Memory & Cognition
https://doi.org/10.3758/s13421-025-01788-9

Syntactic knowledge does support working memory for serial order: A comparison across French and German languages

Hong Xiao[1] · Judith Schweppe[2,3] · Pauline Querella[1] · Lucie Attout[1,4] · Friederike Contier[5] · Steve Majerus[1]

Received: 19 March 2025 / Accepted: 11 August 2025

Abstract
A large body of research demonstrates robust interactions between verbal working memory (WM) and phonological and lexico-semantic language knowledge, particularly for item recall. The role of syntactic knowledge, involving knowledge about word positioning in a verbal sequence, has been explored to a lesser extent but is of theoretical interest given that this type of knowledge may also support serial order recall in multi-item sequences. This hypothesis has not been supported so far, either by French or German language studies. The present research reexamines the impact of syntactic knowledge on WM for lists of adjective–noun pairs by controlling for short-term syntactic predictability of memoranda. Contrary to previous studies, across two experiments involving French and German, we observed enhanced order recall accuracy for adjective–noun pairs presented in legal syntactic order, as well as more order migration errors for pairs in illegal syntactic order. This study is the first to demonstrate that long-term syntactic structures do support recall of serial order information in WM, consistent with hybrid and full linguistic accounts of verbal WM.

Keywords Verbal working memory · Long-term memory · Syntactic knowledge · Serial order

Introduction

There is increasing consensus that verbal working memory (WM) interacts with long-term language representations (Baddeley, 2000; Cowan, 1999). Numerous studies have demonstrated that verbal WM performance is enhanced for word lists compared with nonword lists, a phenomenon known as the lexicality effect in verbal WM (Brener, 1940; Gathercole et al., 2001; Hulme et al., 1991, 1995; Kowialiewski & Majerus, 2018; Majerus & Van der Linden, 2003; Patterson et al., 1994; Saint-aubin & Poirier, 2000). Furthermore, recall performance improves for words with richer or more accessible lexical and semantic representations, such as high-frequency words versus low-frequency words (Gregg et al., 1989; Hulme et al., 1997; Kowialiewski & Majerus, 2018; Poirier & Saint-Aubin, 1996; Roodenrys et al., 1994; Watkins & Watkins, 1977) or concrete words versus abstract words (Bourassa & Besner, 1994; Hulme et al., 1997, 2003; Kowialiewski & Majerus, 2018; Majerus & Van der Linden, 2003; Walker & Hulme, 1999). These findings underscore the robust interactions between verbal WM and language knowledge, particularly regarding item memory at phonological and lexico-semantic levels. They have led to models assuming that verbal WM is determined by access to language knowledge at least for item-level aspects (encoding, retrieval, and/or completion of a word within a list of words to be memorized; e.g., Baddeley et al., 2017; Burgess & Hitch, 1999, 2006; Cowan, 1999; Gupta & MacWhinney, 1997; Majerus, 2013). Other theoretical accounts consider that linguistic knowledge should support both item and sequence-level representation of information in WM (e.g., Acheson & MacDonald, 2009; Majerus, 2019; Martin et al., 1994; Schwering & MacDonald, 2020). The present study focuses on the potential impact of syntactic knowledge on

✉ Hong Xiao
hong.xiao@uliege.be

1 Department of Psychology, Psychology & Cognitive Neuroscience Research Unit, University of Liège, Place Des Orateurs 1 (B33), 4000 Liège, Belgium

2 Department of Educational Sciences, University of Erfurt, Erfurt, Germany

3 Department of Psychology, University of Passau, Passau, Germany

4 Department of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

5 Department of Psychology, University of Potsdam, Potsdam, Germany

Published online: 01 September 2025

verbal WM. This type of knowledge is of particular interest given that it concerns the serial positioning of words within a sentence, and hence, according to linguistic accounts of verbal WM, should also support recall of serial order information in verbal WM (Perham et al., 2009; Querella & Majerus, 2024; Schweppe et al., 2022).

Regarding the distinction between item and serial order aspects in WM, a significant number of studies have revealed dissociations between these two aspects. Most empirical studies have shown that phonological and lexical-semantic knowledge support item recall, but not order recall (Gathercole et al., 2001; Hulme et al., 1991, 1997; Majerus & D'Argembeau, 2011; Nairne & Kelley, 2004; Poirier & Saint-Aubin, 1996; Saint-aubin & Poirier, 1999, 2000; Walker & Hulme, 1999). This is also supported by neuropsychological studies revealing the possibility of selective item versus serial order impairment, with item WM impairment often being linked to language deficits (Majerus, 2008; Majerus et al., 2015). For example, patients with semantic impairment may show poor item recall in word serial recall tasks, while serial order recall remains intact (Majerus et al., 2007). Additionally, neuroimaging studies comparing item and order recall tasks have found that WM for serial order recruits specific fronto-parietal networks, including the intraparietal sulcus, while item recall tasks engage fronto-temporal networks involved in linguistic processing, including the supramarginal gyrus (Henson et al., 2000; Majerus et al., 2006, 2010; Marshuetz et al., 2000, 2006). These studies suggest that encoding and retention of item information in WM engage access to phonological and semantic aspects of words, while encoding and retention of serial order information involve a more specific mechanism, whose nature still remains to be determined (Hartley et al., 2016; Majerus, 2019).

While semantic information is generally considered to exert a stronger influence on item memory than on serial order memory, a growing body of evidence suggests that semantic processing can nonetheless interact with serial order representations, albeit in a complex manner (Guérard & Saint-Aubin, 2012; Murdock, 1976; Neale & Tehan, 2007; Saint-aubin & Poirier, 1999). For instance, Saint-Aubin et al. (2005) reported that semantic relatedness impaired serial order memory, even when the number of recalled items was equated across semantically related and unrelated lists. Similar effects were observed by Tse (2009) and Tse et al. (2011), who demonstrated that both categorical and associative forms of semantic relatedness disrupt order memory, independent of item recall performance. These findings align with the hypothesis that semantic proximity among items increases the likelihood of order errors, such as item migrations (Poirier et al., 2015; Saint-Aubin et al., 2023). Consistent with this view, Guitard et al. (2025) recently showed that participants produced more conditional order errors when recalling semantically related lists compared with unrelated lists, further supporting the notion that semantic relatedness exerts a destabilizing effect on serial order representations. Other studies have observed opposite findings, by observing improved serial order recall for semantically related versus unrelated word lists, but only when the similar items were presented in a grouped manner (Kowialiewski et al., 2024).

