Title
Strategic allocation of memory resources through the lens of dependency locality: Evidence from a controlled reading experiment

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Xu, Weijie
Hao, Hailin
Futrell, Richard

Publication Date
2026

Strategic allocation of memory resources through the lens of dependency locality: Evidence from a controlled reading experiment

Weijie Xu[1]**, Hailin Hao**[2] **& Richard Futrell**[1]
[1]Department of Language Science, University of California Irvine
[2]Department of Linguistics, University of Southern California

Abstract

Human working memory resources are strategically allocated in sentence processing. Recent studies have argued that less predictable linguistic units are prioritized for memory resources to be encoded with higher precision in sentence processing, as a principled solution to minimize the overall memory error given noisy encoding and resource limitations. The investigation of this claim has been operationalized from the perspective of memory robustness, predicting that representations of unexpected linguistic units are more robust against memory interference since they are more precisely encoded. In this study, using a controlled reading experiment in the A-Maze paradigm, we examine this prediction through the lens of the dependency locality effect in comprehension, a classic sentence-processing effect arising from memory decay and interference. Focusing on the subject–verb dependency in English, we find that the locality effect only occurs when the retrieval target (i.e., left codependent) is of high predictability, suggesting that high-predictability linguistic units are less prioritized for memory resources, thus encoded with less robust representation against the interference from the intervening material between codependents. Interestingly, we also observe a mild anti-locality pattern for unexpected retrieval target, pointing to a potential competition between memory- and expectation-based sentence processing mechanisms under the impact of strategic allocation of memory resources.

Keywords: strategic resource allocation; memory encoding and retrieval dependency locality; sentence processing; A-Maze

Introduction

Human working memory has limited resources, yet it supports language processing in a seemingly effortless way. From a domain-general perspective, considerable empirical evidence has shown that the limited working memory resources are flexibly distributed, in the sense that some kinds of information are strategically prioritized for memory resources so they are better memorized (e.g., Bays & Husain, 2008; Bruning & Lewis-Peacock, 2020; Ma et al., 2014; van den Berg et al., 2012). In language processing specifically, a recent proposal by Xu and Futrell (2026) provides a resource-rational account of strategic resource allocation (SRA) for memory encoding and memory efficiency. They argue that the encoding resources of working memory are dynamically allocated to prioritize unexpected and surprising linguistic inputs, an efficiency strategy that arises as a principled solution to minimize the overall memory error. This mainly arises from the noisy nature of memory encoding and working memory deals with this representational noise. Put simply, the true form of a past input is inaccessible due to representational noise, and memory retrieval is effectively an inferential process that reconstructs a past input using the statistical structure of long-term knowledge (e.g., C. J. Bates & Jacobs, 2020; Bays et al., 2024; Brady et al., 2024; Futrell, Gibson, & Levy, 2020; Hahn et al., 2022). One property of this inferential process is that more surprising inputs are less likely to be successfully reconstructed, resulting in higher retrieval error. Therefore, as a rational solution to minimize the total memory error across multiple linguistic inputs, more surprising inputs need to be encoded with higher precision using more memory resources.

Empirically, the argument for strategic resource allocation has been operationalized from the perspective of memory robustness, predicting that the memory representation of more surprising inputs, since they are encoded with higher precision, should have enhanced robustness against memory interference (Xu & Futrell, 2025, 2026). Indirect evidence for this prediction can be found in previous sentence processing literature. For example, semantically complicated expressions, despite being more difficult to encode, are easier to be retrieved at a later time point (Hofmeister, 2011; Hofmeister & Vasishth, 2014; Troyer et al., 2016). Moreover, through a binary erasure of memory representation, the resource-rational sentence processing model by Hahn et al. (2022) demonstrates a dynamic where linguistic units that are highly predictable from the context (such as function words) are more likely to be erased.

