Order Recall in Verbal Short-Term Memory:

The Role of Semantic Networks

Marie Poirier

City University London

Jean Saint-Aubin

Moncton University

Ali Mair

City University London

Gerry Tehan

University of Southern Queensland

Anne Tolan

Australian Catholic University

Correspondence concerning this article should be addressed to Marie Poirier, Department of Psychology, City University London, Northampton Square, London, United Kingdom, EC1V 0HB.

Contact: M.Poirier@city.ac.uk

Abstract

In their recent paper, Acheson, MacDonald, and Postle (2011) made an important but controversial suggestion: they hypothesised that a) semantic information has an effect on order information in short-term memory (STM) and b) that order recall in STM is based on the level of activation of items within the relevant lexico-semantic long-term memory (LTM) network. However, verbal STM research typically has led to the conclusion that factors such as semantic category have a large effect on the number of correctly recalled items and little or no impact on order recall (Poirier & Saint-Aubin, 1995; Tse, 2009; Saint-Aubin, Ouellette, & Poirier, 2005). Moreover most formal models of short-term order memory currently suggest a separate mechanism for order coding – that is one that is separate from item representation and not associated with long-term memory lexico-semantic networks. Both of the studies reported here tested the predictions we derived from Acheson et al. The findings show that as predicted, manipulations aiming to affect the activation of item representations significantly impacted order memory.

Keywords: Short-term memory; working memory; order recall; immediate memory; activated long-term memory.

Order Recall in Verbal Short-Term Memory:

The Role of Semantic Networks

We are all familiar the experience of reading a paper in our field of expertise. Expressions are recognised, some arguments and ideas are anticipated, and grasping the experimental logic is facilitated by our understanding of the strategies in the area. Our previous knowledge of the constituents of the paper significantly supports our understanding of the work. In important ways, this example illustrates one of the most fundamental functions that memory performs: allowing the past to support and guide our present interactions with the world. This is the issue that motivated the current work; the studies reported here examine the interaction between semantic knowledge and the last few seconds of our most recent past – the content of verbal short-term memory (STM).

Here, STM is viewed as a less general system than working memory. More specifically, STM is defined as the system that carries out the temporary maintenance of information necessary for many mental or cognitive operations and tasks (Baddeley, 1986). Generally, STM is recognised and playing an important role in everyday cognition (Majerus, 2009; Cowan, 1999). Moreover, the role of STM for order has also been highlighted in cognitive development and in particular in learning new words (Cowan, 1999; Majerus & Boukebza, 2013). One of the roles of STM that is regarded as central is the short-term maintenance of the order of events (Majerus, 2009). As a simple example, consider keying in a new security code, address, or phone number. These can of course be written down, but even in order to do so, they must be maintained in order long enough for the writing down to take place.

Short- and Long-Term Memory

Until relatively recently, the literature examining how the lexical/semantic properties of verbal items affect performance in STM tasks was sparse. However, current work bears witness to the growing interest in this area, with recent research systematically exploring the relationship between language organisation in long-term memory (LTM) and verbal short-term recall (e.g. Acheson, MacDonald & Postle, 2011; Hamilton & Martin, 2007; R. C. Martin, 2006; Majerus, 2009; Tehan, Humphreys, Tolan and Pitcher, 2004; Thorn & Page, 2009). Nevertheless, there has been less work on factors typically associated with semantic LTM. The studies reported here tested a controversial hypothesis which suggests that semantic LTM plays an important role in verbal STM and more specifically in short-term order memory.

The Role of LTM in Short-Term Recall

The study of LTM contributions to verbal short-term recall—as well as the study of STM in its own right—have typically relied on a classic task: immediate serial recall. In this task, a small number of items are presented —usually between 5 and 7—and participants must attempt to recall them, in their order of appearance, immediately after list presentation. It is well established that multiple factors associated with long-term knowledge of the language have a significant impact of the performance of this task. Word frequency/familiarity has a positive effect on immediate serial recall (Poirier & Saint-Aubin, 1996), as have concreteness (Walker and Hulme, 1999), and lexicality (Hulme, Maughan, & Brown, 1991; Saint-Aubin & Poirier, 2000; for a review, see Saint-Aubin & Poirier,1999a). This is also true at a sub-lexical level (Roodenrys, 2009); for example, when trying to remember non-words, items containing more familiar phonemic components are better recalled (Thorn & Frankish, 2005). Currently, it can be argued that there are two general classes of views that address these findings. The first are

typically known as redintegration accounts while the second suggest that verbal STM relies more directly on long-term representations.

Redintegration. From the redintegration perspective, immediate recall is a two-step process. It is assumed that participants first encode verbal material into phonological forms, as suggested by the seminal multi-component model first proposed by Baddeley & Hitch (1974; Baddeley, 1986). In the absence of rehearsal, these representations are thought to rapidly become degraded either through decay or interference. At the point of recall, a retrieval mechanism produces a phonological representation as a candidate for output. The memory trace may or may not be degraded (but see Roodenrys & Miller, 2008). If the trace is intact then recall will not be problematic. However, if the trace is degraded a second step is initiated. Long-term lexical/phonological information is accessed in an attempt to reconstruct the item (e.g. accessing knowledge of words to complete a fragmented trace, somewhat like filling in the gaps in cr_ _odi_e). This reconstruction process is often referred to as redintegration (Hulme, et al., 1991; Schweickert, 1993). It has been used to explain lexicality, word frequency, concreteness and imageability effects upon serial recall. However, recent ideas about the contribution of long-term representations to STM have started to move away from dual process accounts (i.e. degradation of phonological short-term memory followed by redintegration). For example, Thorn, Frankish and Gathercole (2009), after reviewing their work on phonotactic and lexical frequency, conclude that long-term knowledge impacts immediate recall accuracy in two ways: by strengthening the representations that support performance and by influencing the reconstruction process. Romani, McAlpine, & Martin (2008) suggest a similar conclusion after a series of studies examining the effects of concreteness on a range of STM tasks.

