

A predictive memory activation model of sentence comprehension

Himanshu Yadav (himanshu@iitk.ac.in)

Indian Institute of Technology Kanpur, 208016 Kanpur, India

Bhaj Gobindh Raj (bhajgobindh24@iitk.ac.in)

Indian Institute of Technology Kanpur, 208016 Kanpur, India

Kanika Sachdeva (kanika25@iitk.ac.in)

Indian Institute of Technology Kanpur, 208016 Kanpur, India

Abstract

A well-established claim in sentence comprehension research is that the subject-verb dependency is resolved via a cue-based retrieval process: The subject noun is searched and retrieved from memory using retrieval cues, such as subject and animate. The cue-based retrieval process can be hindered by decay in the accessibility of the subject noun when the subject and the verb are kept away from each other, causing a processing difficulty at the verb, called the locality effect. The locality effect is robustly observed in reading studies from English, Spanish, Danish, and Russian. However, in verb-final languages like German and Hindi, an opposite effect is observed: When the distance between the subject and the verb increases, a speedup is observed in reading times at the verb. This effect has been called anti-locality and has been attributed to the strong predictability of the upcoming verb in rich case-marking, verb-final languages. We implement this idea as a computational model within the cue-based retrieval architecture. Our model assumes that pre-verbal nouns may preactivate the upcoming verb phrase in memory, and if the preactivation of the verb is strong enough, it can override the cost of retrieval, causing an anti-locality effect. Under the assumption that preactivation is stronger and more relied upon in verb-final languages, the model can account for locality as well as anti-locality effects across languages. We find strong evidence for the predictive activation model compared to the retrieval-only model, given data from five published studies on locality.

Keywords: Prediction; cue-based retrieval; working memory constraints; locality

Introduction

Comprehending a sentence requires the reader to correctly identify linguistically-related words in the sentence. For instance, to comprehend the sentence (1a), one must figure out *who* did the supervision and *what* was being supervised. This process of identifying relationships between pairs of words in a sentence is called dependency completion. Dependency completion requires the comprehender to temporarily maintain, retrieve, and integrate pieces of sentential information in real time. Therefore, working memory constraints have been argued to play a central role in sentence comprehension (Just & Carpenter, 1992; Gibson, 2000; Lewis & Vasishth, 2005). One of the well-established working memory-based models is the cue-based retrieval model (Lewis & Vasishth, 2005). The model assumes that dependency completion is driven by a content-addressable search in memory. For example, in the sentence (1a), when the verb *supervised* is encountered, a search is triggered in memory for the target subject noun using retrieval cues such as *subject* and *animate*;

the nouns that match these cues receive a certain amount of activation. A noun with the highest activation is retrieved to complete the subject-verb dependency. The retrieval cost is determined by the activation of the target noun to be retrieved from memory.

1. (a) **The nurse** who was from the clinic **supervised** the patient.
(b) **The nurse supervised** the patient.

A key prediction of the cue-based retrieval model is the locality effect: When the linear distance between the subject and the verb increases, the activation of the subject noun decays over time; consequently, it becomes difficult to retrieve the subject noun at the verb. Thus, the model predicts a slowdown in processing times at the verb *supervised* in the long subject-verb distance condition (1a) compared to the short distance condition (1b). The locality effect is consistently observed in reading studies on English (Grodner & Gibson, 2005; Bartek, Lewis, Vasishth, & Smith, 2011), Russian (Levy, Fedorenko, & Gibson, 2013), Spanish (Nícenboim, Logachev, Gattei, & Vasishth, 2016), and Danish (Balling & Kizach, 2017). These data provide strong evidence for the cue-based retrieval model and other working memory-based models of sentence comprehension (Gibson, 2000; Lewis, 1996).