More recently, a growing body of work have found more direct support for strategic resource allocation in two empirical domains of long-distance dependency processing, namely the agreement attraction effect (Tung & Brennan, 2023; Xu & Futrell, 2025) and the dependency locality effect (Xu & Futrell, 2024, 2026). In Xu and Futrell (2025), the authors manipulate the predictability of the retrieval target in subject–verb dependencies (i.e., the subject head noun), and find that the number feature on less predictable subject nouns exhibits weaker agreement attraction induced by the distractor noun during retrieval at the main verb. This reduced agreement attraction effect suggests that less predictable information is encoded in more robust representation, consistent with the prediction of strategic resource allocation. A similar reduced attraction effect is also find in Tung and Brennan (2023) in the case of noun-phrase ellipsis of Chinese classifier–noun agreement, where the retrieval of the legitimate antecedent noun at the classifier is less interrupted by a distractor when

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

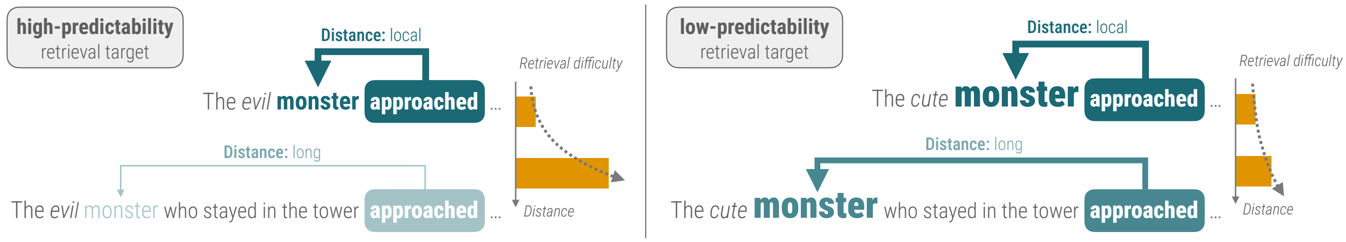

Figure 1: Empirical prediction of strategic memory-resource allocation in dependency locality in comprehension (Xu & Futrell, 2026). The low-predictability retrieval target (i.e., the left codependent) is prioritized for memory resources, encoded with more robust representation against memory interference, thus exhibiting a reduced locality effect on retrieval difficulty.

the antecedent noun is less predictable.

Another line of strategic resource allocation comes from the empirical domain of dependency locality, where the retrieval of less predictable left codependents have been found to exhibit a reduced locality effect. As shown in (1), at the main verb "approached", the left codependent "monster" needs to be retrieved in order to establish this subject–verb dependency. The Dependency Locality Theory (Gibson, 1998, 2000) holds that this retrieval difficulty is increased with longer dependency length (e.g., from (1a) to (1c)) due to memory decay and the interference from intervening materials. This positive correlation between retrieval difficulty and dependency length has been empirically established by a large number of work (e.g., Bartek et al., 2011; Demberg & Keller, 2008; Grodner & Gibson, 2005; Traxler et al., 2002).

(1) The dependency locality effect: longer dependency length induces higher retrieval difficulty

   a. The monster *approached...*
   b. The monster in the tower *approached...*
   c. The monster who stayed in the tower *approached...*

Through the analysis of large-scale reading corpora, Xu and Futrell (2026) find that the locality effect in the comprehension of subject–verb dependencies is reduced for less predictable left codependents, in the sense that their retrieval difficulty increases more slowly as the dependency length gets longer, as illustrated in Figure 1. Moreover, a similar pattern is also observed in the text corpus, where less predictable left codependents tend to associate with longer dependency length, suggesting that they can tolerate more intervening materials before the retrieval site at the right codependent (Xu & Futrell, 2024, 2026). These findings indicate that strategic resource allocation is not only a memory efficiency principle that affects real-time sentence processing, but also shape language structure in a way that relaxes the dependency minimization pressure (Futrell, Levy, & Gibson, 2020; Gibson et al., 2019; Hawkins, 2004).

The current study

Previous studies mentioned above have only examined the effect of strategic memory-resource allocation on dependency locality corpus analysis. Despite reflecting language use in naturalistic contexts, observational data poses challenges for controlling confounding factors, potentially leading to spurious correlations that undermine the internal validity of empirical examination. Moreover, the analysis largely relies on the annotation quality of the corpora, and may involve automatized parsing that sometimes can be unavoidably erroneous. Therefore, in the current study, we conduct a controlled reading experiment to examine strategic resource allocation in the context of dependency locality, providing additional support for the effect observed in the corpus analysis of previous studies. To preview the result, consistent with the observation of Xu and Futrell (2026), we find that the locality effect in subject–verb dependencies is indeed modulated by the predictability of the left codependent. Specifically, the retrieval difficulty indeed increases with longer dependency length, but only for highly predictable left codependents. For low-predictability left codependents, however, this locality effect is mildly reversed. This interesting divergence between locality and anti-locality effect raises the potential impact of strategic resource allocation on the competition between memory- and expectation-based sentence processing mechanisms, as we discuss in General Discussion.

