Memory & Cognition
https://doi.org/10.3758/s13421-023-01496-2

Lexico-syntactic constraints influence verbal working memory in sentence-like lists

Steven C. Schwering[1] · Cassandra L. Jacobs[2] · Janelle Montemayor[3] · Maryellen C. MacDonald[1]

Accepted: 11 November 2023

Abstract

We test predictions from the language emergent perspective on verbal working memory that lexico-syntactic constraints should support both item and order memory. In natural language, long-term knowledge of lexico-syntactic patterns involving part of speech, verb biases, and noun animacy support language comprehension and production. In three experiments, participants were presented with randomly generated dative-like sentences or lists in which part of speech, verb biases, and animacy of a single word were manipulated. Participants were more likely to recall words in the correct position when presented with a verb over a noun in the verb position, a good dative verb over an intransitive verb in the verb position, and an animate noun over an inanimate noun in the subject noun position. These results demonstrate that interactions between words and their context in the form of lexico-syntactic constraints influence verbal working memory.

Keywords Verbal working memory · Long-term memory · Serial recall · Language

Theories of working memory often address the ways in which long-term memory (LTM) affects the brief storage of information. Emergent theories suggest working memory is the activated portion of LTM under the focus of attention (Cowan, 1993, 1999), with verbal working memory (VWM) at least in part supported by language comprehension and production processes and strongly integrated with linguistic LTM (Acheson & MacDonald, 2009; MacDonald, 2016; Majerus, 2013; Schwering & MacDonald, 2020). By contrast, buffer accounts assume memoranda are maintained in a dedicated temporary store or buffer, separate from linguistic LTM, though the two systems may limitedly interact to support one another (Baddeley, 2000; Baddeley & Hitch, 1994; Norris, 2017). Buffer and emergent accounts of VWM therefore offer divergent perspectives on the architecture of verbal memory systems; any progress in distinguishing these perspectives should further our understanding of the interactions between VWM, LTM, and the language comprehension and production system.

Here, we sought to test the viability of an emergent perspective on VWM by manipulating sentence-likeness. We leverage the sentence superiority effect, the finding that memory lists with similarities to natural language are recalled better than random word lists (Allen et al., 2018; Baddeley et al., 2009; Jones & Farrell, 2018; Lombardi & Potter, 1992; Miller & Selfridge, 1950; Perham et al., 2009; Potter & Lombardi, 1990). Buffer and emergent theories of VWM diverge in how they explain the effects of sentence-likeness on memory using different mechanisms. These theories have implications for characterizing both the contents of linguistic LTM and how linguistic LTM influences VWM. We first review how inter-word relationships are characterized in verbal working memory and psycholinguistics respectively. We then present two studies that test how psycholinguistic definitions of sentence-likeness can refine our understanding of the characterization of linguistic LTM and its interactions with VWM.

Sentence-likeness in VWM research

Sentence-likeness is typically defined through reference to long-term knowledge of (linear) word order (Allen et al., 2018; Baddeley et al., 2009; Jones & Farrell, 2018; Perham et al., 2009). Lists are sentence-like in two ways: the list of words obeys abstract grammar rules or exactly matches

✉ Steven C. Schwering
schwering@wisc.edu

1 University of Wisconsin–Madison, Madison, WI, USA

2 University at Buffalo, SUNY, Amherst, NY, USA

3 University of California, Berkeley, CA, USA

Published online: 21 December 2023

previously encountered sequences, such as in the sentence-like list *FAT WAITER HELPS TALL SOLDIER NOT OLD BISHOP*. Across many studies, recognition rates and recall of words were higher for words from sentence-like lists relative to scrambled orders (e.g., Allen et al., 2018). Prior word co-occurrence is not necessary for the sentence superiority effect to emerge, because unattested but grammatical adjective–noun pairs of English words (e.g., *WATERY LION*) are remembered better than ungrammatical ones (e.g., *WINDOW ITCHY;* Perham et al., 2009), showing the importance of abstract relationships between words absent prior experience with specific word co-occurrences.

Buffer theories of VWM account for sentence superiority findings via domain-general chunking and redintegration repair processes (Allen et al., 2018; Baddeley et al., 2009; Jones & Farrell, 2018). Importantly, buffer accounts hypothesize two kinds of chunking: one based on strings that have previously been encountered, like *TALL SAILOR*; and the other chunking novel, unseen parts-of-speech sequences such as nouns and verbs that follow grammatical rules of the language (e.g., *WATERY LION*; Jones & Farrell, 2018). Chunks enhance memory because these two types of sentence chunks from linguistic LTM resemble and repair sentence-like lists in VWM. While most tests of buffer accounts have focused on part-of-speech manipulations, buffer accounts can be modified to accommodate phonological features and lexico-semantic properties (e.g., Martin et al., 1994). However, lexico-syntactic properties, which partially encode word meaning and word order in languages like English, are understudied in the memory literature.

Sentence-likeness in psycholinguistic research

Language comprehension and production research characterize linguistic LTM and sentence-likeness as more than word or part-of-speech sequences. Part-of-speech refers to the category of syntactic role a word fills in a sentence, such as a noun, verb, or adjective. Sequences of parts of speech, such as the tendency for adjectives to precede nouns in English, are one of many cues to sentence comprehension. However, many words have multiple parts of speech in English, such that a word's part-of-speech depends on the surrounding context. For example, context changes whether the ambiguous word *fires* is a verb (*the corporation fires…*) versus a noun (*the canyon fires…*). This context need not vary in part-of-speech co-occurrences (here, both *corporation* and *canyon* are nouns), and disambiguating context need not be immediately adjacent to the ambiguous word (MacDonald, 1993). Thus, part-of-speech is not a clear route to chunking, as ambiguities must be determined at the sentence level. Linguistic knowledge from LTM, such as a word's typical parts-of-speech (Levelt, 1993; MacDonald, 1993), sentence structures in which that word typically occurs (Stallings et al., 1998; Trueswell et al., 1993), thematic roles (e.g., agent, patient) that word fills in a sentence (MacDonald, et al., 1994; Tanenhaus & Trueswell, 1995), all influence disambiguation. Long-term knowledge of word meanings and word order is essential for efficient language use, because they jointly constrain possible interpretations in sentence comprehension and possible utterances in language production (MacDonald & Seidenberg, 2006; Schwering & MacDonald, 2020).

Using computational models, language processing researchers have also characterized language comprehension quite differently than part-of-speech sequence tracking. For example, connectionist neural networks employed in modeling language comprehension and production (e.g., Chang et al., 2006) track a bevy of linguistic regularities that span both word-specific (lexical) and sequential or combinatorial (syntactic) features, which we term *lexico-syntactic constraints*, following common terminology in language comprehension research. Critically, these algorithms use joint information about both words and their contexts to inform efficient processing (Dell & Chang, 2013; MacDonald & Christiansen, 2002; see Schwering & MacDonald, 2020, for review); lexico-syntactic constraints impact language processing as a function of both the properties of individual words as well as the context in which they are embedded. These mechanisms differ dramatically from how linguistic LTM is typically defined in chunk- and part-of-speech sequence-based models of VWM (e.g., Jones & Farrell, 2018; Page & Norris, 2009). Recent attempts at modeling VWM as an emergent process built on experience and language comprehension and production mechanisms have shown promise in this vein (Schwering, 2023). Furthermore, lexico-syntactic constraints are not merely phonological or semantic features of words, which have been extensively studied in the memory literature (e.g., Martin et al., 1994). Rather, lexico-syntactic constraints affect processing due to both qualities of individual words as well as their sentence context.

Potential role for language processing systems in VWM

The emphasis on interactions between words and syntax in language research suggests that the VWM literature's focus on exact word and part-of-speech sequences may underestimate how linguistic LTM bears on VWM. Emergent theories of VWM provide one way to apply the multitude of lexico-syntactic constraints that affect sentence processing to VWM. Emergent theories of VWM posit that language comprehension and production processes directly govern behavior in VWM tasks (Acheson & MacDonald, 2009; MacDonald, 2016; Majerus, 2013). In this view, manipulations that go beyond the traditional notions of sentence-likeness in VWM research (familiar word or part-of-speech sequences) should affect the sentence-likeness of a memory list and thereby influence VWM. Evidence for these accounts comes from recent work by Schwering and MacDonald (2023), who manipulated combinations of English nouns in memory lists according to the tendency for some nouns (e.g., *gold*) to serve as modifiers of other nouns (e.g., *watch*, where *gold watch* specifies a kind of *watch*). Despite the fact that all sequences were novel and part-of-speech did not vary, recall was higher for sequences that followed the statistical patterns of usage for individual nouns over those in the reverse order. Schwering and MacDonald interpreted this result as evidence that joint lexical and syntactic information influence VWM performance. According to language emergent theories of VWM, a multitude of lexico-syntactic constraints should influence sentence-likeness, even outside of minute noun-noun compounds.

Emergent accounts posit that mechanisms supporting language use also fully support VWM tasks. Emergent theories state that when the language comprehension system encounters a memory list, it creates a representation of the grammatical relationships between words in addition to their linear order (Fig. 1, Panel 1). The language production system takes this representation as a "message" (e.g., Levelt, 1993) that must be maintained and ultimately articulated (Panel 2). When repeating a studied list, such as during serial recall, the recalled list will be similar to the studied list to the extent that the studied list conformed to linguistic constraints, gleaned from past experience with language (Panels 2 and 3). Both language comprehension and production mechanisms are sensitive to sentence-likeness. Sentence-likeness is a function of many lexico-syntactic constraints and are not limited to previously attested word orders or parts of speech. On this view, comprehension and production processes are better able to encode and maintain study lists that conform to linguistic constraints than lists that do not conform to linguistic constraints.