Psycholinguistic and LTM Network Models. Over the past two decades, the redintegration hypothesis was the dominant view of LTM effects on short-term recall. Currently however, another class of models is becoming increasingly influential. Although the models in this group are more heterogeneous, they suggests that the LTM representations and the systems involved in language processing are more closely related to short-term recall than the redintegration hypothesis suggests (e.g. Acheson & MacDonald, 2009). In its typical form, the redintegration hypothesis restricts the influence of LTM representations to the retrieval stage of short-term recall. The psycholinguistic and LTM network models we refer to here propose that there is considerable overlap between STM tasks and language processing; hence the semantic, lexical, and sublexical networks that are widely thought to underlie language representations are viewed as supporting STM. In essence, these models are mostly moving away from the classic suggestion that verbal STM relies on a separate system. Rather, the premise is that processing linguistic information for recall involves the activation of the relevant long-term networks; in turn, the characteristics of these networks will influence performance.

Burgess and Hitch (2006), for example, offer a computational / network model of verbal STM where items are represented within lexical and phonological inter-connected networks. More recently, in order to explain the effects of a number of lexical and sub-lexical variables, Roodenrys (2009) proposed that an interactive network model was necessary where various levels of representation, including letter, phonemic, and lexical levels are activated and compete with each other. Other recent models explicitly include semantic levels of representation also. This group includes the computational model proposed by Gupta (2003, 2009), the conceptual models proposed Cowan (1999; Cowan & Chen, 2009) and Majerus (2009), the psycholinguistic

models proposed by Martin & Gupta (2004) and R.C. Martin (2006) and from cognitive neuroscience, the proposals of Acheson, et al. (2011) and Buchsbaum and D'Esposito (2008).

Choice amongst the models described above depends on a number of developments, one of which is a better understanding of how semantic memory influences STM performance. Assuming these models are appropriate, then semantic LTM should influence STM performance in predictable ways. As of yet however, there has been little detailed investigation of semantic LTM effects in short-term recall in healthy adults. Exceptions include the work on categorical similarity, the work of Romani et al. (2008) on concreteness and the recent work of Acheson et al (2011).

Categorical Similarity. Poirier and Saint-Aubin (1995; Saint-Aubin & Poirier, 1999a; 1999b; Saint-Aubin, et al., 2005) re-examined the widely held idea that similarity amongst list items in immediate serial recall had an adverse effect upon STM for *order* recall. While this finding is highly reliable when phonological similarity is manipulated, Poirier and Saint-Aubin argued that this was not necessarily the case with semantic similarity. In their experiments, they explored semantic similarity effects on both item and order memory; participants studied lists of items that were either from one semantic category or unrelated to each other. They found that categorical similarity was advantageous to item memory but had little effect upon order memory; in effect, across conditions, order errors were proportional to the number of items recalled (although see Saint-Aubin et al., 2005). As there are more items recalled for categorised lists, there is a proportional increase in order errors. In explaining their results, they suggested that the taxonomic category could be used as an extra retrieval cue supporting recall; this lead to better item recall and a stable level of order errors per item.

However, assuming semantic LTM underpins STM performance suggests another explanation of the semantic category effect and generates further predictions. The latter relate to the widespread idea of mutual activation between semantically related items such as those that belong to the same semantic category (see Atkins & Reuter-Lorenz (2008) for evidence of spreading activation effects in short-term memory tasks).

For instance, Saint-Aubin et al (2005) suggested that increased access to same-category items might depend on their long-term associative links (see also Hulme, Stuart, Brown, & Morin, 2003). Items from the same category tend to co-occur more frequently than items taken from different categories and this is thought to strengthen their associative links in memory (Deese, 1960; Stuart & Hulme, 2000). This is in line with many conceptualisations of lexical/semantic memory in other fields, which often depict semantic/lexical memory in terms of a network of associatively related items; activation in one part of the network can spread and influence recall of other items in the network. It seems plausible that activating multiple items in an associative network might produce higher levels of activation and support recall.

A related idea was put forward by Acheson et al. (2011) although coming from a somewhat different perspective. Importantly, their particular proposal led us to develop novel, specific and testable predictions. A quote from their paper makes their view clearer (emphasis ours): "After initial encoding, lexical activation is determined by repeated interaction with semantic and phonological representations. *Serial ordering errors occur when the relative activation levels of the lexical items change because of this interaction*. (…). If the maintenance of information in verbal WM is achieved by virtue of activation of language-production architecture, this leads to the prediction that disrupting semantic processing should influence the relative activation of lexical-level representations, thus influencing serial ordering." (Acheson et

al., 2011, p. 46). Acheson et al. used a dual-task strategy to show that when the interference task involves semantic processing, more order errors are produced than with a spatial task. Interestingly, this effect disappeared with non-words, i.e. there was no differential disruption by the semantic dual task when the primary task involved items with no meaning.

There are a number of reasons why this hypothesis is important:

(1) Knowledge-based effects have typically been considered as affecting item recall rather than order recall; this is especially true of semantic effects. Establishing that semantic factors influence order recall would be significant for extant theories of serial and short-term memory.

(2) Formal models of serial STM typically do not pay attention to long-term memory contributions, even though many instantiations (connectionist models) require that the hypothetical networks that sustain performance be trained before STM for order can be modelled (Botvinik & Plaut, 2006; Lewandowsky & Farrell, 2008; Page, 2005).

(3) Finally, although some reviews of serial order coding have typically discarded an activation-based account of order representation, this was founded on logical argument rather than empirical verification. The Acheson et al. (2011) data obviously argues in the other direction; moreover, as we now turn to, here we present further tests of the idea that order coding relies at least in part on sematic networks.

One established way of "disrupting" semantic processing is by using associates that are highly related to a target item. This is the strategy we adopted in the first experiment reported here. At first glance, the Acheson et al. (2011) quote above could be taken to imply that semantically related lists should generate more order errors than control lists, as the latter have reduced levels of inter-item activation. There are multiple studies that suggest this is not the case – but there is controversy surrounding this point (see Saint-Aubin, et al., 2005 and Tse, 2009).

As mentioned earlier, order errors are proportional to item recall and semantically related lists produce better item recall.

According to the hypothesis just reviewed [hereafter ANet for Activated Network view] manipulating the semantic activation level of item representations within a list can influence serial ordering in predictable ways. Before we turn to the specifics of Experiment 1, we wish to outline a basic model that calls upon principles that have broad empirical and theoretical support in the field (Hurlstone, et al. 2014). This is no way a full-fledged model; our aim is to suggest as simple as possible an architecture but one that a) relies on principles / mechanisms that are broadly agreed upon when it comes to immediate serial memory and b) makes specific, testable predictions in relation to manipulations of semantic activation and order. Apart from a semantic network that can support activation, our suggestion is that the following elements are required, namely: 1) encoding that produces a primacy gradient, 2) a response selection mechanism that relies on competitive cueing.