Experiment

In this experiment, using the A-Maze paradigm (Boyce et al., 2020), we examine whether there is a reduced dependency locality effect for low predictability left codependents as predicted by strategic allocation of memory resources. We focus on the subject–verb dependency, with the subject being the retrieval target and the main verb (MV) being the retrieval site. As shown in Table 1, in a 2×2 within-subject design, we manipulate the subject–verb dependency DISTANCE and the subject noun PREDICTABILITY.

Predictions. We first expect to replicate the classic dependency locality effect, where the retrieval difficulty manifested as the reading time (RT) at the main verb should increase from local to long condition. Second, if unexpected linguistic inputs are prioritized for memory resources and are encoded with more robust representation against interference, we should observe a reduced locality effect for low predictability condition. That is, as illustrated in Figure 1, since less predictable

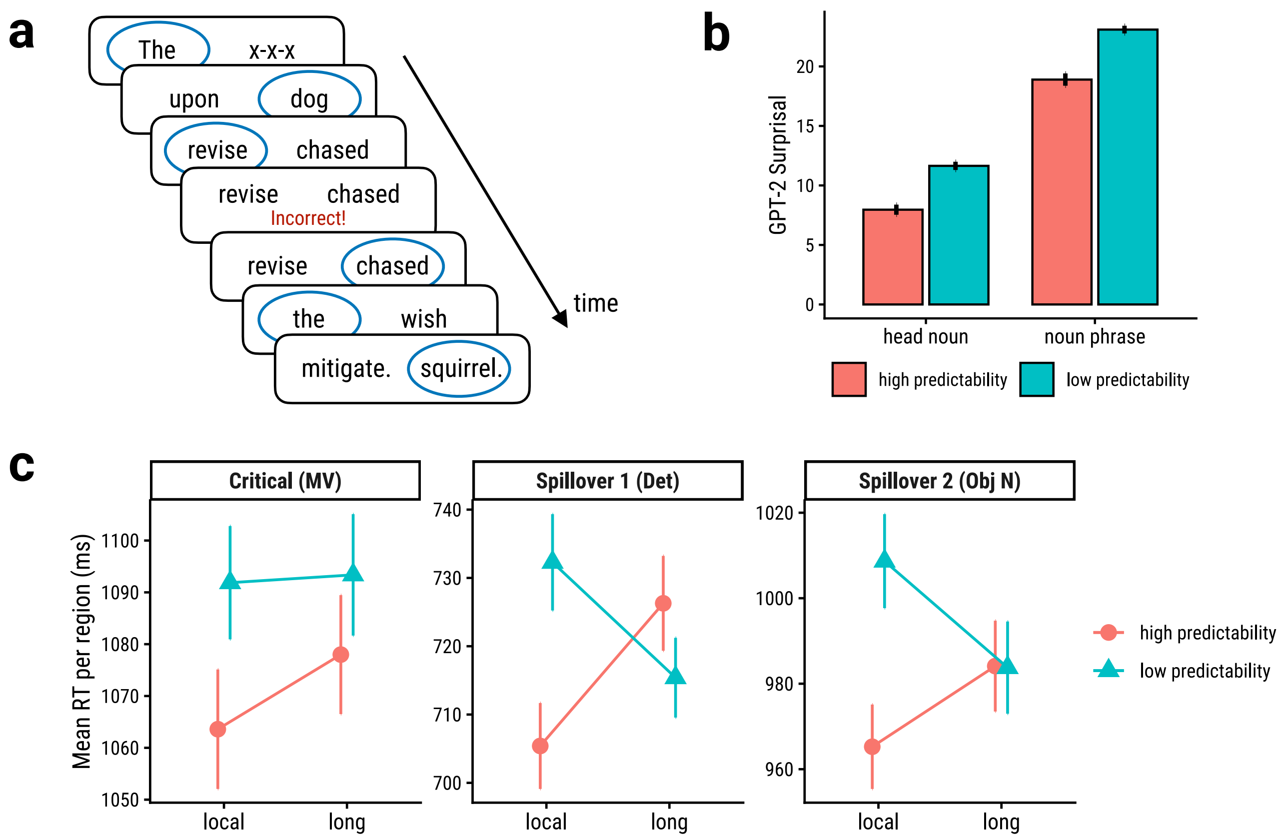

Figure 2: **a.** The procedure of A-Maze paradigm. **b.** Mean GPT-2 surprisal estimates of the subject head noun and the subject noun phrase as a whole per condition, for the norming of subject PREDICTABILITY. **c.** Mean reading times (RTs) with standard error for each analysis region; the critical region is the main verb (MV); the spillover regions immediately follow the MV.

subject nouns are more robust against interference, they should be able to tolerate more intervening materials before the retrieval site, thus their retrieval difficulty of the less predictable subject noun should increase more slowly from local to long condition. Therefore, on top of the DISTANCE main effect, we also expect to observe a DISTANCE × PREDICTABILITY two-way interaction for RTs at the main verb.

Methods

Participants. 100 English native speakers living in the U.S. are recruited via Prolific and are paid $4 for participation.

Material. The sample stimuli are in Table 1, with manipulation on the subject–verb dependency DISTANCE and subject-noun PREDICTABILITY. For DISTANCE, the *local* condition has the subject head noun and the main verb adjacent to each other, whereas the *long* condition has a relative clause inserted before the main verb. For PREDICTABILITY, following Xu and Futrell (2025), we manipulate the expectancy of the left-codependent of the subject–verb dependency into two conditions (i.e., high vs. low) through a prenominal adjective that precedes the subject head noun. We verify our manipulation of subject predictability using the surprisal estimates from GPT-2 (Radford et al., 2019), such that the surprisal of both the subject head noun and the subject noun phrase as a whole are higher in the low predictability condition, as shown in Figure 2b. There are 32 experimental items and 32 filler items.