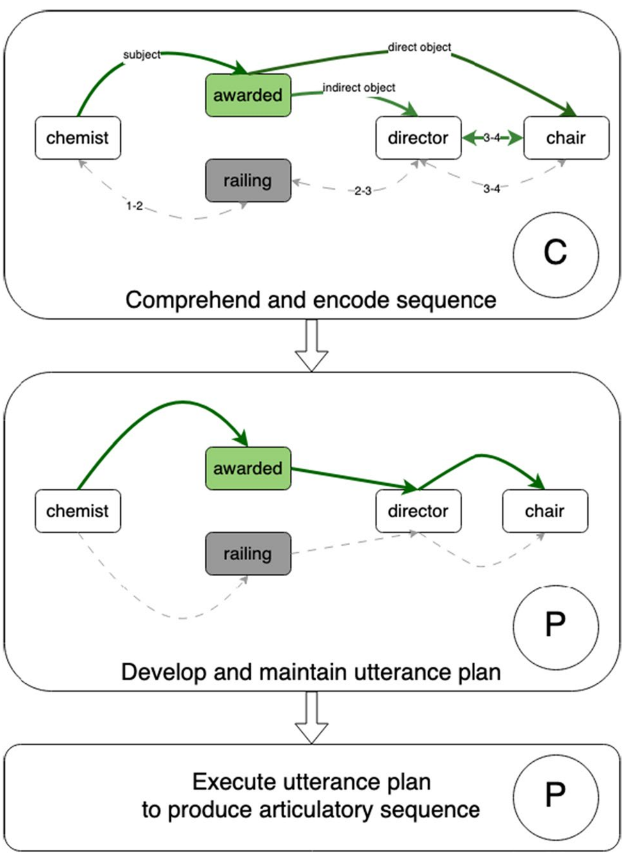

Fig. 1 Language comprehension (C) and production (P) processes in emergent accounts and how they support storage and maintenance of linguistic content. Thick arrows represent lexico-syntactic relationships that bind words and word order together in sentence-like sequences. Dashed arrows indicate sequences without these LTM-based relationships, leading to weaker representations of words and their orders

In two experiments, we consider how joint lexical and structural regularities, i.e., lexico-syntactic constraints, support VWM. We compare immediate serial recall of sentence-like lists versus non-sentence-like lists that differ by a single word, either a change of part-of-speech (Experiment 1) or changing the subcategory of an item (e.g., a transitive or intransitive verb) (Experiments 2A and B). This highly focused single-word manipulation allows greater precision in understanding interconnections between words. According to language emergent theories of VWM, recall should be higher for words supported by lexico-syntactic constraints that make the memory list more sentence-like.

Experiment 1: Dative verb versus noun

In this experiment, we constructed lists building on the structure of the English dative alternation (also known as ditransitive) sentences, such as *The teacher handed the student a pencil*. In this and other double object (DO) dative sentences, the indirect object (*the student*) occurs after the verb, and the direct object (*a pencil*) occurs last. Dative sentences are common in English and are interesting from a language emergent approach to VWM because dative verbs have distinguishable semantics from other English verbs (Levin, 1993), and their syntax diverges from standard English sentences, where direct objects immediately follow verbs. Given these unique properties, we manipulated the presence or absence of a dative verb to change the sentence-likeness of memory lists.

English speakers are sensitive to the statistical, semantic, and morphophonological lexical qualities that affect the goodness of specific dative verbs in sentences (Ambridge et al., 2014). In Experiment 1, our single-word manipulation created a change in part-of-speech in sentence-like lists resembling DO datives. In a position in which the sentence-like list contained a verb, the list-like list contained a noun. Importantly, neither version of the list contained common

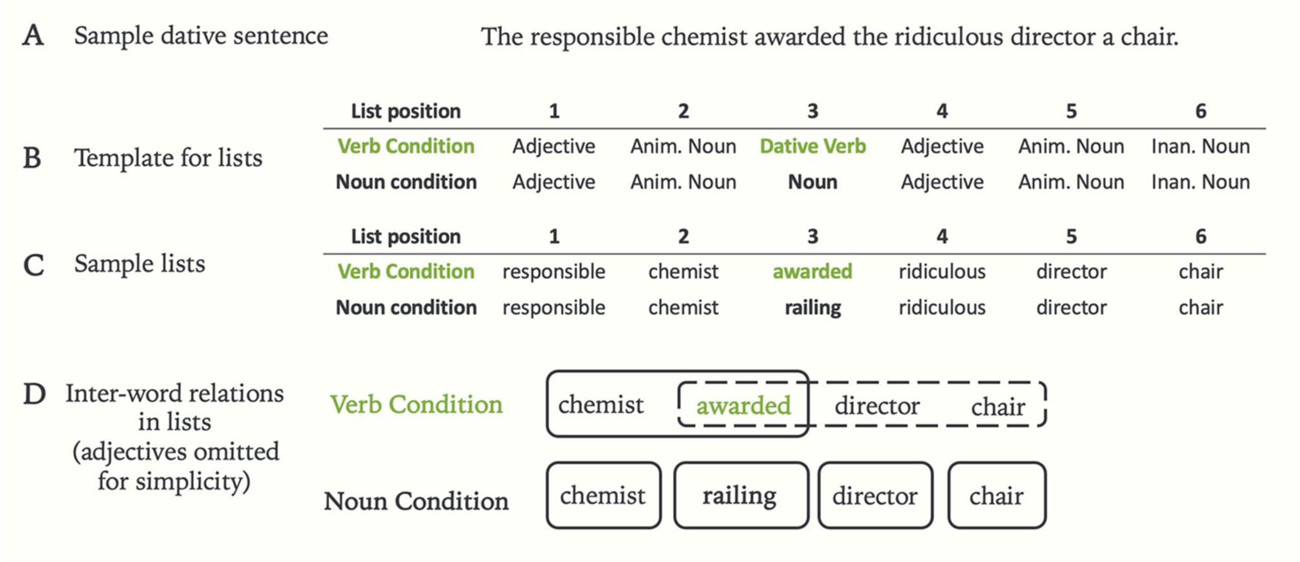

Fig. 2 Construction of sentence-like lists for Experiment 1. Participants were presented with six-word lists in which a single word was manipulated. **A** demonstrates a typical dative sentence, which is translated into a sentence-like list in **C** following the format in **B**. **D** demonstrates how participants are expected to form relations among constituents in a sentence-like dative list and a list without the dative verb. Anim. = Animate, Inan. = Inanimate

word co-occurrences or meaningful word relations. We began with a part-of-speech manipulation to test our single-word method, because previous studies have found part-of-speech sequence effects on serial recall (Jones & Farrell, 2018; Perham et al., 2009). Obtaining a similar result in Experiment 1 with the single-word change across lists would replicate and extend previous findings. Thus, changing a single word should influence recall throughout the list, including at nonadjacent positions to the manipulated word.

Figure 2A–C shows the basic steps for transforming DO dative sentences into lists for the memory task. Figure 2D illustrates hypothesized relationships between list words in the verb (sentence-like) and noun (list-like) conditions. In the verb condition, the animate first noun can be interpreted as the subject of the verb, creating a grammatical relationship that we predict will promote recall of the noun. The verb also promotes interpretation of the subsequent animate noun and inanimate noun as part of a dative relation because the verb occurs in dative sentences in English, and the order of the animate and inanimate nouns is consistent with typical direct and indirect objects of dative sentences. By contrast, having a noun in Position 3 does not integrate the nouns into a sentence-like representation. Even though the two lists differ only in Position 3, we predict poorer recall throughout the lists in the noun condition compared with the verb condition. Moreover, because the lists vary in their similarity to sentences via long-term knowledge of the conjunction of both positional and lexical information, the predicted result of a language emergent approach would support integrated lexical and positional information guiding immediate serial recall performance.

Method

Participants

We recruited 68 undergraduate students (M_{age} = 19.2, SD_{age} = 2.08, 41 women) from the University of Wisconsin–Madison course credit participant pool who self-reported as having English as spoken in the home before the age of 5. This experiment and all following experiments were approved by the University of Wisconsin–Madison IRB and all participants gave their informed consent prior to participation.

Materials

We constructed 65 six-item lists with two variants for an immediate serial recall task. In all experiments, lists varied in the extent to which they were like English double object dative sentences. Dative sentences minimally contain four key words: two animate nouns, one inanimate noun, and a verb. To ensure that participants could not trivially remember the entire list while also retaining grammatical coherence, an adjective was placed before both animate nouns to create six-word lists, as shown in Fig. 1. Sentence-like lists (Table 1) had the structure Adjective–Animate

Noun–Verb–Adjective–Animate Noun–Inanimate Noun, while list-like lists contained a control noun at the third position, replacing the verb, but were otherwise identical. No word was repeated in any other trial for a given participant. All lists for this experiment and subsequent experiments are available alongside analysis scripts on OSF (https://osf.io/6gtnk/?view_only=2f41a6dffdbd46a882fcd41f717604a7).

The selection methods for each type of word in the dative lists are described next. Because our aim is to test the effect of participants' long-term linguistic knowledge on immediate serial recall, we constructed memory lists with the help of corpus analyses, which provide an estimate of participants' linguistic knowledge. The number of lists was limited by the number of suitable dative verbs in English.

Target verbs Using the past tense forms of 200 dative-accepting verbs from the DAIS dataset (Hawkins et al., 2020, based on Levin, 1993). We extracted and parsed all sentences containing these forms in the 1990–2015 portion of COCA (Davies, 2008) comprising 560+ million words. COCA is a balanced multi-genre corpus and has broad coverage for our target verbs due to its size. We then computed the frequency of different realizations of English dative sentence types for each verb.

Using the Python programming language package spaCy, we heuristically identified sentences containing a dative relation, defined as the presence of exclusively both *dobj* and *dative* dependencies (de Marneffe et al., 2021) for our verbs of interest. We then heuristically classified verbs for which the *dative* argument occurred to the left of the double object (*dobj*) tag within the sentence. All verbs that occurred at least once were included, resulting in a total of 65 critical verbs (Column 1 of Table 1).