However, when it comes to verb-final languages such as Hindi, German, and Japanese, there is no clear evidence for the locality effect (Konieczny, 2000; Vasishth & Lewis, 2006; Husain, Vasishth, & Srinivasan, 2014; Nakatani & Gibson, 2008). Surprisingly, on similar constructions in verb-final languages, an opposite effect is observed: A speedup in reading times at the verb in the long dependency distance condition compared to the short distance condition. This effect has been called the anti-locality effect. The anti-locality effect observed in verb-final languages poses a serious challenge to existing working memory-based models of sentence comprehension. If working memory constraints are central to dependency completion, why would they not affect comprehension in verb-final languages? It is reasonable to assume that working memory (WM) does play a central role during comprehension, but the existing WM-based models may not be sufficient to account for data across languages. Thus, existing data demand a new model of sentence comprehension

that can explain processing behavior in verb-final as well as verb-medial languages.

The anti-locality effect in verb-final languages has been attributed to the strong predictability of the upcoming verb in these languages (Levy et al., 2013; Husain et al., 2014). Arguably, in verb-final, rich case-marking languages, the upcoming verb phrase is strongly predicted due to rich informative case markers and extensive exposure to head-final structures (Vasishth, Suckow, Lewis, & Kern, 2010). Consequently, intervening material between arguments and the verb can increase the predictability of the verb, which can facilitate processing at the verb. Such prediction facilitation, if strong enough, can override the retrieval cost posed by long argument-verb distance. However, this predictability proposal has never been systematically implemented and tested in a computational model.

Given the strong implication of predictive processing in verb-final languages and robust evidence for working memory constraints in verb-medial languages, it is reasonable to assume that both prediction and working memory mechanisms together influence sentence comprehension. How do we integrate working memory and prediction in a single model that can explain data from verb-final as well as verb-medial languages?

We propose the preactivation retrieval model, which integrates a predictive memory activation mechanism and the cue-based retrieval process. The model assumes that intervening material between the arguments and the verb preactivates the upcoming verb phrase in memory, which causes a facilitation in processing the verb. This prediction facilitation may override the cost of retrieving the arguments from memory. The model further assumes that predictability is stronger and more relied upon in verb-final, rich case-marking languages compared to verb-medial languages. We show that the model can capture locality effects observed in verb-medial languages as well as anti-locality effects observed in verb-final languages.

We evaluate the preactivation retrieval model against the cue-based retrieval model using data from five published studies on (anti-)locality effects. We find strong evidence in favor of the preactivation retrieval model under different prior assumptions about how strongly the upcoming verb is predicted in verb-final languages.

The paper is organized as follows. We first present the assumptions of the preactivation retrieval model. Next, we generate model predictions for locality studies on different languages. Then, we quantify evidence for the model using data from five published studies. We then discuss the broader implications of the work and conclude.

The preactivation retrieval model of sentence comprehension

We attempt to computationally implement and test the *predictability proposal*: In verb-final, rich case-marking languages, the upcoming verb phrase is strongly predicted, and

intervening material increases the predictability of the verb, which may facilitate processing at the verb. We implement the predictability idea within the cue-based retrieval framework of Lewis and Vasishth (2005) such that the upcoming verb phrase is preactivated in memory through spreading activation from pre-verbal nouns (nouns that appear before the verb).¹ We call it the preactivation retrieval model.

The model assumes that the activation of a pre-verbal noun i stored in memory decays with time, such that the activation of the item i at time t is given by:

$$A_{i,t} = A_{i,0}e^{-t} \quad (1)$$

where $A_{i,0}$ is the activation of the noun phrase at the time t_0 when it was first accessed for the current sentence from the declarative memory. The initial activation $A_{i,0}$ can be given by

$$A_{i,0} = B_i + WS_i \quad (2)$$

where B_i is the baseline activation of the noun i determined by its past retrievals, S_i is the associative strength between a noun's visual cues and its memory trace, and W is the weighting of the noun's visual cues.

