the null model (H_0) —while classical frequentist statistical techniques test only evidence against the null model. Bayesian statistics also are not sensitive to the time of analysis (i.e. planned *versus* post hoc analysis) and do not suffer from increased false alarm rates when conducting multiple comparisons (see Dienes, 2016).

In the present study we report results with a relative measure of evidence called the Bayes factor (BF) that quantifies the extent to which data are more likely to be observed under a model relative to another (e.g., H_1 *versus* H_0, see Jeffreys, 1961, Morey 2011). For example, if after looking at the data H_1 is preferred over H_0 by a factor of 20, this indicates that the data are 20 times more likely under H_1 than under H_0. When interpreting the strength of evidence associated to BF values, we used the following terminology (see M. D. Lee & Wagenmakers, 2014; Wagenmakers et al., 2017): BFs lesser than three, between three and 10, between 10 and 30, between 30 and 100, and higher than 100, were considered as representing anecdotal, moderate, strong, very strong, and decisive level of evidence, respectively. For simplicity, we report here the $BF_{inclusion}$ values when ANOVAs are performed. The $BF_{inclusion}$ averages the evidence of all models including a given factor, relative to all other models not including this factor. In other words, it provides the evidence in favor of the presence or the absence of a given effect across all the models. We report $BF_{inclusion.10}$ and $BF_{inclusion.01}$ for evidence supporting the alternative or the null hypothesis, respectively. Likewise, when T-Tests were performed, we use the labels "BF_{10}" and "BF_{01}" to indicate that the data support the alternative or the null hypothesis, respectively. All the analyses were performed using JASP (Version 0.8.5.1, JASP Team, 2017) and all parameters were set to the default Cauchy prior distribution. By default, the r scale of the Cauchy distribution in t-tests is set to .707. For ANOVAs, the r scale is set to .5, 1, and .354, for fixed effect, random effect, and covariates, respectively.
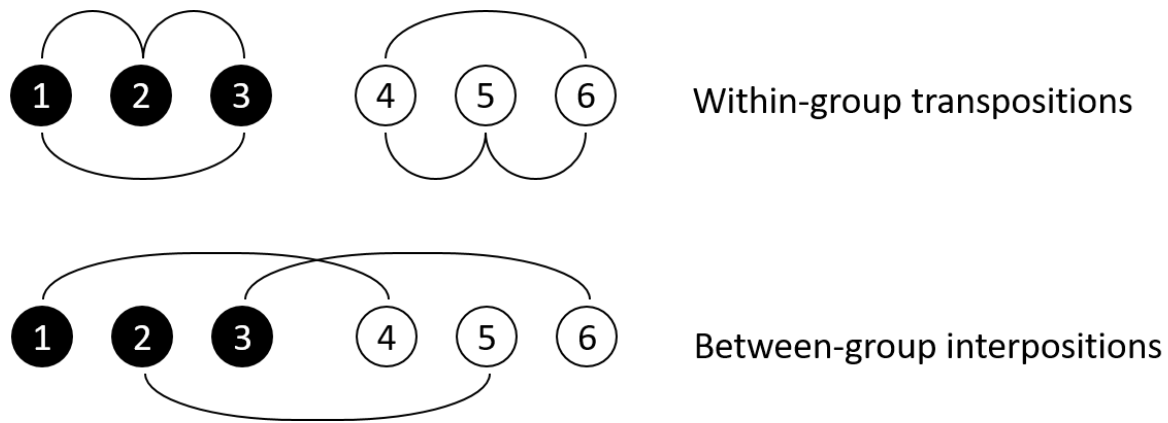
Figure 6.1. Graphic illustration of between-group interposition errors (top) and within-category transposition errors (bottom) for Experiment 1. Items drawn the same semantic category have the same color. Items in different temporal groups are spatially separated.

When presenting the data graphically, we used 95% confidence intervals as error bars. In repeated measures designs, the data were first corrected for between-subject variability (Cousineau, 2005; Morey, 2008). We then computed the standard error on these corrected data, times the critical T-Test value associated to it, with N-1 degree of freedom (Baguley, 2012).

Scoring procedure. To determine the relative impact of the different grouping variables on WM performance, we applied two different scoring procedures. First, we performed a *strict serial recall scoring* procedure, considering an item as correct only if it was recalled at the correct serial position. For instance, given the target sequence "Item1 – Item2 – Item3 – Item4 – Item5 – Item6" and the output sequence "Item1 – Item2 – Item4 – Item3 – blank – Item6", items 1, 2 and 6 would be scored as correct. Second, we used an item scoring procedure by considering an item correct if recalled, regardless of its serial position in the list. In the example described above, items 1, 2, 3, 4 and 6 would be scored as correct.

We characterized the pattern of serial order errors by determining within-group transposition and between-group interposition errors. Within-group transposition errors reflect transposition errors occurring within a temporal or semantic group (e.g. transposing D and F in ABC-DEF, see **Figure 6.1**). Between-group interposition errors reflected between-group transposition errors that kept their initial, within-group position (e.g. transposing C and F in ABC-DEF, see **Figure 6.1**). We determined the proportions of within-group transposition and between-group interpositions, as a function of semantic versus temporal group, by dividing them by the total number of transposition errors observed for a given participant.

Results and discussion

A first analysis assessed recall performance as a function of semantic relatedness (related vs. unrelated), temporal grouping (grouped vs. ungrouped) and serial position (position 1 through 6). Descriptive results are displayed in **Figure 6.2**. Using a Bayesian Repeated Measures ANOVA, we found for both item and strict serial recall criteria, decisive evidence supporting the presence of semantic relatedness, temporal grouping, and serial position effects (all $BF_{inclusion.10} \rightarrow +\infty$). The interaction between semantic relatedness and temporal grouping was strongly supported when using the two criteria ($BF_{inclusion.10} = 27.33$ and $BF_{inclusion.10} = 86.39$ using an item and strict serial recall criterion, respectively). As can be seen in the right panels of **Figure 6.2**, the effect of temporal grouping was more pronounced in the unrelated as compared to the related semantic condition. Similarly, decisive evidence supported the interaction between semantic relatedness and serial position ($BF_{inclusion.10} \rightarrow +\infty$), and the interaction between temporal grouping and serial position ($BF_{inclusion.10} > 100$). Using the item recall criterion, the triple interaction received anecdotal evidence ($BF_{inclusion.10} = 2.461$), while it was supported by decisive evidence when using a strict serial recall criterion ($BF_{inclusion.10} = 106.89$). As expected, the temporal grouping manipulation induced a specific pattern of performance as function of temporal groups, with items within the same temporal group being recalled at similar performance levels. Critically, this was not observed in the semantic relatedness condition without temporal grouping, as indicated by the triple interaction for the strict serial recall criterion.
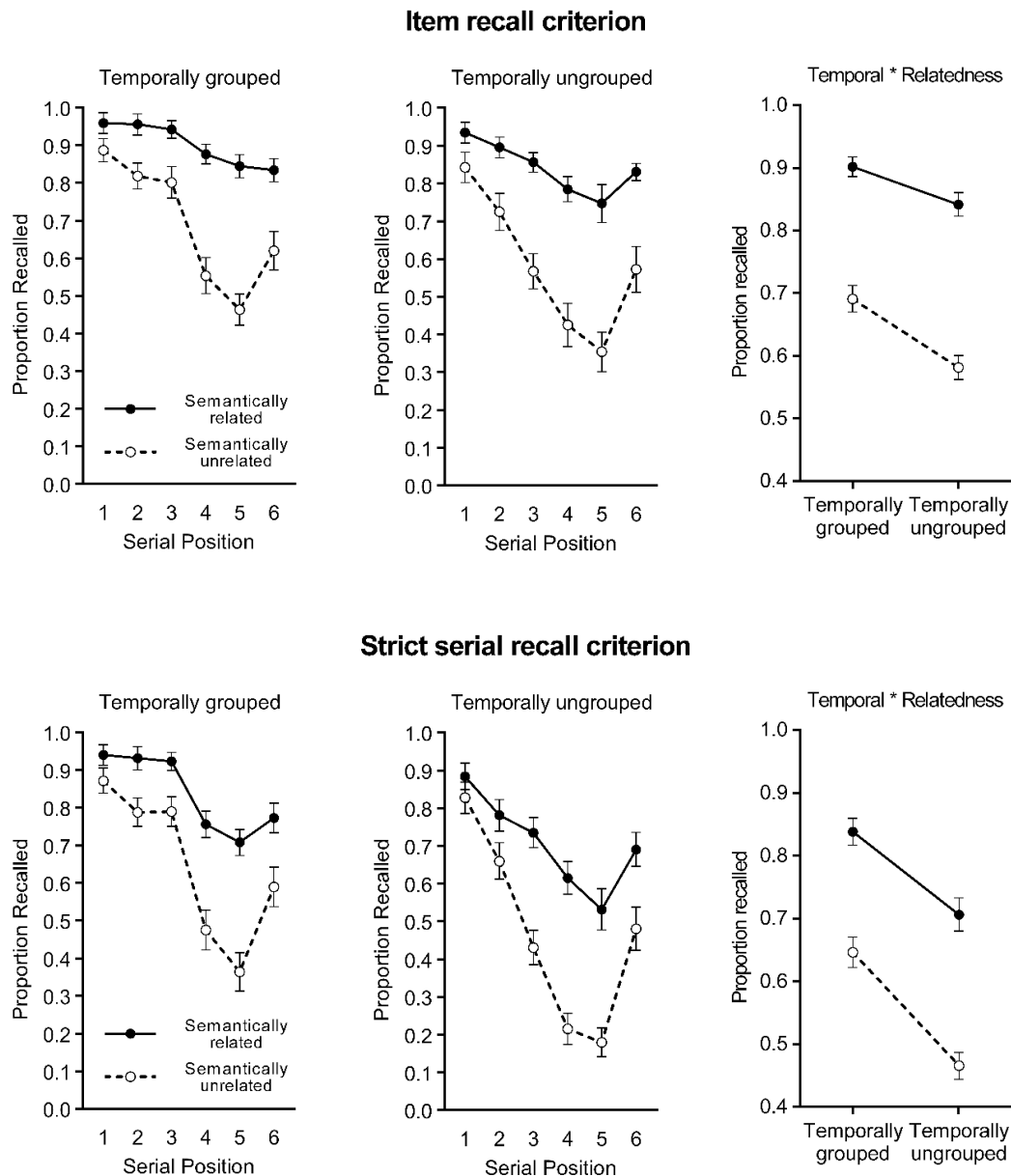
Item recall criterion

Strict serial recall criterion

Figure 6.3. Proportion of items correctly recalled (y axis) as a function of serial position and semantic relatedness (x axis), for temporally grouped (left panels) and ungrouped (middle panels) lists in Experiment 1. The right panels show the interaction between the grouping and semantic conditions without the serial position factor. Upper panels: item recall criterion. Lower panels: strict serial recall criterion. Error bars represent 95% confidence intervals, corrected for between-subject variability (Baguley, 2012; Cousineau, 2005; Morey, 2008).

The most critical analyses were those assessing the effects of semantic relatedness and temporal grouping on the pattern of transposition errors. For these analyses, several participants had to be discarded because they produced zero transposition

errors in at least one of the four experimental conditions, leading to a final sample of 33 participants. First, we assessed the effect of the two variables on the proportion of within-group transposition errors (see **Figure 6.3**, middle panel). We found decisive evidence supporting the presence of both semantic relatedness ($BF_{inclusion.10}$ = 2.974e+8) and temporal grouping ($BF_{inclusion.10}$ = 4610.2) effects. In other words, when a transposition error occurred in grouped sequences, the vast majority of these errors involved the transposition of an item within its semantic or temporal group, as compared to equivalent positions in ungrouped sequences. The interaction term was supported by very strong evidence ($BF_{inclusion.10}$ = 85.18). This interaction was further explored using Bayesian Paired Samples T-Tests. The effect of semantic relatedness was moderately supported in the temporal grouping condition (BF_{10} = 3.41) and was associated with decisive evidence in the condition without temporal grouping (BF_{10} = 3.425e+6). Finally, the temporal grouping effect was supported in the semantically ungrouped condition (BF_{10} = 134.166), but not in the semantically grouped condition (BF_{01} = 4.219). This lack of temporal grouping effect on semantically grouped lists (no additive effect) may however be due to a ceiling effect, as can be seen in **Figure 6.3**, middle panel. We also assessed the effect of the semantic relatedness and temporal grouping variables on the proportion of between-group interposition errors (see **Figure 6.3**, right panel), and found strong evidence supporting a semantic relatedness effect ($BF_{inclusion.10}$ = 10.66), and moderate evidence supporting an absence of temporal grouping effect ($BF_{inclusion.01}$ = 4.237). More specifically, between-group interposition errors almost never occurred in the semantically related condition. The absence of interaction was supported by moderate evidence ($BF_{inclusion.01}$ = 4.83). Note that the overall the proportion of interposition errors was very low.

To sum up, we observed an impact of both semantic relatedness and temporal grouping on item and serial order recall performance. Furthermore, the semantic category strongly constrained the pattern of transposition errors. When a transposition occurred, it involved serial positions associated to semantically related words up to ~95% of the time (see **Figure 6.3**, middle panel). These results support approaches considering that linguistic knowledge information also interacts with the processing of serial order information.

It could be argued however that the semantic relatedness effect actually reflected also a temporal grouping effect as the semantically related items were presented in two successive groups of 3 items. One element of the results does not support this hypothesis. When sequences were temporally grouped, items within the same temporal group were recalled at similar performance levels, while this was not observed for items stemming from similar semantic groups, with recall performance
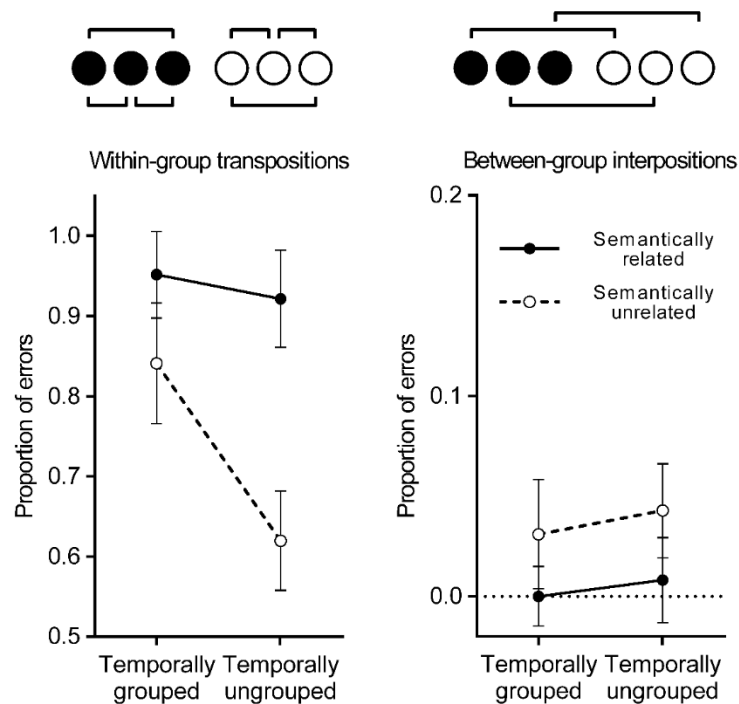
Figure 6.3. Proportion of order errors (y axis) as a function of semantic relatedness and temporal grouping. Left panel: proportion of within-group transpositions. Middle panel: proportion of between-group interposition errors. Filled and dashed lines represent performance for related and unrelated lists, respectively. Error bars represent 95% confidence intervals, corrected for between-subject variability (Baguley, 2012; Cousineau, 2005; Morey, 2008).

gradually decreased over the whole set of items in the list. On the other hand, in order to contrast the effects of temporal grouping and semantic relatedness, we had predicted increased interposition error rates for temporal grouping condition but not the semantic relatedness condition. This prediction could not be tested as the rate of interposition errors war very low (2.8% of total errors). In Experiment 2, we aimed at further dissociating the effects of semantic relatedness and temporal grouping on serial order recall performance in WM.

Experiment 2

Experiment 1 manipulated the semantic relatedness effect by using semantic categories presented in subgroups (e.g. leaf – three – branch – *cloud – sky – rain*). In Experiment 2, the same manipulation was performed, but this time by presenting the categories in an interleaved fashion (e.g. leaf – *cloud* – tree – *sky* – branch – *rain*). If the increase of serial position errors between semantically related words observed in Experiment 1 was generated solely by the grouping of semantically related words,

then these errors should be no more observed when semantically related words are presented in an interleaved fashion.

Furthermore, the temporal grouping manipulation this time involved three groups composed of two items (e.g. AB-CD-EF). The purpose of this manipulation was twofold. First, we aimed at optimizing the opportunity for interposition errors. With temporal groups of two items, interposition errors will correspond to transpositions of distance two or four; particularly transpositions of distance two should be more likely to observe than the transpositions of distance three that defined interposition errors in Experiment 1. In addition, we explored the possibility that the temporal grouping manipulation could have additive effects with the semantic relatedness in terms of transposition errors, by allowing us to determine whether within-semantic category transpositions occur more often when they also share the same relative serial position as defined by the temporal grouping variable. For example, for the target sequence "leaf – sky – tree – cloud", the semantically related words "leaf" and "tree" might be transposed more often, particularly when also sharing confusable relative serial order codes, such as when presented in a temporally grouped sequence such as "leaf – sky – *pause* – tree – cloud". Finally, in line with the results Experiment 1, we expected that items within the same temporal groups should be recalled at similar performance levels.

The changes in the manner semantic relatedness and temporal grouping were manipulated in this experiment imply that there is no distinction between within-semantic group transposition errors and between-temporal group interposition errors anymore. Indeed, given the sequence "leaf – sky – *pause* – tree – cloud", transposing "tree" and "leaf" is a within-semantic category transposition, but also a between-temporal group interposition. From now on, we will refer to these errors simply as interposition errors, which reflect the impact of the semantic variable when comparing the semantically related to the unrelated conditions and which reflect the impact of the temporal grouping variable when comparing the temporally grouped to the temporally ungrouped conditions.