Simply put, a primacy gradient means each successive item presented is encoded with diminishing strength (Grossberg, 1978a, 1978b). Most formal models of short-term serial recall include or imply a primacy gradient as such a mechanism is necessary to account for the typical form of the typical serial recall curve (Hurlstone et al., 2014). This curve plots correct in position recall as a function of presentation order. In the case of immediate serial recall, the curve shows pronounced primacy, and a small recency effect for the last item(s). The said recency depends on materials and testing conditions. How proposed primacy gradients are conceptualised and justified varies across models. For example, in the Primacy Model, Page and Norris (1998, 2009) suggests that the primacy gradient could be produced by the association of each incoming item with a start of sequence context, with the strength of the

association diminishing with distance from the said context (i.e. the fourth word in a sequence would be farther from the start of list context than the first). A number of other systems for producing primacy gradients have been suggested; Hurlstone et al. (2014) provide a review of the various implementations of the principle. Here, the most parsimonious view would be that the said primacy gradient is represented within activation levels in a semantic network; however, other architectures could also be envisaged. The important point is that to account for immediate serial recall performance, an encoding that generates a primacy gradient appears as a reasonable assumption.

Another mechanism that has broad support was also put forward by Grossberg (1978a, 1978b) and is usually known as competitive queueing (CQ; Houghton, 1990). Competitive queuing can be thought of as a noisy competition between activated response candidates; the system is important as it can transform the parallel activation of items captured by the primacy gradient into a serial sequence of responses. One way of describing the operation of a generic CQ mechanism is as follows. The activations represented within the primacy gradient are fed forward to the CQ response selection mechanism; there, items compete for selection based on their activation levels; mutual inhibition and noise make the process error prone. The most activated item is typically selected, unless activation levels are too low or competition leads to the wrongful selection of another item. For instance, if noise makes it difficult for an item to be selected when appropriate (i.e. there is no winner of the competition within a threshold number of iterations / attempts) then this item as well as all the remaining ones become less likely to be selected because of the increased pool of candidates and increased mutual inhibition (reducing activation). Importantly, CQ systems typically suppress the activation of any selected response, preventing perseveration.

These relatively simple building blocks provide the needed architecture for the ANet predictions tested in Experiment 1. In essence, the suggestion is as follows. Presenting a list of items for immediate serial recall generates activation in the lexico-semantic network with activation following a primacy gradient. At the point of recall, the dynamics of the QC mechanism predict that the first item is very likely to be output first. It will then be suppressed, removing it from the competition for the next response. The second item is then the most likely winner of the competition for response selection, and so forth.

Experiment 1 manipulated the level of activation of a target item to test the prediction that this would increase order errors for that item, making it likely that the CQ mechanism would select this item earlier because of its heightened activation; this early selection would mean that activation affected the order in which items were recalled. Lists of six visually presented items were used; experimental lists contained a target item, presented in position 5. The three first items of these lists were strong associates of the target. Control lists contained the same three associates in positions 1 to 3, but the item in position 5 was unrelated [see Table 1 for list examples].

Table 1: Sample experimental and control lists

Experimental list examples					
officer	badge	siren	fence	police	tractor
band	record	concert	yellow	music	tourist
Control list examples					
officer	badge	siren	music	tourist	yellow
band	record	concert	tractor	fence	police

For the experimental lists, it is expected that the first three items will activate the target (5th item) within LTM networks, making its activation level seem more like that of earlier list items. Based on the ANet view and the summary model described above, the prediction is that the target fifth item will migrate towards earlier positions more often than a non-target item studied in the same position.

Basically, the prediction from this version of the ANet hypothesis involves one of the characteristics of order errors in immediate serial recall known as the *locality constraint* (Henson, Norris, Page, & Baddeley, 1996). It is well established that when a list item is recalled in an incorrect position, it is more likely to migrate to a neighbouring list position. So, the third item is more likely to be be recalled in the second or fourth output position than in the first or seventh. In other words, order errors obey a rule whereby displacements are increasingly unlikely as one moves away from the actual presentation position of the item. The main prediction of Experiment 1 is that the locality constraint will still apply to the target item, but not as strictly as it applies to the comparable control item. The target item is expected to be recalled in earlier positions more often than what is observed for the corresponding item in the control condition.

Experiment 1

Method

Participants. A total of 40 adults took part (14 men and 26 women, age range from 18 to 57, mean 27); they were offered a small fee (£7) for participating.

Materials. The experiment comprised 32 lists, with 16 experimental and 16 control lists. We first generated a set of 16 lists where the first three items were strong associates of a target word, based on the University of South Florida norms (Nelson, McEvoy & Schreiber, 2004).

These words, when used as cue words in a semantic association / production task, generate the target as a strong associate. More specifically, cue words had to have a forward association strength with the target that was above 0.2; also, they were excluded if they had a backward association strength above of 0.1. The target was placed in the 5th position of each list, the cue words were placed in positions 1, 2 and 3, and the remaining positions (4 & 6) were filled with unrelated words. The same words were then used again to create a further set of 16 control lists, so each word was used twice within the experiment; more specifically, each participant encountered the words. However, the condition in which they encountered the words for the first time (and the second) was counterbalanced across participants. Control lists had the same three associates in the first positions, in the same order. The last three words were a random selection from the filler words and from targets associated with other lists. The 32 lists thus created were then mixed to create 4 sets, with a different, quasi-random order of lists. This was done such that a given trio of related words was presented once in the first block of 16 lists and once in the second block of 16 lists. Also, each block of 16 lists contained 8 experimental and 8 control lists. Each participant was only presented with one set of 32 lists, with sets counterbalanced across participants. To be clear, each participant studied each word twice, once in the first block of 16 lists and once in the second; however, the order of the condition encountered first was counterbalanced across participants. A bespoke computer program controlled stimulus presentation and response collection.

