How linguistic boundaries form encoding contexts in memory: Evidence from temporal order effects

Lalitha Balachandran[a] (lalithab@ucsc.edu), Stephanie Rich[b] (skrich@ucsc.edu), & Matt Wagers[c] (mwagers@ucsc.edu)

[a]Department of Ecology & Evolutionary Biology, UC Santa Cruz,
[b]Department of Psychology, Concordia University, [c]Department of Linguistics, UC Santa Cruz

Abstract

While temporal contiguity and order effects have been shown to be highly robust in the word list memory literature, little is known about how the presence of boundaries signaling higher-order linguistic units affects temporal order memory for items within those units. Here, we present the results of two sentence memory experiments that show that linguistic boundaries block contiguity effects through a temporal context mechanism, whereby words encoded across distinct linguistic environments are also encoded as contextually dissimilar. Two consequences of this encoding mechanism are that (i) retrieval of an item results in reactivation of that item's linguistic context alone due to co-activation of contextually similar content and (ii) significant linguistic boundaries reduce encoding interference between proximal, semantically similar items. We take these results to suggest that linguistic groupings map onto encoding contexts, which constrain the effect of item-to-item associations in sentence memory.

Keywords: linguistic boundaries; temporal context; temporal contiguity; temporal order memory; encoding contexts

1. Introduction

The temporal organization of items in memory strongly influences retrieval dynamics for word lists; successful recall is, in part, guided by proximity, serial order, and group membership (Frankish, 1995; Healey, Long, & Kahana, 2019; Hitch, 1996; Howard, Youker, & Venkatadass, 2008; Unsworth, 2008). While early work in sentence memory explored the role of temporal groupings in organizing subsentential linguistic units in memory (Carroll & Tanenhaus, 1978; Jarvella, 1971, 1979; Marslen-Wilson & Tyler, 1976; Slowiaczek, 1981), the effects of proximity and serial order have largely been underexplored. Instead, modern models of linguistic memory have turned their attention to similarity-based interference between linguistically relevant word-level features (e.g., number, gender, animacy, etc.) during real-time sentence comprehension. For example, encountering the verb *was complaining* in (1) triggers a search for a feature-matching subject. This search becomes more costly in the presence of multiple featurally overlapping +ANIMATE noun phrases (bolded) that may receive activation during the retrieval process (Van Dyke & McElree, 2006).

(1) The **worker** was surprised that the **resident** who was living near the dangerous **neighbor** *was complaining* about the investigation.

Models of linguistic memory have strayed away from investigating issues of temporality for good reason: sentences are inherently structural in nature, and the properties of sentence memory and word list memory diverge in notable ways. Sentences are better remembered than word lists, and sentence recall does not exhibit the characteristic bowing of serial position functions (i.e., the recall advantage for initial and final positions relative to intermediate ones) in word list recall (Baddeley, Hitch, & Allen, 2009).

Nevertheless, recent work highlights that issues of proximity and order do matter in sentential contexts. For instance, in narrative contexts, temporally contiguous events (*e*+/-1) are accurately recalled with greater probability (Raccah, Chen, Gureckis, Poeppel, & Vo, 2024), suggesting that serial order information and temporal relationships between contiguous higher-order linguistic units are encoded in memory. At the item level during online comprehension, in (2), proximal encoding of two similar items (*knife* and *sword*) leads to greater difficulty at a distant retrieval site (*sharpened*) compared to a less similar pair (e.g., *knife* and *stick*), due to interference during the encoding phase (Rich, 2024).

(2) It was clear that the **knife** that the {**sword**, stick} was placed by had been recently *sharpened* in the kitchen.

Because research on linguistic memory has focused so deeply on item-level features, we are missing a thorough understanding of (i) how and when proximal item-to-item associations affect sentence memory, and (ii) how information about a particular item's membership within a larger linguistic unit is encoded and utilized by the memory system. Here, we offer a unified account of how (i) and (ii) may interact by exploring another source of featural similarity between items, that of shared *encoding context*. We contend that models of memory that incorporate a notion of *temporal context* (Howard & Kahana, 2002) particularly lend themselves to understanding how the presence of a significant linguistic boundary may serve to group linguistic items in memory (Wagers, 2008), thereby increasing contextual similarity between within-group items while also increasing contextual distinctiveness between items that span boundaries.