Method

Participants. Thirty-nine undergraduate students (26 females) aged between 18 and 30 years ($M = 20.67$, $SD = 2.86$) were recruited from the university community of the University of Liège. All the participants were French-native speakers and reported no history of neurological disorder or learning difficulties and gave their written informed consent before starting the experiment. The experiment had been approved
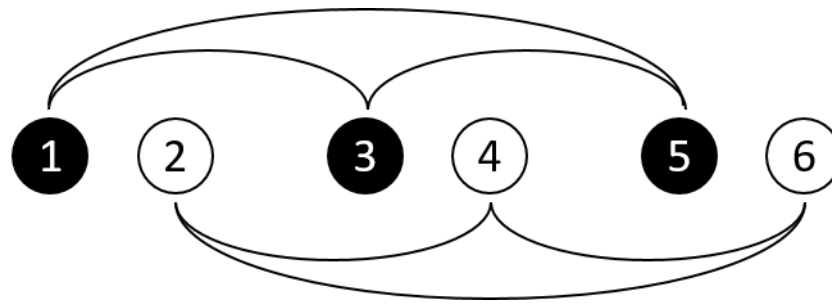
Figure 6.4. Graphic illustration of interposition errors for Experiment 2. Items from the same semantic category share the same color. Items in different temporal groups are spatially separated.

by the Ethics committee of the Faculty of Psychology, Speech and Language Therapy, and Education of the University of Liège.

Material. The material used was identical as in Experiment 1.

Design & procedure. The semantic relatedness manipulation involved now two different semantic categories presented in an interleaved fashion (e.g. leaf – *cloud* – tree – *sky* – branch – *rain*) in the related condition. In addition, there were three instead of two temporal groups in the temporally grouped sequences. As in Experiment 1, words in the ungrouped lists were presented with an ISI of 250 ms. For grouped lists, the ISI was 100 ms inside each temporal group. A longer ISI of 375 ms separated the second and the third words, as well as the fourth and the fifth words, leading to three two-stimulus temporal groups.

Scoring procedure. As already noted, the changes in the manner semantic relatedness and temporal grouping were manipulated in Experiment 2 imply that there is no distinction anymore between within-semantic group transposition errors and between group interposition errors identified in Experiment 1 (see **Figure 6.4**). In Experiment 2, these errors will all be referred to as interposition errors reflecting either the impact of the semantic relatedness variable (when comparing the semantically related to the unrelated condition) or the temporal grouping variable (when comparing the temporally grouped to the temporally ungrouped condition). Interposition errors were considered to occur between positions 1 – 3 – 5 or positions 2 – 4 – 6. All other aspects of scoring procedures, including assessment of item and order recall performance were identical to Experiment 1.
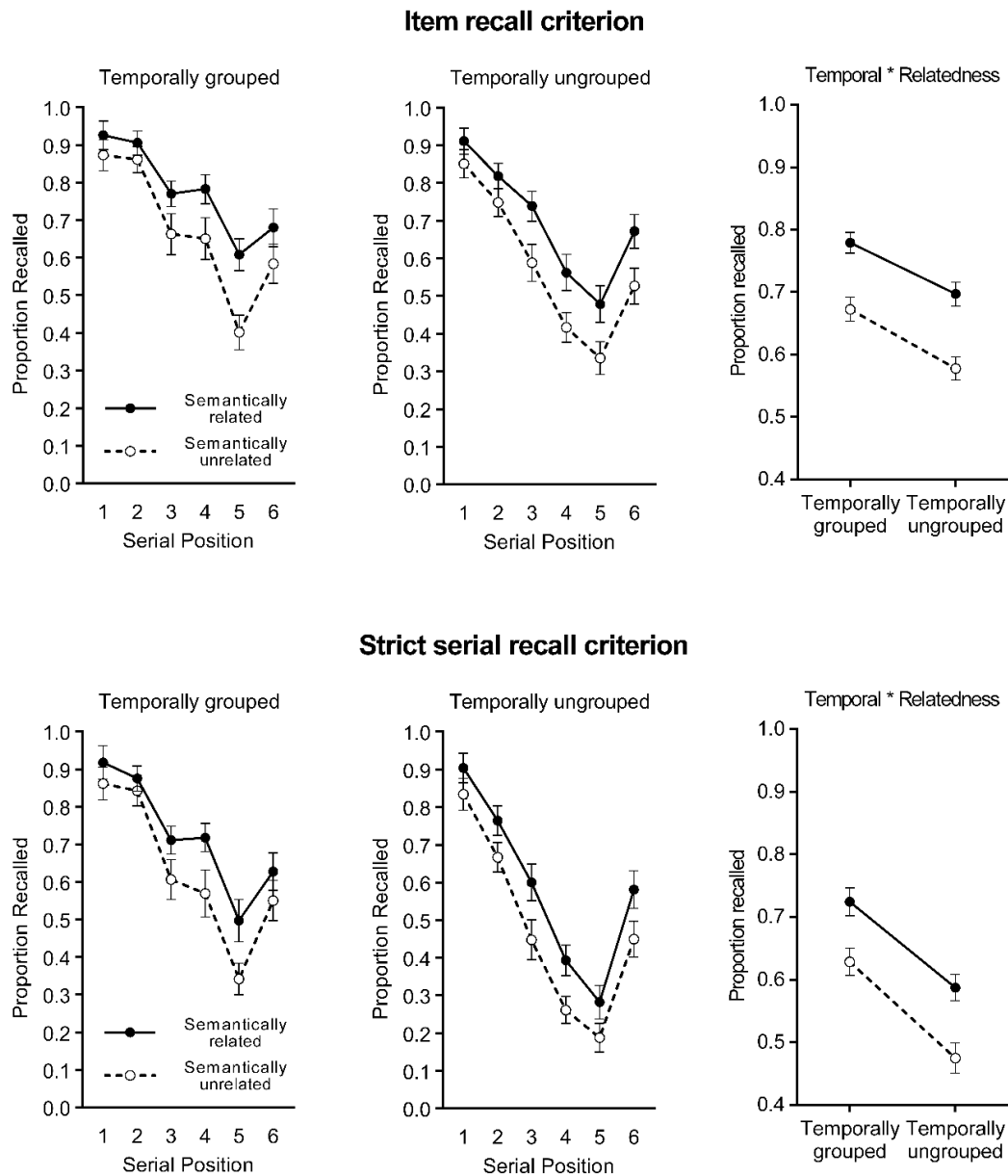
Item recall criterion

Strict serial recall criterion

Figure 6.5. Proportion of items correctly recalled (y axis) as a function of serial position and semantic relatedness (x axis), for temporally grouped (left panels) and ungrouped (middle panels) lists in Experiment 2. The right panels show the interaction between the grouping and semantic conditions without the serial position factor. Upper panels: item recall criterion. Lower panels: strict serial recall criterion. Error bars represent 95% confidence intervals, corrected for between-subject variability (Baguley, 2012; Cousineau, 2005; Morey, 2008).

Results and discussion

First, we assessed the impact of semantic relatedness (related vs. unrelated) and temporal grouping (grouped vs. ungrouped) on the proportion of items recalled across

serial positions (1 through 6). Using Bayesian Repeated Measures ANOVAs we observed decisive evidence supporting the three main effects of semantic relatedness, temporal grouping, and serial position (item recall criteria: all $BF_{inclusion.10}$ = 3.217e+15; strict serial recall criteria: all $BF_{inclusion.10} \rightarrow +\infty$). As can be seen in **Figure 6.5**, both semantic relatedness and temporal grouping manipulations led to increased recall performance. The interaction between semantic relatedness and temporal grouping provided moderate evidence using an item recall criterion ($BF_{inclusion.01}$ = 3.049), and anecdotal evidence using a strict serial recall criterion ($BF_{inclusion.01}$ = 2.833). The interaction between semantic relatedness and serial position was supported by decisive evidence using an item recall criterion ($BF_{inclusion.10}$ = 128.533), but was anecdotal using a strict serial recall criterion (BF_{01} = 2.262). The interaction between temporal grouping and serial position was supported by decisive evidence using an item recall ($BF_{inclusion.10}$ = 2.306e+9) and a strict serial recall criterion ($BF_{inclusion.10} \rightarrow +\infty$). Evidence against the triple interaction was found using both recall criterions ($BF_{inclusion.01}$ = 12.5 and $BF_{inclusion.01}$ = 37.037, for item and strict serial recall criteria, respectively). As shown in **Figure 6.5**, the temporal grouping manipulation produced the expected pattern of performance, with items within the same temporal groups being recalled at similar levels of performance.
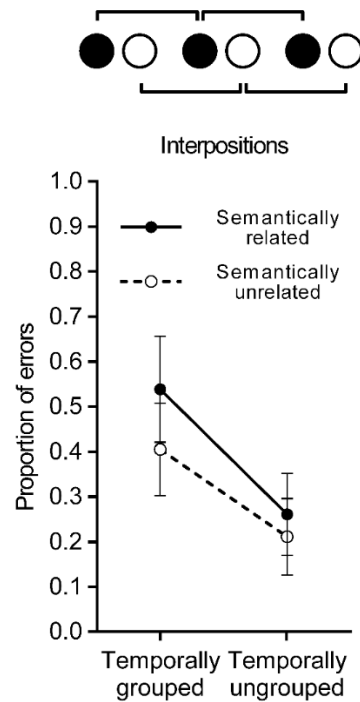
Figure 6.6. Proportion of interposition errors (y axis) as a function of semantic relatedness and temporal grouping. Filled and dashed lines represent performance for related and unrelated lists, respectively. Error bars represent 95% confidence intervals, corrected for between-subject variability (Baguley, 2012; Cousineau, 2005; Morey, 2008).

Next, we assessed the impact of semantic relatedness and temporal grouping manipulations on the proportion of interposition errors and this using a Bayesian Repeated Measures ANOVA. We have to note that, as in Experiment 1, some participants produced zero transposition errors in at least one of the experimental conditions. Twelve of them were therefore excluded from this analysis, leading to a final sample size of 27 participants. Contrary to Experiment 1, the effect of interleaved semantic relatedness on interposition errors was ambiguous ($BF_{inclusion.10}$ = .993). Thus, there was no clear evidence supporting an increase of serial order errors between related words. However, we observed this time decisive evidence supporting the presence of a temporal grouping effect ($BF_{inclusion.10}$ = 4082.484). As can be seen in **Figure 6.6**, right panel, interposition errors were more frequent in temporally grouped sequences. Ambiguous evidence was also observed for the interaction term ($BF_{inclusion.10}$ = .755), although numerically speaking, interposition errors were more frequent for semantically related lists in the grouped condition (see **Figure 6.6**, right panel).

To sum up, in Experiment 2, when presenting semantically related words in an interleaved manner, no increase of transposition errors was observed between

204

semantically related items, contrary to Experiment 1. These results suggest that the increase of serial order errors observed between semantically related words observed in Experiment 1 was caused by the grouped presentation of semantically related items.

General discussion

This study assessed the impact of semantic knowledge on WM for serial order information by manipulating the semantic relatedness of words presented for immediate serial recall. In Experiment 1, we showed that when related lists contained two successive groups of semantically related items, a high proportion of transposition errors occurred for recalling words within each group. This effect disappeared when the semantically related words were presented in an interleaved manner in Experiment 2.

Does semantic knowledge impact serial order processing?

Whether semantic knowledge has a (detrimental) effect on serial order processing remains a controversial question (e.g. Saint-aubin & Poirier, 1999). As already noted by Saint-Aubin et al. (2005), the proportion of order errors usually points towards a detrimental effect although these effects tend to be small and are not always reliable at the statistical level (Tse et al., 2011). The present study partially supports these findings, by showing that when recalled at a wrong serial position, semantically related words migrate more often towards other related words, rather than migrating towards another unrelated word, but this was only observed when the semantic categories were presented in semantic sub-groups. Once the semantic categories were presented in an interleaved fashion, there was no increase of serial order transposition errors anymore between semantically related items. At the same time, the effect size of semantically driven serial order transpositions observed in Experiment 1 was much stronger ($\eta_p^2 = .669$) than the mean effect size observed in previous studies ($\eta_p^2 = .07$ in Tse et al., 2011 study), showing that semantic variables can also have a strong influence on serial order processing when assessed using more fine-grained methods.

At the same time, it could be argued that the strong impact of semantic relatedness on serial order recall errors observed in Experiment 1 reflects a more general temporal grouping effect rather than a genuine semantic effect. Although we can currently not fully reject this possibility, several observations indicate that the effects of semantic grouping and temporal grouping may stem from different sources. Indeed, the pattern of results observed in the semantic grouping as compared to the temporal grouping conditions tend to differ. First, from a temporal grouping perspective, we would have expected an increase of between-group interposition

errors while the reverse was observed for the semantic relatedness manipulation: semantically related items were transposed most of the time within the same semantic subgroup. Note however that the direct manipulation of temporal grouping did not lead to increased interposition errors in Experiment 1 either. Second, the temporal grouping manipulation led to a specific shape of the serial position curve, with items stemming from similar temporal groups being recalled at similar performance levels. Critically, this pattern was not observed when the semantic categories were presented in subgroups. These results suggest that temporal and semantic grouping effects may reflect the intervention of different mechanisms.

How does semantic knowledge influence the retention of serial order information?

If the effect of semantic grouping on serial order error patterns does not (only) arise from general temporal grouping effects, then how can this effect be explained? First, according to the trace confusion hypothesis, items tend to be confused because participants maintain the shared semantic content of items throughout the list and then reconstruct the degraded items based on these general semantic features (Saint-Aubin & Poirier, 1999a). This hypothesis could easily account for the overall better recall performance for semantically related lists as well as for the increased transposition errors within semantically related lists observed in Experiment 1. However, this hypothesis, in its actual state, cannot account for the absence of a semantic relatedness effect on serial order errors observed in Experiment 2: trace reconstruction should lead to erroneous but semantically related item reconstructions also when items are presented in an interleaved manner. Note however that the theory is difficult to test, because there is no formal definition of the way semantic reconstruction is implemented when semantically related items occur in non-adjacent serial positions.

According to the chunking hypothesis, semantically related items are compressed by activating and maintaining a broader conceptual representation (Chekaf et al., 2016; N. Martin, Minkina, Kohen, & Kalinyak-Fliszar, 2018; Mathy et al., 2018), leading to a reduction of WM load and an increase of overall better recall performance when items are presented in different subgroups. This mechanism will also lead to increased serial order errors within each subgroup, because information about the arbitrary serial order of items will get lost during the compression-decompression process of the different items from the semantic category. This is however different when related items are presented in an interleaved fashion; in this case, the compression mechanism will occur to a lesser extent, because the broader semantic category will be less obvious at least for earlier presented items, or might
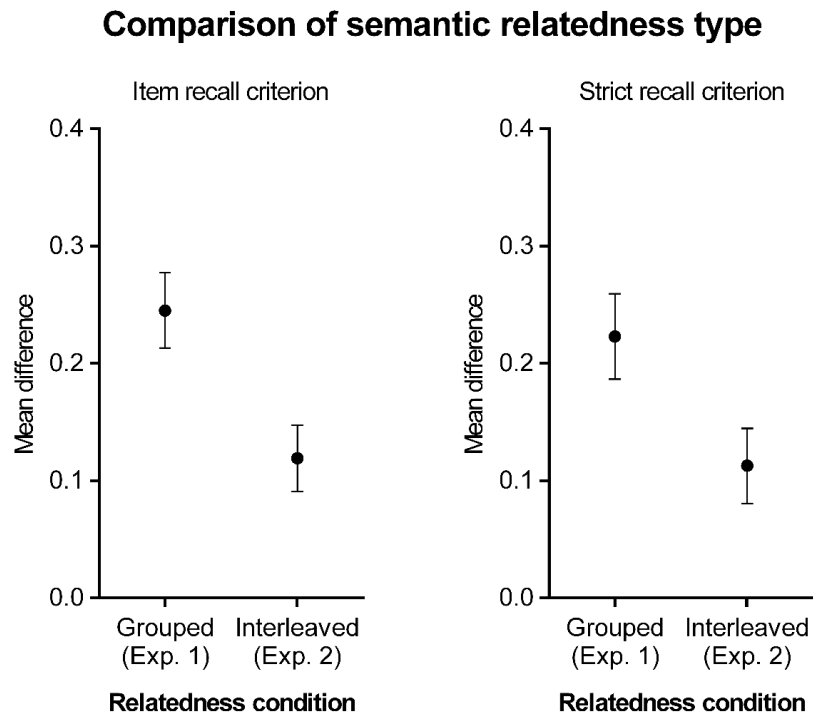
Comparison of semantic relatedness type

Figure 6.7. Direct comparison of recall performance between Experiment 1 and Experiment 2. The y axis represents the mean difference between related and unrelated lists. Left panel: item recall criterion. Right panel: strict serial recall criterion. Error bars represent 95% credible intervals of the mean.

require additional controlled mechanisms in order to be detected, leading to a reduced relatedness effect on overall recall accuracy, but also a reduction of serial order errors between semantically related items as compared to semantically grouped presentation. This is what had been observed: as can be seen in **Figure 6.7**, the impact of the semantic relatedness dimension on recall performance was overall smaller as compared to the semantically grouped condition. In addition, in Experiment 1, where semantically related items were presented in sub-groups, serial order errors occurred most of the time between semantically related words, but this increase was not observed in Experiment 2 when semantically related items were presented in an interleaved manner.

According to the interactive activation hypothesis (Dell et al., 1997; McClelland & Rumelhart, 1981), semantically related words will reactivate each other via their shared semantic features, leading to overall better recall accuracy, but also more transposition errors due to the fact that the different semantically related words will be all co-activated at the same time. A similar interpretation has been proposed to account for the refractory effect in so-called blocked semantic task (Belke, Meyer, &

Damian, 2005). This effect refers to slowed naming latencies for semantically related pictures as they are repeated across successive trials. It has been proposed that repeated activation of words from the same semantic category leads to continuous co-activation of these words which will make more difficult the selection of an individual item for response output (Python, Fargier, & Laganaro, 2018). In the interleaved semantic condition, related items may also reactivate each other, but to a lesser extent due to their separation by other items as compared to the situation where semantically related items follow each other directly. This should lead to an overall weaker impact of the semantic relatedness variable on recall performance when using an interleaved presentation format as in Experiment 2, and reduced serial order transpositions between semantically related words as compared to a grouped presentation condition. The reduced impact of the semantic relatedness dimension in Experiment 2 as compared to Experiment 1, and the absence of increased transposition errors between semantically related items in Experiment 2 support this assumption.

Conclusion

By using a semantic relatedness manipulation, we observed that semantic knowledge had a strong impact on the pattern of serial order errors, but only when semantically related items were presented in subgroups. These results could reflect the intervention of general grouping mechanisms, but also stem from direct interactions between the processing of serial order mechanisms and linguistic knowledge. Further investigations are required in order to tease apart these possibilities.

General discussion

Discussion

Overview of results

In this PhD thesis, we confronted different accounts explaining the impact of linguistic knowledge on verbal WM, and this in order to achieve a better understanding of the nature of verbal WM and its interactions with the language system. We investigated three different accounts: (1) a strategic, controlled account of the influence of linguistic knowledge on verbal WM, (2) a post-memory reconstructive account, and (3) a fast and automatic linguistic account. In addition, we further assessed the breadth of the impact of linguistic knowledge on WM, by assessing the role of linguistic knowledge in protecting memoranda against interference, as well by exploring the interactions between linguistic knowledge and the maintenance of serial order information. In this General discussion section, we start with a synthesis of the empirical studies conducted in this PhD thesis.

