

Slow readers capitalize on predictive processing

Himanshu Yadav (himanshu@iitk.ac.in)

Indian Institute of Technology Kanpur, 208016 Kanpur, India

Samar Husain (samar@hss.iitd.ac.in)

Indian Institute of Technology Delhi, 110016 Delhi, India

Abstract

Working memory constraints have been shown to be pervasive in human sentence comprehension. Working memory-based theories of sentence comprehension predict the locality effect: when the distance between two linguistically related words (e.g., a subject and a verb) increases, the processing difficulty increases. While the locality effect has been consistently observed in English, there is no sufficient evidence for locality in verb-final languages such as Hindi and German. The absence of locality in verb-final languages is attributed to robust prediction of the verb, which may override certain working memory demands. We take an individual differences approach to study how working memory and prediction interact. In four self-paced reading experiments in Hindi, we consistently find an anti-locality effect: Speedup at the verb when the argument-verb distance is increased. However, individuals differ in their extent of locality effect. Why do certain individuals in a language population show stronger anti-locality effects than others? We test two competing hypotheses: (i) Varying prediction-strength hypothesis: Individuals with stronger predictability of the verb would show larger anti-locality effects and faster reader times, on average; (b) Varying lexical-access hypothesis: Individuals with poorer lexical accessibility are, on average, slower in reading and may capitalize more on predictability, showing larger anti-locality effects. The two hypotheses make opposite predictions about the correlation between the individual-level locality effect and reading speed. We find strong evidence that slow readers exhibit stronger anti-locality effects. The results support the varying lexical access hypothesis, implying that individuals with slowed lexical access compensate with predictive processing.

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

Introduction

For successful sentence comprehension, one must identify and link together related pairs of words, e.g., the subject and the verb, in a sentence. This process is called dependency resolution. Humans possess a remarkable ability to resolve dependencies in real time. What are the cognitive constraints that underlie real-time dependency resolution? A well-established proposal is working memory constraints: Since dependency resolution between two chunks requires temporary maintenance, retrieval, and integration of pieces of information in real time, working memory has been argued to play a critical role in sentence comprehension (Gibson, 1998; Just & Carpenter, 1992; Lewis, 1996; Gibson, 2000; Lewis & Vasishth, 2005).

A key prediction of working memory-based theories (Gibson, 2000; Lewis & Vasishth, 2005) is the *locality effect*: As the distance between two linguistically related words

(e.g., a subject and a verb) increases, it becomes costlier to resolve the dependency between them. For instance, a well-known account based on working memory constraints—the dependency locality theory (Gibson, 2000)—assumes that a limited, shared resource is available to store chunks in memory and to integrate them into a larger structure. As the distance between two co-dependents increases, more resources are consumed in the intervening region, making it difficult to maintain and integrate the co-dependents in memory. The theory predicts that reading times at the verb *supervised* in sentence (1a) will be slower than in sentence (1b) because in (1a), the distance between the subject *the nurse* and the verb is larger than in (1b).

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

The locality effect has been consistently observed in reading studies on verb-medial languages, including English (Grodner & Gibson, 2005; Bartek et al., 2011), Russian (Levy et al., 2013), Spanish (Nicenboim et al., 2016), and Danish (Balling & Kizach, 2017). The locality phenomenon provides strong evidence for working memory-based theories of sentence comprehension. However, a serious challenge to existing working memory-based theories comes from verb-final languages such as German and Hindi. The verb-final languages do not show the locality effect, they exhibit an opposite effect, the *anti-locality effect*: Reading times at the verb are faster when the subject and the verb are kept apart compared to when they are kept close (Konieczny, 2000; Vasishth & Lewis, 2006; Husain et al., 2014; Nakatani & Gibson, 2008).

The anti-locality effect has been attributed to the strong predictability of the upcoming verb due to frequent exposure to head-final structures (Yamashita, 2000; Kamide et al., 2003; Konieczny & Döring, 2003), such that strong predictability of the verb may override working memory demands imposed by increased dependency distance (Husain et al., 2014).