1.1 Temporal contiguity & order

The Temporal Context Model (TCM; Howard and Kahana, 2002) assumes that items are encoded along with information about the spatiotemporal context in which they occurred; thus, a slowly evolving context vector binds a "timestamp" of

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

sorts to each item, which carries information about the external environment or the cognitive system's internal context at the time of encoding. These contextual features drive the formation of associations between contiguously encoded items. Sudden shifts in the context (e.g., in physical location, goal, perceptual input, or discourse relationships) lead to a corresponding shift in the context vector, which distinguishes the contextual states associated with upcoming items from previous input (see Clewett, DuBrow, & Davachi, 2019). This model has been used to account for two patterns of temporal proximity that are central to the current studies: (i) temporal contiguity and (ii) temporal order memory effects.

The temporal contiguity effect is a robust property of human memory: after correctly recalling item *i* from a list, participants are most likely to recall the item presented either directly before or after item *i*. The effect is typically asymmetrical in the forward direction, where recall of item *i*+1 is more likely than that of *i*-1 (see Healey et al. (2019) for examples and an explanation under the TCM). Across-list contiguity applies as well: participants are more likely to make recall transitions between adjacent lists, usually in the forward direction (Howard et al., 2008; Unsworth, 2008). In Experiment 1, we investigate whether such contiguity effects are evident in cross-clause transitions in sentence recall.

Temporal order memory (TOM) refers to memory for the relative order of pairs of items, and is highly sensitive to context shifts. When list items are temporally distant, TOM is more accurate than when they are proximal, because distant items are contextually distinguishable (e.g., Pu, Kong, Ranganath, & Melloni, 2022). Items in boundary positions also evidence better memory for item-context associations (Heusser, Ezzyat, Shiff, & Davachi, 2018). Somewhat counterintuitively, TOM for items *i* and *i*+1 is impaired when those items span a context boundary compared to when they occupy the same group (DuBrow & Davachi, 2013). This is because contextual associations render immediately adjacent items more accessible; given a perceived temporal discontinuity, the contextual state associated with a cross-boundary item will be less accessible. However, there are notable exceptions to this generalization. When items are meaningfully associated with their contexts, and these associations are made salient during recall, TOM for cross-boundary items is better (Wen & Egner, 2022). Similarly, if adjacent contexts are featurally overlapping, cross-boundary TOM also benefits (Gurguryan, Dutemple, & Sheldon, 2021). In Experiment 2, we examine the interaction between linguistic boundaries and temporal order in sentence memory, by investigating whether intervening boundaries enhance or reduce encoding interference between pairs of proximal, semantically similar nouns.

1.2 Temporal context in linguistic groupings

Under a view where contextual (dis)similarity mediates membership within a linguistic group, we might imagine that the contents of Box A of (3) are encoded as contextually more similar to one another than they are to those in Box B, because significant linguistic boundaries, like the prosodic and clausal boundary in (3), induce abrupt shifts in the temporal context. Throughout the remainder of the paper, we refer to the contents of each box as a distinct *encoding context*.

(3) [$_A$ The child fed the cat,]
[$_B$ the mom fed the dog,]
[$_C$ and the student fed the hamster.]

Intuitively, we might imagine that item-level proximity is not weighed heavily in the context of a larger structural representation. In this case, successful recall of the subject of context A (*child*), for example, may provide a good contextual retrieval cue to the remainder of context A, but the presence of a clause-final boundary may not extend this benefit to context B (e.g., from *cat* → *mom*). We term this the **Linguistic Sensitivity Hypothesis**. Alternatively, linguistic relationships between higher-order units may serve to create associative links between the last item of one context and the first item of the next, or between entire adjacent contexts, akin to the cross-list contiguity effect (Howard et al., 2008; Unsworth, 2008). In this case, successful recall of context B should provide a good retrieval cue to the subject of context C (cross-context item-level contiguity), or context C in its entirety (context-level contiguity). Either state of affairs would provide evidence for temporal contiguity in the sentential domain.