In **Study 1**, we assessed the strategic versus non-strategic nature of psycholinguistic effects. Several studies have shown that, in standard immediate serial recall tasks, participants spontaneously implement strategic processes such as verbal elaboration in order to enhance their WM performance. These strategic processes could be responsible for the presence of at least some of the psycholinguistic effects. In Study 1, we assessed the presence of the most frequently studied psycholinguistic effects (lexicality, lexical frequency, semantic similarity and imageability) in a specific running span procedure aimed at preventing the intervention of strategic processes during encoding. Overall, we observed that most of the psycholinguistic effects could still be observed when using this specific immediate serial recall procedure. The amplitude of these effects appeared to be of similar size to what has been found in previous studies using standard immediate serial recall procedures. Only the imageability effect could not be observed when using a running span procedure. We also showed that the imageability effect was absent in a running span procedure using a slow stimulus presentation rate. The imageability effect was however observed again when assessed under standard immediate serial recall conditions. The results of Study 1 suggest that surface-level linguistic knowledge impacts WM performance in a non-strategic manner, as assessed by the lexicality, lexical frequency and semantic relatedness effects. This is however different for the deeper semantic features associated with the imageability dimension, which may require more elaborative processes to appear.

In **Study 2**, we tested the redintegration account as an exclusive explanatory factor of psycholinguistic effects by manipulating the lexicality effect in a procedure strongly limiting the opportunities for reconstruction processes. We used again a running span procedure but coupled it with a fast item probe recognition procedure requiring participants to judge as fast as possible whether a probe item had been presented in the memory list. This study demonstrated a robust lexicality effect for both recognition performance and response times. This study allows to rule out a strong version of the redintegration hypothesis which considers that no lexicality effect should appear when standard reconstruction cannot be implemented during inter-stimulus intervals at encoding or at retrieval. This study does however not rule out the possibility of a more automatic and very rapid reconstruction process already occurring during encoding. This phenomenon, also called predictive coding, has been studied extensively in the language processing domain as the predictive coding. This type of fast and automatic reconstruction process, embedded within the language system, is more generally consistent with language-based models of verbal WM.

In **Study 3**, we used a functional neuroimaging study in order to assess both redintegration-based and language-based accounts, by measuring participants' brain activity while they were presented with lists composed of words or nonwords in a running span procedure. After encoding the items, the participants were required either to maintain the items, or not to maintain them, which allowed us to distinguish the different stages at which the word/nonword distinction could appear during WM processing. Using a multivariate voxel pattern analysis approach, we observed that the word versus nonword status could be reliable decoded during WM encoding but also during WM maintenance. Critically, the word/nonword distinction was reliably observed in linguistic cortices involved in semantic and phonological processing. Interestingly, reliable word versus nonword decoding was also observed in the intraparietal sulcus where items are supposed to be maintained via attentional focalisation. Overall, this study shows that linguistic cortices represent the verbal type of WM content during both encoding and maintenance, as predicted by language-based models.

In **Study 4**, we explored the impact of the lexical cohort effect on WM recall performance, a psycholinguistic effect whose impact on WM tasks has not yet been investigated. In linguistic tasks, the lexical cohort effect is characterized by the fact that words drawn from large cohorts are responded to more slowly than words drawn from smaller cohorts. In a first pilot experiment, we did not observe a consistent impact of the lexical cohort manipulation on WM performance. In a second experiment, using a purely linguistic task (lexical decision), a reliable lexical cohort effect was observed.

In a third experiment, we replicated again the absence of lexical cohort effect in the context of a WM task. This result suggests that speed of lexical access which characterizes the lexical cohort dimension does not reliably impact WM performance. Instead, we suggested that when psycholinguistic effects are observed in WM, they stem from differences in the robustness or richness of linguistic representations in the language system.

In **Study 5**, we went one step further by exploring the extent to which linguistic knowledge not only supports basic storage of memoranda, but also can protect memoranda against the deleterious effect of a secondary task interference. We explored this possibility by manipulating semantic relatedness and word imageability effects in an immediate serial recall tasks requiring participants to perform a secondary interfering task (backward counting) directly after the encoding phase. In Experiments 1a and 1b, the semantic relatedness dimension reliably impacted WM performance. Importantly, semantically related lists were less strongly impacted by the interfering task than were unrelated lists. In Experiments 2a and 2b, the imageability dimension also reliably impacted WM recall performance, but this time high and low imageability lists were similarly impacted by the interfering task. These results demonstrate that semantic knowledge can have a protective effect on WM representations, but only for verbal items characterized by stronger inter-item associations. We propose that this protective effect may stem from inter-item reactivations of memoranda occurring within the linguistic system, or from the reduction of cognitive load for semantically related items via conceptual chunking mechanisms.

In **Study 6**, we investigated a final critical theoretical question about the interactions between WM and linguistic processing, by assessing the impact of semantic knowledge more specifically on the retention of serial order information as compared to the retention of item information. In a first experiment, we presented lists of words composed of two semantic categories, with semantically related words being presented in sub-groups of 3 items. We showed that the semantic grouping manipulation strongly constrained the pattern of serial order errors: when serial position migrations occurred, they involved most of the time serial position exchanges between semantically related words, rather than between other unrelated words of the same list. In a second experiment, when the semantically related items were presented in an interleaved fashion, we did not observe increased serial position migrations between semantically related items. We interpreted these findings by suggesting that the grouped presentation allows semantically related items to be compressed into a broader conceptual representation. At the moment of recall, the exact serial position of each item within the same semantic subgroup is difficult to retrieve as items are

directly decompressed from the broader semantic category (which does not code serial order information). The results of this study could also be explained by interactive activation accounts, with the impact of interactive activation between memoranda being exacerbated when items are presented in semantic sub-groups. More generally, these results provide important evidence for the existence of interactions between serial order processing and semantic knowledge while many language-based WM accounts do not consider these interactions.

Overall assessment of the strategic account

What are the overall implications of our results for the validity of the strategic account of psycholinguistic effects in WM? According to the strategic account, participants have the opportunity to implement slow, controlled strategic processes in standard immediate serial recall tasks (Morrison et al., 2016), and these processes such as semantic elaboration might be responsible for the presence in WM tasks of at least some of the psycholinguistic effects. This account is supported by studies showing that leaving more opportunity for these processes to occur (i.e. by decreasing presentation rate) increases the impact of semantic effects on WM performance (Campoy et al., 2015; H. G. Shulman, 1970). In this PhD thesis, we adopted an opposite experimental procedure, by determining whether psycholinguistic effects can still be observed when the implementation of encoding strategies is maximally prevented.

In Study 1, we observed that when the psycholinguistic effects were assessed in a fast-encoding running span task, the large majority of psycholinguistic effects were still observed. Even though the purpose of Study 2 was not to directly assess the strategic account, the observation of a lexicality effect in a combined running span–fast probe recognition procedure also does not favour a purely strategic account of the lexicality effect. This is furthermore supported by Study 3, in which linguistic cortices reliably decoded word and nonword stimuli during the encoding and maintenance of memoranda in a running span procedure, further suggesting that linguistic knowledge is accessed in a fast, direct and automatic manner.

Several criticisms could however be addressed to our studies. First, although we assessed the most frequently studied psycholinguistic effects (lexicality, lexical frequency, semantic relatedness), we did not investigate all psycholinguistic effects using a running span encoding condition. This is not a trivial comment as we have observed in Study 1 that at least one psycholinguistic effect, the word imageability effect, does not appear when using a running span procedure, and hence its appearance may require the implementation of strategic verbal elaboration processes. Second, it is still conceivable that in standard immediate serial recall paradigms,

strategic processes intervene in addition to more direct and automatic activation of language knowledge, although the comparison of effect sizes for studies using immediate serial recall paradigms and our study does not indicate reduced psycholinguistic effects when a running span paradigm is used. The only exception appears to be the imageability effect, which was almost completely abolished in the running span task. We should note here that the imageability effect had been reported to be preserved in other studies using dual-task paradigms (Campoy et al., 2015) and concurrent visual dynamic noise (Castellà & Campoy, 2018; Chubala et al., 2018), which are thought to prevent strategic processing, or at least the strategic use of mental imagery. The results of these studies indicate that strategies in the form of mental imagery are not likely to support the imageability effect. Study 1 of this thesis however suggests that the imageability effect may require the intervention of other strategies such as semantic elaboration. This suggestion could be tested more directly by requiring participants to perform a concurrent semantic judgement task during encoding rather than a visual mental imagery task, and to assess whether the imageability effect would be preserved in such conditions.

To sum up, our results do not support a strong version of the strategic account which considers that all psycholinguistic effects would be the result of strategic, controlled processes during encoding. However, the imageability effect appears to be an exception, but further investigations are required in order to better understand its precise nature.

Overall assessment of the redintegration account

A further major part of this PhD thesis was dedicated to the assessment of the redintegration hypothesis. Through two different studies, we assessed a strong version of the redintegration hypothesis (Hulme et al., 1991; Schweickert, 1993) which assumes that psycholinguistic effects are the result of a post-encoding reconstruction mechanism cleaning up the degraded phonological traces stored within a WM buffer. In Study 2, this hypothesis was assessed by using an experimental WM task setup aimed to prevent the intervention of reconstruction mechanisms. Contrary to the redintegration hypothesis, lexicality effects could still be reliably observed. In addition, in Study 3 we showed reliable decoding of neural patterns associated with word and nonword stimuli in linguistic cortices, and this already during the encoding and maintenance stages of the running span procedure, while the redintegration account would predict that words and nonwords are only differentiated at the moment of retrieval. These studies allow to rule out a strong, post-memory version of the redintegration account.

We should however note that the redintegration account is somewhat difficult to falsify. This is so because of the poor theoretical specification of this account. Different authors will provide different definition of the redintegration mechanism. According to Hulme et al., reconstruction occurs at the moment of recall directly via the language production system. For instance, given the degraded phonological traces "P_G", the word "PIG" might be directly reactivated. However, there was no effort made in order to specify how this re-activation might occur, except that is considered to occur in the language production system. For Schweickert's multinomial processing tree however, the redintegration mechanism acts as a comparison mechanism between degraded phonological traces and stored lexical representations. Again, how this comparison process is performed remains elusive. This is problematic when considering, for instance, the fact that psycholinguistic effects might be expected to increase or decrease in strength, depending on the amount of degradation of phonological traces at the moment of recall (Ritchie et al., 2015). Is there a specific threshold beyond which a trace is so degraded that it might be considered uninformative for the comparison mechanism? Importantly, how does the comparison mechanism deal with missing information? Is the phonological trace "_L_P_A__" processed as "LPA" and then compared to 3-phoneme items, or is there some form of mechanism also keeping in memory the absolute position of each individual phoneme? Due to these conceptual imprecisions, the model has a lot of degrees of freedom, and is only able to make general predictions.

Furthermore, the redintegration framework does not appear to be a parsimonious explanation of the different psycholinguistic effects observed in WM tasks, as this framework struggles to account for all these effects in a coherent manner. Even though the redintegration framework is able to deal with the presence of lexical knowledge, such as the lexicality and lexical frequency effects, it struggles to account for the neighborhood density effect by predicting opposing results to those that have been observed in the literature. The redintegration framework is also limited when it comes to explain the influences of semantic knowledge. Although the semantic similarity effect is explained by assuming that the semantic category of the to-be-remembered items sharpens the search process during redintegration, no precise explanation is given of how this could be achieved. How does the system know whether the next to-be-remembered items will be of the same semantic category? Without assuming the possibility of semantic categorization processes already at the moment of encoding, the system has no a priori knowledge of the list's semantic content, and hence has no means to know whether it would be appropriate for the current trial to initiate a semantically constrained search process. The imageability

effect is also difficult to explain within a strong version of the redintegration account, and this problem had already been pointed out by early investigations of this effect (Walker & Hulme, 1999).

These different issues could still be solved if we assume that even within reconstructive accounts, semantic knowledge can be accessed and maintained already at the moment of encoding (Ritchie et al., 2015; Tse, 2009), and that semantic knowledge is subject to redintegration itself. Furthermore, we could also assume the possibility of very fast reconstructive mechanisms already occurring during or shortly after encoding of a given stimulus. However, if we extend the reconstructive account in such a way, then we are already situating this account within language-based accounts. Indeed, language-based models assume that lexico-semantic knowledge provides very fast stabilizing feedback activation already at the moment of encoding. In addition, fast reconstruction mechanisms are also an inherent property of language-based models: it has been shown that language processing is characterized by an automatic completion process of the speech signal during the stage of perception through the use of pre-existing phonological and lexico-semantic knowledge, and this phenomenon has been termed predictive coding (Hannemann et al., 2007; Heald & Nusbaum, 2014; Leonard et al., 2016).

In sum, although the reconstructive account is difficult to falsify as such due to its loose theoretical definition, it does not seem to offer a parsimonious and complete explanation of the different psycholinguistic effects that have been shown to influence verbal WM. The different facets that would need to be added to the reconstructive account to make it compatible with the empirical findings of this PhD thesis will have the effect of moving the reconstructive account actually towards language-based accounts.

Overall assessment of language-based accounts

The third and final account we assessed is the language-based account considering that verbal WM directly draws on the resources of the linguistic system. With the sole exception of the imageability dimension (see above), the results of Study 1, 2 and 3 appear to support most strongly the language-based account, by showing that linguistic knowledge is accessed in a direct, obligatory and very fast manner, and this during the different stage of WM processing. According to interactive activation models (Dell et al., 1997), once lexical representations are activated, they will send stabilizing feedback activation to phonological levels of representations, but will also critically activate the semantic representations associated to the target words. This explains the presence of lexicality and lexical frequency effects in the running span

task of Study 1, but also the presence of the semantic similarity effect. If lexical and semantic representations were not activated very quickly and automatically during encoding, it would have been very difficult to observe these effects. This is also reflected in Study 3, where neural patterns reliably decoded the activation elicited by word and nonword stimuli during both the encoding and maintenance stages of the running span procedure. However, contrary to what could have been expected, the manipulation of the cohort competition effect in Study 4 did not reliably impact WM performance. This result however does not challenge language-based models as such. They simply show that the mechanisms underlying the lexical cohort effect, namely speed of lexical access rather than robustness of lexical access, do not have an impact on the robustness of time-based maintenance of activated linguistic representations in a WM context. In sum, the language-based account appears to provide the most parsimonious explanation of the different psycholinguistic effects that have been shown to influence verbal WM performance for the specific experimental paradigms used in this work.

Again, the only exception appears to be the imageability effect, which was abolished in the running span procedure. As already explained in the previous section, this effect may depend on more strategic semantic processes such as semantic elaboration. These processes may involve more precisely the chunking of different items into a unified semantic representation, thereby reducing cognitive load, and this process might be easier to perform for high versus low imageability words, when controlling for other confounding factors. This could also explain the increased imageability effect as presentation rates decrease (Campoy et al., 2015): as more time is available during the inter-item interval, there is more opportunity for semantic elaboration to occur. More generally, the imageability effect appears to be context-dependent also in non-memory, linguistic tasks. It has been shown that when words are embedded in a lexical decision task where the nonwords can be detected easily (e.g. "ZXK"), an imageability effect for word stimuli is rarely observed. However, as the nonwords are very word-like, an imageability effect progressively appears (Evans et al., 2012). These results show that the semantic features associated to the imageability dimension of a given item are not always accessed and/or used in an obligatory manner, but may depend on linguistic task requirements, and this phenomenon may also characterize verbal WM processing.

At the same time, most theoretical models within the language-based account remain incomplete or imprecise when it comes to not only explain the general occurrence of psycholinguistic effects in WM, but also to explain how these effects will be deployed over the multiple items of a WM list. As already mentioned in the
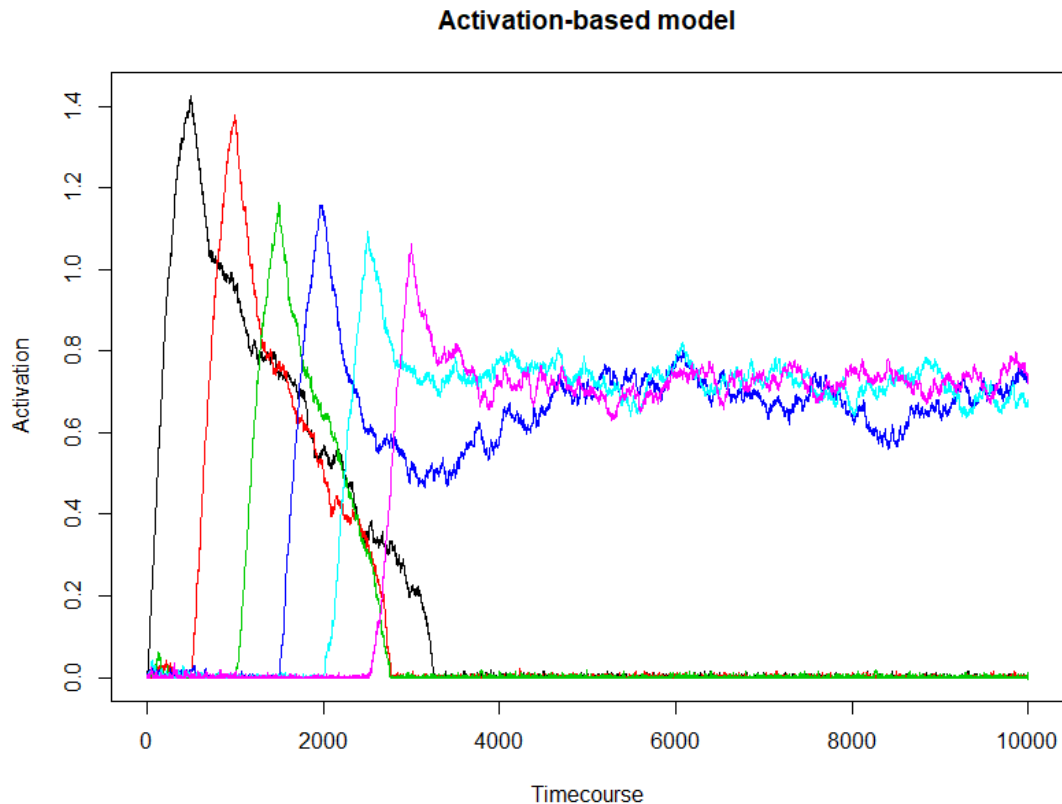
Activation-based model

Figure 1.1. Timecourse of activation within the prefrontal cortex system, as proposed by Haarmann & Usher (2001).

introduction, an early proposal from N. Martin and Saffran considered that WM maintenance is performed via *sustained* activation within the linguistic system (N. Martin & Saffran, 1997; N. Martin et al., 1996). According to this extreme version of activation-based models, items are directly activated within our linguistic long-term memory system, with this activation slowly decaying over time, and WM recall performance depends on the amount of activation available at the moment of recall; the greater the activation, the greater the probability that an item will be correctly recalled. But one might wonder: how likely is this approach to account for WM performance? Indeed, the approach was originally implemented in order to account for the profile of aphasic patients in different language comprehension and production tasks, and has been very successful in doing so (Foygel & Dell, 2000; Majerus et al., 2001; N. Martin & Saffran, 1997). However, their approach could be problematic when considering WM tasks in which multiple items need to be recalled.