Less ambiguous evidence in favor of the supporting influence of language knowledge on memory for serial order may be obtained when considering syntactic aspects of language knowledge. Syntactic knowledge, which governs the order of words in a sentence, should be most strongly expected to exert an impact on the retention of serial order information in verbal WM tasks. Basic syntactic knowledge structures (such as the rule that a verb is generally followed by a noun, or a noun is generally followed by an adjective in a large range of languages; Bock, 1995; Garrett, 1980) constrain the possible positions of words within a sentence. These positional rules should also be expected to support WM for verbal sequences. The impact of syntactic knowledge on serial order memory has not yet been extensively explored. Some evidence comes from studies that have investigated the sentence superiority effect (SSE; e.g., Bonhage et al., 2014; Scheerer, 1981). These studies showed that regular sentences are recalled more accurately than scrambled word lists (Allen et al., 2018; Baddeley et al., 2009; Brener, 1940; Cattell, 1886; Jefferies et al., 2004; T. Jones & Farrell, 2018; Marks & Miller, 1964; Miller & Isard, 1963; Miller & Selfridge, 1950; Poirier & Saint-Aubin, 1996; Savill et al., 2015, 2018). This pattern of findings can be accounted for within the conceptual regeneration hypothesis (Potter & Lombardi, 1990, 1998), according to which sentence recall is based on a conceptual representation of the sentence that is generated while processing the to-be-recalled sentence. At recall, this conceptual representation is regenerated, and lexical and syntactic priming of words and syntactic structures presented in the sentence are particularly available when regenerating the sentence. Based on this, lexical and syntactic representations in long-term memory are key to the immediate recall of sentences. However, although Potter and Lombardi (1998; see also Lombardi & Potter, 1992) already emphasize the importance of syntactic structure for ordering words in sentence recall, their studies do not focus on order memory. Another framework that can account for the sentence superiority effect is the Construction-Integration (CI) model of discourse comprehension (Kintsch, 1988), which highlights the dynamic interplay between working memory, prior knowledge, and textual input in the construction of meaning. According to this model, sentence processing involves the initial extraction of multiple levels of linguistic information—phonological, lexical, semantic, and syntactic—which are used to construct a preliminary mental representation. This representation is subsequently integrated with relevant long-term knowledge structures, enabling the formation of a coherent and contextually grounded mental model. In the context of immediate serial recall tasks, such integrative processes may facilitate both the comprehension and recall of verbal sequences by enriching item representations and establishing associative links that support retrieval. The results of these studies and associated models are, however, difficult to interpret in terms of a specific influence of syntactic knowledge on serial order memory, given that sentence recall not only involves access to syntactic structures but also involves semantic and conceptual knowledge. More recently, T. Jones and Farrell (2018) demonstrated that recall accuracy for verbal sequences is improved if they conform to legal syntax, even for meaningless sequences. Similar findings had already been observed (Perham et al., 2009) for recall of lists for semantically unrelated adjective–noun pairs, the pairs occurring in either legal or illegal syntactic order. However, both of these studies did not dissociate the contribution of syntactic structure to item versus serial order retention, as either no specific measure of item-level working memory performance was included or only overall serial recall performance was determined. Another study by Schwering and MacDonald (2023) demonstrated that specific lexico-semantic and syntactic constraints, such as word context and part of speech, influence verbal WM, particularly concerning item recall performance, with less clear evidence for an impact on serial order recall.

So far, only two studies have aimed more directly at examining the impact of syntactic knowledge on item vs. order memory. Schweppe and her colleagues (2022) presented lists of syntactically legal (adjective before noun) or illegal (noun before adjective) word pairs to German-speaking participants for immediate serial recall and specifically determined item and order recall scores. While observing a robust impact on item recall, they observed no impact on order recall. A similar pattern was observed in a French language study (Querella & Majerus, 2024). French has the specificity of allowing for both adjective anteposition (i.e., adjective before noun) and postposition (i.e., noun before adjective), but the specific type of position is determined in a probabilistic manner by the nature of the adjective. Despite these more flexible syntactic rules, Querella and Majerus (2024) still found no specific impact of adjective–noun syntactic legality ([1]) on order recall, the effects being restricted to item recall. This impact of syntactic legality on item recall, but not order recall, is surprising given that, by nature, syntactic knowledge about the sequential organization of words should impact memory for serial order. However, the manner in which the lists were constructed in these two studies (as well as in the study by Perham et al., 2009) might have contributed to this pattern of results. More specifically, for a given list, all pairs were either presented in adjective–noun order or noun–adjective order, and hence the ordering of adjectives and nouns became predictable. This is likely to have allowed for the creation of temporary syntactic frames during the syntactic parsing process of list encoding, which may have constrained the order in which the adjectives and nouns were output, irrespective of the syntactic legality of the sequences, hiding the potential impact of long-term syntactic knowledge about adjective–noun output order.

The present study reexamines the impact of syntactic knowledge about adjective–noun order on verbal WM performance by using lists mixing adjective anteposition and postposition. This should prevent the build-up of short-term syntactic predictability rules during list encoding and allow for a specific assessment of the impact of long-term syntactic knowledge on order recall for lists of adjective–noun pairs. Experiment 1 tested this hypothesis for adjective–noun pairs in French, which allows for a full crossing of adjective–noun position (anteposition vs. postposition) with syntactic legality*. Experiment 2 used a cross-language design by adapting the lists used in Experiment 1 for an administration in both French and German.

Experiment 1

Experiment 1 was conducted in French and was based on a 2 × 2 within-subject design with the factors, syntactic legality* (lists consisting of syntactically legal* pairs versus lists consisting of syntactically illegal* pairs) and adjective position (anteposition vs. postposition). Critically, a single trial involved both postposition and anteposition pairs to avoid the build-up of temporary syntactic predictability of adjective position within a list. The key dependent variables were serial recall scores, order recall scores, and within-pair order migration errors. In addition, we recorded item recall scores and omission errors.

Method

Transparency and openness

We reported how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The data and materials for all experiments are available online (https://osf.io/kuj6h/?view_only=b5a91544211941c3b7ff9e0cee455e76). The design, hypotheses, and analysis plan for Experiment 1 were preregistered (https://osf.io/x7sgk).

[1] When referring to French, we will append a '*' symbol to the term "legality" to acknowledge that French adjective–noun orderings cannot be categorized as exclusively legal or illegal, but legality is determined in a probabilistic manner.

Participants

Ninety-three participants (49 women, 44 men; average age = 21.39 years, *SD* = 1.90) were recruited (see Scoring and Analysis Procedure section for justification of sample size) from the Université de Liège staff and student community. All participants were native French speakers without any history of language, learning, neurological, or psychiatric disorder, and reported no current drug use (e.g., cannabis) or alcohol abuse. Participants who were not monolingual or spoke another language during half-time in their daily life were not eligible for inclusion in this study. Seventy-six participants were right-handed and seven were left-handed. One participant was excluded because of misunderstanding the task procedure. The study had been approved by the ethics committee of the Faculty of Medicine of the University of Liège. Participants were informed that no financial compensation was provided. All participants have given their informed consent to participate in the study before testing.

Materials

To select adjective–noun pairs and determine their syntactic preference in terms of adjective position, 57 additional young adults from the University of Liège community were recruited for an online syntactic preference judgment test (26 men and 31 women, 18 to 28 years old, average age = 21.84 years, *SD* = 2.11). Participants were presented with 130 French adjective and noun pairs with minimized semantic plausibility: 36 pairs taken from Querella and Majerus (2024) and 94 additional newly created pairs. For each pair, the two possible orders (adjective before noun or reverse) were displayed, one on each side of a switch bar. Participants needed to slide the bar according to their preferred order. The switch bar was divided into a scale with five different levels (1 = *strongly prefer left pair*, 2 = *prefer left pair*, 3 = *no preference,* 4 = *prefer right pair*, 5 = *strongly prefer right pair*). To control for the effect of inflectional morphology, all pairs involved masculine nouns that are not inflected.

Based on the judgment results, we kept the 48 pairs that showed the strongest judgment preference in either direction. We recoded the responses to mean 1 = *strongly preferring postposition* and 5 = *strongly preferring anteposition*. Twenty-four pairs with preferred anteposition (mean preference = 4.24, *SD* = 0.49, Minimum = 3.46, Maximum = 4.83) and 24 with preferred postposition (mean preference = 1.44, *SD* = 0.23, Minimum = 1.21, Maximum = 1.93) were used to create 16 lists containing each three adjective–noun pairs and by mixing adjective anteposition and postposition in each list. These lists were called the legal* lists. The 16 illegal* lists were created by simply reversing the order of the noun and adjective within each pairing of the legal* lists. This procedure ensured that legal* and illegal* conditions were matched for phonological similarity and the lexico-semantic status of the words. In addition, the number of syllables was kept constant at 11 syllables per list.

The material was recorded by a French-native female speaker using a neutral voice. Each item was recorded separately and then combined to form adjective–noun/noun–adjective pairs. The full stimulus lists are presented in Appendix Tables A7 and A8, the judgment results for each pair are in Table A9, and the characteristics descriptions of stimuli are in Table A10.