The model further assumes that a cue-based retrieval process triggers at the verb to search for its arguments. Each noun phrase that matches a retrieval cue receives a certain amount of activation. The total activation of a noun phrase i at time t can be given by

$$A_{i,t} = (B_i + WS_i)e^{-t} + \sum_{j=1}^n W_j S_{ji} + \epsilon_i \quad (3)$$

where the $(B_i + WS_i)e^{-t}$ term represents the residual activation of the noun i after decay at time t and the $\sum_{j=1}^n W_j S_{ji}$ term means that the noun phrase i receives activation from all matching cues j depending on the associative strength S_{ji} between the cue j and the noun i , and the cue's weight W_j . The term ϵ_i is activation noise, such that $\epsilon_i \sim Normal(0, \sigma)$.

The activation of a chunk can also increase due to preactivations. The pre-activation of the upcoming verb phrase k in memory increases due to spreading activation from each pre-verbal noun that associates with the verb k :

$$A_{k,t} = ((B_k + G_p W_1 S_{1,k})e^{-(t-t_1)} + G_p W_2 S_{2,k})e^{-(t-t_2)} + \dots \quad (4)$$

The pre-activation level of verb phrase k at time t due to a pre-verbal noun m can be given by:

$$A_{k,t} = (B_{k,t_m} + G_p W_m \cdot S_{m,k})e^{-(t-t_m)} + \epsilon_k \quad (5)$$

where B_{k,t_m} is the residual pre-activation of the verb phrase k at time t_m when the noun m is encountered; $G_p W_m S_{m,k}$ is the amount of pre-activation spread from the preverbal noun

¹For simplicity, we use the term pre-verbal noun, but the assumption applies to any kind of material, nouns, adverbs, or others, that appear before the verb.

m to the verb phrase k . G_p is the source preactivation parameter; $(t - t_m)$ is the time elapsed between the noun m and the verb k , which appears at t . The associative strength $S_{m,k}$ between noun m and verb k is determined by their history of co-retrievals, which can be captured using pointwise mutual information between the two; the value of this parameter can be estimated from a corpus. The W_m term represents weighting for the noun's features for prediction; certain nouns with features like an Ergative case marker might be more robust and reliable at predicting the upcoming verb.

The model further assumes that prediction is stronger and more relied upon in rich case-marking, verb-final languages compared to verb-medial languages. We use the source preactivation parameter G_p to operationalize prediction strength. The model assumes that the value of parameter G_p is higher in verb-final languages compared to verb-medial languages. We have made this simplified assumption here about the prediction strength parameter G_p in order to model cross-linguistic differences in the locality effect. However, this parameter needs to be measured independently using a task that can estimate the extent of predictability of the upcoming material. We envisage that prediction strength can differ across individuals within and across language populations and will have to be measured experimentally. However, in this paper, we make a reasonable assumption based on existing speculations in the literature that verb-final language speakers have stronger predictability than verb-medial language speakers.

The reading time at the verb is the summation of time taken to lexically access and encode the verb chunk k and the time taken to retrieve its argument noun i :

$$T = Fe^{-A_{k,t}} + Fe^{-A_{i,t}} \quad (6)$$

where the latency factor F is a scaling parameter and reflects the overall reading speed of the comprehender. The latency factor is a free parameter in the model.

Model predictions

To assess whether the model can account for (anti-)locality effects observed in verb-medial as well as verb-final languages, we generate model predictions for five published studies on locality: three classic locality experiments from English (gg05), German (koin2000), and Hindi (vl06) and two *predictability x locality* experiments from Persian (safavi) and Hindi (Husain). We assess whether the model can qualitatively capture the locality effect in English as well as the anti-locality effect observed in verb-final languages, such as German and Hindi, which is a long-standing empirical challenge.

Model predictions for the English experiment (Grodner and Gibson 2005, Experiment 2)

For the locality experiment in English, Grodner and Gibson (2005) used the embedded object relative clause (RC) construction shown in Example sentences (1a-1b). In sentence (1a), the dependency between the subject and the RC verb

is interrupted by another relative clause, which increases the linear distance between the subject and the verb *supervised*. The working memory-based models predict that resolving the subject-verb dependency should be difficult in sentence (1a) compared to (1b) due to a longer subject-verb dependency distance. The difficulty in resolving this RC subject-verb dependency should reflect in the reading times at the RC verb.