In order to illustrate the problems that their approach may face, we will take the computational approach of Haarmann & Usher (2001), who tested the theoretical plausibility of activation-based models of WM. In this model, items are activated within the "parietal cortex" which is supposed to be the locus where the long-term

memory representations remain. Once items are activated in the parietal cortex, they send activation to a "prefrontal cortex" system, whose function is to maintain the items active, thanks to self-excitatory connections. The model also includes lateral inhibitions between items, which naturally produces a capacity limitation of about 4 items. For illustration purpose, we implemented the model and run a simulation for one trial with basic parameters, whose behaviour across time is illustrated in **Figure 1.1**. This figure represents very well the typical time-course of activation occurring in the model. The model was initially implemented in order to account for semantic relatedness effects in free recall paradigms, and successfully does so thanks to inter-item excitatory connections between semantically related items. However, this model, in its actual form, is unable to produce realistic serial position curves when assessed in immediate serial recall conditions. This is due to the fact that all items tend to reach a similar stable activation level, a stable point which acts as an attractor. Hence, at the moment of recall, any active item has a similar chance to be equally recalled, because selection is based on item's activation level (i.e. the most active item is recalled first).

The model proposed by Haarmann and Usher faces several issues that models based on sustained activation need to address more generally. First, as already mentioned, items tend to be recalled in random order, because their activation level is very rapidly confounded, and because no serial order mechanism is available to track the position of each item. This first aspect underlies the importance of serial order coding for WM processing. Second, as can be seen in the figure, even though items at the beginning of the list receive more initial activation (this is a natural consequence of the inter-item inhibitions within the frontal cortex system), each newly activated item nonetheless receive a *stronger* amount of activation than the previous one. This is because during the time that one item is getting activated, the previous one decays to some extent. In this case, an item presented earlier has a higher probability to be forgotten, and this is due to the inter-item inhibitions within the frontal cortex system: items more strongly activated inhibit the less strongly activated ones. This observation contradicts the claim of N. Martin and Saffran, who proposed that the maintenance of serial order information may also be performed via an activation gradient which progressively encodes items with decreasing strength. In order for this statement to be sustainable, it must be assumed that each newly encoded item is *less strongly activated* than the currently decaying items previously presented, which is actually a strong assumption: when an item B is presented, can its activation level be weaker than the activation level of item A, which is currently decaying?

Obviously, these issues might be specific to the Haarmann & Usher approach, and the situation could be different in another implementation, or it may *not*. But as

can be seen, once formally implemented, the model produces behaviours that go beyond our mere intuitions. Importantly, note that this model is not strictly speaking a pure activation-based model; the prefrontal cortex system is indeed supposed to keep items active, otherwise they rapidly decay. To the best of our knowledge, there is no such things as a purely activation-based model in the literature, and this is probably because such models are not self-sufficient. Overall, activation-based models assuming sustained activation need to explicitly describe how items are maintained, because due to the temporal dimension of the sustained activation and constant feedback activation, the behaviour of these model is very complex.

At a more conceptual level, a purely activation-based position is interesting, because it minimizes the need for additional WM systems specialized for the retention of serial order, and is therefore a simpler and more parsimonious explanation to account for WM performance. However, this parsimony has also a cost in terms of flexibility; this type of model would be strictly limited to explain WM performance in immediate serial recall procedures. It would have difficulties to explain WM performance when using probed recall procedures in which participants are invited to recall an item cued from a specific serial position. In this situation, the system would require to recall or retrieve each item successively before recalling the probed item (e.g. retrieving items "1 2 3 4" before retrieving the cued item located in fifth position). Critically, the model would also struggle to account for a paradigm requiring participants to retrieve a position when presented with an item. For instance, given the sequence "ABCD", when presented with "B", the system would have no means to retrieve the information "position 2". If we assume that the probed letter "B" would simply reactivate its corresponding representation in long-term memory, "B" would be the most strongly activated item, and the system would assume that this item has been presented first. Another problem pointed out by Norris (2017) but also by R. C. Martin et al. (1999), and as already mentioned in the introduction, is the capacity of WM to represent items occurring more than one time in a given memory list. A pure activation-based model of WM would be unable to recall sequences composed of repeated items. Instead, the repeated item would receive the strongest activation and would be recalled only once, and this invariably at the beginning of the list. These examples show that a viable language-based WM account needs some additional mechanisms than mere language activation.

One important aspect that these models need to consider is serial order. Indeed, in Study 6, we have shown that linguistic knowledge interacts with the ability to maintain items in their serial order. As mentioned by Majerus (2019) serial order is a complex mechanism, involving several levels of representation simultaneously

(phoneme level, syllable level, word level, list level). It is therefore very challenging to account for serial order processing in an exhaustive manner. A good start could however be to represent order first at the whole list level, coding serial position for each individual item within a sequence composed of multiple memoranda. Many models providing detailed mechanisms for the coding of serial order information at the item level have been proposed (Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Page & Norris, 1998). One of the most interesting models here is the model by Burgess & Hitch (1999, 2006) which considers both specific serial order coding mechanisms and an at least rudimentary structure for the representation of linguistic knowledge. The model uses positional marking for coding serial position information, by assuming that each item, represented in the language structure of the system, is temporarily encoded to specific positional markers in a dynamic contextual system via Hebbian learning. This type of approach could potentially solve the issues that current activation-based models face, by considering that items are directly associated with (extra-linguistic) positional markers. However, such models, and this is also the case for the Burgess and Hitch model, do not yet represent in a detailed manner the activation occurring within the linguistic system. In the following section, we will present an outline of a computational model that incorporates both specific serial position coding mechanisms and a detailed interactive activation language system for accounting for the presence of psycholinguistic effects in WM tasks requiring the recall of multiple words.

A computational approach

In this section, we propose a preliminary draft of an architecture, whose purpose is to account in an explicit manner for the presence of psycholinguistic effects in WM, while also taking into account the problem of coding for serial order information. Following the conceptual framework proposed by Majerus, we will take into consideration the activation occurring within the linguistic system, via the inclusion of an adapted version of Dell's interactive activation model, but also serial order mechanisms, by adopting the key principles of the Burgess and Hitch model.

The model we propose is composed of three layers, including the phonological, lexical and semantic levels of language processing, with each node coding for a representation (i.e. one unit = one phoneme, one lexical representation or one semantic feature = localist connectionist architecture). The model also includes a fourth layer which contains the positional markers to which the items activated in the linguistic system will be associated to (see **Figure 1.2**). The model starts with the presentation of a word at the phonological level within the linguistic system, and this initial activation
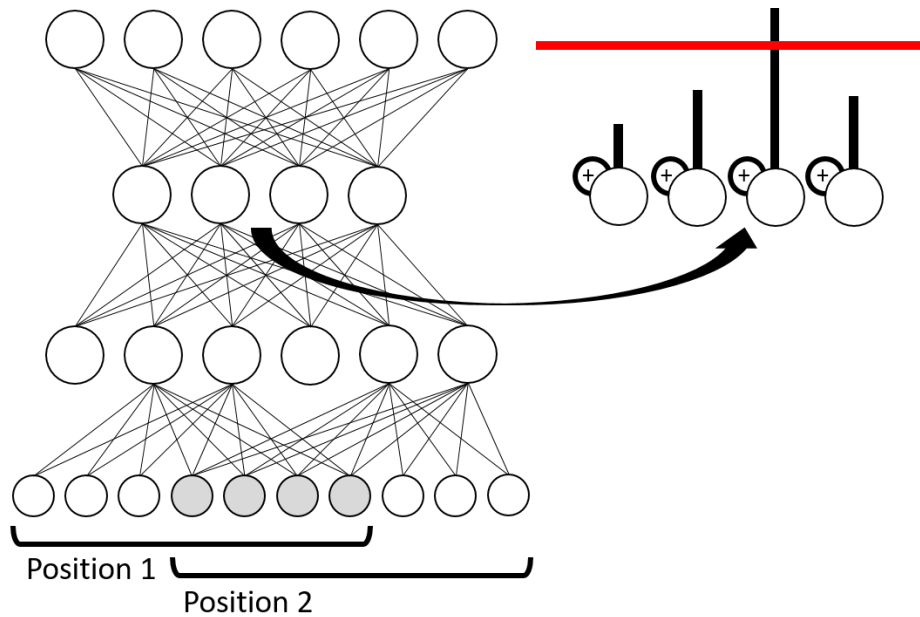
Figure 1.2. The general computational architecture proposed in this PhD thesis.

at the phonological level spreads throughout the system according to the interactive activation principles of Dell's model. As in the original model of Dell, the architecture is a hard-wired neural network limited to 6 items, with the connection weights connecting the nodes from one layer to another layer via a matrix multiplication (Kosko, 1988). Each time that an item is presented to the system, a specific set of positional markers representing the current position is activated. As in the Burgess and Hitch model, adjacent positions have overlapping positional markers. Each time an item and a set of positional markers are simultaneously activated, their associations are kept in memory via a Hebbian learning procedure. In the present implementation, the associations link the positional markers to the phonemes of the phoneme layer. Furthermore, across serial positions, the different items are associated to the positional markers with decreasing strength, via the implementation of a primacy gradient, and this in order to model realistic serial position curves. The recall phase is modelled by reactivating successively each set of positional markers, and the items associated to them are automatically re-activated thanks to the previously learned associations. For each position, an item is selected for output at the lexical level, and this via a simplified competitive cueing mechanism, whose function is to accumulate the activation over multiple iterations, and then select the most activated item after it reached a sufficient amount of activation. After an item has been recalled, it is suppressed via response suppression, by setting the weights that connect the item to the contextual nodes to a value of zero. See **Appendix** for a detailed and technical presentation of the computational architecture.
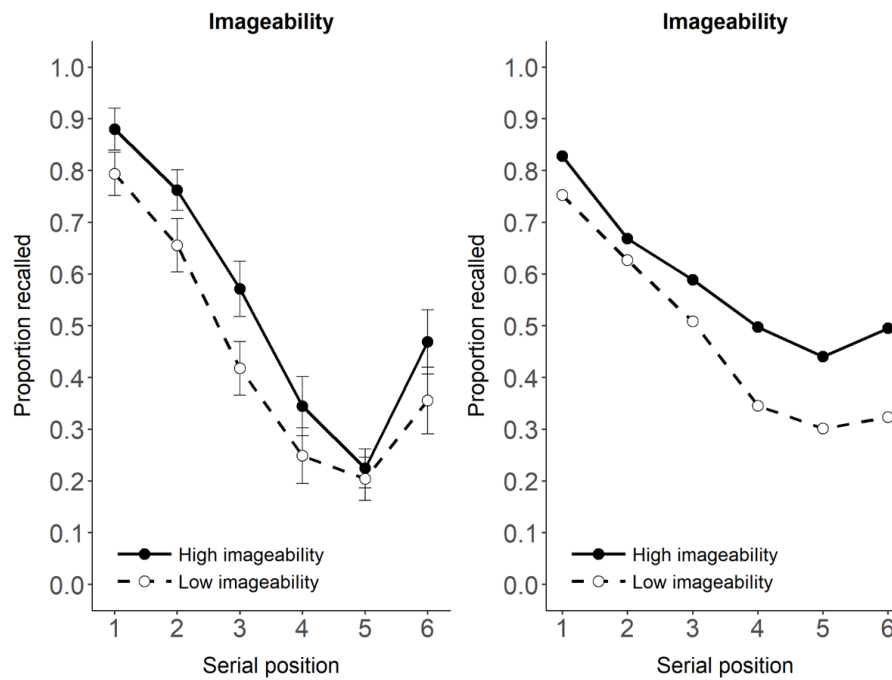
Figure 1.3. Recall performance for high vs. low imageability words across serial position, and this in the empirical data (left panel) and in the simulation (right panel).

In a concept-of-proof study, we determined the ability of this architecture to produce data that are comparable to those observed in empirical studies with human subjects. In order to address this question, we assessed the model's behaviour for two psycholinguistic effects: the imageability and semantic similarity effects. In a first step, we simulated performance for a neutral list condition (words not differing in imageabilty or semantic relatedness). The different parameters of the model were adjusted so that it produced realistic performance across serial position. In a second step, we changed the nature of items presented to the system, without changing the other different parameters used in the neutral condition. Hence, the only free parameters we manipulated in order to model psycholinguistic effects was the type of list that the system was presented with. Note that all simulations were performed over $N_{iterations}$ = 10,000 in order to minimize the random variability produced by the system. To model the imageability effect, the architecture was presented with lists composed of items whose lexical representations were associated to a larger number of semantic features. As can be seen in **Figure 1.3**, when compared to empirical data (left panel), the model successfully predicted that high imageability words would be better recalled than low imageability words (right panel). This is a direct consequence of the additional semantic features associated with high imageability words: when activated, they send more feedback activation to lower levels of representations, thereby increasing recall performance.

Next, the same simulation over recall performance was performed on the semantic similarity effect, by using a semantic grouping manipulation. This manipulation is equivalent to the semantically related condition as assessed in Study 6, in which semantically related items were presented by subgroups of three (e.g. leaf – tree – branch – cloud – sky – rain). In the simulation, semantic similarity was modelled by manipulating the amount of overlapping semantic features between semantically related and unrelated items. As can be seen in **Figure 1.4**, upper panel, the model successfully predicted the recall advantage for semantically related lists, as compared to neutral lists. This is due to the interactive activation occurring between related items, which reactivate each other via their shared semantic features. One further prediction is that since related items reactivate each other, they are also supposed to have similar activation levels at the moment of recall, as discussed in Study 6. Hence, according to these models, when a transposition error will occur, it should involve more often the serial position of a semantically related item than of an unrelated item. We further analysed the within-group transpositions and compared the results of the simulation with those observed in Study 6. As can be seen in **Figure 1.4**, lower panel, as more items have overlapping semantic representations, an increase of within-group transposition errors is observed. Note however that the impact of the semantic conditions was substantially smaller than in the empirical data, suggesting that other mechanisms might also be at play, in addition or instead of the mechanism we proposed.
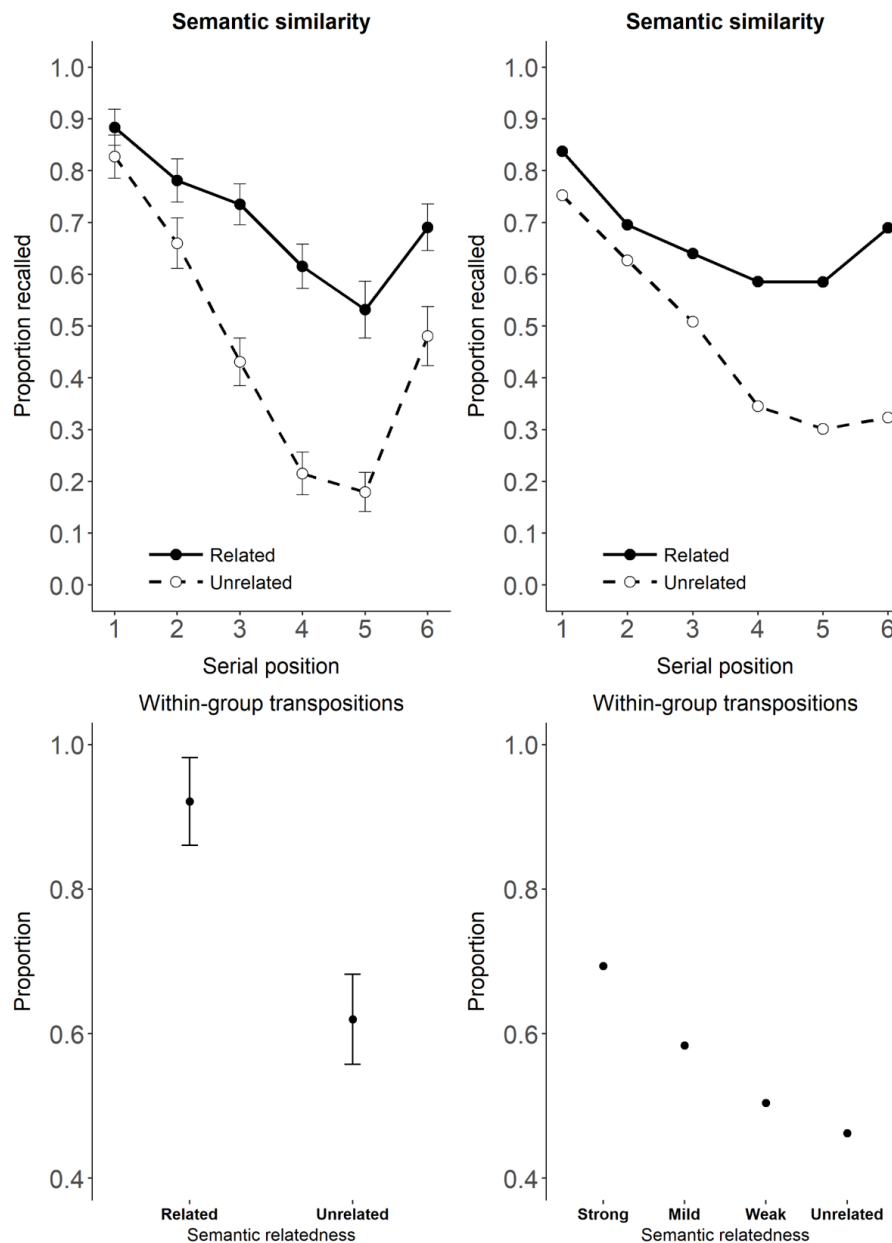
Figure 1.4. Upper panel: Recall performance for semantically related vs. semantically unrelated lists across serial position, and this in the empirical data (left panel) and in the simulation (right panel). Lower panel: Proportion of within-group transpositions for semantically related vs. semantically unrelated lists, and this in the empirical data (left panel) and in the simulation (right panel).

To sum up, the results of this simulation show that, when modelled according to the assumptions made by interactive activation models, the qualitative pattern of imageability and semantic similarity effects can be successfully captured in a simulation of a WM recall task. Critically, the model produces imageability effects not only for for overall recall performance, but also at the level of transposition error patterns. Despite these encouraging results, the architecture we propose here needs to be considered as preliminary as several important WM components are still missing.