2. Experiment 1

Experiment 1 tested the Linguistic Sensitivity Hypothesis by having participants memorize multi-clausal lists followed by a "reactivation question" (RQ) task prior to sentence recall. By prompting retrieval of a previous contextual state associated with the answer to RQs like in Table 1, we sought to investigate whether a recall benefit would extend to the contents of an adjacent encoding context (i.e., for the RQ in C2Q, an additional benefit for *enemy* and/or *groom*).

The guest$_{\text{C1-Subj}}$ loved the voter$_{\text{C1-Obj}}$, the miner$_{\text{C2-Subj}}$ loved the guide$_{\text{C2-Obj}}$, the enemy$_{\text{C3-Subj}}$ loved the groom$_{\text{C3-Obj}}$, and the boxer$_{\text{C4-Subj}}$ loved the artist$_{\text{C4-Obj}}$.		
NoQ	——	
C2Q	Who did the miner$_{\text{C2-Subj}}$ love?	(A: the guide)
C3Q	Who did the enemy$_{\text{C3-Subj}}$ love?	(A: the groom)

Table 1: Example stimulus from Experiment 1.

Under Linguistic Sensitivity, we predicted no recall benefit beyond the targeted context. A contiguity-based account would predict an extended benefit for the subsequent subject or the subsequent clause as a whole in the C2Q condition. Due to the forward asymmetry, we expected that if contiguity did obtain, any recall benefit would be reduced for the contents of clause 2 in the C3Q condition.

2.1 Method

2.1.1 Participants 48 participants were recruited via Prolific (https://www.prolific.com/) to participate in the experiment. Participants were located within the USA, were at least 18 years old, had attained at least a high school diploma or equivalent, self-reported their first language as English, and self-reported no reading or attention differences. Each experimental session took 60-90 minutes, and participants were compensated $12/hour for their time.

2.1.2 Materials 48 items modeled after Table 1 contained four comma-separated transitive clauses with animate nouns. By hypothesis, we assume that each clause comprises a distinct encoding context. Prior to the final recall stage, we manipulated the presence of a free-response reactivation question, which prompted recall of the object of Context 2 or Context 3, or included no question.

Within one item, all nouns were distinct. We also controlled for a number of lexical characteristics within each item. All nouns were within 1-2 characters of one another in length; pairwise differences in their log word frequency, measured using SUBTLWF from the English Lexicon Project (Balota et al., 2007), had a threshold of 1; and pairwise similarity between nouns, measured using `word2vec` scores (Mikolov, Chen, Corrado, & Dean, 2013), had a threshold of .35. In addition, we verified that mean lexical frequency was comparable for each sentence position across all items.

2.1.3 Procedure The experiment was presented using PCIbex (Zehr & Schwarz, 2018). Each trial consisted of three phrases: an exposure phase, in which participants read the sentence using clause-by-clause self-paced reading with cumulative presentation; a reactivation phase, in which participants were prompted to respond to one of the reactivation questions (or no question); and finally a recall phase, in which participants were instructed to type the full sentence verbatim into four text boxes (one per clause). Participants were permitted to input a '?' for words they were unsure of.

2.2 Results

Participants' recall accuracy by position (i.e., the proportion of correct responses) is plotted in Figures 1a and 1b, which represent trials following correct and incorrect reactivation question responses, respectively; the baseline NoQ condition is identical across the two plots. Results were analyzed with Bayesian logistic mixed effects models using the `brms` package (Bürkner, 2017) in R (R Core Team, 2021), with random intercepts and slopes for participants and items. The RQ predictor was treatment coded, with the NoQ question as the reference group.