Procedure

The main experiment was conducted via OpenSesame software (https://osdoc.cogsci.nl/4.0/). Subjects were pseudorandomly assigned to one of two presentation conditions (Condition A: block of legal* lists followed by block of illegal* lists; Condition B: block of illegal* lists followed by legal* lists). The order of the lists in each block was the same for all participants. They were tested individually in silent testing rooms. Participants were instructed to carefully listen to each list of six items, presenting the items in three pairs via the specific timing of presentation, followed by immediate oral serial recall, asking the participants to recall all the words, from first to last word. If participants could not remember a word at a particular position of the list, they had to say "oublié" (forgotten) for that position. Before the presentation of the experimental lists, participants had to complete two practice trials to ensure that they had correctly understood the task instructions. In each trial, every word was presented for 1,000 ms. The interstimulus interval was 350 ms between nouns and adjectives in the same pair. Between each adjective–noun or noun–adjective pair, the interstimulus interval was longer (1,000 ms) to clearly mark the boundaries between the different pairs. By default, stimuli were presented via the speakers of the presentation PC. Headphones were also available if preferred by participants. Experimenters ensured that participants could clearly and comfortably perceive the stimuli during practice trials. There was no time limit for the participants to respond. The experiment lasted about 20 min per participant. All responses were digitally recorded for transcription and scoring.

Scoring and analysis procedure

Two standard recall accuracy scores were determined: serial recall score (number of items correctly recalled in correct serial position) and item recall score (number of items correctly recalled independently of serial position). A specific order recall score was also computed by dividing the sum of the serial recall score by the sum of the item recall score, as preregistered.

The following error types were identified: Within-pair order migration errors (the adjective and noun for a specific pair exchange their position) and omission errors (items for which the participant said "oublié" [forgotten] or items not recalled).

All analyses were conducted using a Bayesian statistical approach (see, e.g., Dienes, 2011; Morey & Rouder, 2011) according to our preregistration. This approach has the advantage of relying on a model comparison rationale to select and quantify the strength of evidence associated with each model, and crucially, allows for testing the strength of evidence for and against an effect of interest (i.e., positive evidence for the null hypothesis). The Bayesian framework does not involve traditional *p* values, thereby avoiding multiple testing problems such as alpha inflation (Wagenmakers et al., 2008). All analyses are based on Bayes factors (*BF*), which can be considered as a relative measure of statistical evidence (Morey et al., 2016). The *BF* represents the degree to which the observed data update the initial belief in favor of one hypothesis relative to another. The *BF* is the likelihood ratio of a given model, the best-fitting model being the one with the highest *BF*. BF_{01} indicates evidence in favor of the null hypothesis, while BF_{10} indicates evidence in favor of the alternative hypothesis. Although there are no fixed thresholds for *BF* values, we used the following categories for describing strength of evidence: A *BF* of at least 1 is considered to indicate anecdotal evidence, a *BF* of at least 3 is considered to indicate moderate evidence, a *BF* of at least 10 is considered to provide strong evidence, a *BF* of at least 30 is considered to provide very strong evidence, and a *BF* of at least 100 is considered to indicate decisive evidence (Jeffreys & Jeffreys, 1998).

The R platform and the *brms* package (Bürkner, 2017) were used for running Bayesian generalized linear mixed models. The dependent measures were the different scores and error types as defined above. Fixed factors included legality* (legal* vs. illegal*), adjective position (anteposition vs. postposition), and their interaction. Serial position was excluded from the analyses to reduce the model's complexity. The random intercept included the subject variable. To determine the random slopes to be included in the model, we first ran four models including the random intercept and different random slopes combinations (none, legality*, adjective position, interaction of legality* and adjective position). Random slopes were included in the subsequent models if the *BF* value for a given model was larger than 3, relative to the null model including only the intercept.

Sample size was monitored via a Bayesian sequential sampling approach (Schönbrodt & Wagenmakers, 2018) combined with an effect size stabilization procedure (Anderson et al., 2022). Previous work has demonstrated that relying on a fixed critical *p* value as the sole criterion for terminating data collection can lead to inflation of the Type I error rate, thereby compromising the validity of the resulting *p* values. When Bayesian methods are employed, the use of a threshold Bayes factor as a stopping rule has been shown to bias the evidential value of the data due to the influence of optional stopping. In contrast, adopting effect size stabilization as a stopping criterion does not introduce such bias and has been found to preserve the integrity of statistical inference under both frequentist and Bayesian frameworks. Beginning from the 20th participant, we computed, for each additional participant, the effect sizes for the effect of interest (syntactic legality*—signed Cohen's *d*) and continued sampling until the effect size, for both item and serial order recall scores, stabilized (a priori defined minimal absolute change of effect size $<.05$ over five consecutive analyses; Anderson et al., 2022). The effect sizes of the item recall score and serial recall score were both stable after the 35th participant. In addition, in order to guarantee similar statistical power between the present study and previous studies by Schweppe and her colleagues (2022) and Querella and Majerus (2024), which involved larger samples, we continued recruitment until reaching a comparable sample size. Also note that Kowialiewski (2024) recently showed that the effect size stabilization procedure does not guarantee the detection of a true effect in the population if the stabilized effect size is small.

Besides, for the display of the data in the figures, we presented the data as proportions. The proportion of item recall scores and serial recall scores was determined by dividing the number of correctly recalled items (or of the items recalled in the correct serial position) by the total number of items. The proportion of within-pair order migration errors was calculated by dividing the number of within-pair order migration errors by the total number of order errors (= difference between item and serial recall scores). Similarly, the proportion of omission errors was obtained by dividing the number of omission errors by the total number of item errors (= difference between item recall score and total number of items in the task).

Results

Serial recall score

A first Bayesian generalized linear mixed-model analysis was run on the serial recall score. The best-fitting model included only legality* ($\beta = 0.46$, 95% CI [0.28, 0.64], $BF_{10} = 1.32 \times 10^4$; see Appendix Table A1). This model was 3.76 times more likely than the model incorporating both legality* and adjective position (legality*: $\eta_p^2 = 0.232$; adjective position: $\eta_p^2 = 0.057$; legality* × adjective position: $\eta_p^2 = 0.013$). This means that serial recall scores were higher for syntactically legal* pairs (mean = 0.667, $SD = 0.159$) than illegal* pairs (mean = 0.603, SD = 0.174), regardless of whether adjectives were antepositional or postpositional (see Fig. 1 and Appendix Table A2).

Item recall score

For item recall scores, the same type of analysis identified the best-fitting and most parsimonious model as including legality* as a fixed factor (β = 0.18, 95% CI [0.09, 0.28], BF_{10} = 87.46; see Appendix Table A1). Although the model including legality* and adjective position is 1.40 times more likely than the model including only legality*, the latter should be retained as the more parsimonious model (legality*: η_p^2 = 0.109; adjective position: η_p^2 = 0.079; legality* × adjective position: η_p^2 = 0.035). The result means that item recall performance was higher for syntactically legal* pairs (mean = 0.780, SD = 0.110) than illegal* pairs (mean = 0.755, SD = 0.107), regardless of whether adjectives were antepositional or postpositional (see Fig. 1 and Appendix Table A2).

Order recall score

For the order recall score**,** given its derivational nature, a Bayesian analysis of variance was conducted on subject-averaged scores per experimental cell. The best model was again including the effect of legality* (β = 0.38, 95% CI [0.23, 0.53], BF_{10} = 1.18 × 10^4; see Appendix Table A1). This model was 5.76 times more likely than the model including both legality* and adjective position (legality*: η_p^2 = 0.242; adjective position: η_p^2 = 0.004; legality* × adjective position: η_p^2 = 0.006). That means order recall scores were higher for syntactically legal* pairs (mean = 0.848, SD = 0.119) than for illegal* pairs (mean = 0.789, SD = 0.154), independently of adjective positions (see Fig. 1 and Appendix Table A2).