A first important missing component of the model are the different mechanisms involved in attentional control and focalization. In Chapter 1, we have seen that a full understanding of WM implies also the consideration of attentional processes that have been shown to support WM performance. Furthermore, a major type of language-based models, Nelson Cowan's embedded processes model, considers that language-activation is not sufficient for accurate WM performance, but that attentional focalization furthermore interacts with activated language representations The role of attentional focalization on specific memoranda is not explicitly modelled or represented, but the role of this mechanism could be involved at different steps of the model: (1) the activation of items, (2) the reactivation of the positional markers at the moment of recall and (3) the selection of the most appropriate response for output. As already mentioned, we kept the architecture as simple as possible. In a further extension of the model, we could propose that items are refreshed and re-encoded during the inter-item intervals at list presentation by reactivating the positional markers, in order to simulate the role of attentional refreshing. It should be noted that in the present model, attention does not serve to keep items active, and the model is nevertheless able to simulate human verbal WM performance with a reasonable level of precision. Also note that in the computational model presented here, WM representations are supposed to be stored in a format that cannot be simply reduced to temporary activation in the linguistic network, as items are encoded via of the updating of connection weights between the linguistic system and the contextual nodes.

Furthermore, we need to acknowledge that we only simulated here two of the many psycholinguistic whose role in WM tasks has been demonstrated. The model could however be easily adapted to take into account other effects such as the lexical frequency effect. In computational terms, it has been proposed that this effect is due to high frequency words having higher resting activation levels or lower activation thresholds (McClelland & Elman, 1986), which means that they need a smaller total amount of input activation to be accessed. However, some models also consider that high frequency words might have stronger connection weights between lexical and phonological nodes (Besner & Risko, 2016), which means that they will be more strongly activated for an equal amount of phonological input. Critically, as already explained, the lexical frequency effect could also be modelled by implementing inter-item excitatory connections at the lexical level given that this effect is considered to stem at least partially from high frequency words co-occuring more often (Hulme et al., 2003). Moreover, Acheson & MacDonald (2009) made the suggestion that the phonological similarity, a hallmark effect for buffer accounts of verbal WM, might also

stem from the interactive activation occuring between phonological and lexical levels of processing directly within the linguistic system. Acheson and MacDonald proposed that the higher transposition errors observed for phonological similar versus dissimilar items may be caused by an over-activation happening at the lexical level, because phonologically similar words are also lexical neighbors (e.g. "CAT – BAT – FAT") and will therefore reactivate each other. At the moment of recall, because each item will be strongly activated, the system will more often fail to choose the right target, leading to increased transposition errors. At the same time, since the overall activation level will be higher across all items, the probability of producing omission errors will also decrease, thereby producing a recall advantage at the item level. In the same vein, the neighborhood density effect could also be potentially modelled in the present architecture, by presenting lists of words which have a higher number of phonological neighbors. Finally, the lexicality effect could not be properly simulated in the current implementation, because the selection of an item at the moment of recall always occurs at the lexical level. Hence, even when presented with a nonword, the system would always recall a word. However, by implementing the selection process at the phonological level, the model could potentially also recall nonwords, and hence should also be able to simulate the lexicality effect.

Conclusion

This PhD thesis aimed at obtaining a better understanding of the structure and functioning of verbal WM, by carrying out a detailed investigation of the role of linguistic knowledge within verbal WM. Through different behavioural and neuroimaging studies, we not only showed that WM interacts reliably with linguistic knowledge, but most critically, these interactions appear to be non-strategic and automatic. Our results allow to reject a pure strategic account of WM, whereby linguistic influences would be the exclusive product of elaborative strategic processes that participants implement during encoding. Our results also reject a post-encoding redintegration account for explaining the presence of psycholinguistic effects in WM. Our data indicate that, if such a redintegration mechanism exists, it must be very fast and already occurs during WM encoding. The outcome of our different studies is most parsimoniously explained by theoretical models considering that language processing is an integral part of verbal WM processing, and that activation occurring within the linguistic knowledge base provides the representational basis for WM maintenance. Our results challenge theoretical models that consider a strict distinction between verbal WM and language processing.

Several questions remain however unanswered. A pervasive problem in past studies has been the definition of the exact nature of the imageability effect. The present work highlights the strategic aspect of this effect. However, the exact form of strategic processes responsible for the imageability effect still remains unknown. Another open question is the problem of serial order, and the extent to which serial order information can be influenced by linguistic knowledge. Although we provided preliminary evidence for such interactions, the exact nature of these interactions warrants further research. The different questions identified at the closing of this thesis remain difficult to interpret within current theoretical framework, and call for the development of more integrated and detailed models of WM by further taking advantage of the theoretical precision of computational modelling techniques. A proof-of-concept, integrative model has been presented in this PhD thesis and should be developed in future research activities.

Appendix – Details of the computational working memory architecture.

The basic assumptions of the architecture is that verbal WM relies on activation within the linguistic system. As in the framework by Majerus and the approach by N. Martin and Saffran, the activation occurs in an interactive activation neural network, and more specifically as in Dell's neural network. The network is a three-layer neural network composed of phonological, lexical and semantic level of processing, and each level is composed of unitary units. The units of adjacent levels of representation connect to each other via bidirectional connection weights. However, only relevant connections are set to positive values, the other being set to 0. For instance, the phoneme nodes "C", "A" and "T" connect to the lexical unit "CAT" with a value of – for instance – .1, but with a value of 0 to the lexical unit "DOG". The network is currently a prototype which tests the overall plausibility of interactive activation models in order to account for the presence of psycholinguistic effects. For this reason, the network is currently limited to a restricted lexicon of 6 words, and this in order to keep the model very simple. The associations between levels of representation are hard-wired, or are simply learned via a matrix multiplication (Kosko, 1988). To begin with, the simulation starts by feeding the network with inputs values of .15 at the phonological level, and this during a fixed number of iterations. As soon as the network receives inputs at the phonological level, activation propagates between levels according to this formula:

$$(1) \; a_{j,t} = a_{j,t-1}\alpha + \sum w_{ij} \cdot a_{i,t-1}$$

In this formula, $a_{j,t}$ is the activation of unit j at time t, $a_{j,t-1}$ is the activation of unit j at time t-1, and α is a decay rate. The network also receives the sum of all activations coming from all units a_i at time t-1, scaled by their connection weights w_{ij}. Implemented in such way, the activated units behave as leaky integrators: their activation level is strictly dependent on the external input, and as soon as the network is no longer fed, their activation level rapidly drops off to a stable resting level. In **Figure A1**, upper panel, the time-course of activation of different words is represented, and this at the lexical level. As can be seen, when the word "CAT" is activated in the network, it also triggers to some extent the activation of the word "DOG". This is because these two words share common features at the semantic level. Even though "DOG" has never been presented, it also receives feedback activations from the semantic level, arising from the original lexical activation stemming from the word "CAT". Another consequence of the interactive activation principle is that semantic neighbors will also reinforce each other. Although the effect is mild in this example, it

can be seen that words with semantic neighbors will decay less rapidly, as compared to words with no or fewer semantic neighbors (i.e. the word "RIP"). This is the observed timecourse for one single item, but multiple items can be presented in a row, as it typically occurs in a working memory task. Two examples are presented in **Figure A1**, middle and lower panel. In the middle panel, the timecourse for two groups of semantically related word is displayed, while the lower panel shows what happens in terms of activation when a list of completely unrelated words is presented. As can be seen, related words, when presented successively, reactivate each other, and this activation is maintained over an extended period of time. At this point, the reader may wonder why activation within the system decays so rapidly, and why no active maintenance process occurs between each successive item. The potential role of attention will be discussed later, but in the series of simulation we will present, each item is assumed to be successively presented without a temporal pause, and this in order to keep the model as simple as possible. Hence, there is no active intervention of attentional refreshing or rehearsal in this situation; participants are just assumed to encode each item one by one.
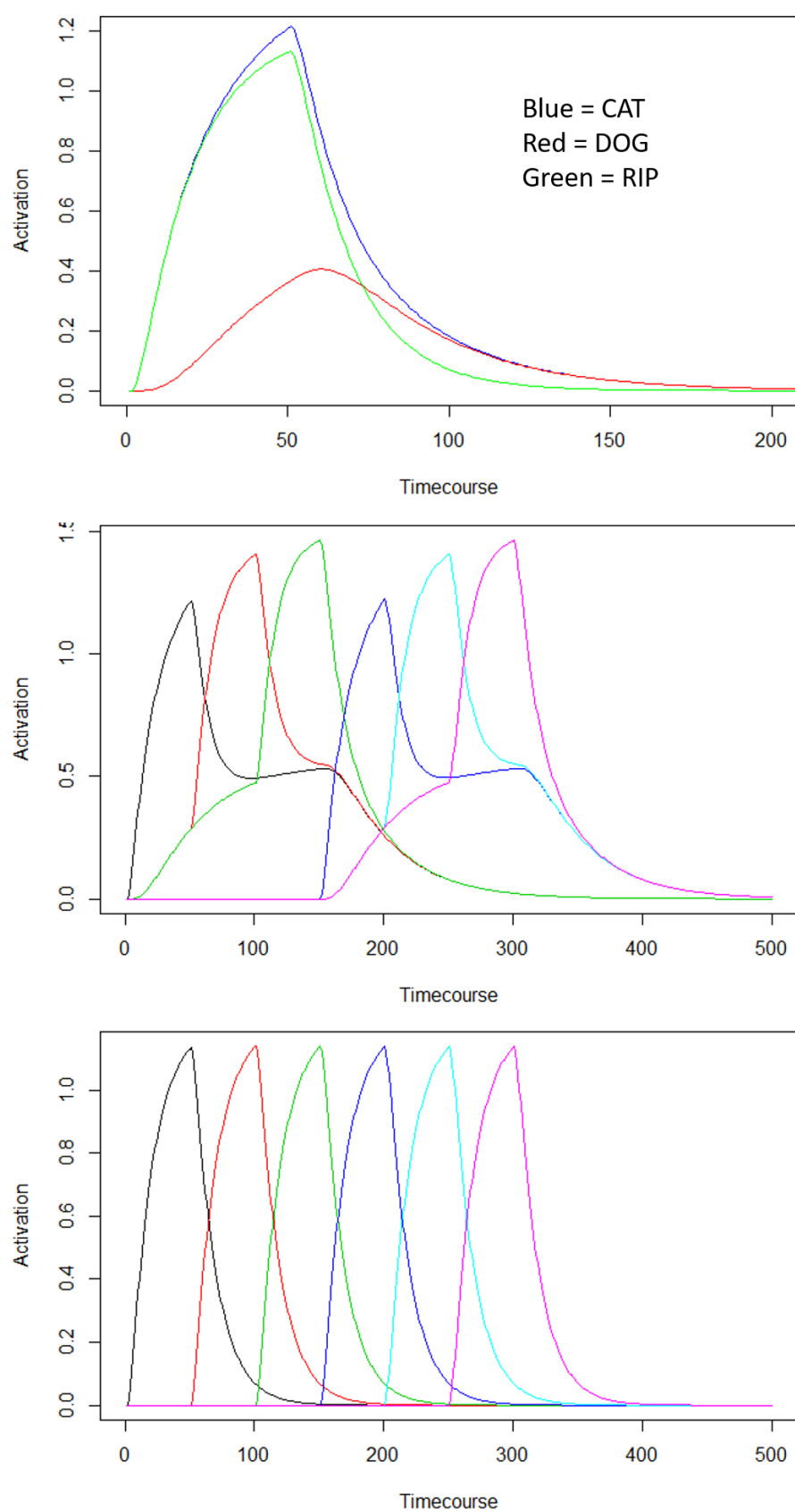
Figure A1. Timecourse of the activation occuring within the interactive activation neural network.

233

With such a strong effect of decay on linguistic representation, how does the system maintain items active? The point of this architecture is precisely that items do not need to remain in a highly activated state in order to be maintained in WM. Instead, as in the original model by Burgess and Hitch, they are kept in memory thanks to associations created between the items and contextual markers. Each time an item is presented to the system, specific positional markers are also activated. These markers are represented by a distributed set of nodes, whose values are set to 1 for a given position. Critically, the positional markers also share some degree of overlap, such that two adjacent positions will be represented by a similar number of positional units they have in common. At a computational level, the associations between the items and their positional markers are kept in memory by using Hebbian learning:

$$(2)\ \Delta w_{ij} = \left(1 - w_{ij}\right)\eta a_{i,t}a_{j,t}$$

Where w_{ij} is the weight connecting item i to the contextual node j, $a_{i,t}$ is the activation of item i at time t, and $a_{j,t}$ is the activation of the contextual node j at time t. The η term is a learning rate parameter, fixed across all simulations. Finally, the term $\left(1 - w_{ij}\right)$ guarantees that the association strength will not go beyond a maximum value of 1. The weights w_{ij} are then updated by adding the delta to the previous values. As can be seen, the learning rate between items and their contextual nodes critically depends on the items' activation level. That is, as items' activation level increase, so is the strength of associations. Similarly, if an item is not active at all, no association will be created. This phenomenon is illustrated in **Figure A2**. In this simulation, we gradually increased the learning rate, as if items were more strongly activated. It can be seen that as the learning rate increases, the more rapidly the associations will reach their asymptotic level. Note that in the current computational implementation, the associations are created between phonemes and the positional markers, because verbal WM is generally considered to maintain primarily phonological information. But the system could also create those associations based on lexical and semantic representations, or even the three levels at the same time without any trouble, if it is theoretically assumed that semantic knowledge are also encoded.

Implementing encoding in this way has two important consequences. First, items that are more strongly activated will also be more strongly associated to their contextual nodes. Hence, the model naturally predicts a recall advantage for items receiving stronger feedback activation. Second, items activated in the linguistic system decay quickly, and this steep decay function is *desired and required* in the model. If an item remains strongly active for too long, it will also be associated with a wrong
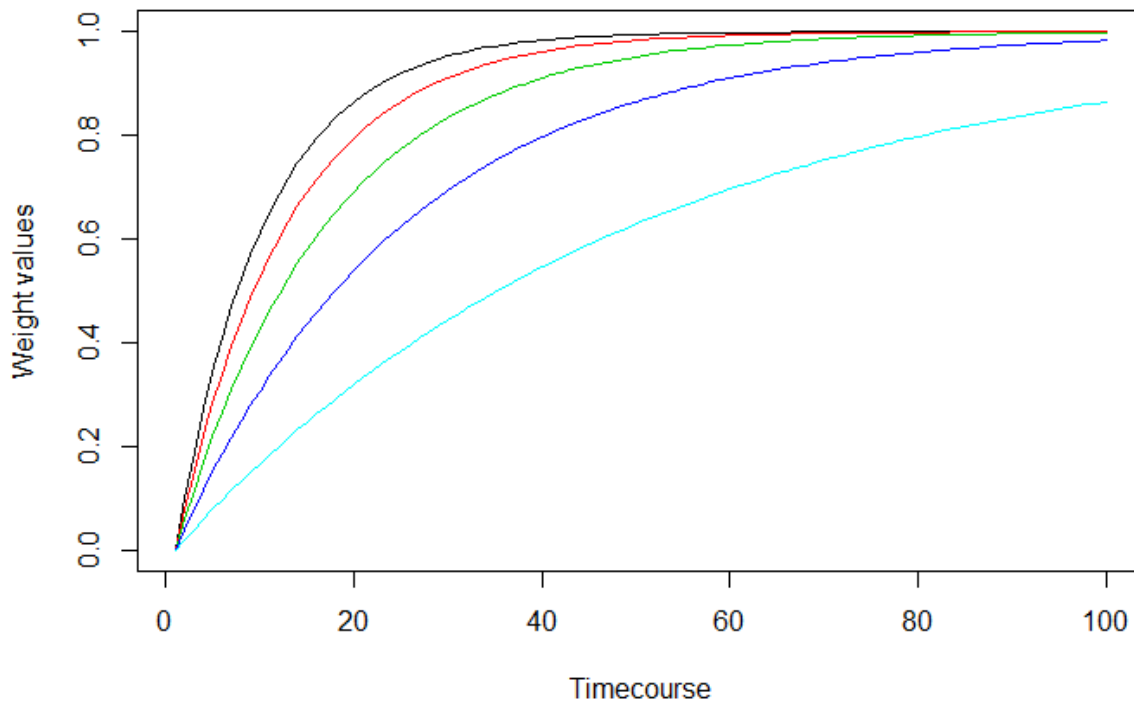
Figure A2. Evolution of the Hebbian learning across iterations. The different lines represent different learning rate.

contextual marker. It appears that, in this model, when items are activated, they are nevertheless already associated with a wrong contextual marker. Indeed, when an item is associated with position p, its activation rapidly decays, but not completely. Hence, because decay is not immediate, the remaining activation of the item will automatically create associations with the serial position p+1.

Now that items are encoded in WM, how is the system supposed to recall them? Recall is performed by activating directly the contextual nodes. This is made possible by successively reactivating the contextual nodes each time a new item needs to be recalled, as in the Burgess and Hitch approach. The activations within the contextual nodes are then directly transmitted to the items linked to them, via the connections weights created at the moment of WM encoding. As soon as one item is reactivated within the linguistic system, the activation is propagated via interactive activation through the different levels of language processing, as it does during encoding. Hence, formula (1) is simply re-used. The contextual layer is activated long as an item has not been selected for output. This means that the system needs a criterion in order to recognize when an item is considered to be recalled. In many models, simply the most activated item is selected (Page & Norris, 1998). But other models also implement a competitive queuing mechanism (Bullock & Rhodes, 2003; Hurlstone et al., 2014), in
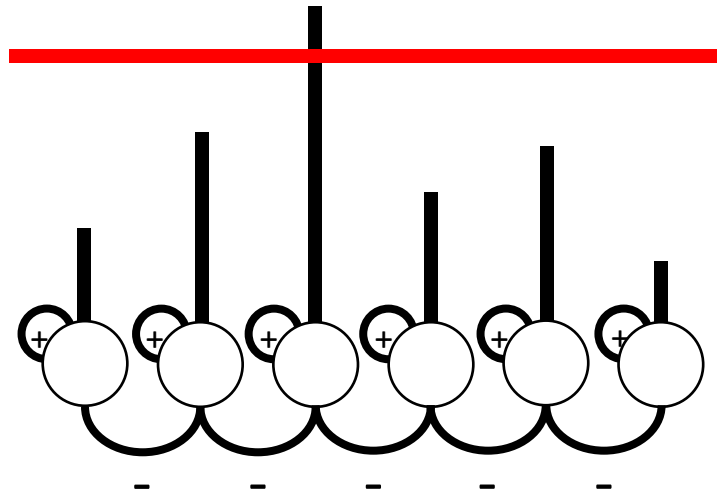
Figure A3. Architecture of a classical competitive queuing mechanism. Each nodes has a self-recurrent excitatory connection, and each node inhibit each other via lateral inhibitory connections. The rend bars represents the threshold required for output.

which the activation of items is transmitted within an accumulator, composed of nodes with self-excitatory connections and lateral inhibitory connections (see **Figure A3**). This mechanism automatically produces a winner-take-all phenomenon, in which the most activated item inhibits the less activated ones. The timecourse of the phenomenon is illustrated in **Figure A4**. The item is considered to be recalled when it reaches a threshold. If the threshold has not been reached after a certain amount of iterations, the item is considered to be forgotten, and the system will produce an omission error. In the best case scenario, the competitive queuing mechanism should be implemented, because it produces expected recall latencies across serial positions as observed in empirical data (Farrell & Lewandowsky, 2004). However, in the current simulation, we simply used an accumulator with self-excitatory connections, without the inter-item inhibitory connections, and this in order to reduce the complexity of the system, and because the focus was not on recall latencies. This implementation still allowed the system to produce omission errors (i.e. the participant recall "blank"), which is an important aspect since omission errors are the most common error observed in immediate serial recall tasks. Now that an item had been recalled, it must be suppressed in order for the system to avoid recalling the same item over again (Lewandowsky, 1999). In the current implementation, this is made possible by setting the connection weights between the contextual nodes and the associated item to 0.