2.2.1 Correct reactivation question trials Model results are in Table 2 for trials following correct RQ responses. In these trials, we observe a pattern of results such that only the targeted context displays a recall benefit relative to the baseline NoQ condition. In the C2Q condition, recall accuracy is above baseline for the clause 2 subject (β = 1.2, CrI = [1.2, 2.5]), which was explicitly provided in the reactivation question, and the clause 2 object (β = 2.3, [1.6, 3.2]). This benefit did not extend to either the subject or the object of the following clause. Similarly in the C3Q condition, recall accuracy was above baseline for the clause 3 subject (β = 1.6, CrI = [1.1, 2.2]), provided by the reactivation question, and the clause 3 object (β = 1.6, CrI = [1.2, 2.1]), but this benefit did not extend to either of the nouns in the preceding (C2) or final (C4) clauses. Interestingly, we do not observe any evidence of primacy or recency benefits (higher recall accuracy for the initial and final clause of a sentence), although this is typical in sentence and word list memory experiments.

Q-Corr	β	95% CrI	β	95% CrI
		C2Q		C3Q
C2-Subj	**1.8**	(1.2,2.5)	0.1	(-0.3,0.5)
C2-Obj	**2.3**	(1.6,3.2)	0.1	(-0.3,0.5)
C3-Subj	-0.09	(-0.4,0.2)	**1.6**	(1.1,2.2)
C3-Obj	-0.2	(-0.5,0.08)	**1.6**	(1.2,2.1)

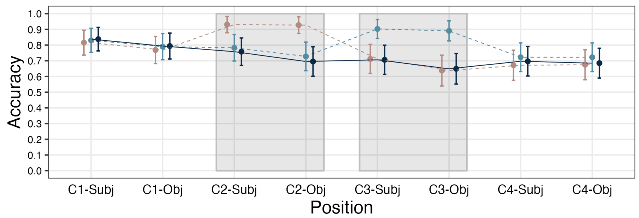
Table 2: Output from brms mixed effects modeling for trials following a correct response to the reactivation question.

Q-Incorr	β	95% CrI	β	95% CrI
		C2Q		C3Q
C2-Subj	-0.7	(-1.4,0.02)	**-1.2**	(-1.8,-0.7)
C2-Obj	**-3.2**	(-5.8,-1.8)	**-1.0**	(-1.6,-0.4)
C3-Subj	**-1.7**	(-2.3,-1.2)	-0.5	(-1.3,0.4)
C3-Obj	**-1.1**	(-1.7,-0.5)	**-3.4**	(-5.1,-2.3)

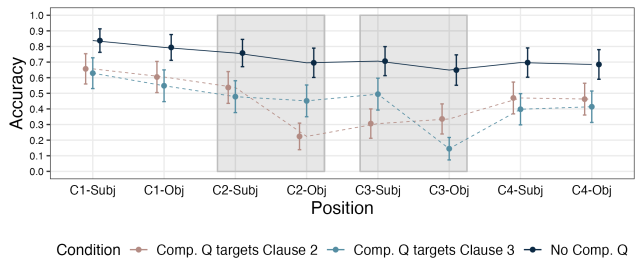
Table 3: Output from brms mixed effects modeling for trials following an incorrect response to the reactivation question.

2.2.2 Incorrect reactivation question trials Model results are in Table 3 for trials following incorrect RQ responses. Incorrect responses consisted of both intrusions and '?'; we delay a discussion of intrusion patterns until §2.3. Overall recall accuracy for all positions was lower than baseline when participants failed to respond correctly to the RQ, with the exception of the nouns explicitly provided in each RQ (the subjects of clauses 2 and 3).

With respect to the positions targeted by reactivation questions, we observe a different pattern in the incorrect trials. In the C2Q condition, incorrect recall of the clause 2 object also leads to a higher proportion of incorrect responses to both the clause 3 subject position (β = -1.7, CrI = [-2.3, -1.2]) and the clause 3 object position (β = -1.1, CrI = [-1.7, -0.5]). This suggests that failure to recall the targeted position degraded memory for the entire following clause. Similarly, failure to recall the clause 3 object did not affect recall of the preceding subject (given by the question), but was associated with lower recall accuracy for both the clause 2 subject (β = -1.2, CrI = [-1.8, -0.7]) and object (β = -1, CrI = [-1.6, -0.4]) positions.