1. (a) The administrator who the nurse who was from the clinic **supervised** ...
- (b) The administrator who the nurse **supervised** ...

We generate prior predictions from the preactivation retrieval model for this English experiment. To calculate how strongly the upcoming RC verb is preactivated in memory, we need to specify the associative strength between the intervening items (e.g., *who, was, from the clinic*) and the RC verb. For simplicity, here, we assume some reasonable real values between 0 and 1 for the associative strength between intervening items and the verb, but these values can be estimated from the corpus based on their pointwise mutual information and semantic similarity.

The model has two free parameters, the latency factor and the source preactivation parameter. On the latency factor F , the prior is chosen such that it does not allow unreasonably fast and unreasonably slow reading times. We use a truncated normal prior with lower bound of 0.05 ($lb=0.05$),

$$F \sim Normal_{lb=0.05,ub=1}(0.45, 0.05) \quad (7)$$

We make the following prior assumption for the source (goal) preactivation parameter G_p for English, which reflects how strongly and reliably the prediction strategy is used for processing. Assuming preactivation strength is weaker in verb-medial languages, we allow small values for source preactivation in English, between 0 and 0.2.

$$G_p \sim Normal_{lb=0,ub=1}(0.1, 0.05) \quad (8)$$

Figure 1 shows the prior prediction of the model for the Groden and Gibson (2005) Experiment 2. The x-axis shows the extent of locality effect predicted by the model: Difference in reading times between condition (1a) and (1b). We observe that the model predicts a locality effect for English: a slowdown in long distance condition compared to short distance condition.

Model predictions for the German experiment (Konieczny 2000, Experiment 2)

Konieczny (2000) tested the dependency distance effect in German using embedded vs. extraposed relative clause (RC) constructions shown in Example (2a-2b). To understand these sentences, the comprehender needs to complete the dependency between the subject *das Buch* and the verb *hingelegt*. In sentence (2a), the subject-verb dependency is intervened by an embedded relative clause (RC) that increases the subject-verb distance in contrast to (2b). The actual experiment used

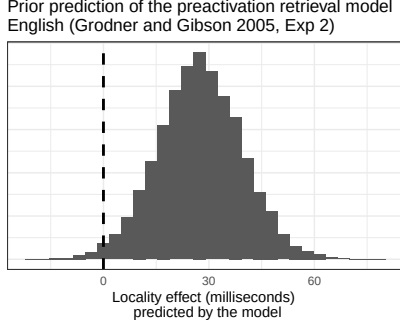


Figure 1: Model predictions for the English experiment.

many other manipulations such as relative clause length and matrix clause length, but we focus on embedded vs. extra-posed RC manipulation.

2. (a) Er hat das Buch, das Lisa gestern gekauft hatte, **hingelegt**.
 (b) Er hat das Buch **hingelegt**, das Lisa gestern gekauft hatte.

We generate predictions from the preactivation retrieval model for the German experiment under the assumption that predictability is stronger and more relied upon in German compared to a verb-medial language like English. Since we operationalize prediction strength as the source preactivation parameter, we assume that source preactivation is higher in German. We allow the source preactivation G_p values to be higher than English, somewhere in the range of 0.3 to 0.5. We choose the following prior on G_p for German,

$$G_p \sim \text{Normal}_{lb=0,ub=1}(0.4, 0.05) \quad (9)$$

For latency factor, we choose the same prior as for English.

Figure 2 (left panel) shows the prior predictions of the model for the German experiment (Konieczny, 2000). We observe that the model predicts an anti-locality effect: A speedup at the verb in the long-distance condition compared to the short-distance condition.