At this point, an important aspect of the model has not been mentioned. What makes the model produce errors and forget? First, the model produces errors because a large quantity of noise ($\mu = 0$, $\sigma = .15$) is introduced within each node's activation level of the interactive activation network at each time step, and this during the
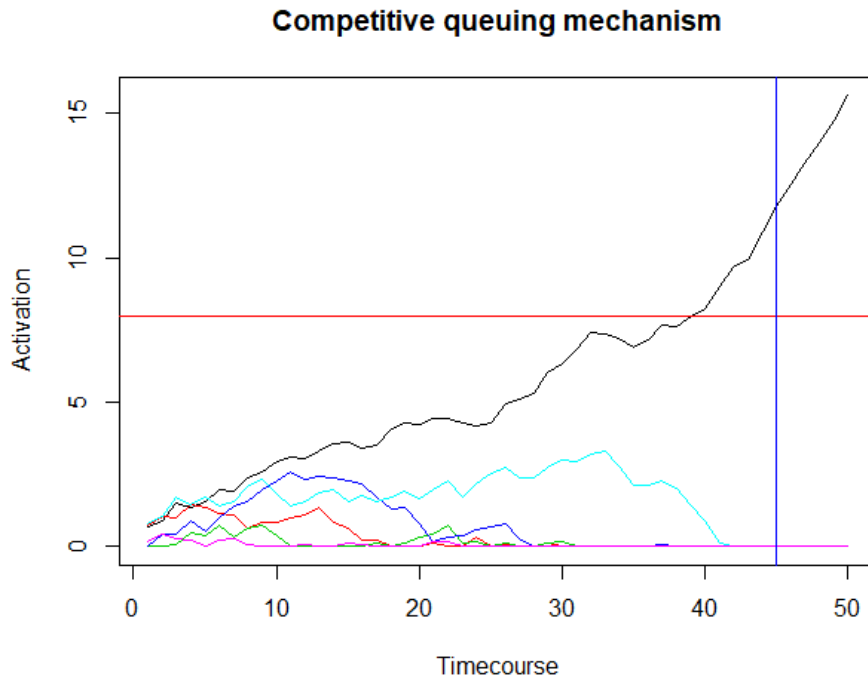
Competitive queuing mechanism

Figure A4. Timecourse of the activation within the competitive queuing mechanism. Over time, the most activated inhibit the other, less activation items, leading to a winner-take-all phenomenon. The red bar represents the threshold required for output, while the blue bar represents the number of iteration maximum required for output.

encoding and recall phases, in agreement with the original Dell's model. Second, in order to mimic forgetting, a primacy activation gradient (Lewandowsky, 1999; Page & Norris, 1998) modulates the learning rate at the moment of encoding, so that successive items are encoded with decreasing strength, according to the following function:

$$(3)\ L_j = cj^{-\lambda}$$

Where L_j corresponds to the value applied to the learning rate for a given position, j is the position of the item within the list, and c and λ are two free parameters, both set to 1. Note however that empirical evidence rather point to output interference as a plausible mechanism responsible for the primacy effect (Cowan et al., 1992, 2002), which could be implemented by adding noise to the weight matrix connecting the items to the positional markers. However, for the purpose of the current simulations, a primacy gradient is a very convenient way to make the model produce realistic serial position curves, while reducing its complexity and sparing computational resources.

References

Abrahamse, E., Dijck, J. Van, Majerus, S., & Fias, W. (2014). Finding the answer in space: the mental whiteboard hypothesis on serial order in working memory. *Frontiers in Human Neuroscience*, *8*. http://doi.org/10.3389/fnhum.2014.00932

Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, *23*(6), 1358–1367. http://doi.org/10.1162/jocn.2010.21519

Acheson, D. J., & MacDonald, M. C. (2009). Verbal Working Memory and Language Production : Common Approaches to the Serial Ordering of Verbal Information. *Psychological Bulletin*, *135*(1), 50–68. http://doi.org/10.1037/a0014411

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. http://doi.org/10.1037/a0017679

Albers, A. M., Kok, P., Toni, I., Dijkerman, C. H., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Current Biology*, *23*(15), 1427–1431. http://doi.org/10.1016/j.cub.2013.05.065

Allefeld, C., Görgen, K., & Haynes, J.-D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, *141*, 378–392. http://doi.org/10.1016/j.neuroimage.2016.07.040

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*(1), 64–88. http://doi.org/10.1016/j.jml.2006.02.002

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., & Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic space - Where and how variable? *NeuroImage*, *11*(1), 66–84. http://doi.org/10.1006/nimg.1999.0516

Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies : An event-related brain potential study. *Psychophysiology*, *32*(2), 177–190.

Andersson, J. L. R., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling Geometric Deformations in EPI Time Series. *NeuroImage*, *13*(5), 903–919. http://doi.org/10.1006/nimg.2001.0746

Andersson, L., Sandberg, P., Olofsson, J. K., & Nordin, S. (2018). Effects of task demands on olfactory, auditory, and visual event-related potentials suggest similar top-down modulation across senses. *Chemical Senses*, *43*(2), 129–134. http://doi.org/10.1093/chemse/bjx082

Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed Neural Representations of Phonological Features during Speech Perception. *Journal of Neuroscience*, *35*(2), 634–642. http://doi.org/10.1523/JNEUROSCI.2454-14.2015

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. http://doi.org/10.1016/j.neuroimage.2005.02.018

Asplund, C. L., Todd, J. J., Snyder, A. P., & Marois, R. (2010). A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nature Neuroscienceeuroscience*, *13*(4), 507–512. http://doi.org/10.1038/nn.2509.A

Atkins, A. S., & Reuter-lorenz, P. A. (2008). False working memories ? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*(1), 74–81. http://doi.org/10.3758/MC.36.1.74

Attout, L., Fias, W., Salmon, E., & Majerus, S. (2014). Common Neural Substrates for Ordinal Representation in Short-Term Memory, Numerical and Alphabetical Cognition. *PLoS One*, *9*(3), e92049. http://doi.org/10.1371/journal.pone.0092049

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, *18*(4), 362–365. http://doi.org/10.1080/14640746608400055

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory ? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D., & Andrade, J. (2000). Working Memory and the Vividness of Imagery. *Journal of Experimental Psychology: General*, *129*(1), 126–145.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, *36*(2), 233–252. http://doi.org/10.1080/14640748408402157

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The Multiple-Component Model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (Cambridge:, pp. 28–61). http://doi.org/10.1017/CBO9781139174909.005

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.

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*(1), 158–175. http://doi.org/10.3758/s13428-011-0123-7

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461. http://doi.org/10.3758/s13421-010-0034-3

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. http://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In *The cognitive neuroscience of working memory* (pp. 59–80).

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

References

http://doi.org/10.18637/jss.v067.i01

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, *58A*(4), 667–692. http://doi.org/10.1080/02724980443000142

Besner, D., & Davelaar, E. (1982). Basic Processes in Reading : Two Phonological Codes. *Canadian Journal of Psychology*, *36*(4), 701–711.

Besner, D., & Risko, E. F. (2016). Thinking Outside the Box When Reading Aloud: Between (Localist) Module Connection Strength As a Source of Word Frequency Effects. *Psychological Review*, *123*(5), 592–599.

Bettencourt, K. C., & Xu, Y. (2015). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150. http://doi.org/10.1038/nn.4174

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall : Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition*, *37*(5), 689–713. http://doi.org/10.3758/MC.37.5.689

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*(4), 1096–1108. http://doi.org/10.3758/s13423-015-0909-1

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex Cortex*, *19*(12), 2767–2796. http://doi.org/10.1093/cercor/bhp055

Binder, J. R., Frost, J. a, Hammeke, T. a, Bellgowan, P. S., Springer, J. a, Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528. http://doi.org/10.1093/cercor/10.5.512

Botto, M., Basso, D., Ferrari, M., & Palladino, P. (2014). When working memory updating requires updating : Analysis of serial position in a running memory task. *Acta Psychologica*, *148*, 123–129. http://doi.org/10.1016/j.actpsy.2014.01.012

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

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*(5), 467.

Brock, J., & Jarrold, C. (2005). Serial order reconstruction in Down syndrome: evidence for a selective deficit in verbal short-term memory. *Journal of Child Psychology and Psychiatry*, *46*(3), 304–316. http://doi.org/10.1111/j.1469-7610.2004.00352.x

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-Based Memory for Serial Order. *Psychological Review*, *107*(1), 127–181.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human Dorsal and Ventral Auditory Streams Subserve Rehearsal-Based and Echoic Processes during Verbal Working Memory. *Neuron*, *48*(4), 687–697. http://doi.org/10.1016/j.neuron.2005.09.029

Bullock, D., & Rhodes, B. (2003). Competitive queuing for serial planning and

performance. In *Handbook of brain theory and neural networks* (pp. 241–244).

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *The Quarterly Journal of Experimental Psychology*, *59*(10), 1691–700. http://doi.org/10.1080/17470210600848402

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167. http://doi.org/10.1023/A:1009715923555

Burgess, N., & Hitch, G. J. (1999). Memory for Serial Order : A Network Model of the Phonological Loop and Its Timing. *Psychological Review, 106*(3), 551–581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*(4), 627–652. http://doi.org/10.1016/j.jml.2006.08.005

Camos, V., & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: the executive and phonological loops. *Frontiers in Human Neuroscience*, *8*(900), 1–11. http://doi.org/10.3389/fnhum.2014.00900

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory (Hove, England)*, *16*(4), 329–340. http://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. http://doi.org/10.1080/17470218.2014.966248

Castellà, J., & Campoy, G. (2018). The (lack of) effect of dynamic visual noise on the concreteness effect in short-term memory. *Memory*, *0*(0), 1–9. http://doi.org/10.1080/09658211.2018.1476550

Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96–107. http://doi.org/10.1016/j.cognition.2016.05.024

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *199*(2), 417–430. http://doi.org/10.1037/a0027175.Competition

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. *The Journal of Neuroscience*, *32*(38), 12983–12989. http://doi.org/10.1523/JNEUROSCI.0184-12.2012

Chubala, C., Surprenant, A. M., Neath, I., & Quinlan, P. T. (2018). Does dynamic visual noise eliminate the concreteness effect in working memory ? *Journal of Memory and Language, 102*(May), 97–114. http://doi.org/10.1016/j.jml.2018.05.009

Clarkson, L., Roodenrys, S., Miller, L. M., & Hulme, C. (2016). The phonological neighbourhood effect on short-term memory for order. *Memory*, *25*(3), 391–402. http://doi.org/10.1080/09658211.2016.1179330

References

Collette, F., Majerus, S., Van der Linden, M., Dabe, P., Degueldre, C., Delfiore, G., … Salmon, E. (2001). Contribution of lexico-semantic processes to verbal short-term memory tasks: a PET activation study. *Memory (Hove, England)*, *9*(4–6), 249–259. http://doi.org/10.1080/09658210143000056

Coltheart, V. (1999). Comparing Short-term Memory and Memory for Rapidly Presented Visual Stimuli. *International Journal of Psychology*, *34*(5–6), 293–300.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word Familiarity and Frequency in Visual and Auditory Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex . Une base de données lexicales informatisée pour le français écrit et parlé. *L'année Psychologique*, *90*(4), 551–566.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). Review The Reorienting System of the Human Brain : From Environment to Theory of Mind, 306–324. http://doi.org/10.1016/j.neuron.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. http://doi.org/10.1038/nrn755

Corbin, L., & Marquer, J. (2009). Individual differences in Sternberg's memory scanning task. *Acta Psychologica*, *131*(2), 153–162. http://doi.org/10.1016/j.actpsy.2009.04.001

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-Based Translation: fMRI Decoding of Spoken Words in Bilinguals Reveals Language-Independent Semantic Representations in Anterior Temporal Lobe. *Journal of Neuroscience*, *34*(1), 332–338. http://doi.org/10.1523/JNEUROSCI.1302-13.2014

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

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). New York: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. http://doi.org/10.1017/S0140525X01003922

Cowan, N., Day, L., Saults, J. S., Keller, Tim, A., Johnson, T., & Flores, L. (1992). The Role of Verbal Output Time in the Effects of Word Length on Immediate Memory. *Journal of Memory and Language*, *31*, 1–17.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. http://doi.org/10.1016/j.cogpsych.2004.12.001

Cowan, N., Johnson, T. D., & Saults, S. (2005). Capacity limits in list item recognition: Evidence from proactive interference. *Memory*, *13*(3–4), 293–299. http://doi.org/10.1080/09658210344000206

Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., & Christ, S. E. (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*, *23*(10), 2852–63. http://doi.org/10.1162/jocn.2011.21625

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding Serial Recall. *Journal of Memory and Language*, *46*(1), 153–177. http://doi.org/10.1006/jmla.2001.2805

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning : neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3773–3800. http://doi.org/10.1098/rstb.2009.0111

Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, *93*(3), 283–321.

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.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. http://doi.org/10.1177/1745691611406920

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781), 1–17. http://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. http://doi.org/10.1016/j.jmp.2015.10.003

Dollaghan, C., Biber, M., & Campbell, T. F. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech, Language, and Hearing Research*, *36*(5), 1051–1054.

Dollaghan, C., & Campbell, T. F. (1998). Nonword Repetition and Child. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136–1146.

Dowd, E. W., Pearson, J. M., & Egner, T. (2015). Mind-reading without the scanner : Behavioural decoding of working memory content Mind-reading without the scanner : Behavioural, *6285*. http://doi.org/10.1080/13506285.2015.1093244

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

Egorova, N., Pulvermüller, F., & Shtyrov, Y. (2014). Neural dynamics of speech act comprehension: an MEG study of naming and requesting. *Brain Topography*, *27*(3), 375–92. http://doi.org/10.1007/s10548-013-0329-3

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. http://doi.org/10.1016/j.neuroimage.2004.12.034

References

Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory, *33*(15), 6516–6523. http://doi.org/10.1523/JNEUROSCI.5732-12.2013

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A Neural Measure of Precision in Visual Working Memory. *Journal of Cognitive Neuroscience, 25*(5), 754–761. http://doi.org/10.1162/jocn

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus Specific Mnemonic Representations during Visual Working Memory. *Neuron, 87*(4), 893–905. http://doi.org/10.1016/j.neuron.2015.07.013

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review, 19*(2), 325–331. http://doi.org/10.3758/s13423-011-0213-7

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2017). Seeing the Meaning : Top – Down Effects on Letter Identification, *8*(April), 1–13. http://doi.org/10.3389/fpsyg.2017.00322

Fallon, A. B., Mak, E., Tehan, G., & Daly, C. (2005). Lexicality and phonological similarity: A challenge for the retrieval-based account of serial recall? *Memory, 13*(3/4), 349–356. http://doi.org/10.1080/09658210344000215

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51*, 115–135. http://doi.org/10.1016/j.jml.2004.03.007

Fiebach, C. J., Friederici, A. D., Smith, E. E., & Swinney, D. (2007). Lateral Inferotemporal Cortex Maintains Conceptual—Semantic Representations in Verbal Working Memory. *Journal of Cognitive Neuroscience, 19*(12), 2035–2049.

Fiore, F., Borella, E., Mammarella, I. C., & De Beni, R. (2012). Age differences in verbal and visuo-spatial working memory updating: Evidence from analysis of serial position curves. *Memory, 20*(1), 14–27. http://doi.org/10.1080/09658211.2011.628320

Flegal, K. E., & Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory & Cognition, 42*(5), 701–711. http://doi.org/10.3758/s13421-013-0391-9

Foygel, D., & Dell, G. S. (2000). Models of Impaired Lexical Access in Speech Production, *216*, 182–216. http://doi.org/10.1006/jmla.2000.2716

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*(5), 262–268. http://doi.org/10.1016/j.tics.2012.04.001

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology, 23*(2), 250–254. http://doi.org/10.1016/j.conb.2012.10.002

Gagnepain, P., Gael, C., Landeau, B., Dayan, J., Eustache, F., & Lebreton, K. (2008). Spoken Word Memory Traces within the Human Auditory Cortex Revealed by

Repetition Priming and Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, *28*(20), 5281–5289. http://doi.org/10.1523/JNEUROSCI.0565-08.2008

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., & Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *NeuroImage*, *132*, 32–42. http://doi.org/10.1016/j.neuroimage.2016.02.033

Gaskell, M. G., & Marslen-Wilson, W. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84–95. http://doi.org/10.1037/0278-7393.25.3.562

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. http://doi.org/10.1080/027

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down Enhancement and Suppression of the Magnitude and Speed of Neural Activity. *Journal of Cognitive Neuroscience*, *17*(3), 507–517. http://doi.org/10.1162/0898929053279522

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of Automatic and Strategic Lexical-Semantics: Functional Magnetic Resonance Imaging Evidence for Differing Roles of Multiple Frontotemporal Regions. *The Journal of Neuroscience*, *26*(24), 6523–6532. http://doi.org/10.1523/JNEUROSCI.0808-06.2006

Gordon, J. K., & Dell, G. S. (2001). Phonological Neighborhood Effects: Evidence from Aphasia and Connectionist Modeling. *Brain and Language*, *79*(1), 21–31. http://doi.org/10.1006/brln.2001.2574

Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-Generality of Timing-Based Serial Order Processes in Short-Term Memory : New Insights from Musical and Verbal Domains. *PloS One*, *11*(12), 1–25. http://doi.org/10.1371/journal.pone.0168699

Graves, W. W., Boukrina, O., Mattheiss, S. R., Alexander, E. J., & Baillet, S. (2016). Reversing the Standard Neural Signature of the Word–Nonword Distinction. *Journal of Cognitive Neuroscience*, *29*(1), 79–94. http://doi.org/10.1162/jocn

Greene, C. M., & Soto, D. (2014). Functional connectivity between ventral and dorsal frontoparietal networks underlies stimulus-driven and working memory-driven sources of visual distraction. *Neuroimage*, *84*, 290–298. http://doi.org/10.1016/j.neuroimage.2013.08.060

Guérard, K., & Saint-Aubin, J. (2012). Assessing the Effect of Lexical Variables in Backward Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 312–324. http://doi.org/10.1037/a0025481

Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word Length, Set Size, and Lexical Factors: Re-Examining What Causes the Word Length Effect. *Journal of Experimental Psychology: Learning Memory and Cognition*. http://doi.org/10.1037/xlm0000551

Guitard, D., Saint-Aubin, J., Tehan, G., & Tolan, A. (2018). Does neighborhood size really cause the word length effect ? *Memory & Cognition*, (October 2017), 244–260.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology*, *56A*(7), 1213–1236. http://doi.org/10.1080/02724980343000071

Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall, and nonword learning. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 108–135). Hove, UK: Psychology Press.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*(6), 1001–1016.