The contrasting profile of locality effects in verb-medial vs. verb-final languages is one of the biggest empirical challenges to existing theories of sentence comprehension. While it is reasonable to assume that working memory constraints are central to comprehension across languages, the existing

theories are incomplete as they fail to explain data from verb-final languages. Second, the anti-locality effect in verb-final languages makes a strong case for the predictability proposal: Prediction of the upcoming sentence material may facilitate dependency resolution in real time. Several experimental studies have accumulated evidence for either working memory-based proposals or prediction-based proposals.

However, the inferences drawn for the (anti-)locality debate are based on the average behavior of the sample. For example, on average, verb-medial language comprehenders show the locality effect, while verb-final language comprehenders show the anti-locality effect. Average behavior may not be sufficiently informative about the universality of cognitive processes (Estes, 1956; Underwood, 1975). An alternative approach is individual differences modeling (Haaf & Rouder, 2018; Mätzig et al., 2018).

For instance, it is possible that most verb-final language speakers exhibit strong predictability of the upcoming verb, and a few might exhibit weak predictability of the verb. Consequently, the average population-level behavior would show an anti-locality effect, but some individuals might show a locality effect or no effect. Investigating why individuals differ in their anti-locality effect could be important for theory building.

In this paper, we present four self-paced reading experiments on Hindi to investigate why individuals differ in their (anti-)locality effect? Particularly, I test the following two hypotheses:

- (a) **The varying prediction strength hypothesis:** Individuals vary in how strongly they predict the upcoming verb. Individuals with strong predictability would show faster reading times at the verb and stronger anti-locality effects.
- (b) **The varying lexical access hypothesis:** Individuals differ in their lexical accessibility. Individuals with slowed lexical accessibility would be slower in reading and to compensate, they rely more on a top-down prediction strategy, hence showing stronger anti-locality effects.

The above hypotheses make opposite predictions about the correlation between individual-level reading speed and (anti-)locality effects. We implement these two hypotheses as individual difference models and test them using data from four anti-locality experiments in Hindi.

The key finding is that individuals with a slower reading speed show stronger anti-locality effects, suggesting that low retrieval-capacity readers compensate with top-down predictive processing. We propose a new model that integrates this varying lexical accessibility assumption with the cue-based retrieval model of Lewis and Vasishth (2005).

The paper is organized as follows. Next, we present Bayesian hierarchical models that implement two individual difference hypotheses. After that, we present results from four Hindi experiments, and then, we discuss the findings and conclude.

Two individual difference models

We implement the varying prediction strength hypothesis and the varying lexical accessibility hypothesis as individual difference models in the Bayesian framework. The models differ in their assumptions about how individuals in the language population differ from each other.

Both the models assume that the reading time at the verb on trial i from subject j comes from a lognormal distribution with parameters μ and σ

$$rt_{ij} \sim \text{Lognormal}(\mu_{ij}, \sigma) \quad (1)$$

where

$$\mu_{ij} = \alpha_j + \beta_j \text{Distance}_i \quad (2)$$

where α_j is the individual-level intercept, which indicates the average reading speed of an individual, and β_j is the individual-level slope, which indicates (anti-)locality effect—difference between long and short distance condition—exhibited by an individual. The predictor variable Distance_i encodes whether the trial i has short dependency distance or long dependency distance.

The individual-level reading speed (α_j) and anti-locality effect (β_j) are assumed to come from a multivariate normal distribution with population-level mean intercept α and population-level anti-locality effect β and a variance-covariance matrix:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \Sigma = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right) \quad (3)$$

where $\mathcal{N}_2$ represents the bivariate normal distribution and Σ is called the variance-covariance matrix; the parameters τ_1 and τ_2 indicate the extent of individual differences in reading speed and anti-locality effects respectively. The parameter ρ is the correlation between the individual-level reading speed and the anti-locality effect. A negative value of ρ would imply that individuals with slower reading speed show a stronger anti-locality effect. The two models differ in their assumptions about the correlation parameter ρ .