Procedure. Participants (Ps) were tested individually, in sound-proofed cubicles, within a session lasting approximately 20 minutes. Following instructions, they completed two practice trials. A fixation cross appeared in the centre of the screen, for two seconds, indicating that the first word was about to be presented. Words appeared sequentially on the screen, for one and a

half seconds each, and were separated by a 500 msec blank. Right after the six words from a list had been presented, participants were to type them into response boxes, in the order in which they had appeared in the list, starting with the word presented first. If they did not remember a word, they were asked to type the letter "b" and proceed to the following position. The program prevented Ps from typing a response if the previous one was not entered or if the enter key had not been pressed. They were not allowed to backtrack to correct a previous response.

Results and Discussion

The hypothesis examined here relates to the recall of the critical word and its control both appearing in the 5th position of their respective lists. The ANet view predicts that there will be more movement toward earlier positions for the 5th item when the first three words presented were strong associates of said target.

Table 2 presents correct-in-position scores (i.e. to be scored correct, the item must be recalled in its presentation position) as well as the item recall score (i.e. item scored correct if it is recalled, irrespective of position). The table also presents means for the critical 5th item. As a perusal of the table shows, overall correct-in-position performance is very similar in both conditions. Table 2 also presents item recall scores; as can be seen, Item 5 was better recalled in the experimental condition. Item scores are usually higher than correct-in-position scores because if an item is recalled in the wrong position it will be given a correct-in-position score of zero; however, it will be considered correct with item scoring. It can be seen for example, that for the control lists, the correct-in-position score for Item 5 is 0.58 while it is 0.68 when item scoring is used. This difference is larger for Item 5 in the experimental condition, where correct-in-position performance is 0.58 and item scoring leads to a 0.75 mean. The implication is that

Item 5 was recalled out of position more often in the experimental condition than in the control condition.

Table 2: Mean recall across positions and for position 5

	All positions	**Position 5**
	Correct in position scores	
Control lists	0.71	0.58
Exp. lists	0.71	0.58
	Item recall scores	
Control lists	0.78	0.68
Exp. lists	0.79	0.75

As would be expected based on the content of Table 2, there was no statistically reliable effect for the correct in position scores. With respect to item scores, paired sample T-tests showed no reliable difference for the overall means, but there was a significant difference for position 5 ($t(39)$=2.5, p=0.017).

Figures 1, below, shows the percentage of trials for which the item studied in position 5 is actually recalled in another position – essentially error frequency per position, for the target 5th item. As can be seen, the rate with which the 5th word is recalled in an incorrect position appears higher for the experimental condition than for the control condition, particularly for positions 2 and 3. A 2 (condition) x 5 (error position) repeated measures ANOVA revealed that errors were significantly more frequent for the experimental condition [$F(1, 39)$= 12.63, MSe = 0.56]. There was also a significant effect of position [$F(4, 156)$= 16.76, MSe =0.85] and a significant

interaction [$F(4, 156=2.75$, MSe$= .52$]. Simple main effect tests showed that Item 5 migrated

more often to positions 2 and 3 in the experimental condition – while there was no evidence of
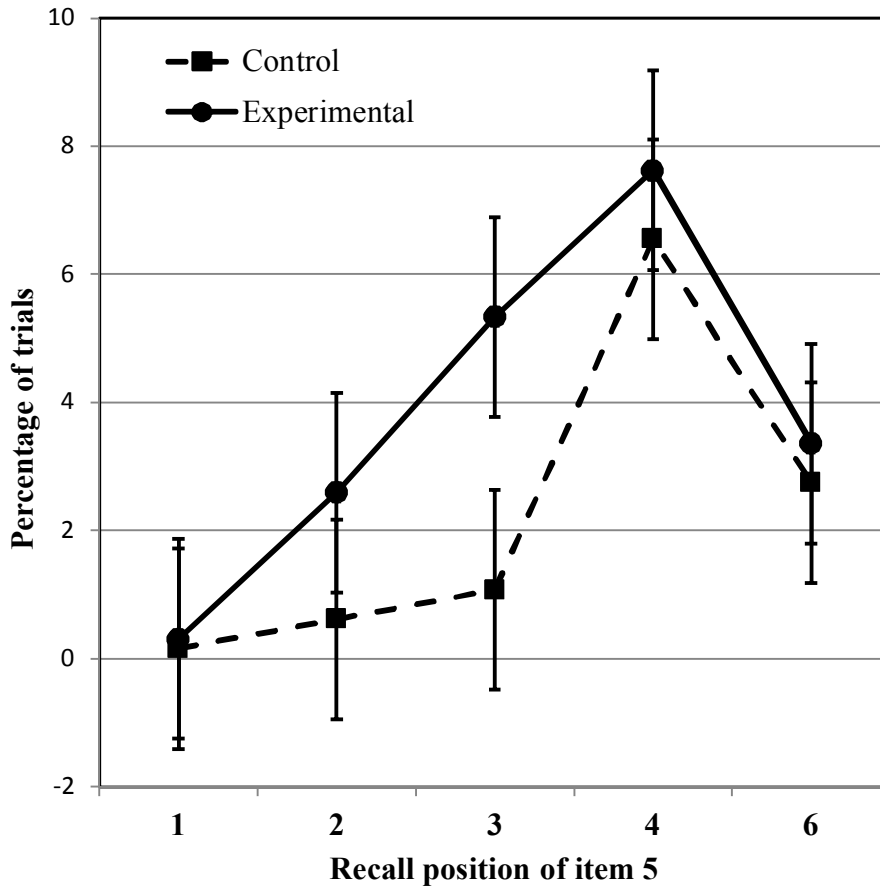
more migrations for position 6.

Figure 1. Percentage of trials showing an error for Item 5 as a function of presentation

position; only the erroneous recall positions are plotted on the *x*-axis. Error bars represent 95%

confidence intervals computed according to the method of Loftus and Masson (1994) for within-

subject factors. When the difference between two means is significant, those confidence intervals

do not overlap by more than half the distance of one side of an interval (Masson & Loftus, 2003).

These findings support the predictions derived from the ANet account: when the first three items in a list are strong associates of the 5th item, the latter tends to migrate more than a control item appearing in the same position; as expected, the target item migrated towards typically better recalled positions rather than towards the posterior position (6).