Gupta, P., & MacWhinney, B. (1995). Is the Articulatory Loop Articulatory or Auditory? Reexamining the Effects of Concurrent Articulation on Immediate Serial Recall. *Journal of Memory and Language, 34*, 63–88.

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bulletin & Review, 8*(3), 568–578.

Hanley, R. J., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity : A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review, 10*(2), 435–444.

Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research, 1153*, 134–143. http://doi.org/10.1016/j.brainres.2007.03.069

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632–635. http://doi.org/10.1038/nature07832

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences : A model and tests of its stimulus-driven mechanism. *Cognitive Psychology, 87*, 135–178. http://doi.org/10.1016/j.cogpsych.2016.05.001

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*(1), 1–15. http://doi.org/10.3389/fnsys.2014.00035

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics, 6*(2), 107–128.

Henson, R. N. A. (1996). *Short-term memory for serial order*.

Henson, R. N. A. (1998). Short-Term Memory for Serial Order : The Start-End Model. *Cognitive Psychology, 36*(2), 73–137.

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition, 27*(5), 915–927.

Henson, R. N. A., 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,   56A(8),   1307–1334. http://doi.org/10.1080/02724980244000747

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402.

Hill, F., Korhonen, A., & Bentz, C. (2014). A Quantitative Empirical Analysis of the Abstract / Concrete Distinction. *Cognitive Science*, *38*(1), 162–177. http://doi.org/10.1111/cogs.12076

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal Grouping Effects in Immediate Recall: A Working Memory Analysis. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 116–139. http://doi.org/10.1080/713755609

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. http://doi.org/10.1080/14640747308400328

Hoffman, P., Jefferies, E., Ehsan, S., Jones, R. W., & Lambon Ralph, M. A. (2009). Semantic memory is key to binding phonology: Converging evidence from immediate serial recall in semantic dementia and healthy participants. *Neuropsychologia*,   *47*(3),   747–60. http://doi.org/10.1016/j.neuropsychologia.2008.12.001

Hoffman, P., Jefferies, E., Ehsan, S., Jones, R. W., & Lambon Ralph, M. A. (2012). How does linguistic knowledge contribute to short-term memory? Contrasting effects of impaired semantic knowledge and executive control. *Aphasiology*, *26*(3–4), 383–403. http://doi.org/10.1080/02687038.2011.581798

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.

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.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527–536.

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), 1217–1232.

Hulme, C., Stuart, G., Brown, G. D. A., & 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. http://doi.org/10.1016/S0749-596X(03)00096-2

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for Serial Order Across Domains: An Overview of the Literature and Directions for Future Research. *Psychological Bulletin*, *140*(2), 339. http://doi.org/10.1037/a0034221

References

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., & Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. *NeuroImage, 16*(1), 217–240. http://doi.org/10.1006/nimg.2001.1054

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, M. (2011). When Does Length Cause the Word Length Effect?, *37*(2), 338–353. http://doi.org/10.1037/a0021804

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect?, 1198–1210. http://doi.org/10.3758/s13421-011-0094-z

Jarrold, C., Cocksey, J., & Dockerill, E. (2008). Phonological similarity and lexicality effects in children's verbal short-term memory: Concerns about the interpretation of probed recall data. *The Quarterly Journal of Experimental Psychology, 61*(2), 324–340. http://doi.org/10.1080/17470210701202210

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006a). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language, 54*(1), 81–98. http://doi.org/10.1016/j.jml.2005.08.001

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006b). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *The Quarterly Journal of Experimental Psychology, 59*(5), 949–964. http://doi.org/10.1080/02724980543000141

Jefferies, E., Grogan, J., Mapelli, C., & Isella, V. (2012). Paced reading in semantic dementia: word knowledge contributes to phoneme binding in rapid speech production. *Neuropsychologia, 50*(5), 723–32. http://doi.org/10.1016/j.neuropsychologia.2012.01.006

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford, NY: Oxford University Press.

Jeffreys, H. (1998). The theory of probability. *Oxford, UK: Oxford University Press*.

Jones, G., & Macken, B. (2017). Long-term associative learning predicts verbal short-term memory performance. *Memory & Cognition*, 1–14. http://doi.org/10.3758/s13421-017-0759-3

Kalm, K., Davis, M. H., & Norris, D. (2012). Neural Mechanisms Underlying the Grouping Effect in Short-Term Memory. *Human Brain Mapping, 33*(7), 1634–1647. http://doi.org/10.1002/hbm.21308

Kalm, K., & Norris, D. (2014). The Representation of Order Information in Auditory-Verbal Short-Term Memory. *Journal of Neuroscience, 34*(20), 6879–6886. http://doi.org/10.1523/JNEUROSCI.4104-13.2014

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*(5), 679. http://doi.org/10.1016/j.surg.2006.10.010.Use

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A Controlled-Attention View of Working-Memory Capacity. *Journal of Experimental Psychology, 130*(2), 169–183.

Kelley, K. (2007). Confidence Intervals for Standardized Effect Sizes: Theory, Application, and Implementation. *Journal of Statistical Software, 20*(8), 1–24.

Kelley, K. (2017). MBESS (Version 4.0.0 and higher). Accessible from http://cran.r-

project.org.

Kocagoncu, E., Clarke, A., Devereux, B. J., & Tyler, L. K. (2017). Decoding the Cortical Dynamics of Sound-Meaning Mapping. *The Journal of Neuroscience, 37*(5), 1312–1319. http://doi.org/10.1523/JNEUROSCI.2858-16.2016

Kosko, B. (1988). Bidirectional Associative Memories. *IEEE Transaction on Systems, Man and Cybernetics, 18*(1), 49–60.

Kotz, S. A., Cappa, S. F., von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the Lexical – Semantic Network by Auditory Semantic Priming : An Event-Related Functional MRI Study. *NeuroImage, 17*(4), 1761–1772. http://doi.org/10.1006/nimg.2002.1316

Kowialiewski, B., & Majerus, S. (2018a). Testing the redintegration hypothesis by a single probe recognition paradigm. *Memory, 26*(9), 1256–1264. http://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. http://doi.org/10.1016/j.jml.2018.03.005

Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science, 6*(3), 299–312. http://doi.org/10.1177/1745691611406925

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package : Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. http://doi.org/10.18637/jss.v082.i13

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863), 1–12. http://doi.org/10.3389/fpsyg.2013.00863

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42–55. http://doi.org/10.1038/nrn.2016.150

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato 's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review, 104*(2), 211–240.

LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2016). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex, 27*(10), 4881–4890. http://doi.org/10.1093/cercor/bhw283

Leavitt, M. L., Mendoza-Halliday, D., & Martinez-Trujillo, J. C. (2017). Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends in Neurosciences, 40*(6), 328–346. http://doi.org/10.1016/j.tins.2017.04.004

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Pratical Course.* Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781139087759

Lee, S., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience, 16*(8), 997. http://doi.org/10.1038/nn.3452

References

Leff, A. P., Schofield, T. M., Crinion, J. T., Seghier, M. L., Grogan, A., Green, D. W., & Price, C. J. (2009). The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. *Brain*, *132*(12), 3401–3410. http://doi.org/10.1093/brain/awp273

Leonard, M. K., Baud, M. O., Sjerps, M. J., & Chang, E. F. (2016). Perceptual restoration of masked speech in human cortex. *Nature Communications*, *7*, 1–9. http://doi.org/10.1038/ncomms13619

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Levy, B. A. (1971). Role of Articulation in Auditory and Visual Short-Term Memory. *Journal of Verbal Learning and Verbal Behavior*, *10*(2), 123–132.

Lewandowsky, S. (1999). Redintegration and Response Suppression in Serial Recall: A Dynamic Network Model. *International Journal of Psychology*, *34*(5/6), 434–446.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*, *59*(2), 200–222. http://doi.org/10.1016/j.jml.2008.04.004

Lewandowsky, S., & Murdock, B. B. (1989). Memory for Serial Order. *Psychological Review*, *96*(1), 25–57.

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in Serial Recall : An Unworkable Solution to the Nonexistent Problem of Decay. *Psychological Review*, *122*(4), 674–699.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2011). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79.

Lewis-Peacock, J. A., & Postle, B. R. (2012). Decoding the internal focus of attention. *Neuropsychologia*, *50*(4), 470–478. http://doi.org/10.1016/j.neuropsychologia.2011.11.006

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. http://doi.org/10.1016/j.jesp.2013.03.013

Linden, D. E. J., Oosterhof, N. N., Klein, C., & Downing, P. E. (2012). Mapping brain activation and information during category-specific visual working memory. *Journal of Neurophysiology*, *107*(2), 628–639. http://doi.org/10.1152/jn.00105.2011

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wyn, V. (1996). Group aggregates and individual reliability : The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305–321. http://doi.org/10.3758/BF03213295

MacGregor, L. J., Pulvermuller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nat Commun*, *3*, 711. http://doi.org/10.1038/ncomms1715

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In

Interactions between short-term and long-term memory in the verbal domain (p. 244).

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

Majerus, S. (2019). Verbal working memory and the phonological buffer : The question of serial order. *Cortex*, 1–12. http://doi.org/10.1016/j.cortex.2018.04.016

Majerus, S., & Attout, L. (2018). *Working Memory for Serial Order and Numerical Cognition : What Kind of Association ? Heterogeneity of Function in Numerical Cognition*. Elsevier Inc. http://doi.org/10.1016/B978-0-12-811529-9.00019-4

Majerus, S., Attout, L., Artielle, M.-A., & Van der Kaa, M.-A. (2015). The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia*, *77*, NSYD1500380. http://doi.org/10.1016/j.neuropsychologia.2015.08.010

Majerus, S., Attout, L., D'Argembeau, A., Degueldre, C., Fias, W., Maquet, P., … Balteau, E. (2012). Attention supports verbal short-term memory via competition between dorsal and ventral attention networks. *Cerebral Cortex*, *22*(5), 1086–1097. http://doi.org/10.1093/cercor/bhr174

Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. *Cerebral Cortex*, *26*(1), 166–179. http://doi.org/10.1093/cercor/bhu189

Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Linden, M. Van Der, & Collette, F. (2010). The Commonality of Neural Networks for Verbal and Visual Short-term Memory. *Journal of Cognitive Neuroscience*, *22*(11), 2570–2593.

Majerus, S., Lekeu, F., de Linden, M., & Salmon, E. (2001). Deep dysphasia: Further evidence on the relationship between phonological short-term memory and language processing impairments. *Cognitive Neuropsychology*, *18*(5), 385–410. http://doi.org/10.1080/02643290126060

Majerus, S., Martinez, T., & Oberauer, K. (2012). Two distinct origins of long-term learning effects in verbal short-term memory. *Journal of Memory and Language*, *66*(1), 38–51. http://doi.org/10.1016/j.jml.2011.07.006

Majerus, S., Norris, D., & Patterson, K. (2007). What does a patient with semantic dementia remember in verbal short-term memory? Order and sound but not words. *Cognitive Neuropsychology*, *24*(2), 131–151. http://doi.org/10.1080/02643290600989376

Majerus, S., Péters, F., Bouffier, M., Cowan, N., & Phillips, C. (2017). The Dorsal Attention Network Reflects Both Encoding Load and Top–down Control during Working Memory. *Journal of Cognitive Neuroscience*, *30*(2), 144–159. http://doi.org/10.1162/jocn

Majerus, S., & Van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, *21*(2), 303–310.

Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological

language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*, *51*(2), 297–306. http://doi.org/10.1016/j.jml.2004.05.002

Majerus, S., Van der Linden, M., Poncelet, M., & Metz-Lutz, M.-N. N. (2004). Can phonological and semantic short-term memory be dissociated? Further evidence from Landau-Kleffner syndrome. *Cognitive Neuropsychology*, *21*(5), 491–512. http://doi.org/10.1080/02643290342000104

Mallett, R., & Lewis-peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention, *1424*, 256–267. http://doi.org/10.1111/nyas.13647

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. http://doi.org/10.1016/0010-0277(87)90005-9

Martin, N. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language*, *43*(2), 240–274. http://doi.org/10.1016/0093-934X(92)90130-7

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of Paraphasias in Deep Dysphasia: Testing the Consequences of a Decay Impairment to an Interactive Spreading Activation Model of Lexical Retrieval. *Brain and Language*, *47*(4), 609–660.

Martin, N., Minkina, I., Kohen, F. P., & Kalinyak-Fliszar, M. (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neurolinguistics*, *48*(February), 199–225. http://doi.org/10.1016/j.jneuroling.2018.02.006

Martin, N., & Saffran, E. M. (1990). Repetition and Verbal STM in Transcortical Sensory Aphasia : A Case Study. *Brain and Language*, *39*(2), 254–288.

Martin, N., & Saffran, E. M. (1997). Language and Auditory- verbal Short-term Memory Impairments : Evidence for Common Underlying Processes. *Cognitive Neuropsychology*, *14*(5), 641–682. http://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., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*(1), 3–29.

Martinez Perez, T., Majerus, S., & Mahot, A. (2012). Evidence for a Specific Impairment of Serial Order Short-term Memory in Dyslexic Children. *Dyslexia*, *18*(2), 94–109. http://doi.org/10.1002/dys.1438

Martinez Perez, T., Poncelet, M., Salmon, E., & Majerus, S. (2015). Functional Alterations in Order Short-Term Memory Networks in Adults With Dyslexia. *Developmental Neuropsychology*, *40*(7–8), 407–429. http://doi.org/10.1080/87565641.2016.1153098

Martinez Perez, T., Steve, M., & Martine, P. (2013). Research in Developmental Disabilities Impaired short-term memory for order in adults with dyslexia.

Research in Developmental Disabilities, *34*(7), 2211–2223. http://doi.org/10.1016/j.ridd.2013.04.005

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame : An open-source , graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. http://doi.org/10.3758/s13428-011-0168-7

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and Complex Working Memory Tasks Allow Similar Benefits of Information Compression. *Journal of Cognition*, *1*(1), 1–12.

Matsuyoshi, D., Ikeda, T., Sawamoto, N., Kakigi, R., Fukuyama, H., & Osaka, N. (2010). Neuropsychologia Task-irrelevant memory load induces inattentional blindness without temporo-parietal suppression. *Neuropsychologia*, *48*(10), 3094–3101. http://doi.org/10.1016/j.neuropsychologia.2010.06.021

Mattheiss, S. R., Levinson, H., & Graves, W. W. (2018). Duality of Function: Activation for Meaningless Nonwords and Semantic Codes in the Same Brain Areas. *Cerebral Cortex*, *28*(7), 2516–2524. http://doi.org/10.1093/cercor/bhy053

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375.

McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia*, *41*(3), 293–303. http://doi.org/10.1016/S0028-3932(02)00162-8

Melnik, N., Mapelli, I., & Özkurt, T. E. (2017). Modulation of alpha oscillations is required for the suppression of semantic interference to avoid false memories in short term. *Neurobiology of Learning and Memory*. http://doi.org/10.1016/j.nlm.2017.05.007

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature encoding in Human Superior Temporal Gyrus. *Science*, *343*(6174), 1006–1010. http://doi.org/10.1126/science.1245994

Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., … Krupa, P. (2008). Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*, *26*(4), 490–503. http://doi.org/10.1016/j.mri.2007.08.006

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, *37*(6), 850–865. http://doi.org/10.3758/MC.37.6.850

Miller, L. M., & Roodenrys, S. (2012). Serial Recall, Word Frequency, and Mixed Lists: The Influence of Item Arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1731–1740. http://doi.org/10.1037/a0028470

Monnier, C., & Bonthoux, F. (2011). The semantic-similarity effect in children:

influence of long-term knowledge on verbal short-term memory. *British Journal of Developmental Psychology*, *29*(4), 929–941. http://doi.org/10.1111/j.2044-835X.2010.02024.x

Moore, M. W., Tompkins, C. A., & Dollaghan, C. A. (2016). Manipulating articulatory demands in non-word repetition: A 'late-8' non-word repetition task. *Clinical Linguistics & Phonetics*, *24*(12), 997–1008. http://doi.org/10.3109/02699206.2010.510917

Moore, T. M., Reise, S. P., Depaoli, S., & Haviland, M. G. (2015). Iteration of Partially Specified Target Matrices: Applications in Exploratory and Bayesian Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *50*(2), 149–161. http://doi.org/10.1080/00273171.2014.973990

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Morey, R. D., & Rouder, J. N. (2014). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.2+).

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*(6), 922–936. http://doi.org/10.3758/s13421-016-0608-9

Moseley, R. L., Pulvermu, F., & Shtyrov, Y. (2013). Sensorimotor semantics on the spot: brain activity dissociates between conceptual categories within 150 ms. *Scientific Reports*, *3*, 1–7. http://doi.org/10.1038/srep01928

Murakami, T., Kell, C. A., Restle, J., Ugawa, Y., & Ziemann, U. (2015). Left Dorsal Speech Stream Components and Their Contribution to Phonological Processing. *The Journal of Neuroscience*, *35*(4), 1411–1422. http://doi.org/10.1523/JNEUROSCI.0246-14.2015

Murdock, B. B. (1976). Item and Order Information in Short-Term Serial Memory. *Journal of Experimental Psychology: General*, *105*(2), 191–216.

Murphy, C., Rueschemeyer, S.-A., Watson, D., Karapanagiotidis, T., Smallwood, J., & Jefferies, E. (2017). Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. *NeuroImage, 147*(August 2016), 19–31. http://doi.org/10.1016/j.neuroimage.2016.11.067

Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition*, *35*(8), 1940–1953.

