A bounded rationality account of dependency length minimization in Hindi

Sidharth Ranjan (sidharth.ranjan03@gmail.com)
University of Stuttgart, 70174 Stuttgart, Germany

Titus von der Malsburg (titus.von-der-malsburg@ling.uni-stuttgart.de)
University of Stuttgart, 70174 Stuttgart, Germany

Abstract

The principle of DEPENDENCY LENGTH MINIMIZATION, which seeks to keep syntactically related words close in a sentence, is thought to universally shape the structure of human languages for effective communication. However, the extent to which dependency length minimization is applied in human language systems is not yet fully understood. Preverbally, the placement of *long-before-short* constituents and postverbally, *short-before-long* constituents are known to minimize overall dependency length of a sentence. In this study, we test the hypothesis that placing only the shortest preverbal constituent next to the main-verb explains word order preferences in Hindi (a SOV language) as opposed to the global minimization of dependency length. We characterize this approach as a least-effort strategy because it is a cost-effective way to shorten all dependencies between the verb and its preverbal dependencies. As such, this approach is consistent with the bounded-rationality perspective according to which decision making is governed by 'fast but frugal' heuristics rather than by a search for optimal solutions. Consistent with this idea, our results indicate that actual corpus sentences in the Hindi-Urdu Treebank corpus are better explained by the least effort strategy than by global minimization of dependency lengths. Additionally, for the task of distinguishing corpus sentences from counterfactual variants, we find that the dependency length and constituent length of the constituent closest to the main verb are much better predictors of whether a sentence appeared in the corpus than total dependency length. Overall, our findings suggest that cognitive resource constraints play a crucial role in shaping natural languages.

Keywords: Hindi; Word order; Syntactic choice; Locality; Production; Cognitive modeling

Introduction

Human working memory is viewed as a limited capacity system, where there is pressure to minimize the cognitive load by not retaining information longer than necessary. Simon (1982, 1990, 1991) proposed the idea of *bounded rationality* in decision making which incorporates the aforementioned cognitive limitation and proposed that human's ability to make rational decisions has adapted to making choices that are satisfactory rather than optimal. In language, speakers' decisions are influenced by the context in which the decision is made—*availability of information, cognitive resources,* and the *limited response time*. This tension naturally exerts a challenge for speakers and constrains them to adopt certain strategies to formulate sentences that are most efficient for communication. For example, choosing word-order patterns that are most common in daily discourse or placing the words/phrases early in the sentence that are more accessible or salient in working memory. In this work, we examine whether word order variation in Hindi can be seen as a result of the pressures for efficient communication within the constraints of bounded rationality.

Prior work has investigated the role of phrase length on the constituent ordering of sentences (Bock & Warren, 1985; Arnold, Wasow, Losongco, & Ginstrom, 2000; Hawkins, 1994; Yamashita & Chang, 2001; Choi, 2007; Faghiri & Samvelian, 2020). The findings suggest that SVO languages (e.g., English) have a preference for *short-before-long* linguistic structure such that shorter phrases are more accessible and get realized before relatively less accessible or longer ones (Arnold et al., 2000). In contrast, for SOV languages (e.g., Japanese), *long-before-short* ordering is prevalent, *i.e.,* longer phrases tend to be shifted ahead of shorter ones with a view that long phrases are semantically rich and conceptually salient, and thus realized before than the shorter ones (Yamashita & Chang, 2001). In this vein, Gibson's Dependency Locality Theory provides a unified explanation for constituent ordering patterns in both SOV and SVO languages (Gibson, 2000). The theory prefers constituent order that minimizes the overall dependency length of the sentence irrespective of language. Preverbally, *long-before-short* and postverbally, *short-before-long* constituent ordering minimize the overall dependency length in the sentence (Hawkins, 2004).

The principle of DEPENDENCY LENGTH MINIMIZATION (DLM), is believed to reduce the cognitive load on working memory by reducing the distance between syntactically related words in the sentence for both speakers and hearers (Gibson, 1991; Hudson, 1995; Gibson, 1998, 2000; Ferrer-i-Cancho, 2004; Temperley, 2007). This has been shown to be the case in natural languages, which tend to have shorter overall dependency lengths than those produced randomly (Futrell, Mahowald, & Gibson, 2015; Liu, Xu, & Liang, 2017; Temperley & Gildea, 2018; Futrell, Levy, & Gibson, 2020). The DLM hypothesis has therefore been demonstrated to be an essential characteristic of efficient word order by virtue of communicative efficiency in human language (Gibson et al., 2019). However, the extent to which the minimization of dependency length is applied to human language systems in not yet fully understood. This serves as our primary motivation, for which we apply the idea of bounded rationality to DLM and investigate how well it explains word-order preferences in Hindi.