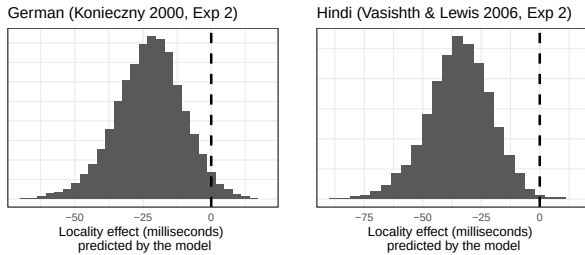


Figure 2: Prior predictions of the preactivation retrieval model for the German and the Hindi experiment.

Model predictions for the Hindi experiment (Vasishth and Lewis 2006)

Vasishth and Lewis (2006) used the subject relative and object relative clause constructions in their Experiment 2. For exposition, we only show object relative clause (RC) construction

here; see sentences (3a) and (3b). In sentences (3a), the dependency between the relative pronoun (RelPro) subject *who* and the RC verb *saw* is intervened by two noun phrases and a verb phrase which modify the RC verb *saw*, while in condition (3b), there is only one intervener noun phrase.

3. (a) that boy who-ERG that paper-ACC table-GEN behind lying **saw** very inquisitive was.
 (b) that boy who-ERG that paper-ACC **saw** very inquisitive was.

We generate predictions from the preactivation retrieval model for the Hindi experiment in the same way as we did for the German experiment, using the same prior assumptions about the parameters of interest. Similar to German, Hindi is also predominantly a verb-final, rich case-marking language; therefore, we assume the same prior on the source preactivation parameter as for German.

Figure 2 (right panel) shows the prior predictions of the model for the Hindi Experiment (Vasishth and Lewis (2006), Experiment 2). We observe that the model predicts a strong anti-locality effect. Interestingly, the effect is predicted to be stronger than the German experiment. This is because, in the Vasishth and Lewis (2006) design, the intervening chunks directly modify the upcoming RC verb, causing stronger preactivation of the upcoming verb phrase; while in German, intervening chunks modify the subject noun, which indirectly preactivates the verb causing weaker prediction facilitation.

Model predictions for the Expectation x Locality experiments (Husain et al. 2014; Safavi et al. 2016)

Husain et al. (2014) and Safavi, Husain, and Vasishth (2016) manipulated dependency distance as well as predictability in a single experiment using complex/simple predicate constructions in Hindi and Persian, respectively. They used sentences like (4a) where a noun-verb complex predicate *khayal rakhna* ('care keep') is used.

4. (a) **Complex predicate:**
 mother-ERG child-ACC asked that she.NOM her **care** without any negligence properly **keep**
 (b) **Simple predicate:**
 mother-ERG child-ACC asked that she.NOM her **guitar** without any negligence properly **keep**

These complex predicates consist of a noun that strictly requires a light verb. In the complex predicate sentence (4a), the noun *care* triggers a strong expectation of the light verb *keep*, while in the sentence (4b), the noun *guitar* does not generate a strong expectation of the verb *keep*, because here *keep* is used as a simple predicate. They tested whether the strong predictability of the upcoming verb attenuates the locality effect.

We generate predictions from the preactivation retrieval model for the complex and simple predicate constructions used in the Hindi and Persian experiment, using the same prior assumptions as for the Hindi and German experiment presented in the previous section.

Figure 3 shows the prior predictions of the model for the Husain et al. (2014) and Safavi et al. (2016) experiments. The

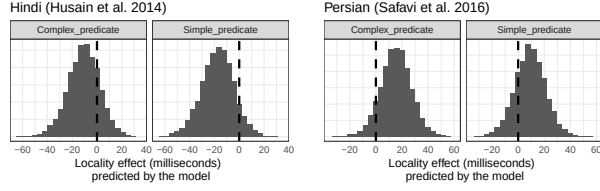


Figure 3: Model predictions for the Expectation x Locality experiments in Hindi and Persian.

model predicts an anti-locality effect for complex and simple predicate constructions in Hindi, and a weak, close-to-zero locality effect in Persian. The anti-locality is not predicted for the Persian experiment because the intervening items between the noun and verb do not directly modify the upcoming verb, causing weak preactivation of the verb.