Nee, D. E., & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory : Evidence for a 3-state model of memory. *NeuroImage, 54*(2), 1540–1548. http://doi.org/10.1016/j.neuroimage.2010.09.002

Nee, D. E., & Jonides, J. (2013). Neural evidence for a 3-state model of visual short-term memory. *Neuroimage, 74*, 1–11. http://doi.org/10.1016/j.neuroimage.2013.02.019

New, B. (2006). Lexique 3: Une nouvelle base de données lexicales. *Actes de La Conférence Traitement Automatique Des Langues Naturelles*.

Newman, S. D., & Twieg, D. (2001). Differences in Auditory Processing of Words and Pseudowords: An fMRI Study. *Human Brain Mapping*, *14*, 39–47. http://doi.org/10.1002/hbm.

Nimmo, L. M., & Roodenrys, S. (2005). The phonological similarity effect in serial recognition. *Memory*, *13*(7), 773–784. http://doi.org/10.1080/09658210444000386

Norris, D. (2017). Short-Term Memory and Long-Term Memory are Still Different. *Psychological Bulletin*, *143*(9), 992.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. http://doi.org/10.1037//0278-7393.28.3.411

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, *49*(4), 469–483. http://doi.org/10.1016/S0749-596X(03)00080-9

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What Limits Working Memory Capacity ? *Psychological Bulletin*, *142*(7), 758–799.

Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. *Psychological Science*, *21*(3), 164–169. http://doi.org/10.1177/0963721412444727

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10–45. http://doi.org/10.3758/s13423-010-0020-6

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *J Exp Psychol Gen*, *142*(2), 380–411. http://doi.org/10.1037/a0029588

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A. R. A., Cowan, N., … Ward, G. (2018). Benchmarks for models of short term and working memory. *Psychological Bulletin*, 1–154. http://doi.org/10.1037/h0021468

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. http://doi.org/10.3758/s13423-012-0272-4

Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural Response Suppression Predicts Repetition Priming of Spoken Words and Pseudowords. *Journal of Cognitive Neuroscience*, *18*(8), 1237–1252.

Page, M. P. A., & Norris, D. (1998). The Primacy Model : A New Model of Immediate Serial Recall. *Psychological Review*, *105*(4), 761–781.

Palladino, P., & Jarrold, C. (2008). Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *The Quarterly Journal of Experimental Psychology*, *61*(3), 392–399. http://doi.org/10.1080/17470210701664989

Papagno, C., Comi, A., Riva, M., Bizzi, A., Vernice, M., Casarotti, A., … Bello, L. (2017). Mapping the Brain Network of the Phonological Loop. *Human Brain Mapping*, *00*, 00–00. http://doi.org/10.1002/hbm.23569

Papagno, C., Vernice, M., & Cecchetto, C. (2013). Phonology without semantics? Good enough for verbal short-term memory. Evidence from a patient with semantic dementia. *Cortex*, *49*(3), 626–636. http://doi.org/10.1016/j.cortex.2012.04.015

Parmentier, F. B. R., Comesaña, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *The Quarterly Journal of Experimental Psychology*, *70*(1), 1–17. http://doi.org/10.1080/17470218.2015.1105268

Patterson, K., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience, 6*(1), 57–69. http://doi.org/10.1162/jocn.1994.6.1.57

Peters, B., Kaiser, J., Rahm, B., & Bledowski, C. (2015). Activity in Human Visual and Parietal Cortex Reveals Object-Based Attention in Working Memory. *The Journal of Neuroscience, 35*(8), 3360–3369. http://doi.org/10.1523/JNEUROSCI.3795-14.2015

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The Impact of Feedback Semantics in Visual Word Recognition: Number of Features Effects in Lexical Decision and Naming Tasks. *Psychonomic Bulletin & Review, 9*(3), 542–549.

Plancher, G., Bernard, C., Tillmann, B., Neuroscience, L., & Bernard, C. (2018). Boosting Maintenance in Working Memory With Temporal Regularities, *44*(5), 812–818.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48*(2), 384–404. http://doi.org/10.1080/14640749508401396

Poirier, M., & Saint-Aubin, J. (1996). Immediate Serial Recall, Word Frequency, Item Identity and Item Position. *Canadian Journal of Experimental Psychology, 50*(4), 408–412.

Poirier, M., & Saint-Aubin, J. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. *Memory, 13*(3–4), 325–332. http://doi.org/10.1080/09658210344000369

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. http://doi.org/10.3758/s13421-014-0470-6

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57*(3), 137–146. http://doi.org/10.1037/h0046137

Poncelet, M., & Van der Linden, M. (2003). L'évaluation du stock phonologique de la mémoire de travail. *Revue de Neuropsychologie*. Retrieved from papers3://publication/uuid/464CDE49-F0A0-4732-99FE-27857DD71D75

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509–522. http://doi.org/10.1037//0278-7393.2.5.509

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics, 76*(2), 270–279. http://doi.org/10.3758/s13414-013-0605-z

Python, G., Fargier, R., & Laganaro, M. (2018). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex, 99*, 1–12. http://doi.org/10.1016/j.cortex.2017.09.008

Quinlan, P. T., Roodenrys, S., & Miller, L. M. (2017). Serial reconstruction of order and serial recall in verbal short-term memory. *Memory & Cognition*, 1–18. http://doi.org/10.3758/s13421-017-0719-y

Quinn, J. G., & McConnel, J. (1996). Irrelevant Pictures in Visual Working Memory. *The Quarterly Journal of Experimental Psychology, 49A*(1), 200–215.

R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org.

Raettig, T., & Kotz, S. A. (2008). Auditory processing of different types of pseudo-words: an event-related fMRI study. *NeuroImage*, *39*(3), 1420–1428. http://doi.org/10.1016/j.neuroimage.2007.09.030

Ravizza, S. M., Hazeltine, E., Ruiz, S., & Zhu, D. C. (2011). Left TPJ activity in verbal working memory: Implications for storage- and sensory-specific models of short term memory. *NeuroImage*, *55*(4), 1836–1846. http://doi.org/10.1016/j.neuroimage.2010.12.021

Restle, J., Murakami, T., & Ziemann, U. (2012). Facilitation of speech repetition accuracy by theta burst stimulation of the left posterior inferior frontal gyrus. *Neuropsychologia*, *50*(8), 2026–2031. http://doi.org/10.1016/j.neuropsychologia.2012.05.001

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related FMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, *15*, 1160–1175. http://doi.org/10.1162/089892903322598120

Ritchie, G., Tolan, G. A. T., & Tehan, G. (2015). Redintegration, Task Difficulty, and Immediate Serial Recall Tasks. *Canadian Journal of Experimental Psychology*, *69*(1), 54–63.

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. http://doi.org/10.1080/17470210601147747

Roodenrys, S., & Hinton, M. (2002). Sublexical or Lexical Effects on Serial Recall of Nonwords ? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 29–33. http://doi.org/10.1037//0278-7393.28.1.29

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-Frequency and Phonological-Neighborhood Effects on Verbal Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019–1034. http://doi.org/10.1037//0278-7393.28.6.1019

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, *8*(2), 71–78. http://doi.org/10.1080/096582100387623

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, *18*(4), 682–689. http://doi.org/10.3758/s13423-011-0088-7

Roxbury, T., McMahon, K., & Copland, D. A. (2014). An fMRI study of concreteness effects in spoken word recognition. *Behavioral and Brain Functions*, *10*(1), 34. http://doi.org/10.1186/1744-9081-10-34

Ruchkin, D. S., Berndt, R. S., Johnson, R., Grafman, J., & Canoune, H. L. (1999). Lexical Contributions to Retention of Verbal Information in Working Memory : Event-Related Brain Potential Evidence. *Journal of Memory and Language*, *364*, 345–364.

Ruiz, R. M., & Elosúa, M. R. (2013). Conditions for positive and negative recencies in

running memory-span recognition. *Acta Psychologica*, *144*(2), 213–223. http://doi.org/10.1016/j.actpsy.2013.07.006

Ruiz, R. M., Elosúa, M. R., & Lechuga, M. T. (2005). Old-fashioned responses in an updating memory task. *The Quarterly Journal of Experimental Psychology*, *58a*(5), 887–908. http://doi.org/10.1080/02724980443000395

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *NeuroImage*, *39*(3), 1444–1456. http://doi.org/10.1016/j.neuroimage.2007.09.052

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, *27*(1), 188–200. http://doi.org/10.1016/j.neuroimage.2005.04.012

Sachs, O., Weis, S., Zellagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., & Kircher, T. (2008). Automatic processing of semantic relations in fMRI: neural activation during semantic priming of taxonomic and thematic categories. *Brain Res*, *1218*, 194–205. http://doi.org/10.1016/j.brainres.2008.03.045

Sadat, J., Martin, C. D., Costa, A., & Alario, F. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. http://doi.org/10.1016/j.cogpsych.2013.10.001

Saint-Aubin, J., Guérard, K., Chamberland, C., & Malenfant, A. (2014). Delineating the contribution of long-term associations to immediate recall. *Memory*, *22*(4), 360–373. http://doi.org/10.1080/09658211.2013.794242

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials. *Psychonomic Bulletin & Review*, *12*(1), 171–177. http://doi.org/10.3758/BF03196364

Saint-Aubin, J., & Poirier, M. (1999a). Semantic Similarity and Immediate Serial Recall: Is There a Detrimental Effect on Order Information? *The Quarterly Journal of Experimental Psychology*, *52A*(2), 367–394. http://doi.org/10.1080/713755814

Saint-Aubin, J., & Poirier, M. (1999b). The influence of Long-term Memory Factors on Immediate Serial Recall: An Item and Order Analysis. *International Journal of Psychology*, *34*(5–6), 347–352.

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords : Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, *7*(2), 332–340.

Savill, N., Ellis, A. W., & Jefferies, E. (2016). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia*, *98*, 1–13. http://doi.org/10.1016/j.neuropsychologia.2016.03.006

Savill, N. J., Cornelissen, P., Pahor, A., & Jefferies, E. (2018). rTMS evidence for a dissociation in short-term memory for spoken words and nonwords. *Cortex*, 1–18. http://doi.org/10.1016/j.cortex.2018.07.021

Savill, N., Metcalfe, T., Ellis, A. W., & Jefferies, E. (2015). Semantic categorisation of a word supports its phonological integrity in verbal short-term memory. *Journal of Memory and Language*, *84*, 128–138. http://doi.org/10.1016/j.jml.2015.06.003

Schleepen, T. M. J., Markus, C. R., & Jonkman, L. M. (2014). Dissociating the effects of semantic grouping and rehearsal strategies on event-related brain potentials. *International Journal of Psychophysiology*, *94*(3), 319–328. http://doi.org/10.1016/j.ijpsycho.2014.09.007

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential Hypothesis Testing With Bayes Factors : Efficiently Testing Mean Differences. *Psychological Methods*, *22*(2), 322. http://doi.org/10.1037/met0000061

Schrouff, J., Kussé, C., Wehenkel, L., Maquet, P., & Phillips, C. (2012). Decoding semi-constrained brain activity from fMRI using support vector machines and Gaussian Processes. *PLoS ONE*, *7*(4), 1–11. http://doi.org/10.1371/journal.pone.0035860

Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, a. F., Chu, C., Ashburner, J., … Mourão-Miranda, J. (2013). PRoNTo: Pattern Recognition for Neuroimaging Toolbox. *Neuroinformatics*, *11*, 319–337. http://doi.org/10.1007/s12021-013-9178-1

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*(2), 168–175.

Schweickert, R., Chen, S., & Poirier, M. (1999). Redintegration and the Useful Lifetime of the Verbal Memory Representation. *International Journal of Psychology*, *34*(5–6), 447–454.

Shtyrov, Y., & Lenzen, M. (2017). First-pass neocortical processing of spoken language takes only 30 msec : Electrophysiological evidence. *Cognitive Neuroscience*, *8*(1), 24–38. http://doi.org/10.1080/17588928.2016.1156663

Shulman, G. L., Mcavoy, M. P., Cowan, M. C., Astafiev, S. V, Tansy, A. P., Avossa, G., … Romani, G. L. (2008). Quantitative Analysis of Attention and Detection Signals During Visual Search, 3384–3397. http://doi.org/10.1152/jn.00343.2003

Shulman, G. L., Pope, D. L. W., Astafiev, S. V, Mcavoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right Hemisphere Dominance during Spatial Selective Attention and Target Detection Occurs Outside the Dorsal Frontoparietal Network, *30*(10), 3640–3651. http://doi.org/10.1523/JNEUROSCI.4085-09.2010

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 499–508.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and Unification of Syntactic Structure in Sentence Comprehension : an fMRI Study Using Word-Category Ambiguity. *Cerebral Cortex*, *19*, 1493–1503. http://doi.org/10.1093/cercor/bhn187

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. http://doi.org/10.1523/JNEUROSCI.5069-11.2012

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering : Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, *136*(3), 370–381. http://doi.org/10.1016/j.actpsy.2011.01.002

Stevens, W. D., Kravitz, D. J., Peng, C. S., Tessler, M. H., & Martin, A. (2017). Privileged Functional Connectivity Between the Visual Word Form Area and the Language System. *The Journal of Neuroscience*, *37*(21), 5288–5297. http://doi.org/10.1523/JNEUROSCI.0138-17.2017

Strand, F., Forssberg, H., Klingberg, T., & Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words — An event related fMRI Study. *Brain Research*, *1212*, 48–54. http://doi.org/10.1016/j.brainres.2008.02.097

Stuart, G., & Hulme, C. (2000). The Effects of Word Co-Occurrence on Short-Term Memory: Associative Links in Long-Term Memory Affect Short-Term Memory Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 796–802. http://doi.org/10.1037//0278-7393.26.3.796

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

Team, J. (2017). JASP (Version 0.8.1.1)[Computer software].

Thalmann, M., Souza, A. S., & Oberauer, K. (2018). How Does Chunking Help Working Memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(4), 729–735. http://doi.org/10.1037/0278-7393.31.4.729

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2002). Language familiarity effects in short-term memory : The role of output delay and long-term knowledge, (4), 1363–1383. http://doi.org/10.1080/02724980244000198

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: an evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology*, *50*(2), 133–58. http://doi.org/10.1016/j.cogpsych.2004.07.001

Todd, J. J., Fougnie, D., & Marois, R. (2005). Visual Short-Term Memory Load Suppresses Temporo-Parietal Junction Activity and Induces Inattentional Blindness. *Psychological Science*, *16*(12), 965–72. http://doi.org/10.1111/j.1467-9280.2005.01645.x

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751. http://doi.org/10.1038/nature02466

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences*, *95*(3), 811–817. http://doi.org/10.1016/j.cca.2007.06.009

Tse, C.-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*(8), 874–891. http://doi.org/10.1080/09658210903376250

Tse, C.-S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects.

Memory, *15*(6), 675–690. http://doi.org/10.1080/09658210701467186

Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *Quarterly Journal of Experimental Psychology*, *64*(12), 2425–2437. http://doi.org/10.1080/17470218.2011.604787

Tubach, J. P., & Boë, L. J. (1990). Un corpus de transcription phonétique(300 000 phones). Constitution et exploitation statistique.

Turner, J. E., & Henry, L. A. (2004). Redintegration and lexicality effects in children : Do they depend upon the demands of the memory task ? *Memory & Cognition*, *32*(3), 501–510.

Tyler, L. K., Moss, H. E., Galpin, A., & Voice, J. K. (2002). Activating meaning in time: The role of imageability and form-class. *Language and Cognitive Processes*, *17*(5), 471–502. http://doi.org/10.1080/01690960143000290

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. *Psychonomic Bulletin & Review*, *7*(2), 320–326. http://doi.org/10.3758/BF03212988

Viechtbauer, W. (2010). Conducting meta-analisys in R with metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Visser, M., Jefferies, E., Embleton, K. V., & Lambon Ralph, M. A. (2012). Both the Middle Temporal Gyrus and the Ventral Anterior Temporal Area Are Crucial for Multimodal Semantic Processing : Distortion-corrected fMRI Evidence for a Double Gradient of Information Convergence in the Temporal Lobes. *Journal of Cognitive Neuroscience*, *24*(8), 1766–1778.

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, *67*(1), 30–44. http://doi.org/10.1016/j.jml.2012.02.008

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *4*(325–329).

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. http://doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-017-1323-7

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). http://doi.org/10.1037/a0022790

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.

Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*,

3(6), 712–718.

Weber, E. M. G., Peters, B., Hahn, T., Bledowski, C., & Fiebach, C. J. (2016). Superior Intraparietal Sulcus Controls the Variability of Visual Working Memory Precision. *The Journal of Neuroscience*, *36*(20), 5623–5635. http://doi.org/10.1523/JNEUROSCI.1596-15.2016

Wen, X., Yao, L., Liu, Y., & Ding, M. (2012). Causal interactions in attention networks predict behavioral performance. *The Journal of Neuroscience*, *32*(4), 1284–1292. http://doi.org/10.1523/JNEUROSCI.2817-11.2012

Whitney, C., Jefferies, E., & Kircher, T. (2011). Heterogeneity of the Left Temporal Lobe in Semantic Representation and Control : Priming Multiple versus Single Meanings of Ambiguous Words. *Cerebral Cortex*, *21*(4), 831–844. http://doi.org/10.1093/cercor/bhq148

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*(1), 1–15.

Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X., & Tan, L. H. (2005). Differential Activity in Left Inferior Frontal Gyrus for Pseudowords and Real Words : An Event-Related fMRI Study on Auditory Lexical Decision. *Human Brain Mapping*, *25*(2), 212–221. http://doi.org/10.1002/hbm.20105

Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic richness effects in lexical decision: The role of feedback. *Memory & Cognition*, *43*(8), 1148–1167. http://doi.org/10.3758/s13421-015-0536-0

Yu, Q., & Shim, W. M. (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage*, *157*, 97–107. http://doi.org/10.1016/j.neuroimage.2017.05.055

Yue, Q., Martin, R. C., Hamilton, A. C., & Rose, N. S. (2018). Non-perceptual Regions in the Left Inferior Parietal Lobe Support Phonological Short-term Memory : Evidence for a Buffer Account? *Cerebral Cortex*, 1–16. http://doi.org/10.1093/cercor/bhy037

Zhuang, J., Randall, B., Stamatakis, E. a., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The Interaction of Lexical Semantics and Cohort Competition in Spoken Word Recognition: An fMRI Study. *Journal of Cognitive Neuroscience*, *23*(12), 3778–3790. http://doi.org/10.1162/jocn_a_00046

Zhuang, J., Tyler, L. K., Randall, B., Stamatakis, E. A., & Marslen-Wilson, W. D. (2014). Optimally Efficient Neural Systems for Processing Spoken Language. *Cerebral Cortex*, *24*(4), 908–918. http://doi.org/10.1093/cercor/bhs366