(a) Correct RQ trials.



(b) Incorrect RQ trials.

Figure 1: Recall accuracy by position.

2.3 Discussion

The pattern of results in correct trials suggested that reactivation of a targeted encoding context did not lead to a cross-context contiguity effect for the following item or clause. In line with the Linguistic Sensitivity Hypothesis, this suggests that contextual reactivation is sensitive to the bounds of linguistic domains. While previous work on free recall of narratives found evidence of contiguity effects between events Raccah et al. (2024), these studies judged recall accuracy on the basis of thematic congruity per event, not verbatim recall as we were concerned with here.

The incorrect trials revealed a different pattern: recall failures were not constrained to the targeted context. Instead, failure to retrieve the target object during the RQ phase led to an overall decrement in recall accuracy for the full sentence. Indeed, participants reported that it was highly cognitively taxing to retain the original order of clauses in memory after attempting to respond to the reactivation question. This is consistent with previous work that evidences retrieval-induced forgetting, where accessing targeted items leads to inhibition of other items in memory (Anderson, 2003), especially when there is an intermediate "practice" phase between study and test (like our RQ manipulation). In addition, recall decrements were especially evident in contiguous (preceding and subsequent) clauses. This suggests that failure to retrieve an item has downstream consequences in serial order recall and provides additional indirect support for the Linguistic Sensitivity Hypothesis as well. If failure to retrieve an item also results in failure to retrieve its associated contextual state, this may disrupt the contextual representation of a sentence in memory, eliminating the benefit ordinarily bestowed by linguistic boundaries.

Rates of intrusion errors, in which a participant provided a different noun (from within the item, within the experiment, or outside of the experiment), were relatively low across the experiment, ranging from 7% to 10% of responses for the four noun positions in clauses 2 and 3. We expected this due to the fact that participants were allowed to write in '?' in cases of uncertainty. As such, we did not conduct a formal analysis of intrusions, but do note a few generalizations: the majority of intrusions were nouns from a contiguous clause (between 50% and 64% across the four positions), but not a contiguous position, meaning that subjects were more likely to intervene in subject positions, and objects to intervene in object positions. Future work should investigate this pattern in more detail, but our preliminary evidence suggests that intrusions are sensitive to contiguity at the level of the clause, and that information about syntactic position may outweigh temporal proximity between words in sentence contexts.

Overall, the results suggest that the encoding context boundaries in language-specific contexts block effects of temporal contiguity. This finding aligns with the idea that top-down control processes (e.g., instructions to focus on meaning-based associations) temper the emergence of temporal effects (Healey & Uitvlugt, 2019). Perhaps, then, item-context binding in sentences involves forming an association between a word and its place in a larger linguistic structure, rather than tracking temporality per se. If so, this leaves open the questions of what type(s) of linguistic boundaries – syntactic, prosodic (see Fodor, 2002), or discourse – drive shifts in context (Balachandran, 2024), and what levels of linguistic representation mediate item-context associations.

3. Experiment 2

Experiment 2, a sentence recognition memory study, uses temporal order memory to further probe the Linguistic Sensitivity Hypothesis. Here, we sought to determine how the position of a prosodic and clausal boundary, underlined in (4), affects encoding interference between semantically similar items, bolded in (4). (4a) exhibits a case where the similar nouns (*tourist* and *visitor*) belong to the same conjoined subject and therefore occupy the same linguistic group, whereas (4b) exhibits a case where these items span a linguistic boundary. We investigated whether the presence of a boundary drives a shift in context, which renders otherwise featurally similar items dissimilar in context, and thus improves memory for their relative ordering.

(4) a. Just as the witness and the vandal did, **the tourist** and **the visitor** hated the traffic.

b. Just as the witness and **the tourist** did, **the visitor** and the vandal hated the traffic.