The varying prediction strength model assumes that individuals differ in how strongly they predict the upcoming verb. Consequently, individuals who strongly predict the upcoming verb would show faster reading times at the verb due to prediction-facilitation. Moreover, individuals with strong predictability would also show stronger anti-locality effects because they use the intervening material to strengthen the prediction of the upcoming verb. Thus, the model assumes a positive correlation between individual-level reading speed and anti-locality effects: Individuals with strong predictability would show faster reading times and stronger anti-locality effects. We can write:

$$\rho > 0 \quad (4)$$

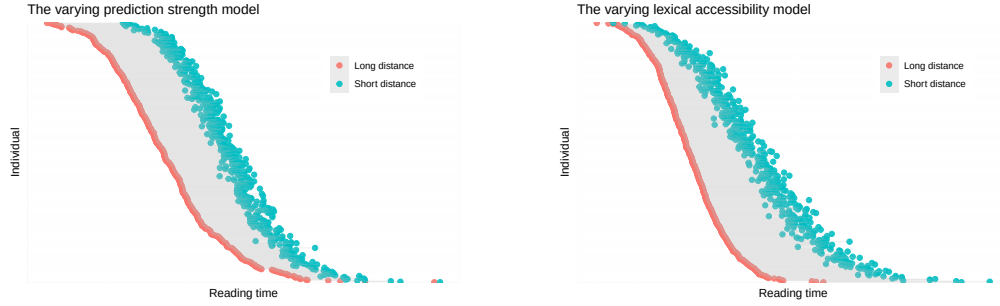


Figure 1: Individual-level reading times in the long-distance vs short-distance condition are represented by red and blue dots, respectively. The gap between the red and blue dots indicates the anti-locality effect for an individual. The reading speed of an individual is represented on the x-axis; larger value on the x-axis means a slower reading speed. Individuals are ordered based on their reading speed from left to right.

The varying lexical accessibility model assumes that individuals differ in how quickly they can access the lexical and syntactic properties of a word when it is encountered in the sentence. Consequently, individuals who are slower at lexical access should show slower reading times at the verb. However, these slow lexical-access readers may compensate through the top-down predictive processing, implying that they would show stronger anti-locality effects compared to others. Thus, the model assumes a negative correlation between individual-level reading speed and anti-locality effects: Individuals with slowed lexical access should show slower reading times, but stronger anti-locality effects. We can write:

$$\rho < 0 \quad (5)$$

Figure 1 shows the individual-level reading times in the long-distance condition (red dots) and short-distance condition (blue dots) predicted by the two models. The larger gap between the red and blue dots indicates the larger anti-locality effect and x-axis represent the reading speed. The varying lexical accessibility model predicts the negative correlation: Individuals with slower reading speed (right side) show a larger gap between red and blue dots (larger anti-locality effect).

We conduct four locality experiment replications in Hindi to test whether Hindi data supports the prediction of the varying prediction strength hypothesis or the varying lexical accessibility hypothesis.

Experiments 1-4: Locality experiments in Hindi

To test our hypotheses about individual differences, we conduct four locality experiments in Hindi. We aim to test (i) whether Hindi consistently shows the anti-locality effect, and (ii) whether individual difference data support the varying prediction strength hypothesis or the varying lexical accessibility hypothesis.

Our four experiments are indirect replications of the Vasishth and Lewis (2006) Experiment 2, which uses embedded relative clause constructions. In our experiments, we use subject relative clause constructions that modify an indefinite

noun phrase (e.g., *a boy*); see example sentences (2a) and (2b).

2. (a) A boy **who** yesterday school-DAT everyone in-front-of song **sing PROG was** outside standing is ...
- (b) A boy **who** song **sing PROG was** outside standing is ...

In sentence (2a), the dependency between the relative pronoun (RelPro) subject *who* and the relative clause (RC) verb *sing PROG was* is intervened by three adjuncts that modify the RC verb, which consequently increase the dependency distance between the RelPro and the RC verb. While in sentence (2b), the RelPro subject and the RC verb are kept close to each other.

Native Hindi speakers (N=69, N=70, N=40, and N=57 for Experiments 1-4) were recruited for each experiment. A self-paced reading paradigm was used to record chunk-by-chunk reading times. The experiment was designed and presented using PCIBex software (Zehr & Schwarz, 2021). The stimuli consisted of 24 critical sentences (like 2a-b) and 72 fillers. Two-thirds of the sentences were followed by yes/no comprehension questions.