Target control nouns We selected 65 nouns as controls that were as similar as possible to the 65 DO verbs, to create a nondative variant of each list. To select nouns, we first computed the standard deviations of concreteness, contextual diversity, and word frequency for the set of verbs using the SUBTLEXUS norms from Brysbaert and New (2009). Control nouns fell within one standard deviation of the verb's value. Then, exact length-matched words that occurred at least 95% of the time as a noun (Brysbaert et al., 2012). We include the control nouns in Column 2 of Table 1. Means and standard deviations of control factors for paired verbs and control nouns are listed in Table 2.

Animate and inanimate nouns We selected additional words for the lists for the animate and inanimate noun categories that occurred at least 95% of the time as nouns (Brysbaert et al., 2012). We determined the animacy of nouns using the VanArsdall and Blunt (2022) norms of mental and physical animacy ratings. We added the mental and physical animacy ratings to calculate a total animacy rating. The final animate noun set was determined by selecting the 130 (65 lists × 2) nouns with the highest total animacy ratings; the final inanimate noun list was determined by selecting the 65 least animate nouns.

Adjectives We selected 130 (65 lists × 2) words that uniquely were observed as adjectives (Brysbaert et al., 2012) and were the most frequent adjectives in the Brysbaert norms.

Design

We employed a two-level factorial design manipulating sentence-likeness (sentence-like; list-like) and counterbalance list (List Series A; List Series B). Sentence-likeness defined whether the list formed a dative-like sentence or did not, corresponding to whether the participant saw the list with the dative verb or the control noun in Position 3. Counterbalance list defined which of two predefined contexts participants saw.

Memory lists were generated randomly, sampling exhaustively from the pool of paired target verbs and control nouns, animate nouns, inanimate nouns, and adjectives. Two counterbalance lists (sets of memory lists) were generated to help control unintended patterns in randomization that may impact recall.

Sentence-likeness was manipulated within-subjects. Each participant saw about half of the lists (either 32 or 33 of the 65 lists) in the sentence-like condition and the rest of the lists in the list-like condition. Order of sentence-likeness was randomized. Each verb/critical noun pair was presented once, with the order of the lists in the experiment and their sentence-likeness randomly assigned for each participant.

Counterbalance list-type was manipulated between-subjects. Participants were randomly assigned either to List Series A or List Series B, which were randomly generated from the same pool of words described above prior to the start of the experiment. All participants assigned to a counterbalance list condition saw the same memory list contexts as the other participants assigned to that counterbalance list condition, and participants in one condition did not see lists from the other condition. This means that, for any given critical pair of target dative verbs and control nouns, participants saw one of two lists. For example, all participants in the List Series A condition saw the list *proper–poet*–[WORD 3]–*brilliant–surgeon–tent*. In contrast, participants in the list series B condition saw the list *serious–lawyer*–[WORD 3]–*regular–warrior–pebble*. In the context of the lists above, participants saw either the critical verb *reimbursed* or the control noun *ionosphere* in Position 3, depending on their sentence-likeness condition for that list. The order of memory lists within the counterbalance list-type condition and the sentence-likeness of lists were random for each participant.

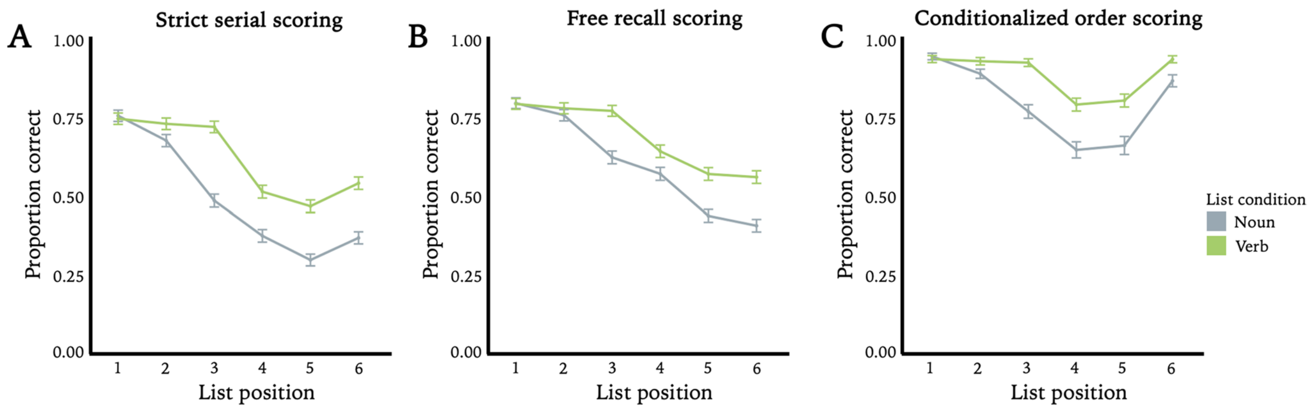

Fig. 3 Recall accuracy in Experiment 1. **A** represents mean accuracy across conditions and list positions using strict serial scoring. **B** represents mean accuracy across conditions and list positions using free recall scoring. **C** represents mean accuracy across conditions and list positions using conditionalized order scoring. Bars represent .95 confidence interval, accounting for by-participant variance estimated from the data

Procedure

All data collection was done remotely. Participants were instructed to complete the study in a quiet location away from distractions.

Demographic survey Following completion of informed consent, participants completed a brief demographics self-report survey run on Qualtrics. Participants were asked to report their gender, age, race/ethnicity, level of education, whether they were a native speaker of English, and any other languages they spoke.

Memory task Upon completion of the survey, participants were redirected to a web-browser-based serial recall experiment. Participants were told that the task was a serial short-term memory task. Participants saw 65 six-word lists, each of which was either sentence-like or list-like. Participants began each trial viewing a fixation cross on screen for 1,250 milliseconds. Participants then viewed each of the 6 words in a list for 750 milliseconds with a 250 millisecond pause between each word. After presentation of the words in each study list, participants were presented with a form on screen corresponding to a specific list position into which they could type the word that they studied at that position.

Participants were instructed to fill in words into the slots in the order that they had appeared on the screen and told to leave a slot empty if they could not remember the word. After participants submitted a response for a position, they were automatically advanced to the next position. Participants did not have the option to alter a response or go back to earlier positions after submitting a response. To reduce participant fatigue, four short breaks were distributed through the experiment. Participants could proceed at their own pace. Following completion of the experiment, participants completed an exit survey asking about their strategies and thoughts about the experiment. Finally, participants were debriefed as to the purpose of the experiment.

Results

Part-of-speech on overall recall Participant responses were scored using strict serial scoring. Responses were marked correct if participants recalled a word in the position in which it was presented. Spelling errors were counted as incorrect. Participants' overall accuracy across list positions using the strict scoring procedure is shown in Fig. 3A.

Results were analyzed using a mixed effects logistic regression with maximal random effects. Fixed effects included sentence-likeness (−0.5 = list-like, 0.5 = sentence-like), counterbalance list (−0.5 = A, 0.5 = B), the interaction between sentence-likeness and counterbalance list, the linear effect of list position, and the quadratic effect of list position. Random effects included a by-participant random intercept, a by-participant random slope for sentence-likeness, a by-verb random intercept, a by-verb random slope for sentence-likeness, and a by-verb random slope for counterbalance list. This model and all subsequent models reported in this paper were initially fit with maximal random effects by participants and by manipulated pair governing list condition following guidelines by Barr et al. (2013). Steps at reducing the model to achieve convergence were followed according to Brauer and Curtin (2018). For this model and all subsequent reported models, model fitting employed the *bobyqa* optimizer. For all models, Type III sum of squares test is reported, determining whether the fixed effect of interest accounts for significant additional variance when added to the model after all other fixed effects. All analyses were conducted in R (Version 3.6.3). Models were fit using the package *lme4* (Bates et al., 2015), and model comparisons were performed using the *Anova()* function from the *car* (Fox & Weisberg, 2019) package. Raw data and analysis scripts for this experiment and all subsequent experiments are available on an Open Science Framework (Foster & Deardoff, 2017) repository (https://osf.io/6gtnk/?view_only=2f41a6dffdbd46a882fcd41f717604a7).

The fit model indicated that participants were more likely to recall a word correctly when presented in the context of a sentence-like list compared with a list-like list, β = 0.65, $\chi^2(1) = 124.18$, $p < .001$. Furthermore, the model did not find a significant effect of counterbalance list, $\chi^2(1) = 2.11$, $p = .15$, indicating that there was no significant difference in the recall of different counterbalance lists. Further, there was no interaction between sentence-likeness and counterbalance list, $\chi^2(1) = 0.04$, $p = .83$. Finally, there was a significant effect of the linear trend of position on recall, β = −5.65, $\chi^2(1) = 496.11$, $p < .001$, as well as a significant effect of the quadratic trend of position on recall, β = 2.84, $\chi^2(1) = 187.23$, $p < .001$.

Part-of-speech on item and order memory Serial recall tasks and the strict scoring procedures that are generally used to measure performance naturally conflate memory for the presence of previously encountered words (i.e., item memory) and memory for the order in which those words occurred (i.e., order memory; Saint-Aubin & Poirier, 1999). As such, researchers have typically conducted separate item memory analyses and order memory analyses to statistically control for these different constructs and thus interrogate behavior further. While our hypotheses primarily concern general serial recall, we also conducted item analyses and order analyses to examine whether lexico-syntactic constraints differentially affect item and order memory.