These predictions were derived based on the idea that item order is coded as an activation primacy gradient within the lexico-semantic network that supports language representation and hence the results lend support to this view. However, there is an alternative interpretation of this pattern of data that is less interesting. This competing interpretation suggests that the 5th item is more frequently recalled with the first three related items because of a grouping strategy. Although the task instructions emphasised ordered recall, participants might have subjectively grouped the related items and this could have generated order errors. Essentially, the alternative hypothesis suggests that the results are an artefact of a study/recall strategy rather than an indication that semantic activation plays a role in order encoding and maintenance. This being said, it is important to note that the said strategy could well originate from the fact that recall relies on activated semantic networks and that this makes maintaining clustered and related items easier. The next study used lists that eliminate any advantage that grouping could involve, making the use of such a strategy useless and hence very unlikely.

Experiment 2

Experiments 2a and 2b were based on a re-analysis of the previously published findings of Saint-Aubin et al. (2005). In their study, the experimental lists contained items that were all from the same semantic category (vegetables, sports, clothing, etc.). They can hence be expected to be reasonably close neighbours within the proposed semantic network. Based on the ANet view, we would expect heightened co-activation for these lists, relative to control lists containing unrelated

items. Importantly, one would not expect any special grouping strategy for the categorised lists as all the items are from the same category. The control lists were constructed by re-organising the items from the semantically related condition so that each word within a list was from a different semantic category. Each condition involved the same items overall. In Experiment 2a lists were studied in silence while in Experiment 2b, participants engaged in articulatory suppression. Both semantic category and suppression were manipulated between participants. There were N=70 in each group for the silent conditions and N=56 in the two suppression conditions (categorised or control lists). All lists were seven items long; there were 14 lists presented in each condition. The details of the methodology are otherwise similar to the study reported above and can be found in Saint-Aubin et al. (2005).

As the lists used in these experiments were seven items long, we examined the recall of items 5, and 6. These seemed like the best candidates as there needs to be a reasonable number of errors made for reliable migration analyses to be possible. In an immediate serial recall task, the highest performance is typically observed for the first few items; the last item (7) is of less interest as it can only migrate in one direction.

What are the predictions for this experiment? When the differences in correct recall between categorised and non-categorised items are examined, what is typically found is that the whole curve moves upwards for the related items, i.e. there is a categorisation advantage that does not interact with serial position (provided ceiling and floor effects are avoided; see Poirier & Saint-Aubin, 1995, and Saint-Aubin & Poirier, 1999b). The number of items recalled is higher for categorised lists and there is a proportional increase in order errors. As before, because of the heightened activation presumed to accompany the presentation of a categorised list, we predict

that the items studied in position 5 and 6 will tend to migrate forwards (up the positions) more than controls.

However, as the reviewers pointed out, with items all taken from the same category, the straightforward expectation would be a similar level of co-activation across items with the result that the entire series would be in a higher state of activation than a control list. Why would items 5 and 6 migrate upwards more than other items in the list?

To clarify this prediction, we need to consider the operation of the basic model described previously in a bit more detail. This is necessary in order to account for a feature of the data obtained in Experiment 1 and to justify the migration prediction made above.

An examination of Figure 1 shows that for the *control condition*, Item 5 moved more often upwards towards position 4 than downwards towards position 6. Hence, even for control items—at least for the less well recalled positions—movement forward, towards earlier positions is more likely. How could the primacy gradient plus CQ mechanism produce this behaviour? In order to answer this question, one must consider how the described system can produce blank responses (i.e. no item recalled in position X) and how the system can lead to an item not being recalled at all (item errors)[1].

In order to illustrate the proposed functioning of the CQ system, let us consider the recall of Item 5. Assume that noise and inhibition from remaining items makes the level of activation of Item 5 drop and its successful retrieval in brought into question. Based on empirical errors rates, overall, the most likely outcome of this situation is a blank response. The second most likely possibility is the retrieval of the strongest competitor based on activation within the

---

[1] In our data (from 3 different laboratories) across multiple experiments, item errors are the most high frequency errors by far when there are different items on each trial, as is the case here.

primacy gradient, Item 6 (assuming previous items were recalled and supressed). If no item is recalled, none of the remaining response choices are supressed, reducing the probability of recall of the non-retrieved Item 5 as well as the other available candidates because of increased competition / mutual inhibition. If Item 6 is recalled, Item 5 remains in the competition, and again, general probability of recall reduces through competition although to a lesser degree as Item 5 is now supressed.

Generally speaking, suppression and competition means that if the correct item is not retrieved and an item is recalled (i.e. the response is not blank), then the most likely item will be the following item (Item 5 in our example) creating the upward migrations observed in the data. If the previous item was not recalled (i.e. Item 4 in our example), it might win the competition and create an item transposition where items 4-5 are recalled as item 5-4. However, in combination with this trend, there is the reduction in retrieval probability associated with moving through the primacy gradient, and a further reduction in the probability of retrieval with every error in recall. To summarise, retrieval difficulties open the window for upward movement and reduced probability of retrieval means this upward movement is not matched by errors in the other direction.

These processes operate for the control lists and will also be at play for the categorised lists. However, increased activation means more items retrieved, more forward movement and proportional difficulties retrieving as the CQ mechanism works its way through the primacy gradient. We hence expect an increase in migrations towards earlier positions for semantically categorised lists in Experiment 2a, where lists were studied in silence. Experiment 2b, where lists were studied under suppression, is thought of as a replication that can help establish the robustness of the findings in Experiment 2a.