(1)   a.   **maa=ne**   **baajaar jaate samaye rote hue bacche=ko**   **toffee**
         mother=ERG market   going while   cry=PROG child=ACC toffee
         di
         di-PFV.F.SG

         The mother gave the crying child a candy while going to market.

   b.   **maa=ne rote hue bacche=ko baajaar jaate samaye toffee** di

   c.   **rote hue bacche=ko baajaar jaate samaye maa=ne toffee** di

Hindi, a Indo-Aryan language from the Indo-European language family, has subject-object-verb (SOV) as the basic word order and has a rich case-marking system (Kachru, 1982). In contrast to English, Hindi exhibits greater flexibility in the arrangement of words within a sentence, as illustrated in Example 1. Example 1a has 4 preverbal constituents and is considered the most preferred syntactic choice since it originally appeared in the corpus as opposed to all variants (4! = 24) that are possible with this sentence.

In this work, we test our hypothesis that speakers adopt a least-effort strategy to determine the word order choices in Hindi as opposed to global minimization of dependency length. We define *least-effort strategy* as the placement of shortest preverbal constituent closest to the main verb as this is the most economical way of reducing all the dependencies between the verb and its preverbal dependents (see Figure 3 for an illustration). We generated different counterfactual variants by randomly rearranging their preverbal constituents (see Examples 1b-1c to list a few) for each reference sentence (1a) in Hindi-Urdu Treebank corpus of written text. Next, we conducted a quantitative analysis of those variants and their corresponding corpus reference sentences. We then deployed a logistic regression classifier to distinguish the corpus reference sentence from the artificially generated variants based on the dependency length and constituent length of preverbal constituents, and overall dependency length of the sentence. The primary motivation behind choosing constituent-level predictors was to test the efficacy of the least-effort strategy (see Figure 3d) against global minimization constraint (see Figure 3b) under the purview of bounded rationality. As the number of preverbal constituents increases (say from 3 → 5), the number of possible variants also increases significantly (3! = 6 → 5! = 120); overburdening the speakers to make choices. A boundedly rational speaker would then apply a heuristic strategy that identifies a satisfactory solution that does not necessarily minimize total dependency length (Newell & Simon, 1972; Gigerenzer & Gaissmaier, 2011).

Our main contribution through this work is that we show the impact of dependency length minimization on word order choices in Hindi using large-scale naturalistic data rather than the constructed experimental stimuli and offer cross-linguistic evidence essential for the advancement of cognitive theories (Jaeger & Norcliffe, 2009).

The remainder of this paper is structured as follows. We first briefly describe the data and methods used to investigate the hypothesis, followed by the results in this work. Towards the end, we discuss the broader implications of our findings with a conclusion and future directions.

Method

Our data set consists of 7,586 declarative sentences from the Hindi-Urdu Treebank (HUTB) corpus of written text belonging to the newswire domain (Bhatt et al., 2009). For each reference sentence in the corpus, we created artificial variants by randomly permuting the preverbal constituents whose head was immediate child to the root verb in their projective dependency tree (see Figure 3). In instances where there were more than 100 variants (an arbitrary cutoff[1] to keep our computation tractable), we randomly selected 99 of them. Our variant generation approach resulted in 184818 competing variants in total, for the reference sentences in our dataset. Refer to Example 1 for an illustration. Figure 1 depicts the percentages of reference-variant sentences for each number of preverbal constituents. Due to insufficient data points for sentences with 7 or more preverbal constituents, we only present the corpus analysis results for sentences with up to 6 preverbal phrases. However, for the binary classification task presented in the later section of this work, we used the entire dataset to compute our results [2].

Corpus analysis

Previous corpus studies in Hindi have revealed a predominance of *long-before-short* constituent orders preverbally, and smaller head-dependent distances (Sharma, Futrell, & Husain, 2020; Ranjan, Rajkumar, & Agarwal, 2022). To test our hypothesis of least-effort strategy, we investigated the role of all preverbal constituents in minimizing the overall dependency length of the sentence. Consistent with previous studies, we also found that *long-before-short* constituent orders are dominant in Hindi (Ranjan et al., 2022). In addition, we also observed a wide distribution of *long-before-short* and *short-before-long* constituents in corpus sentences with varying preverbal constituents (see Figure 2). Crucially, the plot in Figure 2 suggests the constituent closest to the main verb has a strong tendency to be the shortest among all the preverbal constituents in the sentence. This is consistent with the predictions of the least effort principle, because moving the shortest constituent to the main verb minimizes not just that constituent's dependency length but also the dependency lengths of all other preverbal dependents of the main verb, therefore striking a favorable cost-benefit ratio.

Interestingly, Figure 2 also shows that the pressure to move shorter constituent to the main verb increases with the number of preverbal constituents, thus suggesting that speakers may dynamically balance production cost and comprehenders' cognitive constraints.[3] To further test support the idea that speakers adopt a least-effort strategy (i.e. place shortest constituent next to the main verb), we compared the total dependency length of reference sentences (attested in the
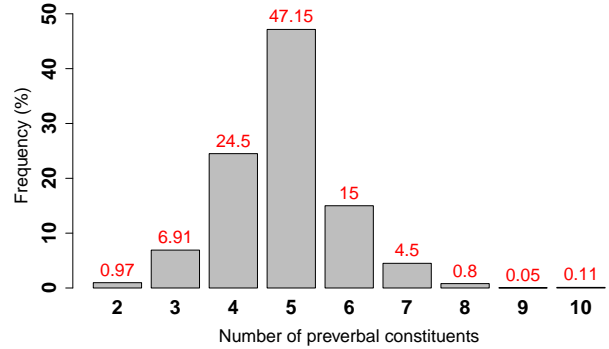
---

[1]Higher or lower cutoffs do not impact our results.