Item memory was assessed through free recall scoring, where a response is marked as correct if it matches any word in the studied list. We fit free recall scoring in the same model framework described above, using the same random effects and optimizer. According to the model, participants were more likely to recall words irrespective of position in the sentence-like list compared with the list-like list, β = 0.49, $\chi^2(1) = 1.42$, $p < .001$, $\chi^2(1) = 2.11$, $p = .15$. Further, there was no effect of an interaction between sentence-likeness and counterbalance condition, $\chi^2(1) = 0.04$, $p = .83$. There was a significant effect of the linear trend of position, β = −5.65, $\chi^2(1) = 496.11$, $p < .001$, as well as a significant effect of the quadratic term for position, β = 2.84, $\chi^2(1) = 187.23$, $p < .001$. We visualize these results in Fig. 3B.

Researchers have previously attempted to measure order memory independent of item memory through specific statistical tests, such as tests comparing the proportion of words recalled in the correct position over the number of words recalled in any position (Kowialiewski et al., 2021a). Here, we analyze order memory using a similar approach, but adapted to a modern mixed effects framework (Jaeger, 2008). Specifically, in our *conditionalized order* analysis, we aim to predict whether each word in each study list was recalled in the correct position (a score of 1) or whether a word was recalled in any other position (a score of 0). This analysis gives us the additional ability to consider list-level and participant-level variability that historically have been ignored in analyses of order memory. Rather than computing the proportion of words recalled in the correct position in each list, we can instead look at order memory at the item (list) level. We can then treat our sentence-level manipulations as fixed effects (and random slopes) in our analyses.

Due to the nature of the analysis, we first excluded from our dataset all those observations where a studied word was not recalled in any position in the recalled list, leaving only those observations where participants had accurate item memory. This also excludes lists for which participants recalled none of the words. These exclusions necessarily reduce the statistical power of this analysis relative to the strict scoring analysis and item memory analysis.

The conditionalized order analysis retained all other fixed effects, random effects, and coding schemes from analyses above. As visualized in Fig. 3C, the model indicated that participants were more likely to recall a word in the correct position in sentence-like lists compared with a list-like lists, β = 0.87. Model comparison showed that the inclusion of a sentence-likeness factor significantly improved model likelihood, $\chi^2(1) = 63.38$, $p < .001$. The counterbalance list factor did not significantly improve model fit to the data, $\chi^2(1) = 0.85$, $p = .36$, nor did model comparison suggest an improvement in model fit by adding an interaction between sentence-likeness and list counterbalance, $\chi^2(1) = 0.23$, $p = .63$. There was also a significant trend for words in later positions to be recalled less often in position, as indicated by a significant effect of the linear trend of position, β = −12.28, $\chi^2(1) = 577.40$, $p < .001$. Finally, there was also a significant effect of the quadratic trend of position, β = 8.79, $\chi^2(1) = 447.62$, $p < .001$.

Discussion

The results of Experiment 1 provide evidence that aspects of syntactic structure in a list of words has substantial impacts on serial recall memory. Specifically, the change of only a single word in the list—the inclusion of a dative verb versus a noun at Position 3—had a robust effect on recall across the list, with higher rates of recall of words in the list when the presence of a verb provided a sentence-like context.

The results of this experiment are broadly consistent with emergent accounts of VWM (Acheson & MacDonald, 2009; MacDonald, 2016; Majerus, 2013; Schwering & MacDonald, 2020, 2023), but they are also consistent with certain buffer models of VWM. In their model of the sentence superiority effect, Jones and Farrell (2018) used part-of-speech tags to index syntactic structure, increasing the probability of list recall in proportion to the probability of the part-of-speech sequence of that memory list. As a result, their account could also explain these results. Our own view is that language users certainly learn the part-of-speech of words they comprehend and produce, but that lexico-syntactic learning does not stop there. Rather, within a single grammatical category (e.g., verb) there may be a high degree of variability in what types of contexts the tokens (words) of that category appear.

In Experiments 2A–B, we explore the ways in which subtle lexico-syntactic constraints may affect VWM. In both experiments, we again change one word across the sentence-like versus list-like conditions, but these changes do not affect part-of-speech. If sentence-likeness in novel word combinations is limited to part-of-speech memory, then there should be no differences in recall between the sentence-like and list-like materials in Experiments 2A-B. However, if more subtle lexico-syntactic manipulations that affect comprehension can also affect performance on VWM tasks, then our Experiment 2A-B manipulations should yield differences in memory performance.

Experiment 2: Manipulations within part-of-speech

Verb manipulation, Experiment 2A Verb bias refers to the statistical properties of verbs to appear in one kind of sentence type versus another (Garnsey et al., 1997; Hare et al., 2003; Trueswell et al., 1993). In Experiment 2A, we replaced Experiment 1's noun versus verb manipulation with a verb bias manipulation, in which all lists have a verb in Position 3, but where the history of a verb's usage in natural language creates statistical properties that modulate the sentence-likeness of the list.

More specifically, we compared lists with dative verbs like *handed* to lists with intransitive verbs like *slept,* which don't occur with direct object noun phrases. For example, *Taylor slept the doctor* is not a grammatical sentence of English because *sleep* cannot take a direct object. This verb type manipulation allows a test lexico-syntactic influences on sentence superiority effects, as part-of-speech is held constant. Because the sentence-likeness of a sequence depends on interactions between items and their orders within a list, the verb manipulation is predicted to influence how participants structure the lists they study.

Rationale for positions analyzed As in Experiment 1, the sentence-like containing dative verbs in Experiments 2A have components of a grammatical (though nonsensical) sentence (e.g., *handsome snail gave tired sheriff notebook*). Replacing dative verbs with intransitive verbs yields lists (e.g., *handsome snail slept tired sheriff notebook*) affects early and late portions of the list differently. The first part of the list is grammatical independent of the dative versus intransitive verb manipulation (*handsome snail gave/slept*). However, the sentence likeness of remainder of the list does vary with the verb (*tired sheriff notebook* is grammatical after *gave* but not after *slept*). If this verb bias manipulation affects VWM performance, then memory for the sentence-like lists should be superior to the list-like lists specifically in the post-verbal positions. Due to these differences in the intransitive and dative verbs' effects, we focus analyses of Experiment 2A on the final three words (the post-verbal region) in the stimulus lists, as these represent the point of divergence between ditransitive and intransitive sentences.

Noun manipulations in Experiment 2B In natural language, subjects in dative sentences tend to be animate, because animate entities tend to engage in transfers of possession events that are described by dative sentences. We retained the sentence-like materials from Experiment 1 and 2A for the sentence-like condition (e.g., *handsome snail gave tired sheriff notebook*) and created the list-like condition by replacing the animate noun in Position 2 with an inanimate noun, e.g., *handsome house gave tired sheriff notebook*. Because the resulting sentence structure was still a (nonsensical) double object dative, the analyses of Experiment 2B data take place over all words in the memory lists.

Previous research has considered how noun properties affect sentence-likeness. Perham et al. (2009) manipulated the order of adjective noun pairs in memory lists, and Schweppe et al. (2021) manipulated morphosyntactic cues to sentence structure, both finding that those patterns consistent with linguistic LTM support for VWM. Unlike our manipulation, these studies relied upon manipulations of order (Perham et al., 2009) or local coherence between an adjective and a modified noun (Schweppe et al., 2021) to support VWM. Experiment 2B tests whether manipulating a single noun feature could also modulate coherence in sentence-like lists, while keeping all other list properties, like order, constant. If the lexico-syntactic constraint of animacy, which affects comprehension also affects VWM task performance, then sentence-like lists with animate subjects should be recalled better than sentence-like lists with inanimate subjects.

Predictions Both the within-verb manipulation in Experiment 2A and the within-noun manipulation in Experiment 2B are much more subtle than the part-of-speech manipulation in Experiment 1. Given the nature of these manipulations, all lists in Experiments 2A and 2B retain some sentence-like coherence, and we accordingly anticipate the overall strength of the manipulations to diminish. We decided to explore these subtle manipulations because they are a natural consequence of considering psycholinguistic claims for the importance of lexico-syntactic information on comprehension. Moreover, these within-word-category manipulations may allow us to

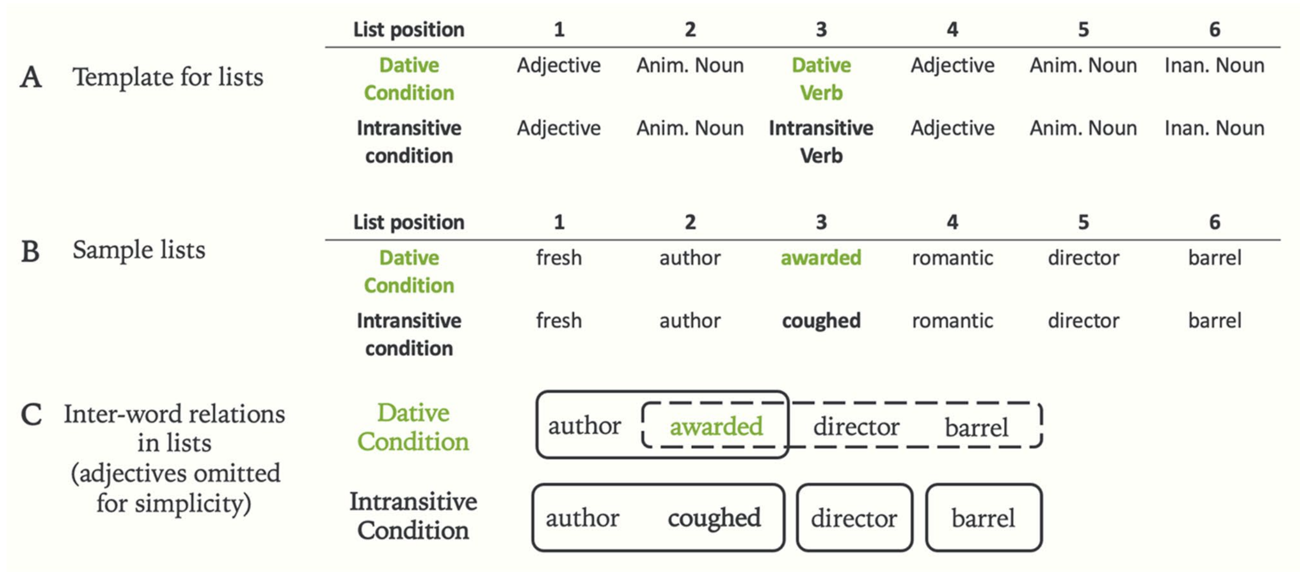

Fig. 4 Construction of sentence-like lists for Experiment 2A. Anim. = Animate, Inan. = Inanimate

disentangle claims in the VWM literature, specifically whether participants are limited to tracking part-of-speech information when remembering sentence-like lists (Jones & Farrell, 2018), or whether these fine-grained lexico-syntactic constraints moderate sentence processing and sentence memory.