Recall from §1.2 that two distinct patterns have been identified in the temporal order memory work on word lists: within-group temporal order is generally better remembered, but when there is a link between contexts, or when items are temporally disparate, temporal order for cross-boundary items typically prevails. In the context of (4), on the one hand, the greater contextual and linguistic coherence between *tourist* and *visitor* in (4a) afforded by the fact that they occupy the same linguistic unit may facilitate greater sensitivity to mismatches in the order of these nouns. We term this the **Within-Group Contiguity Hypothesis**: that when linguistic items occupy the same unit, temporal order memory should be enhanced because the contextual associations between contiguous items should be relatively more accessible than those of contextually isolated items. Thus, proximal items should display a TOM advantage, predicting greater accuracy in detecting changes to (4a) than (4b).

On the other hand, either temporal distance between boundary-spanning items, or the linguistic relationship between adjacent clausal units, may facilitate greater sensitivity to mismatches in the order of cross-boundary pairs compared to within-group ones. This is in line with the Linguistic Sensitivity Hypothesis, because the contextual distinctiveness between nouns spanning a boundary should better support order memory. This predicts greater accuracy in detecting order mismatches of similar nouns for (4b) compared to (4a).

3.1 Method

3.1.1 Participants 48 participants were recruited via Prolific (https://www.prolific.com/) using the same restrictions as in Experiment 1. Each experimental session took 45-60 minutes, and participants were compensated $12/hour.

3.1.2 Materials 48 item sets crossing the Position of similar items (WITHIN, ACROSS) with exposure-target Match (+/-MATCH) were presented to participants using a full-sentence recognition memory paradigm. An example item set is provided in Table 4, where conceptually similar items are bolded. Position manipulated whether similar nouns occupied the same subject (WITHIN), or were separated by an ellipsis site and clausal/prosodic boundary (ACROSS). Match manipulated whether the order of the critical (similar) nouns in the target sentence was the same as in the exposure (+MATCH), or whether it was reversed (-MATCH).

Position	Match
WITHIN	+MATCH: Just as the witness and the vandal did, **the tourist** and **the visitor** hated the traffic.
	-MATCH: Just as the witness and the vandal did, **the visitor** and **the tourist** hated the traffic.
ACROSS	+MATCH: Just as the witness and **the tourist** did, **the visitor** and the vandal hated the traffic.
	-MATCH: Just as the witness and **the visitor** did, **the tourist** and the vandal hated the traffic.

Table 4: Example item set for Experiment 2. Semantically similar nouns are bolded.

The subjects of the subordinate and main clauses always contained conjoined nouns. The syntactic structure of the items involved a sentence-initial subordinate clause (signaled by the connectives *whenever*, *just as*, *right after*, *before*, or *just like*) followed by a prosodic and clausal boundary that marked the site of cataphoric ellipsis. This was done to ensure identical argument structure across the initial and final clauses, and to control the distance (in words) between the critical nouns across all experimental conditions.

Item similarity was controlled using `word2vec` scores (Mikolov et al., 2013). Conceptually similar pairs had scores $>= 0.6$, and these items were controlled such that they were semantically dissimilar to the remaining nouns in the sentence (scores $<= 0.3$). Filler items (N = 48) contained similar connectives and syntactic structures, and varied the syntactic position of items that reversed positions between exposure-target pairs.

3.1.3 Procedure The experiment was presented using PCIbex (Zehr & Schwarz, 2018). Exposure sentences were displayed using word-by-word self-paced reading with cumulative presentation. After reading the exposure sentence, participants were instructed to solve a simple distractor arithmetic problem, involving addition or subtraction of randomly generated numbers between 0-100. Then, a target sentence was presented in its entirety. Participants were asked to determine whether the target sentence was identical to the exposure (responding to the question *Is this the same sentence you saw?*), providing combined recognition decisions and confidence ratings on a 6-point scale from -3 (*definitely different*) to 3 (*definitely the same*).

3.2 Results

Overall judgment accuracy was 80%, with 76% accuracy on experimental items and 84% accuracy on fillers. Math accuracy was also high (97%). The analysis excluded trials for which the response time for math problems exceeded +/- 2 standard deviations from the mean response time.