Experiment 2a: Results and Discussion

Figures 2a and 2b summarise the main findings for this data set. As can be seen, there were more migrations for the categorised items relative to the control lists. The results for each position were analysed with two mixed ANOVAs; the between-subject factor was list type (categorised or not) and the within-subject factor was error position. For position 5, there was a main effect of list type [$F(1, 138)= 10.05$, MSe= 0.516], of position [$F(5, 290)= 82.0$, MSe= 0.514], as well as a significant interaction [$F(5, 690)= 4.45$, MSe= 2.29]. The same effects were obtained for position 6, with list type [$F(1, 138)= 24.69$, MSe= 0.626], error position [$F(5, 290)= 86.81$, MSe= 0.718], and the interaction [$F(5, 690)= 14.10$, MSe= 0.718] producing reliable effects. Simple main effect tests revealed the following: for the words studied in the 5th position, the difference between conditions was only significant for recall errors in position 4. For the items studied in the 6th position, this difference was significant for the errors observed in positions 4 and 5.
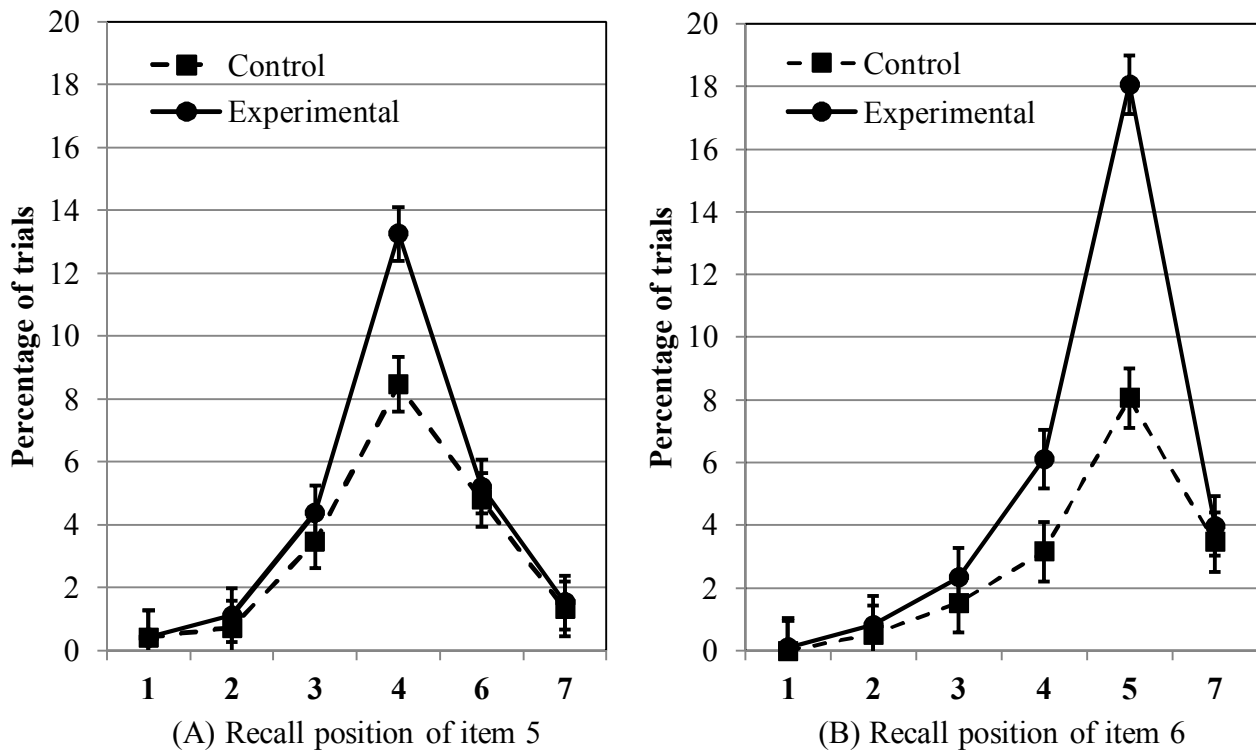
Figure 2. (A) Error frequency for item 5 and (B) item 6 as a function of recall position. Error

bars represent 95% confidence intervals computed according to the method of Loftus and

Masson (1994) for the between-subjects factor of similarity. When the difference between two

means is significant, those confidence intervals do not overlap by more than half the distance of

one side of an interval (Masson & Loftus, 2003).

These findings fit nicely with those of Exp. 1; in both experiments, an increase in order

errors / migrations for semantically related lists was observed, relative to control list, as predicted

by the ANet account. Again the increase in migration towards earlier position and *not* towards

later positions is found in this experiment.

Experiment 2b: Results and Discussion

Figures 3a and 3b summarise the main findings for the immediate serial recall data under suppression. As in experiment 2a, it is clear that there are more migrations for the categorised lists than there are for the lists containing unrelated items. Here also, the results for each position were analysed with two mixed ANOVAs with one between-subject factor, list type (categorised or not), and the within-subject factor was error position. For position 5, there was a main effect of list type $[F(1, 110)= 15.77, MSe= 6.1]$, of position $[F(5, 550)= 45.3, MSe= 14.8]$, as well as a significant interaction $[F(5, 550)= 3.57, MSe= 1.16]$. The same effects were obtained for position 6, with list type $[F(1, 110)= 28.7, MSe= 14.0]$, error position $[F(5, 550)= 28.72, MSe= 11.56]$, and the interaction $[F(5, 550)= 1.97, MSe= 0.79]$ producing reliable effects. Simple main effect tests revealed that migrations were more pervasive for this experiment; for the words studied in the 5th position, the difference between conditions was significant for recall errors in positions 3 and 4. For the items studied in the 6th position, this difference was significant for the errors observed in positions 3, 4 and 5.
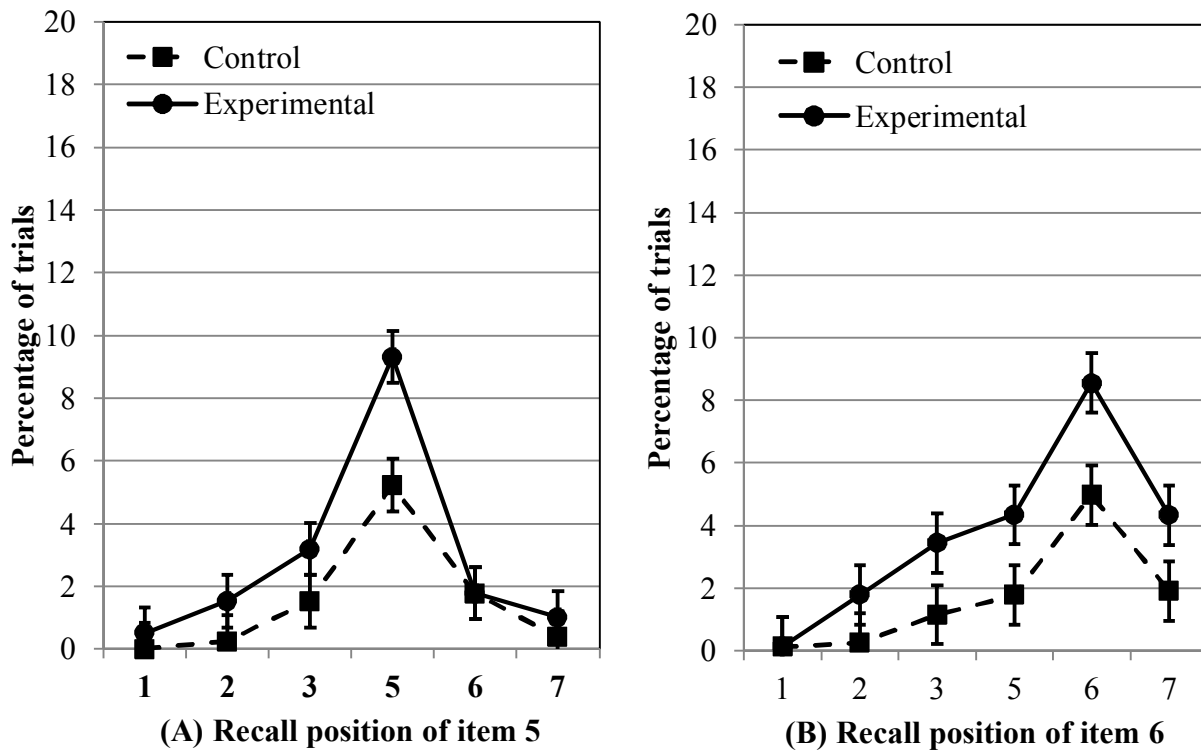
Figure 3. (A) Error frequency for item 5 and (B) item 6 as a function of recall position. Error bars represent 95% confidence intervals computed according to the method of Loftus and Masson (1994) for the between-subjects factor of similarity. When the difference between two means is significant, those confidence intervals do not overlap by more than half the distance of one side of an interval (Masson & Loftus, 2003).