Method

Participants

Experiment 2A recruited 103 undergraduate students (M_{age} = 19, SD_{age} = 0.87, 69 women) from the University of Wisconsin-Madison course credit participant pool. All participants indicated native experience with English (spoken in the home before the age of five). In analyses reported below, three participants were removed from analyses for correctly recalling less than .10 of words in the correct position across all trials (no participant scored below this threshold in Experiment 1).

Experiment 2B was pre-registered on the Open Science Foundation (Foster & Deardoff, 2017) (https://osf.io/3v7km/?view_only=b951bab3aa814a03919acf7d7dbcda70). Participants (M_{age} = 25.08, SD_{age} = 6.15, 83 women) were sampled from the online web service Prolific. Participants were required to be native speakers of English (spoken English in the home before the age of 5) and reside in the United States. Participants were compensated $5 ($10/hour) for their participation. 100 participants were recruited; 97 completed the task. Two participants were removed from analyses for scoring below 0.10 total accuracy in strict scoring on the serial recall task, as in Experiment 2A.

Materials

Verb manipulation, Experiment 2A Lists were composed of six words with the same constraints as in Experiment 1, except that in Position 3, we compared dative verbs (e.g., *awarded*) to intransitive verbs (*coughed*). Otherwise, the selection of adjectives, animate nouns, inanimate nouns, and counterbalancing were all conducted in the same manner as Experiment 1. An example list and list construction logic are demonstrated in Fig. 4. We note that this manipulation supports participants' spontaneous synthesis of the words into either a longer sentence (|*author awarded director barrel*|) or a shorter sentence that is followed by two apparently unrelated final words (|author coughed| |director| |barrel|).

Selection of intransitive verbs We first extracted intransitive verbs from four prior sentence comprehension studies (Lee & Thompson, 2004; Liu, 2008; Shetreet et al., 2010; Thompson et al., 2007), which provided us with an initial set of 112 candidate verbs. We describe the selection methods for Experiment 2A in Appendix A4. Limits on the number of clearly intransitive verbs in English resulted in a smaller number of sampled items per participant (52) than in Experiment 1 (65).

Pairing intransitive and dative verbs As described in Appendix A5, we sought to align the candidate intransitive verbs that we identified with the dative verbs in Experiment 1 such that transitive-intransitive verb pairs were as similar as possible along the four dimensions that we used to control the noun-verb pairs in the previous experiment: concreteness,

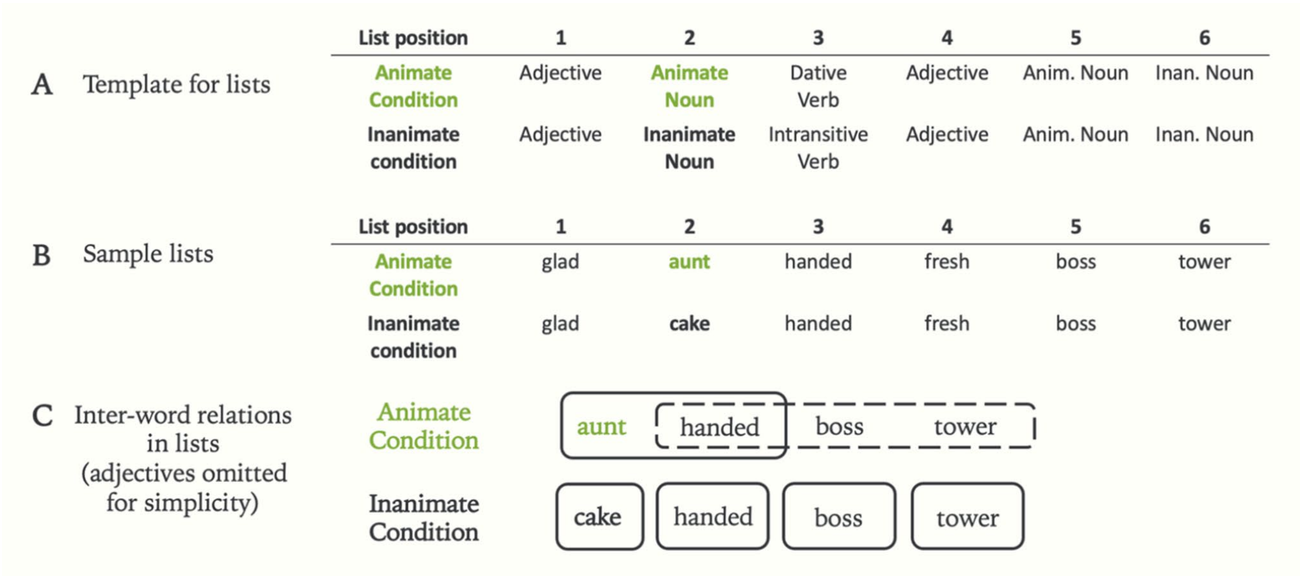

Fig. 5 Construction of sentence-like lists in Experiment 2B

contextual diversity, word length, and word frequency. The final list of matched intransitive verbs is in column 3 of Table 1. Means and standard deviations of control factors for paired transitive verbs and intransitive verbs are listed in Table 2.

Noun manipulation, Experiment 2B Lists followed the same general construction constraints as in Experiment 1 and 2A. However, the manipulated word was the word in the second position, the subject noun. Adjectives, ditransitive verbs, indirect object animate nouns, and direct object inanimate nouns were the same as employed in Experiment 1 and 2A. The list design is illustrated in Fig. 5.

Animate nouns and control inanimate nouns Candidate animate and inanimate nouns were pulled from the VanArsdall norms, starting with the 500 most animate nouns and 300 most inanimate nouns. As described in Appendix A6, the animate and inanimate nouns were then matched to each other based upon their frequency, contextual diversity, length, and semantic similarity, yielding 65 controlled pairs of animate and inanimate nouns. These pairs are in Table 3, with information about control factors for the pairs listed in Table 2.

Procedure

The experimental procedure of Experiments 2A and 2B were identical to that of Experiment 1, as the only difference between the two experiments is the list materials. Slight differences in instruction were provided to Experiment 2B to accommodate participants from Prolific.

Results

Verb manipulation, Experiment 2A We focus our analysis here on the final three words in the memory lists as these words are either within a sentence (ditransitive) or outside one (intransitive) making them maximally different in our manipulation. Analyses across all positions can be found in Appendix A7; the results show the same pattern overall.

As described in the introduction to Experiments 2A-B, the verb manipulation in Experiment 2a is predicted to affect the last three words of the list. A mixed effects logistic regression model was fit on strict serial scoring for words studied in these positions, predicting recall from verb type, counterbalance list, their interaction, the linear trend of position, and the quadratic trend of position. Participants had higher recall when presented with a dative verb relative to an intransitive verb in Position 3, as indicated by a significant effect of verb type, $\beta = 0.19$, $X^2(1) = 12.55$, $p < .001$, see Fig. 6A. Further, there was no effect of counterbalance list condition on recall, $X^2(1) = 0.82$, $p = .36$, and there was no interaction between verb type and counterbalance list, $X^2(1) = 0.11$, $p = .74$. There was a significant effect of the linear trend of list position on recall, $\beta = -14.45$, $X^2(1) = 41.87$, $p < .001$. There was also a significant effect of the quadratic trend of list position on recall, $\beta = 8.80$, $X^2(1) = 43.27$, $p < .001$.

Verb biases on item memory Item memory was predicted from verb bias condition, counterbalance list condition, the interaction between these two effects, the linear trend of list position, and the quadratic trend of list position. The fit model indicated participants were more likely to recall

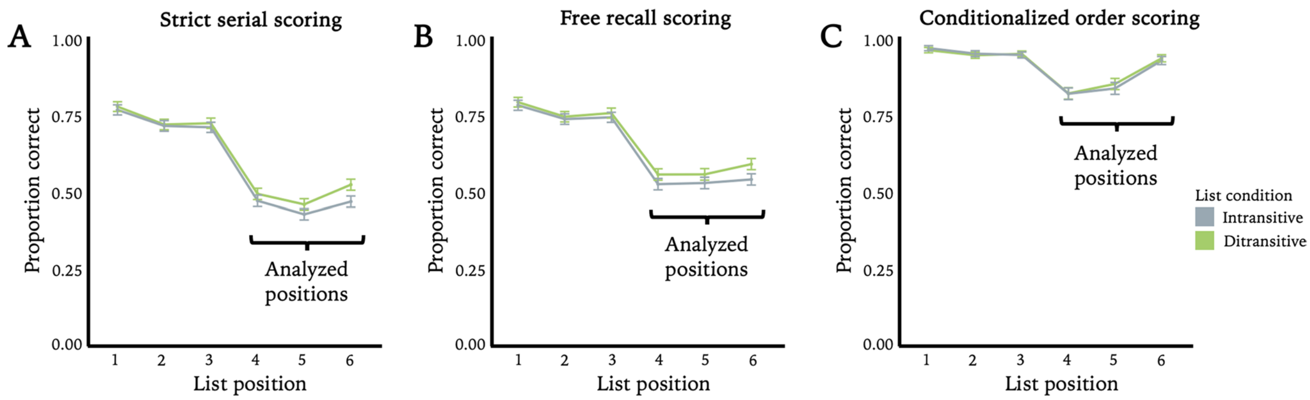

Fig. 6 Recall accuracy in Experiment 2A. **A** represents mean accuracy across conditions and list positions using strict serial scoring. **B** represents mean accuracy across conditions and list positions using free recall scoring. **C** represents mean accuracy across conditions and list positions using conditionalized order scoring. Bars represent .95 confidence interval, accounting for by-participant variance estimated from the data. NB: Analyses were conducted only over recall in Positions 4, 5, and 6, though all list positions are included in the figures

words presented in the context of a ditransitive verb in the last three positions irrespective of position than in the context of an intransitive verb, β = 0.19, $X^2(1) = 17.79$, $p <$ .001, Fig. 6B. There was no significant effect of counterbalance list condition, $X^2(1) = 1.10$, p = .30, nor was there an interaction between verb type and counterbalance list condition, $X^2(1)$ =0.57, p = .45. Finally, there was neither an effect of the linear trend of list position on item memory, $X^2(1)$ – 1.44, p = .23, nor was there an effect of the quadratic trend of list position on item memory, $X^2(1) = 1.87$, p = .45.