Results were analyzed using Unequal Variance Signal Detection Theory (SDT; Hautus, Macmillan, and Creelman, 2021) in order to determine participants' sensitivity to detecting changes in the order of the critical nouns. We used participants' judgments to calculate Hit-False Alarm pairs at each confidence level (+/-3, 2, 1) in order to construct empirical *Receiver Operating Characteristic* (ROC) curves, as in Figure 2. These ROCs are plotted with False Alarm Rates on the x-axis and Hit Rates on the y-axis. Each curve represents a constant sensitivity level, while each point represents a distinct confidence level (i.e., response bias). In the results, we report two mathematically related SDT measures of sensitivity: d_a, a measure of the perceptual distance between target (MATCH) and lure (MISMATCH) response distributions, and the Area Under the Curve (AUC) of the empirical ROC, which quantifies participants' divergence from chance performance (the diagonal in Figure 2).

Area Under the Curve values for the critical conditions were calculated using the `pROC` package (Robin et al., 2011) in R (R Core Team, 2021), using 2000 bootstrap replicates. Model results are presented in Table 5. AUC values suggest that participants performed well above chance (> 0.5) in the recognition task. In addition, the data reveal that sensitivity to ACROSS conditions (d_a = 1.4; AUC = 0.84) was significantly greater (D_{boot} = -3.3; p = 0.001) than sensitivity to WITHIN conditions (d_a = 1.1; AUC = 0.78).

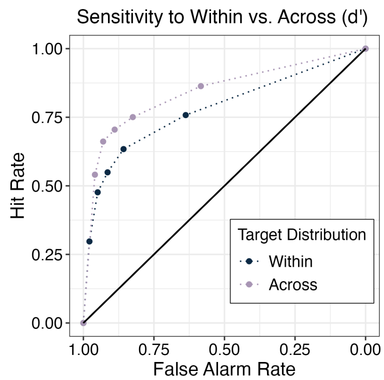


Figure 2: ROCs for WITHIN vs. ACROSS conditions of Experiment 2.

Table 5: pROC model results.

	d_a	AUC	2.5%	97.5%
WITHIN	1.1	0.78	0.75	0.81
ACROSS	1.4	0.84	0.82	0.87
	D_{boot} = -3.3		p = 0.001	

3.3 Discussion

Results corroborate the Linguistic Sensitivity Hypothesis. We found greater sensitivity to changes in temporal order when similar nouns spanned a linguistic boundary. This suggests that boundary-induced contextual shifts effectively separate linguistic content in memory, reducing encoding interference between proximal items. It also provides evidence that item-to-context associations are strengthened at linguistic boundaries (c.f. Heusser et al., 2018; Wen & Egner, 2022), consistent with the idea that privileged memory operations occur at these positions (Miller & Stine-Morrow, 1998).

While we take the results to be consistent with the predictions of the Linguistic Sensitivity Hypothesis, we consider two alternative explanations for the pattern observed. Experiment 2 utilized conjoined subjects. This means that the ACROSS-MISMATCH condition also involved a mismatch in the identity of the individuals in each complex subject across exposure and target pairs (e.g., *witness and tourist* vs. *witness and visitor*; see Table 4). This was not true of the WITHIN-MISMATCH condition, where the complex subject that was manipulated always contained the nouns *tourist* and *visitor*, although their relative ordering was reversed.

This raises the concern that higher sensitivity to ACROSS conditions may not be driven by contextual shifts alone, but may in part be due to accessing a conceptual memory representation where order information is irrelevant. That is, the subject may be accessed as the complex unit *visitor*$\oplus$*tourist*, as opposed to the individual nouns $visitor_i ... tourist_{i+3}$.

This re-evokes a question posed in §2.3: from the current study, we cannot determine whether participants are recruiting a level of representation in memory that contains information about the temporal order of nouns, or a more enriched linguistic representation (e.g., an event representation) that encodes information about the identity of participants involved in distinct events. This is also relevant to the ACROSS-MISMATCH condition, where there was a change in the pairs of participants involved in each event for some items (n = 38/48), like in (5). There, the first clause involves the uncle and the student in the exposure sentence, but the uncle and the teacher in the target. In contrast, the example from Table 4 contains the stative predicate *hate* (n = 10/48 items), which describes a shared emotion experienced by all participants. To control for this dimension of variability, Experiment 2 could be replicated using only stative predicates.