Within-pair order migration errors

Next, we analyzed error scores. Generalized linear mixed analyses of within-pair order migration errors also supported the best-fitting model as including only legality* (β = −0.57, 95% CI [−0.77, −0.37], BF_{10} = 8.75 × 10^5; see Appendix Table A1). This model was 3.00 more likely than the model

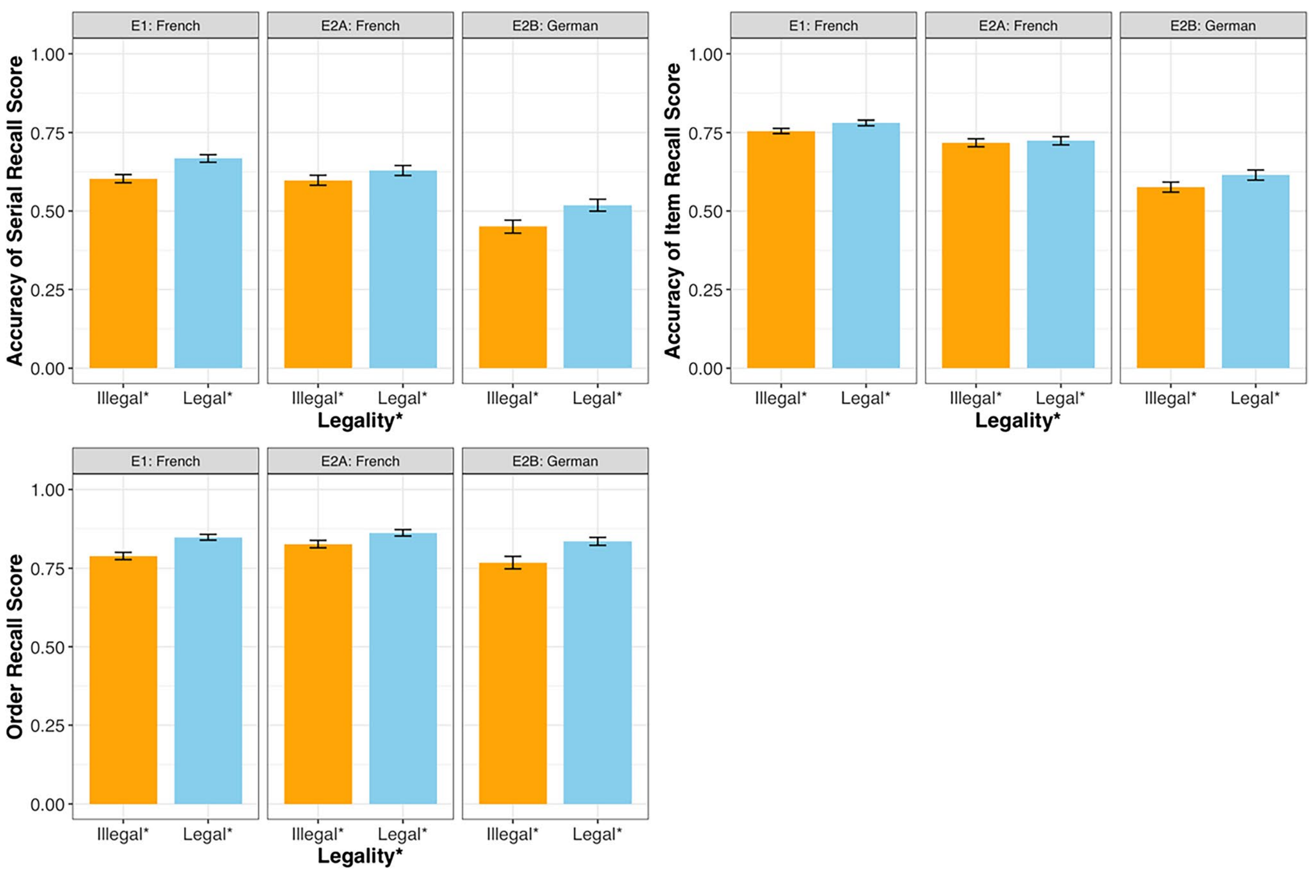

Fig. 1 Accuracy of serial recall score, item recall score, and order recall score, as a function of legality.*, across Experiment 1 (French), Experiment 2 A (French), and 2B (German)

including both legality* and interaction of legality* and adjective position (legality*: $\eta_p^2 = 0.182$; adjective position: $\eta_p^2 = 5.731 \times 10^{-4}$; legality* × adjective position: $\eta_p^2 = 1.672 \times 10^{-4}$). This result means that there were more order migration errors for syntactically illegal* pairs (mean = 0.245, *SD* = 0.093) compared with legal* pairs (mean = 0.161, *SD* = 0.096), independently of adjective position, indicating that for illegal* pairs, adjectives and nouns more frequently exchanged position to occupy their legal* position (see Fig. 2 and Appendix Table A2).

Omission errors

Finally, for omission errors, the best model included only legality* (β = −0.23, 95% CI [−0.37, −0.10], $BF_{10} = 43.78$; see Appendix Table A1). This model was 7.08 times more likely than the model including both legality* and adjective position (legality*: $\eta_p^2 = 0.090$; adjective position: $\eta_p^2 = 0.004$; legality* × adjective position: $\eta_p^2 = 0.002$). This result means that more omission errors occurred for syntactically illegal* pairs (mean = 0.190, *SD* = 0.086) than for legal* pairs (mean = 0.158, *SD* = 0.077), regardless of whether adjectives were antepositional or postpositional (see Fig. 2 and Appendix Table A2).

Discussion

On the one hand, Experiment 1 replicated the results of Querella and Majerus (2024) as well as those by Schweppe and her colleagues (2022) by showing an impact of syntactic legality* on the overall serial recall and item recall scores. Critically, however, we also observed an effect of syntactic legality* on order recall, with reduced order recall accuracy for syntactic illegal* pairs accompanied by an increase in order migration errors. For illegal* pairs, these errors correspond to order regularization errors, where an adjective or a noun in an illegal* position moves to a legal* position. The present results thus show that, when the position of the adjectives (relative to the nouns) cannot be easily predicted based on temporary syntactic parsing of the list at encoding, then long-term syntactic knowledge structures can be shown to exert an impact on order recall. To determine the robustness of these novel findings, we aimed to replicate these results as well as to test their generality by extending them to the German language, in which adjective–noun order is deterministic rather than probabilistic.

Experiment 2

Experiment 2 aimed to replicate the impact of syntactic legality on order recall in WM and to extend the findings to the German language. Given that in German, only adjective anteposition is legal, mixing adjective anteposition and postposition in the same list but still comparing entirely legal and illegal lists is not possible. Therefore, reducing order predictability based on adjective position is only possible by mixing legal and illegal pairs in the same list (as adjective postposition will always be illegal in German). As this change in itself may affect the findings, we also used the same type of list setup for the replication study in the French language. Experiment 2 was thus divided into two subexperiments. Experiment 2 A aimed at replicating the findings of Experiment 1 in the French language by using an adapted list setup that could also be compared with a German version of the task. Experiment 2B then aimed at extending the findings to the German language. In order to ensure maximal comparability between the two language versions, we used for the French version only adjective postposition stimuli as legal* stimulus pairs, given that adjective postposition is more frequent than adjective anteposition and thus has a status more similar to the adjective anteposition