Verb biases on order memory Order memory was predicted from verb type, counterbalance list, the interaction between these two effects, as well as the linear and quadratic trend of list position. The fit model found no effect of verb type, $X^2(1) = 0.49$, p = .49, meaning that participants were not more or less likely to recall words from the last three positions (4, 5, or 6) in the correct position given the type of verb presented in Position 3, see Fig. 6C. Further, the model found no effect of counterbalance list, $X^2(1) = 0.07$, p = .79, nor was there any interaction between counterbalance list and verb type, $X^2(1) = 1.55$, p = .21. There was a significant effect of the linear trend of list position in predicting order memory, β = -40.35, $X^2(1) = 85.42$, $p <$ .001, and there was a significant effect of the quadratic trend of list position to predict order memory, β = 23.92, $X2(1) = 85.17$, $p <$ .001.

Animacy effects, Experiment 2B Strict scoring was fit using a binomial mixed effects logistic regression predicting correct recall of words in position from animacy of the subject noun, counterbalance list condition, the interaction between the two variables, the linear trend of position, and the quadratic trend of position. Participants were more likely to recall words in the correct position if the sentence-like list contained an animate noun in the subject position (Position 2) than if that same list contained an inanimate noun, β = 0.10, $X^2(1) = 4.71$, p = .03. There was no main effect of list counterbalance, $X^2(1) = 0.08$, p = .77, nor was there an interaction between animacy of the subject and list counterbalance, $X^2(1) = 0.03$, p = .85. There was a significant effect of the linear trend of list position on memory, β = -3.90, $X2(1)$ = 238.80, $p <$ .001, and there was a significant effect of the quadratic trend of list position on memory, β = 1.39, $X2(1)$ = 49.10, $p <$ .001. A summary of these results is visualized in Fig. 7A.

Animacy on item memory Item memory was predicted for all studied words from animacy of the subject noun, counterbalance list condition, the interaction between these two effects, the linear trend of list position, and the quadratic trend of list position. Participants were not more likely to have stronger item memory depending on the animacy of the subject noun, $X^2(1) = 3.61$, p = .06, see Fig. 7B. Counterbalance list condition was not a significant predictor of item memory, $X^2(1) = 0.21$, p = .64, nor was there an interaction between animacy of the subject noun and counterbalance list condition, $X^2(1) = 0.06$, p = .80. There was a significant linear trend of list position on item memory, β = -2.00, $X^2(1)$ = 0.93, $p <$ .001, though there was not a significant effect of the quadratic trend of list position on item memory, $X^2(1)$ = 0.93, p = .34.

Animacy on order memory Order memory for all words in the list was predicted from animacy of the subject noun, counterbalance list, the interaction between these two effects, as well as the linear and quadratic trend of list position. Participants were not more likely to recall lists in the correct order with an animate noun in the subject position than if there was an inanimate noun in the subject position, $X^2(1)$ = 1.20, p = .27. Furthermore, there was no main effect of list

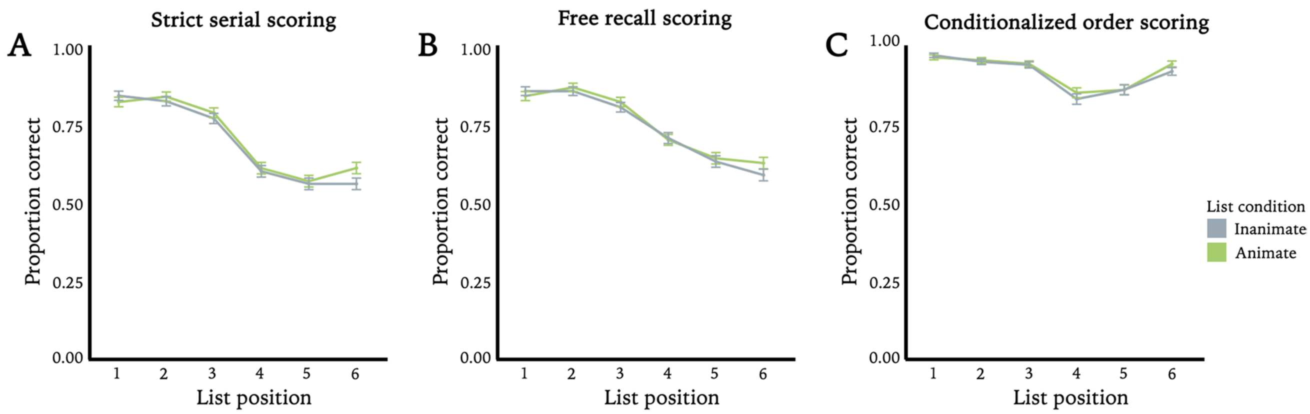

Fig. 7 Recall accuracy in Experiment 2B. **A** represents mean accuracy across conditions and list positions using strict serial scoring. **B** represents mean accuracy across conditions and list positions using free recall scoring. **C** represents mean accuracy across conditions and list positions using conditionalized order scoring. Bars represent .95 confidence interval, accounting for by-participant variance estimated from the data

counterbalance, $X^2(1) = 0.19$, $p = .66$, nor was there an interaction between animacy of the subject noun and list counterbalance, $X^2(1) = 0.31$, $p = .58$. There was a significant effect of position on order memory, β = -10.81, $X^2(1) = 328.88$, $p < .001$. There was also a significant effect of the quadratic trend of position on order memory, β = 7.25, $X^2(1) = 239.56$, $p < .001$. A summary of these results is visualized in Fig. 7C.

Discussion

The results of Experiments 2A and 2B offer partial support for the hypothesis that participants are sensitive to lexico-syntactic constraints in remembering sentence-like lists. There was a reliable effect of the verb type manipulation in the strict serial scoring measure (Fig. 6A) and item memory (Fig. 6B), but the effect of verb type in the conditionalized order analysis (Fig. 6C) was not reliable. There was also a reliable effect of subject noun animacy on strict serial scoring (Fig. 7A), but no effect on item memory (Fig. 7B) or order memory (Fig. 7C). The strict scoring results, which capture the conjunction of item and order memory, are consistent with emergent accounts (Schwering & MacDonald, 2020). However, there were mixed effects of lexico-syntactic constraints on analyses that aim to isolate item and order memory, suggesting that lexico-syntactic constraints may only partially support memory as tapped by these analyses.

The effects in Experiments 2A and 2B were much smaller than in Experiment 1. This pattern was not unexpected, given the subtle contrast between two verbs in Experiment 2A and the more obvious contrast in part-of-speech in Experiment 1, and given the subtlety of many verb bias manipulations in the language comprehension literature (Garnsey et al., 1997; Hare et al., 2003; Trueswell et al., 1993). Similarly, for Experiment 2B, while grammatical subjects are more likely to be animate, inanimate noun subjects do occur reasonably frequently in fantasy and metaphorical senses with dative verbs, as in *The desk gives the student a place to work*. Comprehenders readily interpret these non-literal scenarios (Nieuwland & Van Berkum, 2006). Participants in the present studies may therefore have been able to make at least some of the lists in the inanimate condition reasonably coherent, reducing the difference between the lists with animate versus inanimate nouns. Nonetheless, the lack of reliable effects of our manipulation, particularly in the order analysis, does limit the reach of emergent accounts of VWM (see Majerus, 2013, for discussion of emergent item and order memory).

General discussion

In this work, we adapted both methods (single-word manipulations across conditions) and theoretical claims (lexico-syntactic constraints existing in LTM) from the psycholinguistic literature to investigate temporary maintenance and ordering of words in a verbal working memory task. Across these experiments, we found some evidence that lexico-syntactic constraints affect immediate serial recall, consistent with emergent accounts of VWM (Majerus, 2013; MacDonald, 2016; Schwering & MacDonald, 2020). First, strict scoring in immediate serial recall showed that participants were more likely to recall sentence-like lists than list-like strings, with reliable effects of the manipulation of a single word across conditions: verb versus noun in Experiment 1, dative versus intransitive verb in Experiment 2B, and animate versus inanimate noun in Experiment 2B. However, these effects were small in Experiments 2A and 2B, compared with a larger effect in Experiment 1. Our conditionalized order measure, assessing order correctness given that a word was recalled, provided more limited support for these claims, with only Experiment 1 showing reliably higher rates of order recall in the more sentence-like condition.

These results illuminate some of the complexities underlying the relationship between linguistic LTM and performance on VWM tasks. Buffer accounts of VWM have considered LTM and VWM as separate, but with mechanisms like redintegration and chunking affording sources of interaction between these putatively separate systems. A satisfying account within this framework requires understanding when these interactions do versus do not occur. Meanwhile, emergent accounts posit rich integration between LTM and performance on VWM tasks, and again, a satisfying account here requires understanding when LTM does and does not have significant effects on VWM. All studies of sentence superiority effects can be seen as working toward illuminating these questions and ultimately clarifying debates concerning buffer and emergent accounts.