Procedure. Participants recruited from Prolific are directed to PCIbex (Zehr & Schwarz, 2018) to take the experiment. In this study, we employ the A-Maze paradigm (Boyce et al., 2020), a word-by-word sentence processing task in which participants choose between the intended next word and a distractor, yielding Reaction Time (RT) measurements for each position, as illustrated in Figure 2a. Word reading times (RTs) are thus the reaction time taken by the participants to make the decision. Incorrect selections trigger an error message and require participants to retry before advancing. Distractors are generated automatically using neural language models (Gulordava et al., 2018), ensuring that the alternative word is low in contextual predictability while matching the target in frequency and length. Since A-Maze task is already cognitively demanding, we do not include any additional task (e.g., comprehension questions) in the experiment, and we use the accuracy of A-Maze decisions as the attention check.

Data exclusion. We first exclude participants whose A-Maze accuracy is lower than 85% (one excluded). We then

Table 1: Sample stimuli with 2×2 manipulation on the subject–verb dependency DISTANCE and subject-noun PREDICTABILITY. Sentences are presented word-by-word using A-Maze. Critical main verbs are bold-faced, with spillover words italicized.

1. *local, high predictability*	The evil monster [∅] **approached** *the princess* although it only wanted a friend.
2. *long-distance, high predictability*	The evil monster [who stayed in the tower] **approached** *the princess* although it only wanted a friend.
3. *local, low predictability*	The cute monster [∅] **approached** *the princess* although it only wanted a friend.
4. *long-distance, low predictability*	The cute monster [who stayed in the tower] **approached** *the princess* although it only wanted a friend.

exclude participants whose average RT of filler items is three standard deviations away from the mean (another one excluded), yielding a dataset of 98 participants. We then adopt a two-step method to exclude trial-level RT outliers. In the first step, we exclude RT observations that are faster than 100ms or longer than 4000ms (approx 0.07%). In the second step, we further exclude RTs that are three standard deviations away from the mean per condition per region (approx 0.4%).