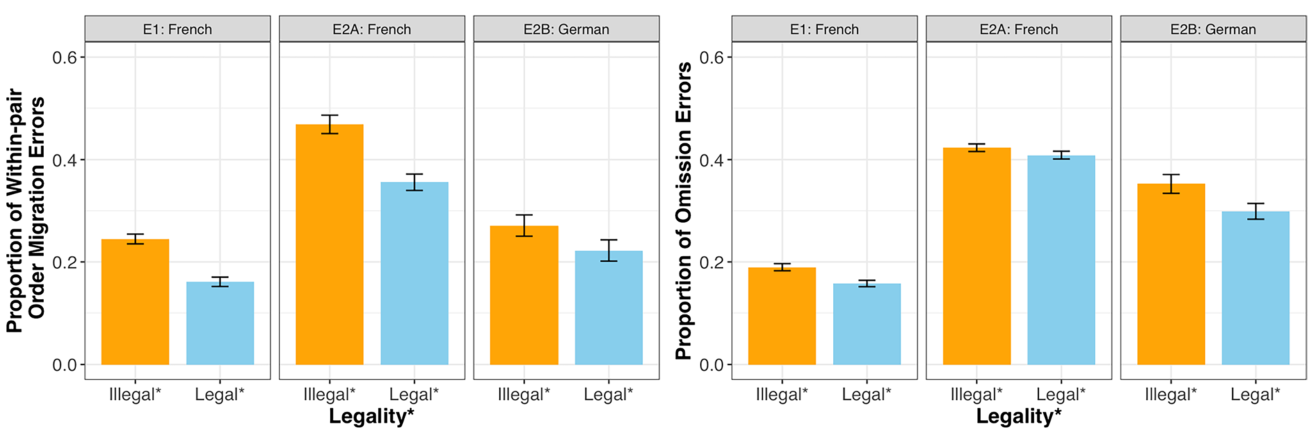

Fig. 2 Proportion of within-pair order migration errors and omission errors, as a function of legality.*, across Experiment 1 (French), Experiment 2 A (French), and 2B (German)

pairs, which are the only legal type of pair in German. Both experiments were based on one-factorial designs with the within-subjects factor syntactic legality* (syntactically legal* vs. syntactically illegal* pairs). We expected effects of syntactic legality* on order recall in both languages.

Experiment 2A

Method

Transparency and openness We reported how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The design, hypotheses, and analysis plan for Experiment 2 were preregistered (https://osf.io/8z7t2).

Participants Eighty-four participants (43 women and 41 men) 18 to 30 years old (average age = 21.6 years, *SD* = 1.87) were recruited (see Scoring and Analysis Procedure section for justification of sample size) from the Université de Liège by advertisements or social media. Seventy-nine of them were right-handed and five were left-handed. The study had been approved by the ethics committee of the Faculty of Medicine of the University of Liège. Participants were informed that no financial compensation was provided. All participants have given their informed consent to participate in the study before testing.

Materials The stimuli were selected from the stimulus pool of Experiment 1, with the selection restricted to preferred postpositional adjective–noun pairs. Consistent with Experiment 1, 16 lists, each containing three adjective–noun pairs, were constructed while controlling for phonological similarity (e.g., avoiding identical rhythmic patterns) and minimizing potential compound formations within each list. Additionally, the total number of syllables per list was controlled at 11. In total, 48 syntactically preferred postposition adjective–noun pairs (mean preference = 1.61, *SD* = 0.31, range: 1.21–2.12) were used to generate the experimental lists, 21 of which were the same pairs as in Experiment 1. To introduce syntactic unpredictability, the order of one or two adjective–noun pairs was reversed within each list (e.g., 2 adj-n + 1 n-adj or 1 adj-n + 2 n-adj per list). We first generated 16 lists (List 1) and then systematically reversed the adjective–noun order within the same pairs to create the corresponding set of lists (List 2). As in Experiment 1, 32 lists in total were presented.

Different from Experiment 1, all stimuli were presented in written format on a computer screen to eliminate potential prosodic-related differences when comparing the French and German subexperiments. The same timing and duration parameters were used as for Experiment 1. Full stimulus lists are provided in Appendix Tables A11 and A12, the judgment results for each pair are in Table A13, and the characteristics descriptions are in Table A14.

Procedure The procedure matched that in Experiment 1. Subjects were pseudorandomly assigned to one of two presentation conditions (Condition A: block of List 1 followed by block of List 2; Condition B: block of List 2 followed by block of List 1).

Scoring and analysis procedure Data scoring was identical to Experiment 1, as preregistered. The statistical design was the following: fixed factors only included legality* (legal* vs. illegal*), and the random-effect structure included random intercept of subjects, with or without random slope of legality* by subject. The same two-step Bayesian model comparison approach was used as in Experiment 1. Likewise, sample size had been determined via a Bayesian sequential sampling approach based on effect size stabilization (Anderson et al., 2022). The effect sizes of the item recall score and serial recall score were both stable after the 35th participant. As in Experiment 1, a higher number of participants was recruited to match the number of participants in the studies by Querella and Majerus (2024), and Schweppe and colleagues (2022).

Results

Serial recall score The first Bayesian generalized linear mixed-model analysis on the serial recall scores indicated that the best-fitting model included legality* ($\beta = 0.30$, 95% CI [0.17, 0.43]; see Appendix Table A3). This model was 1.86×10^3 times more likely than the model excluding legality* (legality*: $\eta_p^2 = 0.224$). That means serial recall performance was significantly better for syntactically legal* pairs (mean = 0.629, *SD* = 0.016) compared with illegal* pairs (mean = 0.598, *SD* = 0.146), the same as in Experiment 1 (see Fig. 1 and Appendix Table A4).

Item recall score For item recall scores, the factor legality* was associated with evidence for its absence ($BF_{01} = 5.79$; $\eta_p^2 = 0.014$; see Appendix Table A3). The result means that item recall performance did not differ between syntactically legal* pairs (*mean* = 0.724, *SD* = 0.120) and illegal* pairs (mean = 0.718, *SD* = 0.117), in contrast to Experiment 1 (see Fig. 1 and Appendix Table A4).

Order recall score For order recall scores, a Bayesian analysis of variance showed that the best-fitting model included legality* ($\beta = 0.28$, 95% CI [0.14, 0.41]; see Appendix Table A3). This model was 5.60×10^2 times more likely than the model excluding legality* (legality*: $\eta_p^2 = 0.198$).

That means order recall performance was superior for syntactically legal* pairs (mean = 0.862, SD = 0.094) compared with illegal* pairs (mean = 0.826, SD = 0.109), consistent with Experiment 1 (see Fig. 1 and Appendix Table A4).

Within-pair order migration errors Generalized linear mixed model analyses on within-pair order migration errors demonstrated that the best-fitting model included legality* (β = −0.33, 95% CI [−0.49, −0.18]; see Appendix Table A3). This model was 7.51×10^2 times more likely than the model excluding legality* (legality*: $\eta_p^2 = 0.135$). That means fewer order migration errors were observed for syntactically legal* pairs (mean = 0.356, SD = 0.146) than for illegal* pairs (mean = 0.469, SD = 0.162). Adjectives and nouns in syntactically illegal* positions were inclined to be recalled in their syntactically legal* positions, in line with prior findings in Experiment 1 (see Fig. 2 and Appendix Table A4).

Omission errors Finally, for omission errors, the factor legality* was associated with evidence for its absence ($BF_{01} = 4.13$; $\eta_p^2 = 0.024$; see Appendix Table A3). The result means that no impact of syntactic legality* was observed for omission errors (mean = 0.409, SD = 0.070 vs. mean = 0.432, SD = 0.068), consistent with the results for item recall scores but different from Experiment 1 (see Fig. 2 and Appendix Table A4).

Experiment 2B

Method

Participants Fifty participants (30 women and 20 men) aged from 20 to 31 (average age = 22.8, SD = 2.12) were recruited (see Scoring and Analysis Procedure section for justification of sample size) from the University of Passau. Forty-seven participants were right-handed and three were left-handed. Participants received course credit for their participation. Participants were treated in accordance with APA ethical standards as well as the guidelines of the German Research Foundation (DFG) and the German Psychological Society (DGPs). As the study was nonmedical, low-risk research, no explicit approval was required from the responsible ethics committee at the University of Passau. All participants gave their informed consent to participate in the study before testing.