(5) EXPOSURE: Right after the uncle and the student did, the teacher and the captain ate the lasagna.
TARGET: Right after the uncle and the teacher did, the student and the captain ate the lasagna.

These additional factors (exposure-target differences in conjoined nouns and event structure) do not change our overall conclusions (see below), but we leave open the possibility that cross-boundary interference might be observable if these additional factors were controlled for.

In sum, we find that linguistic boundaries delimit contextual domains such that contiguity effects between proximal but across-group items are eliminated. This suggests that boundaries introduce a means of increasing perceived temporal and contextual distance between lingusitic items encoded in close proximity.

5. Conclusion

We've argued here that a boundary-sensitive context encoding mechanism serves to shift the contextual representation bound to linguistic items in sentences, thereby delineating linguistic domains in memory. Across two experiments, we show that this has the consequence of blocking cross-boundary contiguity effects in sentence recall, and enhancing cross-boundary temporal order memory in sentence recognition. We interpret these findings as evidence for item-context associations in sentential memory. Against this background, we highlight two avenues for future work: understanding which level(s) of linguistic representation mediate these associations, and which cues to linguistic boundaries are prioritized in shifting the temporal context. We suggest that the notion of shared encoding context should be further investigated as a potential source of similarity-based interference both in online sentence processing and offline sentence memory.

Acknowledgments

We'd like to acknowledge undergraduate research assistants Jenna Maas and Josh Lieberstein for assistance with stimulus creation for Experiment 1. We thank audiences at AMLaP 2023 and 2024, as well as members of the UC Santa Cruz s/lab reading group, for feedback and discussion.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445.

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.

Balachandran, L. (2024). *Memory knows its bounds: Encoding contexts in sentence comprehension*. Unpublished doctoral dissertation, UC Santa Cruz.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

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

Carroll, J. M., & Tanenhaus, M. K. (1978). Functional clauses and sentence segmentation. *Journal of Speech and Hearing Research*, *21*(4), 793–808.

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–183.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Proc. speechprosody 2002* (pp. 83–90).

Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, *9*(7), S5–S22.

Gurguryan, L., Dutemple, E., & Sheldon, S. (2021). Conceptual similarity alters the impact of context shifts on temporal memory. *Memory*, *29*(1), 11–20.

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide*. Routledge.

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720.

Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory & Cognition*, *47*, 719–737.

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1075.

Hitch, G. J. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 116–139.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Howard, M. W., Youker, T. E., & Venkatadass, V. S. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review*, *15*, 58–63.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 409–416.

Jarvella, R. J. (1979). Immediate memory and discourse processing. In *Psychology of Learning and Motivation* (Vol. 13, pp. 379–421). Elsevier.

Marslen-Wilson, W., & Tyler, L. K. (1976). Memory and levels of processing in a psycholinguistic context. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(2), 112.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, L., & Stine-Morrow, E. A. (1998). Aging and the effects of knowledge on on-line reading strategies. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *53*(4), P223–P233.

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications*, *13*(1), 622.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Raccah, O., Chen, P., Gureckis, T. M., Poeppel, D., & Vo, V. A. (2024). The "naturalistic free recall" dataset: four stories, hundreds of participants, and high-fidelity transcriptions. *Scientific Data*, *11*(1), 1–9.

Rich, S. (2024). *The features we use and the features we lose: encoding, maintenance, and retrieval*. Unpublished doctoral dissertation, UC Santa Cruz.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77.

Slowiaczek, M. L. (1981). *Prosodic units as language processing units*. Unpublished doctoral dissertation, University of Massachusetts Amherst.

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, *59*(2), 223–236.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Wagers, M. W. (2008). *The structure of memory meets mem-*

ory for structure in linguistic cognition. University of Maryland, College Park.
Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition*, *225*, 105145.
Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).