Data analysis. We run Bayesian multilevel linear models with `brms` package in `R` (Bürkner, 2017) on log-transformed reading times (RTs). We analyze three RT regions: the critical main verb where the left-codependent of the subject–verb dependency is retrieved, and two spillover words that immediately follow the main verb. All categorical variables are effect coded (DISTANCE: long 1 vs. local -1; PREDICTABILITY: high 1 vs. low -1). We include a DISTANCE × PREDICTABILITY two-way interaction as well as its two lower-order terms as predictors in the statistical models. We also include by-item and by-participant random effects with the maximal random structure (Baayen et al., 2008; Barr et al., 2013; D. Bates et al., 2015). The regression formula is given by[1]

$$\begin{aligned} \text{logRT} \sim\ & \text{logRT.previous} + \text{distance} * \text{predictability} \\ & + (\text{distance} * \text{predictability} \mid \text{item}) \\ & + (\text{distance} * \text{predictability} \mid \text{participant}) \end{aligned} \tag{1}$$

Moreover, since the region before the critical main verb has different lexical content across the local and long conditions in DISTANCE manipulation, we also include the log RTs of the previous region as a control variable in all statistical models to account for the potential spillover effect.[2]

Results

The mean reading times (RTs) in the regions of interest are shown in Figure 2c, and the main result of statistical models is summarized in Table 2.

[1] logRT.previous is the log reading times of the previous region before the region of analysis.

[2] Although a threshold to reject the null hypothesis is strongly discouraged for Bayesian statistics, for expository purposes we interpret the statistical evidence as *reliable* if the posterior probability P > 0.95, *preliminary* if P > 0.9, and unreliable if P < 0.9.

Table 2: Main result of statistical models per region

	Critical Region (MV)		
	Est	95% CrI	Posterior Prob
DISTANCE	0.014	[-0.002, 0.032]	P(β>0)=0.954
PREDICTABILITY	-0.005	[-0.018, 0.008]	P(β<0)=0.76
DIST × PRED	-0.001	[-0.014, 0.013]	P(β<0)=0.533
	Spillover 1 Region (Det)		
	Est	95% CrI	Posterior Prob
DISTANCE	0.002	[-0.008, 0.012]	P(β>0)=0.648
PREDICTABILITY	-0.004	[-0.014, 0.007]	P(β<0)=0.75
DIST × PRED	0.009	[-0.001, 0.019]	P(β>0)=0.962
	Spillover 2 Region (Obj N)		
	Est	95% CrI	Posterior Prob
DISTANCE	-0.005	[-0.022, 0.013]	P(β<0)=0.725
PREDICTABILITY	-0.008	[-0.021, 0.005]	P(β<0)=0.905
DIST × PRED	0.008	[-0.004, 0.021]	P(β>0)=0.909

Critical region. We observe a statistically reliable DISTANCE main effect in the critical main-verb region, where long-distance subject–verb dependency results in higher RTs at the main verb ($\hat{\beta}$=0.014, CrI=[-0.002, 0.032], P(β>0)=0.954), replicating a classic dependency locality effect. However, neither the PREDICTABILITY main effect ($\hat{\beta}$=-0.005, CrI=[-0.018, 0.008], P(β>0)=0.76) nor the interaction ($\hat{\beta}$=-0.001, CrI=[-0.014, 0.013], P(β>0)=0.533) is observed at the main verb.

Spillover 1 region. There is no evidence either for the DISTANCE main effect ($\hat{\beta}$=0.002, CrI=[-0.008, 0.012], P(β>0)=0.648) or for the PREDICTABILITY main effect ($\hat{\beta}$=-0.004, CrI=[-0.014, 0.007], P(β>0)=0.75). However, importantly, there is a statistically reliable DISTANCE × PREDICTABILITY interaction ($\hat{\beta}$=0.009, CrI=[-0.001, 0.019], P(β>0)=0.962) that is driven by a reversed pattern of locality effect across the two predictability conditions (Figure 2c middle), where the high predictability condition exhibits a locality effect while the low predictability conditions points to an anti-locality effect. Given this interaction effect, we run an additional analysis with nested contrast coding (Nicenboim et al., 2024), as specified in Table 3. The result (Table 4) shows some statistical evidence for the locality effect within the high predictability condition ($\hat{\beta}$=0.022, CrI=[-0.006, 0.050], P(β>0)=0.937), whereas the effect within the low predictability, despite numerically in the direction of an anti-locality effect, is not statistically reliable ($\hat{\beta}$=-0.015, CrI=[-0.044, 0.014], P(β<0)=0.849).