Materials and procedure The stimuli comprised 48 adjective–noun pairs. To align with the materials in Experiment 2A, only masculine nouns were used. To maintain maximal comparability with Experiment 2 A, the stimuli were directly translated into German when the German noun was also of masculine gender (e.g., "nervöser–Busch" translated from "bus–nerveux" in French). In some cases, adjective–noun pairs were recombined to minimize phonological similarity within pairs and to equate list length. In addition, pairs with masculine nouns from Schweppe et al. (2022) were included. Unlike in French, the adjectives also needed to be inflected with masculine nouns. This resulted in a slightly longer list length than in Experiment 2A, with 13 or 14 syllables per list (mean length = 13.56 syllables). As in Experiment 2 A, the order of one or two adjective–noun pairs was reversed within each of the 16 lists (List 1). A corresponding set of 16 lists (List 2) was created by systematically reversing the adjective–noun order within each of the pairs, resulting in 32 lists in total. The complete set of stimulus lists is provided in the Appendix (Tables A15 and A16), and the characteristics descriptions of stimuli are in Table A17.

The procedure was the same as in Experiment 2A except that the experiment was conducted via JsPsych (Version 6.3; de Leeuw et al., 2023) with the psychophysics plugin (Kuroki, 2021).

Scoring and analysis procedure The data scoring and analysis procedures were the same as in Experiment 2A.

Sample size had been determined via a Bayesian sequential sampling approach based on effect size stabilization (Anderson et al., 2022). The effect sizes of the item recall score and serial recall score were both stable after the 25th participant. As in the other experiments, a higher number of participants was recruited. Given the clear-cut results in Experiments 1and 2A, recruitment, however, stopped at around 50 in order to avoid unnecessary participant recruitment and time investment.

Results

Serial recall score We first conducted a generalized linear mixed model analysis on the serial recall scores. Results showed that the best-fitting model included legality (β = 0.36, 95% CI [0.14, 0.59]; see Appendix Table A5). This model was 27.26 times more likely than the model excluding legality (legality: $\eta_p^2 = 0.474$). That means serial recall performance was better for syntactically legal adjective–noun pairs (mean = 0.519, SD = 0.135) compared with illegal pairs (mean = 0.450, SD = 0.146), consistent with findings from Experiment 1 and Experiment 2A (see Fig. 1 and Appendix Table A6).

Item recall score Next, the same analysis was performed on item recall scores. The best generalized mixed linear model included legality (β = 0.17, 95% CI [0.09, 0.25]; see Appendix Table A5). This model was 2.79×10^2 times more likely than the following model, excluding legality (legality: $\eta_p^2 = 0.303$). The result means that there was superior item recall performance for syntactically legal pairs (mean = 0.615, *SD* = 0.114) compared with illegal pairs (mean = 0.576, *SD* = 0.113). Interestingly, while differing from the results of Experiment 2A, this finding aligned with Experiment 1 (see Fig. 1 and Appendix Table A6).

Order recall score For order recall scores, a Bayesian analysis of variance indicated the best-fitting model included legality (β = 0.39, 95% CI [0.16, 0.61]; see Appendix Table A5). This model was 53.71 times more likely than the model excluding legality (legality: $\eta_p^2 = 0.220$). That means that order recall performance of adjectives and nouns was better in syntactic legal order (mean = 0.835, *SD* = 0.090) than in syntactic illegal order (mean = 0.768, *SD* = 0.140), in line with Experiments 1 and 2A once again (see Fig. 1 and Appendix Table A6).

Within-pair order migration errors Moreover, generalized linear mixed-model analyses were conducted to examine errors. Analyses of within-pair order migration errors indicated that the best-fitting model included legality (β = −0.80, 95% CI [−1.36, −0.28]; see Appendix Table A5). This model was 51.37 times more likely than the model excluding legality (legality: $\eta_p^2 = 0.038$). That means more within-pair order migration errors were observed with adjectives and nouns in syntactically illegal order (mean = 0.271, *SD* = 0.147) than legal order (mean = 0.222, *SD* = 0.147), indicating a higher frequency of adjectives and nouns in syntactically illegal order recalled at legal positions, consistent with Experiment 1 and 2A (see Fig. 2 and Appendix Table A6).

Omission errors Finally, the generalized linear mixed model analysis for omission errors identified the best-fitting model as including legality (β = −0.23, 95% CI [−0.34, −0.12]; see Appendix Table A5). This model was 2.99×10^2 times more likely than the model excluding legality (legality: $\eta_p^2 = 0.290$). This result means that more items were forgotten when recalling syntactically illegal adjective–noun pairs (mean = 0.353, *SD* = 0.130) compared to legal pairs (mean = 0.299, *SD* = 0.109), aligned with Experiment 1, while contrasting to Experiment 2A (see Fig. 2 and Appendix Table A6).

Discussion

In sum, in both Experiment 2A and 2B, we observed a clear advantage for syntactically legal adjective–noun pairs over illegal pairs on order recall performance. Furthermore, we found the higher incidence of within-pair order migration errors for syntactically illegal pairs, where an adjective or a noun in an illegal* position has a greater possibility to move to a legal* position. One difference between Experiments 2A and 2B needs, however, to be noted. While in Experiment 2B (as well as in Experiment 1) we also observed recall performance differences on item recall and omission errors between legal* and illegal* pairs, this was not the case in Experiment 2A. This discrepancy may be explained by language-specific syntactic constraints and the specific setup of the lists in Experiment 2A. In French, as already mentioned, adjective–noun order is more flexible, allowing for greater variability in parsing and recall strategies, which may reduce the impact of syntactically illegal pairs on recall performance, particularly when only a subset of syntactic rules is being manipulated as opposed to Experiment 1. This idea is supported by an overall smaller difference between legal and illegal conditions in Experiment 2A for most of the scores, relative to Experiment 1, as well as to Experiment 2B.

General discussion

The present study replicated but also critically extends the experiments of Querella and Majerus (2024) and Schweppe et al. (2022), while introducing a crucial methodological modification: we intermixed adjective–noun and noun–adjective pairs within the same list to prevent the build-up of list-level temporary syntactic predictability rules. This methodological modification was associated with the observation of not only an impact of syntactic legality on item recall, but also, crucially, on order recall and order migration errors, in both French and German languages.