Again the overall pattern of effects as well as the details of the findings conform to what would be expected based on the ANet view. In all both Experiments, there were substantial changes in the error patterns for related items; as predicted these migrations were towards earlier positions. Why would suppression lead to migrations across more positions than silent conditions? What would be the effect of suppression on the primacy gradient plus CQ system outlined here? Perhaps the most parsimonious suggestion is that suppression reduces the

resources available for encoding; this would lead to lower activation levels, or a flattened primacy gradient. This would correctly predict more omissions and a greater number of omissions could also open the door to migrations across more positions, especially for the somewhat more activated items from the experimental condition.

General Discussion

In the introduction to this paper we briefly reviewed a group of models that have been increasingly influential. These views insist on the importance of long-term knowledge in producing the behaviour that is typically analysed when studying short-term memory. Within this category of models, the proposal put forward recently by Acheson et al. (2011) makes a controversial suggestion: order recall in STM should be considered as the results of activation perturbations within existing semantic networks. Based on this view, dubbed the ANet model in the current paper, a series of specific predictions were derived. More precisely, we tested the prediction that words for which the semantic activation is heightened by items within the same list would be more likely to migrate towards earlier positions within the list.

The findings of all three experiments plainly support the ANet perspective and predictions. In the first Experiment, we manipulated the content of the first part of the list so that in 50% of the trials, assuming the operation of a semantic network, the fifth item's activation was increased. This was predicted to lead to a specific increase in migrations of this item towards earlier positions at the point of recall. The results of Experiment 1 showed precisely that pattern. Experiments 2a and 2b examined the same predictions, while eliminating a more trivial alternative interpretation of the first set of findings (i.e. that the migration of the target fifth item towards earlier positions was due to a grouping strategy). In both these cases, the hypotheses derived from the ANet model were unequivocally supported.

Taken together the results presented here are in line with the models that suggest that short-term serial recall relies on the activation of the long-term memory networks that are associated with language processing (e.g. Acheson & MacDonald, 2009; Cowan, 1999; Cowan & Chen, 2009; Gupta, 2003, 2009; Majerus, 2009; Martin & Gupta, 2004; R.C. Martin, 2006; Roodenrys, 2009).

The findings may also prove important for more formal models of serial order. More specifically, one way of looking at the present work is that it provides an empirical test for one of the most frequently proposed mechanisms within these models: primacy gradients (Hurlstone et al., 2014). In effect, many recent models of serial STM successfully account for the serial position curve that is typical of STM recall; the said curve is typified by strong primacy and a diminutive recency effect (typically only involving the last item or so). In order to account for the better recall of the first items, these proposals almost invariably include what is referred to as a primacy gradient, i.e. they assume there is a decreasing strength in the encoding of successive items (although other mechanisms are also brought to bear in some instances, e.g. Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves., 2012). The predictions that were tested here assumed that order recall was guided by a primacy gradient such that the most activated item was recalled first, followed by the second most activated item, and so on.

A related point relates to the suggestion, included in the Acheson et al. (2011) model, that the order coding mechanism is integrated into the network that allow item-level representation (i.e. activation within the lexico-semantic network). Recent quantitative models typically involve a separate mechanism for coding order and item information. Consider for example the interference-based model of Oberauer, et al. (2012); they offered a model of complex span that represented order in the same fashion as two previous models of immediate recall (Farrell &

Lewandowsky, 2002; Farrell, 2006). In all three models, the authors call upon a distributed neural-network that has a two-layer structure, with one layer representing serial positions and the other representing items. Items are encoded through Hebbian associations between item and position representations: The first list item is associated with the first position representation (a.k.a. a position marker), the second item is associated with the second position marker, and so on (see also Henson, 1998). Memory for order is maintained by the patterns of association in the weight matrix that connects position markers to item representations. The links from position markers to items are unidirectional (going from the position markers to the items); at the point of recall, the position marker is used as the cue and it leads to the retrieval of a blurry representation of the target. If one focusses on these aspects of the model, the results presented here can seem problematic. This is because it is not clear that a change in the activation of item representations could lead to perturbation of the associations between position markers and items: the activation runs from the position markers to the items and not the other way around. This being said, it is of course likely that this could be addressed by some reasonably slight tweak of the model's architecture. Importantly also, in their review of the formal models of serial STM, Hurlstone et al. (2014) note that perturbing the activations in one or both layers (i.e. item and order layers) predicts transposition errors akin to those observed in serial recall.