Table 3: Nested contrast coding to test the DISTANCE effect within each PREDICTABILITY condition

	DISTANCE (low PRED)	DISTANCE (high PRED)	PRED main effect
high PRED, local	0	-1	1/2
high PRED, long	0	1	1/2
low PRED, local	-1	0	-1/2
low PRED, long	1	0	-1/2

Table 4: Statistical result of the DISTANCE effect within each PREDICTABILITY condition in the Spillover 1 region, using nested contrast coding

	Est	95% CrI	Post Prob
DIST (low PRED)	-0.015	[-0.044, 0.014]	P(β<0)=0.85
DIST (high PRED)	0.022	[-0.006, 0.050]	P(β>0)=0.94
PRED (main)	-0.007	[-0.029, 0.014]	P(β<0)=0.75

Spillover 2 region. Similar to the Spillover 1 region, the data does not support a DISTANCE main effect ($\hat{\beta}$=-0.005, CrI=[-0.022, 0.013], P(β<0)=0.725). However, there is preliminary evidence for a PREDICTABILITY main effect ($\hat{\beta}$=-0.008, CrI=[-0.021, 0.005], P(β<0)=0.905), suggesting that the main verbs are read faster when the left codependent is highly predictable. We also observe preliminary evidence for an interaction effect ($\hat{\beta}$=0.008, CrI=[-0.004, 0.021], P(β>0)=0.909), which as in the Spillover 1 region is driven by a reversed locality pattern in the low-predictability condition (Figure 2c right).

Discussion

We first replicate the classic dependency locality effect in the critical main verb region, supporting the Dependency Locality Theory in the sense that the left-codependent of a syntactic dependency (the subject in our case) is more difficult to retrieved when it is more distant from its retrieval site at the right codependent. Importantly, the result also reveals that the locality effect is dependent on the predictability of the left-codependent. As predicted by the strategic allocation of memory resources, an increased retrieval difficulty is only observed for highly predictable retrieval targets, indicating that the processing of highly predictable information is more susceptible to the interference of intervening materials between the encoding and retrieval sites.[3] However, surprisingly, the locality effect for unpredictable retrieval target is not only reduced, but also reversed, manifested as an anti-locality effect although there is only weak statistical support for this. We discuss the potential methodological concern for this anti-locality effect and its theoretical implications below in General Discussion.

General Discussion

Consistent with previous studies, our result reveals that the dependency locality effect is indeed modulated by the predictability of the left codependent, in the case of subject–verb dependency. As predicted by the strategic allocation of memory resources, the classic locality effect is only observed for high-predictability left codependents, suggesting that their representation is less robust against the interference from other linguistic inputs. However, surprisingly, the locality effect for low-predictability left codependents, instead of being reduced, is mildly reversed, exhibiting a weak anti-locality pattern. Why is there a divergence between locality and anti-locality effect corresponding to different levels of the predictability of left codependents? Here we discuss two potential explanations, one methodological and one theoretical.

Methodologically, one potential confound is that the critical main-verb region in the *local* condition of our stimuli immediately goes after the subject head noun, where the manipulation of predictability takes place. Although this design is also used in previous studies of dependency locality (Bartek et al., 2011; Grodner & Gibson, 2005), it is possible that the high processing difficulty at the low-predictability subject noun may spill over to the following critical main-verb region, resulting in a spuriously high reading time for its corresponding local condition. To address this issue, in the statistical analysis for each region of interest, we have included the reading times of the previous region as a control variable, regressing out the potential impact of this confounding spillover effect. The critical interaction effect remains statistically reliable, and this anti-locality pattern within low-predictability condition becomes statistically weak as noted above in our analysis with nested contrast coding. Moreover, the critical DISTANCE × PREDICTABILITY interaction effect in our result mostly emerges at the first region that goes after the main verb, and there is still an interaction pattern at the second post-verb region, which is already three words away from the subject head noun in the local condition. This persisting interaction effect suggests that it is unlikely to be merely caused by the spillover confound from the predictability manipulation at the subject head noun, further mitigating this methodological concern. Nevertheless, follow-up work should address this spillover confound in their stimuli design by inserting some neutral low-informative lexical items before the critical region.

