Predictive memory activation explains locality and anti-locality effects in Hindi

Bhaj Gobindh Raj (bhajgobind24@iitk.ac.in)[1], Himanshu Yadav (himanshu@iitk.ac.in)[2]
[1]Department of Cognitive Science, IIT Kanpur, [2]Department of Cognitive Science, IIT Kanpur

Comprehending a sentence requires temporary storage and maintenance of words in memory to establish relations between them. Working memory constraints are argued to be central to sentence comprehension [3, 6]. A key prediction of working memory-based theories is the *locality effect*: The processing difficulty at a word (e.g. a verb) increases as its codependent (e.g. the subject) is placed further apart in the sentence [3]. The locality effect is consistently observed in reading studies in languages like English [2, 4], Russian [5], Spanish [7], and Danish [1]. However, data from verb-final languages like Hindi and German pose a serious empirical challenge to existing working memory-based theories: the *anti-locality effect*. When the distance between a verb and its argument increases, it causes a speedup in reading time at the verb. The observed anti-locality in verb-final languages is arguably caused by the strong predictability of the upcoming verb phrase due to rich-case marking and extensive exposure to head-final structures. The strong predictability of the verb may cause a facilitation in processing, overriding certain memory-based constraints. Given the strong implications of predictability in verb-final languages and the well-established role of working memory, the cross-linguistic data demand a new model of sentence comprehension that can integrate working-memory constraints and prediction in a single set of assumptions.

Here, we computationally implement one such model that can account for both locality and anti-locality effects observed across languages. We call it *the predictive memory activation model*. The model draws its assumptions from cue-based retrieval theory [6] and the prediction reactivation proposal [8]. The model assumes that intervening items between an argument and the verb preactivates the upcoming verb phrase in memory, which causes a facilitation in processing when the verb is encountered. This prediction facilitation may override the cost of retrieval at the verb. The model further assumes that intervening items that directly syntactically modify the upcoming verb cause stronger preactivation of the verb phrase compared to words that do not directly modify the verb.

The model predicts that in a verb-final language, the anti-locality effect at the verb will be stronger when the argument-verb dependency is intervened by direct syntactic modifiers of the verb, compared to when it is not intervened by direct modifiers of the verb. To test this model prediction, we conduct two self-paced reading experiments in Hindi, where the distance between a relative pronoun (RelPro) and a relative clause (RC) verb is manipulated. In Experiment 1, the interveners between the RelPro and the RC verb are syntactic modifiers of the RC verb. In Experiment 2, the intervening material modifies a noun within the relative clause and not the RC verb directly. The model predicts a stronger anti-locality effect in Experiment 1, compared to Experiment 2; see Figure 1.

We estimated the effect of the dependency distance on reading times at the RC verb in both experiments using Bayesian regression models. The estimated anti-locality effects are consistent with the model predictions: Experiment 1 showed stronger anti-locality (95% credible interval [-42,5]) as compared to Experiment 2 (95% credible interval [-14,19]). We found strong evidence for the predictive memory activation model compared to the cue-based retrieval model (Bayes factors $>$10). These findings demonstrate that (anti-)locality effects in a rich case-marking, verb-final language can be explained using a predictive reactivation component within a cue-based retrieval architecture.

References

[1] L. W. Balling and J. Kizach. Effects of surprisal and locality on danish sentence processing: An eye-tracking investigation. *Journal of Psycholinguistic Research*, 46(5):1119–1136, 2017.

[2] B. Bartek, R. L. Lewis, S. Vasishth, and M. R. Smith. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5):1178–1198, 2011.

[3] E. Gibson. Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O'Neil, editors, *Image, Language, Brain*, pages 95–126. MIT Press, Cambridge, MA, 2000.

[4] D. Grodner and E. Gibson. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2):261–290, 2005.

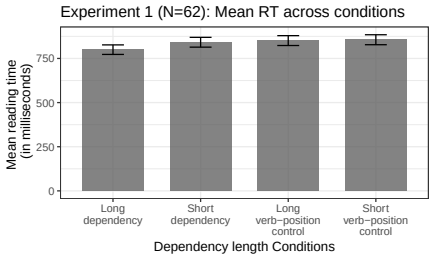
[5] R. Levy, E. Fedorenko, and E. Gibson. The syntactic complexity of russian relative clauses. *Journal of Memory and Language*, 69(4):461–495, 2013.

[6] R. L. Lewis and S. Vasishth. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419, 2005.