The varying prediction strength hypothesis predicts a positive correlation between individual-level reading speed and the anti-locality effect. In contrast, the varying lexical accessibility hypothesis predicts a negative correlation: Individuals with slower reading times would show stronger anti-locality effects. See Figure 1 for the prediction of two hypotheses.

Results

Figure 2 shows the estimated (anti-)locality effect in four experiments. We consistently find the anti-locality effect: On average, faster reading times at the RC verb in the long subject-verb distance condition compared to the short distance condition. The meta-analysis using data from all four experiments also shows an anti-locality effect (95% credible interval: $[-75, -5]$ milliseconds). The average behavior is consistent with findings from other verb-final languages and goes against the prediction of the working memory-based theories.

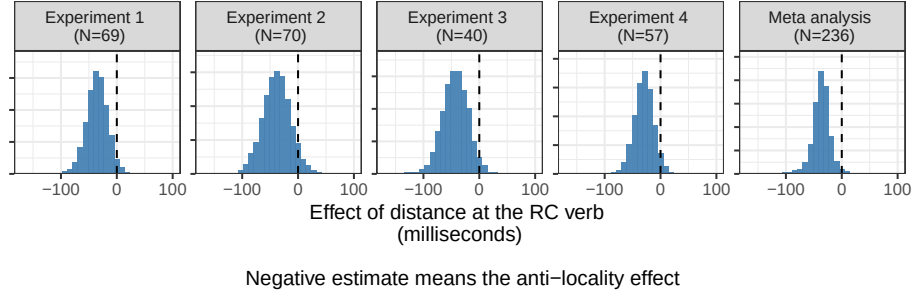


Figure 2: The estimated (anti-)locality effect in four experiments and their meta-analysis is shown on the x-axis. The estimated effect is the difference in reading time between the long-distance condition and the short-distance condition.

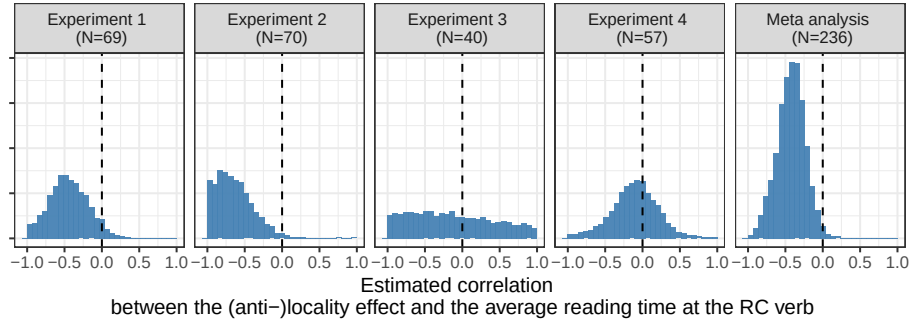


Figure 3: The estimated individual-level (anti-)locality effect in four experiments and their meta-analysis. The estimated effect is the difference in reading time between the long-distance condition and the short-distance condition.

Figure 4 shows the observed individual-level reading times in the long-distance condition (red dots) and short-distance condition (blue dots). The larger gap between the red and blue dots indicates a larger anti-locality effect and x-axis represents the reading speed. We observe that individuals differ in anti-locality effects as well as reading speeds.

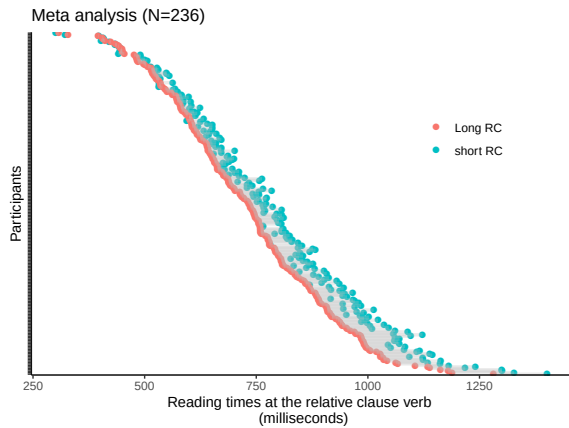


Figure 4: The observed individual-level reading times in the long distance condition, the short distance condition, and overall reading times across words

The key result is that individual-level anti-locality effects

and reading speeds are negatively correlated: Individuals with slower reading speed (right side dots) exhibit a stronger anti-locality effect (bigger gap between red and blue dots). The observed individual difference pattern is consistent with the prediction of the varying lexical accessibility hypothesis. The estimated correlation between the intercept (average reading times) and the slope (anti-locality effect) is negative, as shown in Figure 3. Thus, the individual differences data support the varying lexical accessibility hypothesis and go against the varying prediction strength hypothesis.