Traditional models of verbal WM, such as that of Baddeley and Hitch (1974), have suggested that serial order information is maintained via phonological mechanisms within a phonological store (Baddeley, 2000). However, this perspective has been challenged as it does not fully account for the complexity of serial order phenomena (e.g., Hurlstone et al., 2014). Recent models of verbal working memory increasingly consider the interaction between lexico-semantic knowledge and working memory processes to explain serial recall. The Activated Network (ANet) model (Poirier et al., 2015) proposes that order information is supported by activation dynamics within long-term semantic memory: when later items are semantically related to earlier ones, their activation is boosted, increasing the likelihood of early recall. Other approaches have attempted to formalize the interface between semantic and episodic memory through computational modeling. For example, Mewhort and colleagues (2018) introduced a holographic memory model (Franklin & Mewhort, 2015) that uses BEAGLE vectors (Bound Encoding of the Aggregate Language Environment; Jones et al., 2006; Jones & Mewhort, 2007) to capture semantic similarity and explain effects such as release from proactive interference in immediate recall tasks. More recently, Guitard et al. (2025) developed the Embedded Computational Framework of Memory (eCFM), which combines Latent Semantic Analysis (Landauer & Dumais, 1997) with an episodic retrieval process based on MINERVA 2 (Hintzman, 1986) to account for the influence of semantic structure in working memory performance. However, these models have primarily focused on lexico-semantic features while largely overlooking the contribution of syntactic information to working memory. Moreover, most of them tend to treat memory performance as a unified outcome, without separating the cognitive mechanisms underlying item recall from those supporting order recall. This limits their ability to fully account for the complexity of language-based memory tasks in their current form. From another aspect, behavioral and neuroimaging evidence has suggested that serial order relies on domain-general mechanisms, such as spatial or temporal coding (e.g., Brown et al., 2000; Burgess & Hitch, 2006; Hartley et al., 2016; Majerus, 2008, 2013; van Dijck & Fias, 2011), whereas item recall is more directly linked to linguistic knowledge, particularly at the lexical-semantic level (Gathercole et al., 2001; Hulme et al., 1991, 1997; Saint-aubin & Poirier, 1999; Walker & Hulme, 1999). However, our findings challenge the notion that the representation of serial order information is fully isolated from language processing. Instead, the observation of an impact of syntactic legality on order recall performance and order migration errors supports emergent models of verbal WM, which propose that WM and language systems are deeply interconnected and interact dynamically (Acheson & MacDonald, 2009; Buchsbaum & D'Esposito, 2019; Cowan, 1993; Hasson et al., 2015; MacDonald, 2016; Postle, 2006; Schwering & MacDonald, 2020). These models emphasize that verbal WM is the activated portion of linguistic long-term memory, and all aspects of verbal WM, including serial order, are grounded in linguistic structures (Schwering & MacDonald, 2020). The present study provides support for this perspective, with syntactic information being crucial for serial positioning, even in word lists. Our findings are also in line with other recent studies that suggest there is an impact of linguistic variables on serial order coding. These findings contribute to a more refined understanding of the interaction between language and the representation of serial order information in verbal WM. While previous models of serial order WM have primarily emphasized the role of temporal context signals (Burgess & Hitch, 1999, 2006; Lewandowsky & Farrell, 2008; Oberauer et al., 2012), our results suggest that serial order encoding and maintenance are also influenced by fundamental linguistic structures, aligning with "limited emergent" approaches. Future models of serial order WM may consider the usefulness of integrating syntactic sequential knowledge structures in order to fully account for the multiple mechanisms that support coding of serial order information in WM, in particular when the memoranda do not simply represent one word class.

The present study further highlights the importance of controlling for the multiplicity of variables that can intervene in a verbal WM task when aiming at isolating the impact of a specific variable. Given the robust impact of syntactic legality on order memory in the present study, and its robust absence in previous studies (Querella & Majerus, 2024; Schweppe et al., 2022), it is indeed very likely that the inclusion of adjectives in the same position for a given list was hiding the contribution of long-term syntactic knowledge in the latter studies. Querella and Majerus (2024) had argued that this situation may have led to the build-up of temporary syntactic predictability rules, based on the creation of a particularly robust syntactic frame during the parsing of the stimuli during the encoding of the list. This temporary syntactic frame may have overcome the impact of syntactic knowledge on order recall and order migration errors, as it will constrain the order in which adjectives and nouns are expected to be output for a given list. Paradoxically, this situation may also have increased the impact of syntactic legality on item recall, as the temporary syntactic frame will predict the occurrence of an adjective for a position in which an item reflecting an adjective is not a suitable candidate based on syntactic legality, leading to an item omission error instead of an order error. In the present study, we prevented the impact of temporary syntactic parsing regularities by alternating adjective anteposition/postposition in the same lists, and we were able to observe an impact of syntactic legality on order recall, and this impact appears to be even slightly more robust than the impact on item recall.

In conclusion, the present study provides direct evidence for the influence of syntactic knowledge on order recall performance in verbal WM, as opposed to item recall performance. Our findings indicate the need for a deeper integration of language and WM architectures, including for the modelling of serial order mechanisms.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-025-01788-9.

Acknowledgements This work was supported by a grant from Chinese Scholarship Council (CSC) Fellowship awarded to Hong Xiao. We would like to thank all the participants for the time invested in this study, and thank Magdalena Pucher, Christina Gerweth, Lara Hellmann, Jessica Wagner, and Philipp Radloff, and other students who helped us with the data collection.

Funding This work was supported by a grant from Chinese Scholarship Council (CSC) Fellowship awarded to Hong Xiao.

Data availability Data and materials are available at: https://osf.io/kuj6h/?view_only=b5a91544211941c3b7ff9e0cee455e76

Code availability Code of generalized linear mixed model analyses are available at: https://osf.io/kuj6h/?view_only=b5a91544211941c3b7ff9e0cee455e76

Declarations

Conflicts of interest No conflict of interest in this study.

Ethics approval This study was approved by the ethics committee of the Faculty of Medicine of the University of Liège, and the responsible ethics committee at the University of Passau.

Consent to participate Informed consent was obtained from all individual participants included in this study.

Consent for publication Not applicable.

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. https://doi.org/10.1037/a0014411

Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2018). Exploring the sentence advantage in working memory: Insights from serial recall and recognition. *The Quarterly Journal of Experimental Psychology, 71*(12), 2571–2585. https://doi.org/10.1177/1747021817746929

Anderson, R. B., Crawford, J. C., & Bailey, M. H. (2022). Biasing the input: A yoked-scientist demonstration of the distorting effects of optional stopping on Bayesian inference. *Behavior Research Methods, 54*(3), 1131–1147. https://doi.org/10.3758/s13428-021-01618-1

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (2017). *Exploring working memory*. Routledge.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic. https://doi.org/10.1016/S0079-7421(08)60452-1

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

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller, & P. D. Eimas (Eds.), *Handbook of perception and cognition. Vol. 11: Speech, language, and communication* (pp. 181–216). Academic.

Bonhage, C. E., Fiebach, C. J., Bahlmann, J., & Mueller, J. L. (2014). Brain signature of working memory for sentence structure: Enriched encoding and facilitated maintenance. *Journal of Cognitive Neuroscience, 26*(8), 1654–1671. https://doi.org/10.1162/jocn_a_00566

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. https://doi.org/10.3758/BF03200768

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*(5), 467–482. https://doi.org/10.1037/h0061096

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

Buchsbaum, B. R., & D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex, 112*, 134–148. https://doi.org/10.1016/j.cortex.2018.11.010

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. https://doi.org/10.1037/0033-295X.106.3.551

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. https://doi.org/10.1016/j.jml.2006.08.005

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80,* 1–28. https://doi.org/10.18637/jss.v080.i01

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11*(41), 63–65.

Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition, 21*(2), 162–167. https://doi.org/10.3758/BF03202728

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 University Press. https://doi.org/10.1017/CBO9781139174909.006

de Leeuw, J., Gilbert, R., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *The Journal of Open Source Software*, *8,* 5351. https://doi.org/10.21105/joss.05351

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Franklin, D. R. J., & Mewhort, D. J. K. (2015). Memory as a hologram: An analysis of learning and recall. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 69*(1), 115–135. https://doi.org/10.1037/cep0000035

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

Garrett, M. F. (1980). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 263–271). Academic.