Model predictions vs. observed data

We assess whether the preactivation-retrieval model can qualitatively account for (anti-)locality effects observed in the above five studies. We compare model predictions against the observed effect estimates from the five studies.

Figure 4 shows a comparison of model predictions and observed effects. The data from all five studies are qualitatively consistent with the preactivation retrieval model: the model correctly predicts the anti-locality effect in two Hindi studies, anti-locality in German, locality effect in English, and a weak locality effect in the Persian experiment. However, the quantitative ranges of model prediction and observed effect do not align in many cases, which could be due to noisy estimates from observed data.

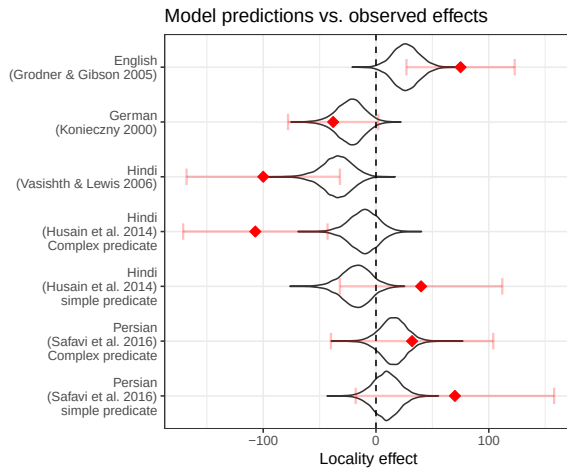


Figure 4: The model predictions for each study are shown as violin plots; the observed effects are shown as red rhombi along with their 95% confidence interval as error bars.

Model evaluation

We quantify evidence for the preactivation retrieval model compared to the cue-based retrieval model using Bayes factors. When comparing two models, say $\mathcal{M}1$ and $\mathcal{M}2$, the Bayes factor is computed by taking the ratio of the marginal likelihoods of the two models, which represents the extent to which Model $\mathcal{M}1$ is more likely than $\mathcal{M}2$ given the data. Following Jeffreys (1939/1998), we interpret a Bayes factor larger than 10 as strong evidence for a model over the other. To compute Bayes factors, we estimated marginal likelihoods for each model using Approximate Bayesian Computation (Palestro, Sederberg, Osth, Van Zandt, & Turner, 2018).

Figure 5 shows Bayes factors in favor of the preactivation retrieval model against the cue-based retrieval model. We find strong evidence (Bayes factors > 10) for the preactivation retrieval model under different prior assumptions on the source preactivation parameter. The results indicate that the model assuming strong predictability in verb-final languages can account for locality as well as anti-locality effects observed across languages.

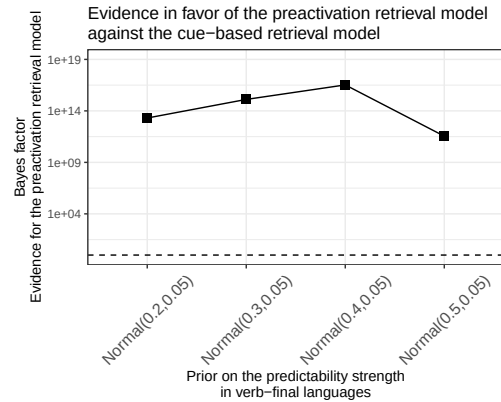


Figure 5: The Bayes factor values on the y-axis show evidence in favor of the preactivation retrieval model against the cue-based retrieval model, under different prior assumptions on source preactivation in verb-final languages.

Discussion

Can a predictive memory activation mechanism explain the (anti-)locality effects observed in verb-final as well as verb-medial languages? To address this question, we implement the preactivation retrieval model that integrates prediction and working memory constraints in a single memory activation framework. The model assumes that when pre-verbal nouns are encoded, they preactivate the upcoming verb phrase in memory, which facilitates processing at the verb. The model further assumes that predictability is stronger and more relied upon in verb-final, rich case-marking languages compared to verb-medial languages. We show that the model can account for both locality and anti-locality effects as emerging from cross-linguistic variation in prediction strength.