Model comparison

The individual differences data support the prediction of the varying lexical accessibility model, which assumes a negative correlation between individual-level reading speed and anti-locality effect. Is this model quantitatively better than a null model that assumes zero correlation between reading speed and anti-locality effects?

We quantify evidence for the varying lexical accessibility model compared to a zero correlation model using Bayes factors. The Bayes factor in favor of a model $\mathcal{M}_1$ over $\mathcal{M}_2$, say BF_{12} , is the ratio of the marginal likelihoods of the two models. BF_{12} represents the extent to which Model $\mathcal{M}_1$ is more likely than $\mathcal{M}_2$ given the data. We estimated Bayes factors using `bridgesampling` package (Gronau et al., 2020) in R.

Figure 5 shows Bayes factors in favor of the varying lexical accessibility model against the null hypothesis model un-

der four different prior assumptions about the correlation parameter. We find moderate evidence ($BF_{12} = 3 - 10$) for the negative correlation between individual-level reading speed and anti-locality effect under the assumption that the correlation lies in the range of $[0, -0.75]$. The result indicates that individuals with slow lexical access compensate for a predictive processing strategy and hence show stronger anti-locality effects.

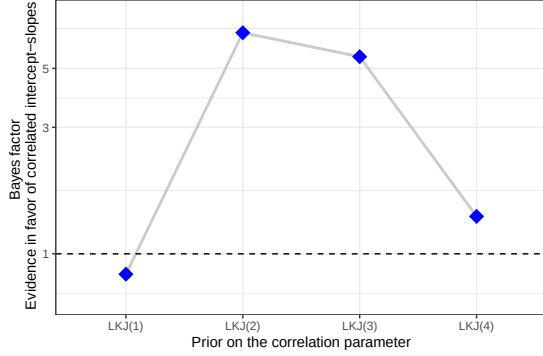


Figure 5: Bayes factors showing evidence in favor of the varying lexical accessibility model under different prior assumptions about the correlation parameter.

Discussion

What explains the individual differences in the anti-locality effect in a verb-final language? We explored two possibilities: (i) *Varying prediction strength*: Individuals differ in how strongly they predict the upcoming verb; consequently, individuals with stronger predictability show stronger anti-locality effects as well as faster reading times at the verb; and (ii) *Varying lexical accessibility*: Individuals differ in how easily they can access the semantic and syntactic properties of a word; consequently, individuals with slower lexical access rely more on top-down prediction of upcoming material, hence they show stronger anti-locality effects but slower reading times.

We tested which of these two proposals is better at explaining individual-level anti-locality effects observed in Hindi, a verb-final language. The experimental results and model comparison reveal two key findings:

1. The anti-locality effect is robustly observed in Hindi at the group-level.
2. Individual-level behavior from the four experiments supports the varying lexical accessibility proposal; Data show evidence for the negative correlation between individual-level reading speed and anti-locality effects.

The results imply that individual-level processing behavior is driven by the interaction of top-down predictability and bottom-up memory retrieval process: Individuals who face difficulty in accessing the linguistic information from memory compensate through top-down prediction of upcoming

material, and hence they produce stronger anti-locality effects but overall slower reading times.

This implication is important for theories of sentence comprehension. It reveals new insights into how bottom-up retrieval capacity constraints and top-down predictability interact: Low retrieval capacity readers rely more on top-down predictive processing.

Can a model integrate predictability and retrieval processes under the above assumption? We propose that a cue-based retrieval model (Lewis & Vasishth, 2005) can integrate a predictability component, such that the upcoming verb phrase is preactivated in memory through spreading activation from pre-verbal nouns (nouns that appear before the verb). We call it the predictive memory activation model.