George A., Miller Stephen, Isard (1963) Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2*(3), 217–228. https://doi.org/10.1016/S0022-5371(63)80087-0

Gregg, V. H., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology, 80*(3), 363–374. https://doi.org/10.1111/j.2044-8295.1989.tb02326.x

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*(2), 267–333. https://doi.org/10.1006/brln.1997.1819

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. https://doi.org/10.1037/a0025481

Guitard, D., Saint-Aubin, J., Reid, J. N., et al. (2025). An embedded computational framework of memory: The critical role

of representations in veridical and false recall predictions. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-025-02669-7
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. https://doi.org/10.1016/j.cogpsych.2016.05.001
Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006
Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia, 38*(4), 426–440. https://doi.org/10.1016/S0028-3932(99)00098-6
Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411
Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F
Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*(4), 527–536. https://doi.org/10.1111/j.2044-8295.1995.tb02570.x
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. https://doi.org/10.1037/0278-7393.23.5.1217
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. https://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), 339–373. https://doi.org/10.1037/a0034221
Jefferies, E., Lambon Ralph, M. A., & Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language, 51*(4), 623–643. https://doi.org/10.1016/j.jml.2004.07.005
Jeffreys, S. H., & Jeffreys, S. H. (1998). *The theory of probability* (3rd ed.). Oxford University Press.
Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language, 55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003
Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1
Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1–13. https://doi.org/10.1016/j.cognition.2015.07.009
Jones, G., & Macken, B. (2018). Long-term associative learning predicts verbal short-term memory performance. *Memory & Cognition, 46*(2), 216–229. https://doi.org/10.3758/s13421-017-0759-3
Jones, T., & Farrell, S. (2018). Does syntax bias serial order reconstruction of verbal short-term memory? *Journal of Memory and Language, 100*, 98–122. https://doi.org/10.1016/j.jml.2018.02.001
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163
Kowialiewski, B. (2024). The power of effect size stabilization. *Behavior Research Methods, 57*(1), 7. https://doi.org/10.3758/s13428-024-02549-3
Kowialiewski, B., & Majerus, S. (2018). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language, 101*, 64–83. https://doi.org/10.1016/j.jml.2018.03.005
Kowialiewski, B., Majerus, S., & Oberauer, K. (2024). Does semantic similarity affect immediate memory for order? Usually not, but sometimes it does. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*(1), 68–88. https://doi.org/10.1037/xlm0001279
Kuroki, D. (2021). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods, 53*(1), 301–310. https://doi.org/10.3758/s13428-020-01445-w
Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 49, pp. 1–48). Academic. https://doi.org/10.1016/S0079-7421(08)00001-7
Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language, 31*(6), 713–733. https://doi.org/10.1016/0749-596X(92)90036-W
MacDonald, M. C. (2016). Speak, act, remember: The Language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science, 25*(1), 47–53. https://doi.org/10.1177/0963721415620776
Majerus, S. (2008). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. S. C. Thorn & M. P. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). Psychology Press.
Majerus, S. (2013). Language repetition and short-term memory: An integrative framework. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00357
Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.04.016
Majerus, S., Attout, L., Artielle, M.-A., & Van der Kaa, M.-A. (2015). The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia, 77*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2015.08.010
Majerus, S., & D'Argembeau, A. (2011). Verbal short-term memory reflects the organization of long-term memory: Further evidence from short-term memory for emotional words. *Journal of Memory and Language, 64*(2), 181–197. https://doi.org/10.1016/j.jml.2010.10.003
Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., . . . Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience*, *22*(11), 2570–2593. https://doi.org/10.1162/jocn.2009.21378
Majerus, S., Martinez Perez, 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. https://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. https://doi.org/10.1080/02643290600989376

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., . . . Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? *NeuroImage*, *32*(2), 880–891. https://doi.org/10.1016/j.neuroimage.2006.03.048

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. https://doi.org/10.1348/026151003765264101

Marks, L. E., & Miller, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior, 3*(1), 1–5. https://doi.org/10.1016/S0022-5371(64)80052-9

Marshuetz, C., Reuter-Lorenz, P. A., Smith, E. E., Jonides, J., & Noll, D. C. (2006). Working memory for order and the parietal cortex: An event-related functional magnetic resonance imaging study. *Neuroscience, 139*(1), 311–316. https://doi.org/10.1016/j.neuroscience.2005.04.071

Marshuetz, C., Smith, E. E., Jonides, J., DeGutis, J., & Chenevert, T. L. (2000). Order information in working memory: FMRI evidence for parietal and prefrontal mechanisms. *Journal of Cognitive Neuroscience, 12*(Supplement 2), 130–144. https://doi.org/10.1162/08989290051137459

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. https://doi.org/10.1006/brln.1994.1061

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *The American Journal of Psychology, 63*(2), 176–185. https://doi.org/10.2307/1418920

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

Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology, 72*, 6–18. https://doi.org/10.1016/j.jmp.2015.11.001

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16*(4), 406–419. https://doi.org/10.1037/a0024377

Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 594–598. https://doi.org/10.1037/0278-7393.28.3.594

Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language, 50*(2), 113–133. https://doi.org/10.1016/j.jml.2003.09.005

Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition, 35*, 1940–1953. https://doi.org/10.3758/BF03192927

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. https://doi.org/10.3758/s13423-012-0272-4

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. https://doi.org/10.1162/jocn.1994.6.1.57

Perham, N., Marsh, J. E., & Jones, D. M. (2009). Short article: Syntax and serial recall: How language supports short-term memory for order. *Quarterly Journal of Experimental Psychology, 62*(7), 1285–1293. https://doi.org/10.1080/17470210802635599

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

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology / Revue Canadienne De Psychologie Expérimentale, 50*(4), 408–412. https://doi.org/10.1037/1196-1961.50.4.408

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*(1), 23–38. https://doi.org/10.1016/j.neuroscience.2005.06.005

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*(6), 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*(3), 265–282. https://doi.org/10.1006/jmla.1997.2546

Querella, P., & Majerus, S. (2024). Sequential syntactic knowledge supports item but not order recall in verbal working memory. *Memory & Cognition, 52*(8), 1737–1761. https://doi.org/10.3758/s13421-023-01476-6

Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., & Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22*(6), 695–701. https://doi.org/10.3758/BF03209254

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology Section A, 52*(2), 367–394. https://doi.org/10.1080/713755814

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. https://doi.org/10.3758/BF03212990

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. https://doi.org/10.3758/bf03196364

Saint-Aubin, J., Poirier, M., Yearsley, J. M., Robichaud, J.-M., & Guitard, D. (2023). Modeling verbal short-term memory: A walk around the neighborhood. *Journal of Experimental Psychology: Learning Memory and Cognition, 49*(2), 198–215. https://doi.org/10.1037/xlm0001226

Scheerer-Neumann, G. (1981). The utilization of intraword structure in poor readers: Experimental evidence and a training program. *Psychological Research, 43*, 155–178. https://doi.org/10.1007/BF00309827

Savill, N., Ellis, R., Brooke, E., Koa, T., Ferguson, S., Rojas-Rodriguez, E., . . . Jefferies, E. (2018). Keeping it together: Semantic coherence stabilizes phonological sequences in short-term memory. *Memory & Cognition*, *46*(3), 426–437. https://doi.org/10.3758/s13421-017-0775-3

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. https://doi.org/10.1016/j.jml.2015.06.003

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schweppe, J., Schütte, F., Machleb, F., & Hellfritsch, M. (2022). Syntax, morphosyntax, and serial recall: How language supports short-term memory. *Memory & Cognition, 50*(1), 174–191. https://doi.org/10.3758/s13421-021-01203-z

Schwering, S. C., & MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production.

Frontiers in Human Neuroscience, *14*. https://doi.org/10.3389/fnhum.2020.00068

Schwering, S. C., & MacDonald, M. C. (2023). Noun sequence statistics affect serial recall and order recognition memory. *Open Mind, 7*, 550–563. https://doi.org/10.1162/opmi_a_00092

Thomas K., Landauer Susan T., Dumais (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. https://doi.org/10.1037/0033-295X.104.2.211

Tse, C.-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory, 17*(8), 874–891. https://doi.org/10.1080/09658210903376250

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. https://doi.org/10.1080/17470218.2011.604787

van Dijck, J.-P., & Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition, 119*(1), 114–119. https://doi.org/10.1016/j.cognition.2010.12.013

Wagenmakers, E.-J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). Springer. https://doi.org/10.1007/978-0-387-09612-4_9

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

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. https://doi.org/10.1037/0278-7393.3.6.712

Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language, 58*(1), 48–65. https://doi.org/10.1016/j.jml.2007.07.002