The model was evaluated using data from three classic locality experiments on English, German, and Hindi, and two Expectation x Locality experiments on Hindi and Persian. We find that: (1) the model can qualitatively explain the (anti-)locality effects observed across typologically different languages; and (2) there is strong evidence in favor of the preactivation-retrieval model compared to the existing cue-based retrieval model.

The results imply that predictability plays a central role in dependency completion along with working memory constraints. A model that integrates top-down prediction and bottom-up retrieval process can explain sentence comprehension behavior in verb-final as well as verb-medial languages. Another key implication is that German and Hindi speakers differ from English speakers in how strongly they predict the upcoming verb, such that verb-final language speakers exhibit stronger predictability of the upcoming sentence material.

These implications are important for theories of sentence processing for two reasons. First, we show that a dynamic memory activation model needs to incorporate a predictive memory mechanism to explain processing behavior across typologically different languages. Second, our findings indicate that there are differences in processing strategy across languages such that the comprehension system adapts to language-specific properties.

Does the human sentence comprehension system adapt to language-specific properties? Our proposal is consistent with the broader *adaptability hypothesis*: The architectural constraints on processing adapt to language-specific experience (Christiansen & MacDonald, 2009). In languages with rich case-marking and dominant verb-final typology, the predictive strategy might be more useful and robust for sentence comprehension because case-markers give away crucial information about the argument structure of the verb phrase. It is reasonable to assume that verb-final language comprehenders adapt to a predictive processing strategy, allocating more resources for preactive activation spread to upcoming sentence material.

A weakness of the current modeling work is that we did not compare the preactivation retrieval model directly against the lossy-context surprisal model (Futrell et al., 2020), which integrates a working memory component within an expectation-based model of sentence processing. The lossy-context surprisal model was recently tested by Hennert et al. (2026), and it failed to account for (anti-)locality effects observed in Russian, Hindi, and Persian. It seems that the preactivation retrieval model would perform qualitatively better than the lossy-context surprisal model. However, we need to verify this using a quantitative comparison of the two models using all existing datasets from typologically different languages. We plan to take this up in future work.

The current work reveals new insights into the underlying cognitive process that underlies real-time dependency resolution: The dependency completion between a verb and its arguments is driven by predictive activation of the verb and

cue-based retrieval of the arguments. To our knowledge, this is the first large-scale quantitative investigation of a predictive memory activation model of sentence comprehension. Our work contributes to understanding how top-down prediction and working memory interact and determine real-time sentence comprehension in humans.

Acknowledgments

We thank the three anonymous reviewers for their detailed and useful feedback. We received funding from the Department of Science and Technology (DST) India under the Cognitive Science Research Initiative (CSRI), Program ID: DST/CSRI/2024/354.

References

- Balling, L. W., & Kizach, J. (2017). Effects of surprisal and locality on danish sentence processing: An eye-tracking investigation. *Journal of psycholinguistic research*, 46(5), 1119–1136.
- 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.
- Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, 59, 126–161.
- 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.
- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O’Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive science*, 29(2), 261–290.
- Hennert, J., Paape, D., Yadav, H., & Vasishth, S. (2026). Can lossy-context surprisal capture both locality and anti-locality effects? a model evaluation using russian, hindi, and persian data.
- Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PLoS ONE*, 9(7), 1–14.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of psycholinguistic research*, 29(6), 627–645.
- Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of russian relative clauses. *Journal of memory and language*, 69(4), 461–495.
- Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of psycholinguistic research*, 25(1), 93–115.

- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics*.
- Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in psychology*, 7, 280.
- Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. Springer.
- Safavi, M. S., Husain, S., & Vasishth, S. (2016). Dependency resolution difficulty increases with distance in Persian separable complex predicates: Implications for expectation and memory-based accounts. *Frontiers in Psychology*, 7.
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4), 767–794.
- Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures. *Language and Cognitive Processes*, 25(4), 533–567.