[3] One caveat is that such an interaction effect is not statistically reliable when we use GPT-2 surprisal, instead of the binary predictability, as the predictor in the statistical models.

Theoretically, the divergence between locality and anti-locality pattern can be driven by a competition between memory-based and expectation-based mechanisms, under the influence of strategic memory resource allocation. There are at least two processes at play when processing the right codependent. On the one hand, the memory-based mechanism retrieves the left codependent in order to establish the dependency, leading to a locality effect. On the other hand, the expectation-based mechanism predicts that longer dependency length should facilitate the processing of the right codependent, since more intervening materials can provide more relevant information that makes it easier to predict the right codependent itself, leading to an anti-locality effect (Konieczny, 2000; Levy, 2008; Levy & Keller, 2013; Nakatani & Gibson, 2010; Vasishth & Lewis, 2006). The competition between these two mechanisms can be mediated by the distribution of memory resources across different linguistic inputs. For example, when the unpredictable left codependent is prioritized for more robust encoding, the interference effect may not be strong enough anymore to cancel out the facilitation brought by the additional information provided by the intervening material in terms of predicting the right codependent. In this case, the expectation-based mechanism may have a stronger influence than the memory-based one, and the anti-locality effect starts to overweigh the locality effect. Future work should investigate in more detail this potential impact of strategic allocation of memory resources on the competition between memory- and expectation-based mechanisms.

Conclusion

In this work, through a controlled reading experiment using the A-Maze paradigm, we find evidence for a strategic allocation of memory resources in the context of the dependency locality effect of sentence processing. Focusing on subject–verb dependencies in English, we show that the locality effect only occurs on highly predictable left codependents, suggesting that these linguistic units are de-prioritized for memory resources, resulting in less robust representations against the interference from additional intervening materials before the retrieval site. We also observe a mild anti-locality pattern for low-predictability left codependents, pointing to a potential competition between memory- and expectation-based sentence processing mechanisms that is mediated by how the memory resources are strategically allocated.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

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.

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*, *127*(5), 891.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Bays, P. M., Schneegans, S., Ma, W. J., & Brady, T. F. (2024). Representation and computation in visual working memory. *Nature Human Behaviour*, 1–19.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082.

Brady, T. F., Robinson, M. M., & Williams, J. R. (2024). Noisy and hierarchical visual memory across timescales. *Nature Reviews Psychology*, 1–17.

Bruning, A. L., & Lewis-Peacock, J. A. (2020). Long-term memory guides resource allocation in working memory. *Scientific Reports*, *10*(1), 1–10.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), e12814.

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371–412.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). The MIT Press.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290.

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *Proceedings of the 2018 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies (NAACL-HLT), 1195–1205.
Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43), e2122602119.
Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press.
Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376–405.
Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, *5*, 1237.
Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627–645.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.
Levy, R., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222.
Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356.
Nakatani, K., & Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science*, *34*(1), 94–112.
Nicenboim, B., Schad, D., & Vasishth, S. (2024). An Introduction to Bayesian Data Analysis for Cognitive Science [Accessed: 2024-12-08].
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.
Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90.
Troyer, M., Hofmeister, P., & Kutas, M. (2016). Elaboration over a discourse facilitates retrieval in sentence processing. *Frontiers in Psychology*, *7*, 374.
Tung, T.-Y., & Brennan, J. R. (2023). Expectations modulate retrieval interference during ellipsis resolution. *Neuropsychologia*, 108680.
van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785.
Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 767–794.
Xu, W., & Futrell, R. (2024). Syntactic dependency length shaped by strategic memory allocation. *Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, 1–9.
Xu, W., & Futrell, R. (2025). Informativity enhances memory robustness against interference in sentence comprehension. *Journal of Memory and Language*, *142*, 104603.
Xu, W., & Futrell, R. (2026). Strategic resource allocation in memory encoding: An efficiency principle shaping language processing. *Journal of Memory and Language*, *146*, 104706.
Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832