The model 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

$$A_{i,t} = A_{i,0} - \ln t^{-d} + G_a \sum_{j=1}^n W_j S_{ji} + \epsilon_i \quad (6)$$

where $A_{i,0}$ is the activation of the noun phrase at the time $t = 0$ when it was first accessed from the declarative memory; $\ln t^{-d}$ indicates decay in activation over time t ; the term $G_a \sum_{j=1}^n W_j S_{ji}$ 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 (Engelmann, Jäger, & Vasishth, 2020). The term ϵ_i is Gaussian noise added to the activation of the noun i , such that $\epsilon_i \sim \text{Normal}(0, \sigma)$.

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 i that associates with the verb k :

$$A_{k,t} = B_k + G_p \sum_{i=1}^m S_{i,k} - \ln(t - t_i)^{-d} + \epsilon_k \quad (7)$$

where B_k is the baseline activation of the verb phrase k ; $G_p S_{i,k}$ is the amount of pre-activation spread from the preverbal noun i to the verb phrase k . G_p is the source preactivation parameter; $(t - t_i)$ is the time elapsed between the noun i and the verb k , which appears at t . The associative strength $S_{i,k}$ between noun i and verb k is determined by their history of co-retrievals, which can be approximated using pointwise mutual information between the two.

When the verb is encountered during reading, the verb phrase's representation k is lexically accessed from memory using visual cues of the verb. This lexical access time at the verb $T_{k,lex}$ will be determined by the verb phrase's preactivation and activation spread through the verb's visual cues.

$$T_{k,lex} = F e^{B_k + G_p \sum_{i=1}^m S_{i,k} - (t - t_i)^{-d} + G_a S_k + \epsilon_k} \quad (8)$$

where the latency factor F is the scaling parameter and is a free parameter in the model; $G_a S_k$ is the activation spread to

the verb phrase's representation using the verb's visual cues, such that G_a is the source activation used for retrieval. The parameter G_p is the source preactivation used for preactivating the upcoming verb phrase.

We assume that the source preactivation G_p and source activation G_a draw from the same pool of resources such that individuals with low source activation for retrieval will have high source preactivation, implying that individuals with low retrieval capacity will rely more on preactivation resource. We can write this assumption as $G_{pj} = 1 - G_{aj}$ where G_{pj} and G_{aj} represent source preactivation and source activation for an individual j . In verb-final languages, most comprehenders would have high G_p and low G_a .

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 :

$$T = T_{k,lex} + Fe^{-A_{it}} \quad (9)$$

This model can potentially explain the cross-linguistic differences in the (anti-)locality effect as well as individual differences in the (anti-)locality effect.

The current work is the first large-scale investigation of the locality effect in Hindi using an individual difference approach. The results reveal new insights into the underlying sentence comprehension process: The dependency between the subject and the verb is identified and resolved by top-down prediction as well as bottom-up retrieval process, such that slow retrieval capacity readers compensate through predictive processing.

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.
- Engelmann, F., Jäger, L. A., & Vasishth, S. (2020). The effect of prominence and cue association in retrieval processes: A computational account. *Cognitive Science*. Retrieved from 10.1111/cogs.12800
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological bulletin*, 53(2), 134.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- 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.
- Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An r package for estimating normalizing constants. *Journal of Statistical Software*, 92, 1–29.
- Haaf, J. M., & Rouder, J. N. (2018). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin and Review*, 1–18.
- 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.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of psycholinguistic research*, 29(6), 627–645.
- Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and srns. In *Proceedings of iccs/ascs* (pp. 13–17).
- 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.
- Mätzig, P., Vasishth, S., Engelmann, F., Caplan, D., & Burchert, F. (2018). A computational investigation of sources of variability in sentence comprehension difficulty in aphasia. *Topics in Cognitive Science*, 10(1), 161–174. (Allen Newell Best Student-Led Paper Award at Math-Psych/ICCM 2017)
- 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.
- Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, 30(2), 128.
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4), 767–794.
- Yamashita, H. (2000). Structural computation and the role of morphological markings in the processing of japanese. *Language and speech*, 43(4), 429–455.
- Zehr, J., & Schwarz, F. (2021). *Pcibex: An on-line platform for running psycholinguistic experiments*. <https://www.pcibex.net/>. (Accessed 2026)