Our own sentence superiority studies focused on the role of lexico-syntactic constraints on memory. Buffer accounts emphasize the role word and part-of-speech patterns play in VWM (Jones & Farrell, 2018). Other accounts point to alternative mechanisms. Semantic coherence (Kowialiewski et al., 2022) or chunking (Thalmann et al., 2019) could increase the probability of sentence recall, as the sentence-like lists may be more semantically coherent or more efficiently chunked than list-like lists, making memory for sentences require fewer resources than scrambled lists. Further, making one word of a list easier to recall may free up memory resources, making other words easier to recall via a shifting of resources (Kowialiewski et al., 2021b). We interpret our results through the lens of emergent theories of VWM, which argue that maintenance and ordering in VWM tasks are subserved by the same processes that underlie everyday language comprehension and production. Nevertheless, we recognize that adjudicating between these different explanations will require future research, as the current results do not identify the specific mechanisms by which lexico-syntactic constraints support memory, only that joint word and context information do support memory.

Regardless of a researcher's theoretical position regarding mechanisms of memory, the psycholinguistic literature may offer interesting methods and insights for memory research. Lexico-syntactic constraints were once a controversial subject in sentence comprehension research several decades ago. Research in the 1970s–1980s suggested that comprehension operated via two modular systems, one using part-of-speech information to interpret word order and sentence structure, and a second system interpreting sentence meaning (Ferreira & Clifton, 1986; Frazier & Rayner, 1982). Studies promoting an alternative view manipulated lexico-syntactic factors and argued for an integrated word and syntax system (e.g., Boland et al., 1990; MacDonald, 1993). The effects of those lexico-syntactic manipulations were fairly subtle, initially limiting their influence, yet refinements in stimulus design and comprehension measurements eventually clarified the nature of comprehension processes, largely in favor of integrated word and syntactic processes over alternative modular accounts (for review, see MacDonald et al., 1994; Tanenhaus & Trueswell, 1995). If language processes support VWM, as argued by emergent accounts of VWM, then subtle lexico-syntactic constraints such as verb subcategory or animacy should also support VWM. The current results support a parallel trajectory between the language and memory literatures, in which subtle lexico-syntactic constraints play a role in processing. Parallels between the language and memory literature could be extended further. For example, individual differences in experience with language, which has informed comprehension and production performance (Acheson et al., 2008), could be brought to bear on VWM research as a route to studying the role of experience and LTM on immediate memory. Previous investigations concerning the representations underlying performance in VWM tasks (Jones & Macken, 2015), naturally fit within this perspective.

In addition to future empirical work, explicit comparison of memory and language mechanisms would help to better understand how language mechanisms may impact memory. A look at the computationally explicit language processing literature would be valuable to specify the mechanisms of language emergent theory and to explicitly model the representation of lexico-syntactic constraints (Dell & Chang, 2014; MacDonald & Christiansen, 2002). Initial work in this area by Schwering (2023) demonstrates that language comprehension, production, and repetition mechanisms instantiated in an artificial neural network can account for some VWM phenomena, suggesting that language experience can account for memory without the need to presuppose buffers or chunking mechanisms. Further exploration of modeling approaches and explicit detail of predictions that differentiate between language emergent and alternative accounts is needed.

Relatedly, our methods and data offer new insights into the sentence superiority effect, including the role of meaningfulness in improved recall of sentence-like lists over less sentence-like lists. Sentence-like notions of meaningfulness, coherence, access to event semantics, and other references to higher order structure in meaningful versus meaningless lists are often invoked as an explanation for why sentence-like lists are recalled better than random lists (e.g., Allen et al., 2018; Potter & Lombardi, 1990). Researchers have typically defined meaningfulness through categorical manipulations defined by experimenters, such as an experimenter's subjective feelings about the quality of their stimuli (e.g., Jones & Farrell, 2018). Language research offers several alternatives for characterizing meaningfulness. Theoretically, psycholinguistic approaches encourage consideration of more abstract relationships, including transition probabilities among neighborhoods of related words (e.g., the class of animate nouns and certain verb types), rather than direct transition probabilities among specific words, which has been shown to be important for characterizing generalization to novel utterances (Juliano & Tanenhaus, 1994). Practically, natural language processing provides one scalable way of characterizing how statistical regularities affect the goodness of novel strings with impacts on processing in a continuous manner (Bresnan & Hay, 2008; Lappin & Lau, 2018; Wasow, 2007), rather than assuming that sentence-likeness is categorical in nature. These developments will likely continue to advance the research and theorizing into the nature of VWM.

Appendix A

Table 1 Critical stimuli of Experiment 1 and Experiment 2A

Double objective dative-accepting verb	Control noun (Experiment 1)	Intransitive verb (Experiment 2A)
advanced	analysis	—
asked	stuff	—
assigned	activity	considered
awarded	railing	coughed
bequeathed	exultation	convulsed
bounced	protein	cringed
brought	trouble	—
carried	fighter	—
cited	saver	yawned
conceded	thinness	sparkled
declared	courtesy	—
delivered	childhood	decreased
dropped	student	crashed
emailed	wheelie	gargled
extended	leverage	withered
faxed	gully	glided
fed	pen	sat
flew	guts	lied
forwarded	retractor	flickered
gave	year	tried
granted	academy	floated
guaranteed	philosophy	dwindled
handed	sunset	boiled
hit	kid	—
lent	dose	swam
lifted	aliens	crawled
loaned	boiler	beeped
lobbed	octane	winked
mailed	timber	blinked
offered	custody	waited
owed	oath	soared
paid	type	—
passed	speech	arrived
phoned	racket	barked
pitched	pasture	drooled
presented	extension	prospered
promised	position	increased
provided	survival	remained
pushed	parent	echoed
quoted	greens	sneezed
raised	brains	—
read	lady	ran
refunded	telepath	perspired
reimbursed	ionosphere	salivated
rented	height	rolled
repaid	enigma	snored
reported	employee	lingered
sang	nest	sank
sent	news	—
shot	game	—
showed	corner	laughed
slipped	curtain	—
smuggled	likeness	seethed
snuck	motto	stood
sold	liar	slept
taught	cousin	lived
threw	space	—
tipped	tomato	shrank
told	hell	died
took	year	came
tossed	poetry	rested
traded	weirdo	dabbled
wired	bench	faded
wrote	brain	—
yielded	bigotry	loitered

Table 2 Means and standard deviations of control factors

Experiment	1		2		3	
List condition	Verb	Noun	Ditransitive	Intransitive	Animate	Inanimate
Mean concreteness	3.16 (±0.77)	3.56 (± 0.94)	3.17 (± 0.78)	3.37 (± 0.78)	4.50 (± 0.50)	4.65 (± 0.52)
Log phonological neighborhood density	4.50 (± 1.97)	4.53 (±1.99)	4.26 (± 1.94)	3.67 (± 2.04)	4.85 (± 2.10)	5.26 (± 2.10)
Contextual diversity	2.55 (± 0.80)	2.47 (±0.80)	2.38 (± 0.78)	1.84 (± 1.00)	2.67 (± 0.56)	2.70 (± 0.56)
Frequency	2.68 (± 0.90)	2.66 (±0.90)	2.49 (± 0.87)	1.92 (± 1.07)	3.0 (± 0.66)	3.0 (± 0.67)

Table 3 Critical stimuli of Experiment 2B

Animate noun	Inanimate noun
attorney	property
aunt	cake
ballerina	nightgown
beast	sword
bird	moon
bride	photo
camel	scarf
chauffeur	limousine
children	school
cobra	venom
community	university
consumer	beverage
cow	egg
crab	bean
dancer	circus
deputy	bureau
designer	bracelet
doctor	blood
dog	hair
donor	tumor
eagle	creek
family	house
frog	cave
goat	pork
guy	thing
horse	truck
husband	dinner
kid	hell
lobster	pudding
lover	couch
maid	doll
octopus	spinach
officer	prison
operator	elevator
panther	emerald
passenger	railroad
plumber	chimney
population	territory
prince	island
professor	medicine
sailor	vessel
salmon	tomato
servant	kingdom
teacher	church
team	club
tenor	flute
tiger	skull
tourist	bicycle
trout	canoe
victim	weapon
warrior	thunder

Appendix A4

Similar to our procedures in Experiment 1, we then extracted and parsed all sentences containing target intransitive verbs across the same portion of COCA. We again extracted coarse-grained dependency information associated with these verbs, heuristically identifying sentences containing intransitive verbs as those that did not contain direct object, dative, or passive dependencies. We also excluded sentences in which the subject of the sentence was the object of a relative clause, such as, *The bill Taylor paid…* because the verb is transitive; the sentence contains a direct object (bill) before the verb. We then tabulated the frequencies for each sentence containing the candidate verbs that met these criteria and excluded those verbs that were not used in strictly intransitive contexts. Finally, the second and fourth authors further excluded those verbs that they judged could accept other kinds of noun phrases in the next position (e.g., *a wink* in *The kids hardly slept a wink*).

Appendix A5

To align transitive and intransitive words, we first estimated the mean Euclidean distance between all dative verbs to their noun controls from Experiment 1, as well as the standard deviation of these distances (mean = 2.18, standard deviation = 1.51) to establish a baseline for dative–intransitive control pairs. For all 65 × 52 (= 3,380) candidate dative-intransitive pairs, we then calculated the Euclidean distance between all four features from the dative verbs to the intransitive verbs. We then excluded all pairs that fell outside plus or minus one standard deviation of the dative-noun mean, which eliminated 50% of the possible dative–intransitive verb pairs. Of the remaining 1,699 verb pairs, we wished to identify sufficiently similar verb pairs in a computationally efficient way, so we applied a greedy allocation algorithm. For each of the intransitive verbs we selected the closest (Euclidean distance) dative verb; this dative verb was then removed from the pool of verbs available as potential matches for the remaining intransitive verbs. This process continued until all verbs were allocated a closest neighboring verb.