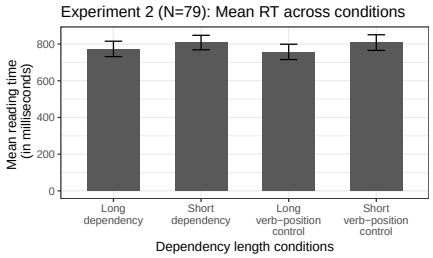
[7] B. Nicenboim, P. Logačev, C. Gattei, and S. Vasishth. When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology*, 7:280, 2016.

[8] S. Vasishth and R. L. Lewis. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4):767–794, 2006.

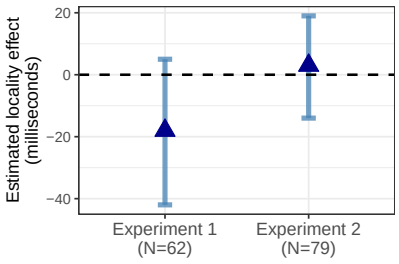
Figures



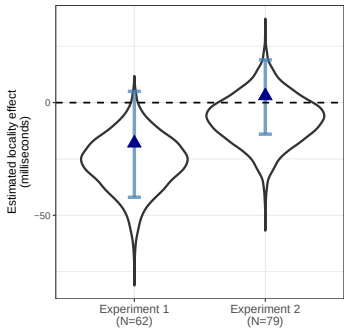
(a) Average reading times in Experiment 1, where interveners directly modify the RC verb.



(b) Average reading times in Experiment 2, where the interveners do not directly modify the verb.



(c) The 95% credible interval of the estimated locality effect (long distance condition minus short distance condition) in each experiment is shown as blue error bars.



(d) The prior predictions of the predictive memory activation model is shown as violin plots for each experiment and the estimated (anti-)locality effect from each experiment is shown as blue error bars.

Figure 1: The average reading times at the RC verb for Experiment 1 and Experiment 2; see (a) and (b). The observed (anti-)locality effect in each experiment; see (c). The prior predictions of the model compared against the observed anti-locality effect for both the experiments; see (d).

Sample Stimuli

Experiment 1 (Interveners modify the verb)

(1) a. *ek ladka* ***jo*** *kal school-mei sabke saamne gaana* ***gaa-rha-tha*** *...*
one boy who yesterday school-LOC everyone-GEN front song sing-PROG-was ...
'A boy who was singing a song in front of everyone at school yesterday.'

b. *ek ladka* ***jo*** *gaana* ***gaa-rha-tha*** *...*
one boy who song sing-PROG-was ...
'A boy who was singing a song.'

Experiment 2 (Interveners do not modify the RC verb but modify a noun)

(2) a. *ek ladka* ***jisko*** *kal school-mei music sikhaane-wale ek-sakht-teacher-ne office-mei*
a boy whom yesterday school-LOC music teach-NFV-PART a-strict-teacher-ERG office-LOC
bulaaya-tha *...*
call-PERF-was ...
'A boy whom a strict teacher who teaches music at school called to the office yesterday.'

b. *ek ladka* ***jisko*** *kal teacher-ne office-mei* ***bulaaya-tha*** *...*
one boy whom yesterday teacher-ERG office-LOC call-PERF-was ...
'A boy whom a teacher called to the office yesterday.'

Supplementary information (The predictive memory activation model)

The model assumes that the activation level of any item in memory decays with time.
At time t, the activation of a pre-verbal noun i is given by

$$A_{i,t} = A_{i,0}e^{-t} \tag{1}$$

where $A_{i,0}$ is the initial activation the noun phrase when it was first accessed from declarative memory for the current sentence.
When the verb is encountered, all nouns which match the retrievals cues at the verb receive a certain amount of activation modulated by the association strength S_{ji} between the cue j and the noun i, and the cue's weight W_j. The total activation of the noun phrase is given by

$$A_{i,t} = (B_i + WS_i)e^{-t} + \sum_{j=1}^{n} W_j S_{ji} + \epsilon_i \tag{2}$$

where $(B_i + WS_i)e^{-t}$ indicates the residual activation of noun i after decay at time t.

Each intervening item m between an argument and the verb spreads a certain amount of pre-activation to the verb phrase in memory, given by $W_m S_{m,k}$. The total activation of the verb phrase k at time t is given by

$$A_{k,t} = \left((B_k + W_1 S_{1,k})\, e^{-(t_2 - t_1)} + W_2 S_{2,k} \right) e^{-(t_3 - t_2)} + \cdots \tag{3}$$

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

$$T_{k,t} = Fe^{-A_{k,t}} + Fe^{-A_{i,t}} \tag{4}$$

where the latency factor F is the scaling parameter and reflects the overall reading speed.