Conclusion

Previous interpretations have insisted that categorised lists have almost all of their effect by increasing item recall (irrespective of position; Saint-Aubin & Poirier, 1999a, 1999b). This increase is accompanied by a proportional increase in order errors. So, if order error proportions are the measure called upon, there is typically no effect of category on order. However, Saint-

Aubin et al. (2005) did report a statistically reliable effect of categorised lists on the *proportion* of order errors.

The ANet framework discussed here offers a straightforward and parsimonious interpretation of this typical pattern of findings: the representation of the words in an immediate serial recall task relies on available language processing systems, including activation within and between phonological, sub-lexical, lexical, and semantic networks. In that sense, our view is well aligned with those suggesting that STM can be conceptualised as activated LTM rather than a separate system (e.g. Acheson et al., 2009, 2011; Cowan, 1999). Categorised lists lead to heightened network activation which produces better item retrieval as well as perturbation of the representation of item order.

Our aim in this paper was to test specific predictions derived from the Acheson et al (2011) proposal; the latter suggests that short-term memory relies on the LTM networks available for language processing. Our findings produced a pattern that was very much in line with the derived predictions. The results support models where STM relies on activated LTM representations and networks.

References

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

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

Atkins, A. S. & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, 36, 74-8

Baddeley, A.D. (1986). *Working memory*. Oxford: OUP.

Baddeley, A.D. & Hitch, G.J. (1974). Working memory, in G.H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. VIII., pp. 47-90, New York: Academic Press.

Botvinick, M. and Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, 113, 201-233.

Buchsbaum B.R. & D'Esposito M. 2008. The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience*, 20 (5), 762-778.

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*, 627-652.

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

Cowan, N., & Chen, Z. (2009). How chunks form in long-term memory and affect short-term memory limits. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 86-101). Hove, East Sussex, UK: Psychology Press.

Deese, J. (1960). Frequency of usage and number of words in free-recall: The role of association. *Psychological Reports, 7,* 337-344.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.), *Coding Processes in Human Memory*. Washington, D.C.: Winston.

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, 55,587–600.

Farrell, S. & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*, 59–79.

Grossberg, S. (1978a). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 17, 199–219.

Grossberg, S. (1978b). A theory of human memory: Self-organization and performance of sensory–motor codes, maps, and plans. In R. Rosen & F. M. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York, NY: Academic Press.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. Quarterly Journal of Experimental Psychology, 56(7), 1213-1236.

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

Hamilton, A. C. & Martin, R. C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex, 43*, 112–123.

Henson, R.N. (1998). Short-term memory for serial order: the Start-End Model. *Cognitive Psychology*, 36, 73-137.

Henson, R. N., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, 49A, 80–115.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London, England: Academic Press

Hulme, C., Maughan, S. & Brown, G.D.A. (1991). Memory for familiar and unfamiliar words: Long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685-701.

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*, 500-518.

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), 339-373

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *The Psychology of Learning and Motivation*, 49, 1-48.

Loftus, G.R. & Masson, M.E.J. (1994) Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, 1, 476-490.

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

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: new evidence from a novel word learning paradigm. *Journal of Experimental Child Psychology*, 116, 811-828

Martin, N. & Gupta, P. (2004). Exploring the relation-ship between word processing and verbal STM: Evidence from associations and dissociations. *Cognitive Neuropsy-chology, 21*, 213-228.

Martin R. C. (2006). Semantic short-term memory, language processing, and inhibition. In Meyer A. S., Wheeldon L. R., Krott A. (Eds.), *Automaticity and control in language processing* (pp. 162-191). Hoboken, NJ: Taylor & Francis.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203-220.

Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. 2004. The University of South Florida Word Association, Rhyme and Word Fragment Norms. *Behavior Research Methods, Instruments, & Computers*, 36, 408-420.

Oberauer, K., Lewandowsky, S.,Farrell, S.,Jarrold, C., & Greaves, M. (2012). Modeling

working memory: An interference model of complex span. *Psychonomic Bulletin & Review*,

19, 779–819. doi:10.3758/s13423-012-0272

Page, M. (2005). Connectionist models of short-term memory for serial order. In G.

Houghton (Ed.), *Connectionist Models in Cognitive Psychology*. (Vol. chapter 7, pp. 215-

219). Taylor & Francis.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate

serial recall. *Psychological Review*, 105(4), 761-781.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb

repetition effect and the learning of phonological word forms. *Philosophical Transactions of

the Royal Society B: Biological Sciences*, 364,3737–3753.

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 A: Human Experimental Psychology*, *48A*, 384-404

Poirier, M. & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item

identity and item position. *Canadian Journal of Experimental Psychology*, 50, 408-412.

Romani C, McAlpine S, Martin RC. (2008). Concreteness effects in different tasks:

implications for models of short-term memory. *Quarterly Journal of Experimental

Psychology*. 61 (2) 292-323

Roodenrys, S. J. (2009). Explaining phonological neighbourhood effects in short-term

memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term

memory in the verbal domain* (pp. 177-197). Hove: Psychology Press

Roodenrys, S. J. & Miller, L. M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Memory and Cognition, 36*, 578-587.

Saint-Aubin, J., Ouellette, D. & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin and Review, 12*, 171-177.

Saint-Aubin, J. & Poirier, M. (1999a). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34*, 347-352.

Saint-Aubin, J. & Poirier, M. (1999b). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology, 52A(2),* 367-394

Saint-Aubin, J. & Poirier, M. (2000). Immediate Serial Recall of Words and Non-Words: A Test of the Retrieval Based Hypothesis. *Psychonomic Bulletin & Review*, 7 (2), 332-340

Saint-Aubin, J., & Poirier, M. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. Memory, 13(3-4), 325-332.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*, 168-175.

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 & Cognition, 26*, 796-802.

Tehan, G., Humphreys, M. S., Tolan, G., and Pitcher, C., (2004). The role of context in producing item interactions and false memories. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30, pp. 107-119. ISSN 0278-7393

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, & Cognition, 31*, 729-735.

Thorn, A., Frankish, C. R., & Gathercole, S. E. (2009). The infuence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 198-219). Hove, UK: Psychology Press.

Thorn, A., & Page, M. (2009). *Interactions between short-term and long-term memory in the verbal domain*. Hove, UK: Psychological Press.

Tse, C.S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*, 874-891.

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