Appendix A6

For every animate noun, the Euclidean distance to every other animate noun on the dimensions of character length, frequency, and contextual diversity was calculated. All dimensions were z-scored to ensure they were weighed equally. The 10 closest inanimate nouns were then selected for each animate noun for further comparison based upon semantic similarity measures. The semantic similarity between target animate nouns and their foil inanimate nouns was defined as the cosine similarity of the word vectors pulled from the spaCy natural language processing Python package (version 2.1.8; specifically the model 'en_core_web_lg'). Animate–inanimate noun pairs were generated by taking the 10 inanimate nouns with the highest cosine similarity for each animate noun and assigning them in a greedy fashion. The final animate–inanimate pairs were identified by selecting the 52 pairs with the highest cosine similarity.

Appendix A7

To allow comparison of the results of Experiment 2A with the other experiments and to test the robustness of the effects in Experiment 2A, a series of mixed effect logistic regression models were fit on recall across all list positions using the data from Experiment 2A. These additional models were fit using the same fixed and random effects as the models reported in the body of the paper. We report the models on three types of scoring: strict serial scoring to assess overall recall, lenient scoring to assess item memory, and conditionalized order scoring to assess order memory.

Verb biases on overall recall. Participants had higher recall when presented with a dative verb in Position 3 relative to an intransitive verb in Position 3, as indicated by a significant effect of verb type, $\beta = 0.11$, $\chi^2(1) = 6.35$, $p < .05$. There was no effect of counterbalance list condition on recall, $\chi^2(1) = 0.23$, $p = .63$, nor was there an interaction between the effect of verb type and counterbalance list condition, $\chi^2(1) = 0.15$, $p = .70$. There was a significant linear trend of list position on recall, $\beta = 1.92$, $\chi^2(1) = 2123.70$, $p < .001$, as well as a significant quadratic trend of list position on recall, $\beta = 1.28$, $\chi^2(1) = 89.11$, $p < .001$.

Verb biases on item memory. Participants were more likely to recall words irrespective of the position in which they were recalled when presented with a dative verb in Position 3 relative to an intransitive verb in Position 3, $\beta = 0.13$, $\chi^2(1) = 13.15$, $p < .001$. There was no effect of counterbalance list condition, $\chi^2(1) = 0.28$, $p = .60$. Further, there was no interaction between type of verb and counterbalance list condition, $\chi^2(1) = 0.04$, $p = .85$. There was a significant linear trend of list position on item memory, $\beta = \chi 1.50$, $\chi^2(1) = 1364.68$, $p < .001$, and there was a significant effect of the quadratic trend of list position on item memory, $\beta = 1.06$, $\chi^2(1) = 61.58$, $p < .001$.

Verb biases on order memory. Participants were no different at recalling all words in a list when presented with a ditransitive verb in Position 3 compared with when an intransitive verb was in Position 3, $\chi^2(1) = 0.08$, $p = .78$. Additionally, there was no effect of counterbalance list condition on recall, $\chi^2(1) = 0.04$, $p = .85$, nor was there an interaction between verb type and counterbalance list condition, $\chi^2(1) = 0.77$, $p = .38$. There was an effect of the linear trend of list condition on order memory, $\beta = -3.86$, $\chi^2(1) = 657.41$, $p < .001$, as well as an effect of the quadratic trend of list position on order memory, $\beta = 4.58$, $\chi^2(1) = 127.17$, $p < .001$.

Open practices statement All materials, data, and analyses from each of the experiments is available in an online repository. The analyses of Experiment 2B were preregistered on the Open Science Framework (https://osf.io/3v7km/?view_only=b951bab3aa814a03919acf7d7dbcda70.).

Funding This research was funded by NSF Grant #1849236, Wisconsin Alumni Research Foundation (WARF) awards, and the Menzies and Royalty Research Award at the University of Wisconsin–Madison. Additional funding was provided by the NSF-supported Psychology Research Experience Program (PREP) at the University of Wisconsin–Madison.

Data availability All stimuli and analyses are available on the Open Science Foundation (https://osf.io/6gtnk/?view_only=2f41a6dffdbd46a882fcd41f717604a7).

Code availability Not applicable.

Declarations

Ethical approval This experiment and the following experiment were approved by the University of Wisconsin–Madison Institutional Review Board, and all participants gave their informed consent prior to participation.

Consent to participate All participants provided informed consent prior to participation.

Consent to publish All participants provided consent to publish their data.

Conflict of interest The authors declare no competing interests.

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

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*(1), 278–289.

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

Ambridge, B., Pine, J. M., Rowland, C. F., Freudenthal, D., & Chang, F. (2014). Avoiding dative overgeneralisation errors: Semantics, statistics or both? *Language, Cognition and Neuroscience, 29*(2), 218–243. https://doi.org/10.1080/01690965.2012.738300

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., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485–493. https://doi.org/10.1037/0894-4105.8.4.485

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29*(4), 413–432. https://doi.org/10.1016/0749-596X(90)90064-7

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389–411. https://doi.org/10.1037/met0000159

Bresnan, J., & Hay, J. (2008). Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua, 118*(2), 245–259. https://doi.org/10.1016/j.lingua.2007.02.007

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

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior research methods, 44*(4), 991–997.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

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

Davies, M. (2008). *The corpus of contemporary American english (COCA): 560 million words, 1990–present*. https://corpus.byu.edu/coca/

Dell, G. S., & Chang, F. (2013). The P-chain: relating sentence production and its disorders to comprehention and acquisition. Philosophical transactions of the royal society of London. *Series B, Bioplogical Sciences, 369*(1634), 20121394. https://doi.org/10.1098/rtsb.2012.0394

de Marneffe, C., Manning, C., Nivre, J., & Zeman, D. (2021). Universal dependencies. *Computational Linguistics, 47*(2), 255–308.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA, 105*(2), 203.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE Publications. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language, 48*(2), 281–303.

Hawkins, R., Yamakoshi, T., Griffiths, T. L., & Goldberg, A. (2020, November). Investigating representations of verb bias in neural language models. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 4653–4663)* Association for Computational Linguistics.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language, 59*(4), 434–446.

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

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1–13.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23*(6), 459–471.

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2022). The semantic relatedness effect in serial recall: Deconfounding encoding and recall order. *Journal of Memory and Language, 127*, 104377. https://doi.org/10.1016/j.jml.2022.104377

Kowialiewski, B., Gorin, S., & Majerus, S. (2021a). Semantic knowledge constrains the processing of serial order information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(12), 1958–1970. https://doi.org/10.1037/xlm0001031

Kowialiewski, B., Lemaire, B., Majerus, S., & Portrat, S. (2021b). Can activated long-term memory maintain serial order information? *Psychonomic Bulletin & Review, 28*(4), 1301–1312. https://doi.org/10.3758/s13423-021-01902-3

Lee, M., & Thompson, C. K. (2004). Agrammatic aphasic production and comprehension of unaccusative verbs in sentence contexts. *Journal of neurolinguistics, 17*(4), 315–330. https://doi.org/10.1016/S0911-6044(03)00062-9

Lappin, S., & Lau, J. H. (2018). Gradient probabilistic models vs categorical grammars: A reply to Sprouse et al. (2018). *Science of Language*.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT Press.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Liu, D. (2008). Intransitive or object deleting?: Classifying English verbs used without an object. *Journal of English Linguistics, 36*(4), 289–313. https://doi.org/10.1177/0075424208317128

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. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32*(5), 692–715. https://doi.org/10.1006/jmla.1993.1035

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676.

Macdonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory; comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54. https://doi.org/10.1037/033-295X.109.1.35

MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 581–611). Academic Press.

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

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33*(1), 83–111.

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

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin, 143*(9), 992–1009. http://dx.doi.org/https://doi.org/10.1037/bul0000108

Page, M. P., & 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 of London. Series B, Biological sciences, 364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

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

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

Saint-Aubin, J., & Poirier, M. (1999). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34*(5/6), 347–352. https://doi.org/10.1080/002075999399675

Schweppe, J., Schütte, F., Machleb, F., & Hellfritsch, M. (2021). Syntax, morphosyntax, and serial recall: How language supports short-term memory. *Memory & Cognition.* Advance online publication. https://doi.org/10.3758/s13421-021-01203-z

Schwering, S. C. (2023). *The Lichtheim-Memory Model: A computational model of language comprehension, production, and verbal working memory* (Doctoral dissertation, The University of Wisconsin-Madison).

Schwering, S. C., & MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14*, 68.

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

Shetreet, E., Friedmann, N., & Hadar, U. (2010). The neural correlates of linguistic distinctions: Unaccusative and unergative verbs. *Journal of cognitive neuroscience, 22*(10), 2306–2315. https://doi.org/10.1162/jocn.2009.21371

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language, 39*(3), 392–417.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Vol. 11: Speech Language and communication (pp. 217–262)*. Academic Press.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 37–55. https://doi.org/10.1037/xlm0000578

Thompson, C. K., Bonakdarpour, B., Fix, S. C., Blumenfeld, H. K., Parrish, T. B., Gitelman, D. R., & Mesulam, M.-M. (2007). Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience, 19*(11), 1753–1767. https://doi.org/10.1162/jocn.2007.19.11.1753

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 528.

VanArsdall, J. E., & Blunt, J. R. (2022). Analyzing the structure of animacy: Exploring relationships among six new animacy and 15 existing normative dimensions for 1,200 concrete nouns. *Memory & Cognition, 50*, 997–1012.

Wasow, T. (2007). Gradient data and gradient grammars. *Proceedings from the Annual Meeting of the Chicago Linguistic Society, 43*, 255–271.