[2]We also carried out a similar analysis with only grammatical variants and found that our findings were consistent with those reported below.

[3]Pearson's correlation coefficient between sentence length and number of preverbal constituents was 0.45.
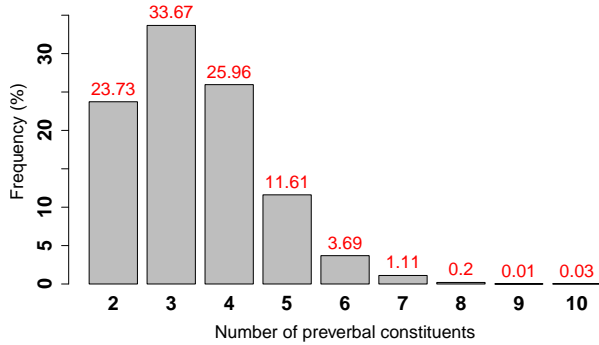
(a) Reference sentences (7586 data points)

(b) Variant sentences (184818 data points)

Figure 1: Percentages of reference and variant sentences with varying preverbal constituents
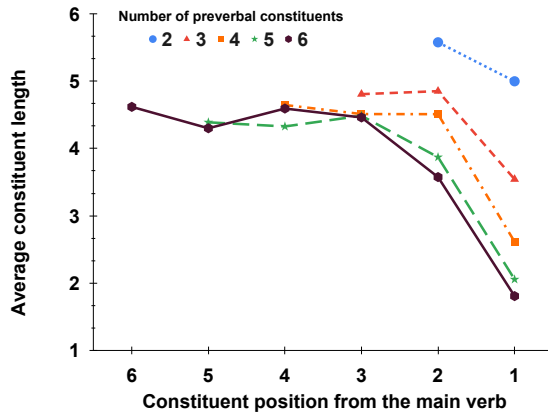

Figure 2: Average constituent length of preverbal constituents for corpus sentences with 2 to 6 constituents separately

corpus) to four alternatives with different constituent orders illustrated in Figure 3 for example sentence 1a:

1. ASCENDING ORDER: Arrange the preverbal constituents in their increasing order of constituent lengths. This arrangement leads to maximal dependency length of the sentence (See Figure 3a).

2. DESCENDING ORDER: Arrange the preverbal constituents in their decreasing order of constituent lengths. This arrangement globally minimizes the dependency length of the sentence (See Figure 3b).

3. RANDOM ORDER: Arrange the preverbal constituents randomly (See Figure 3c).

4. LEAST-EFFORT ORDER: Start with a random order of preverbal constituents and then simply move the shortest constituent next to the main verb (See Figure 3d).

Our results in Figure 4 show that the dependency length of sentences in the corpus generally tracks the dependency length of the least-effort solution for sentences with three or more preverbal constituents. Only sentences with two preverbal constituents deviate from the least-effort order. This may be because the memory pressure is so low in these cases that the system does not optimize dependency length at all, not even with the least effort heuristic. Consistent with previous

studies, the plot also confirms that corpus reference sentences on average have lower dependency length than the sentences with random word orders (Liu, 2008; Gildea & Temperley, 2010; Futrell et al., 2015, 2020).

These results further corroborate the idea that dependency length minimization in Hindi, and thus constituent order, are governed by a least-effort strategy. This is consistent with the bounded rationality view of decision making, based on which we would expect that speakers consider only a limited search space when planning sentences (Gigerenzer, Hertwig, & Pachur, 2011).

In the next section, we further test and validate the least-effort hypothesis using a decision task with a goal to distinguish reference sentences from random variants (Gildea & Temperley, 2010; Temperley & Gildea, 2018).

Computational simulation

We set up a binary classification task to distinguish the reference sentence amidst the artificially generated variants. Our original dataset had a significant class imbalance, with 184818 variants compared to only 7586 reference sentences. To address this imbalance, we applied the technique suggested by Joachims (2002) which transforms the task of classifying a sentence as either reference or variant into a task of ranking reference sentence against each of its variants (pairwise ranking) by training the classifier on the difference between the feature vectors of variant sentence and corresponding paired reference sentence (see Equations 1 and 2).

$$w \cdot \phi(reference) > w \cdot \phi(variant) \tag{1}$$

$$w \cdot (\phi(reference) - \phi(variant)) > 0 \tag{2}$$

Equation 1 represents the objective of a standard binary classifier where corpus reference sentence is preferred against its corresponding variant sentence. The classifier must determine a feature weight (w) such that the dot product of w and the reference feature vector ($\phi(reference)$) is greater than the dot product of w and the variant feature vector ($\phi(variant)$). This objective can be rephrased as Equation 2 where the dot

(a) Ascending order of constituent lengths

(b) Descending order of constituent lengths

(c) Random order of constituent lengths

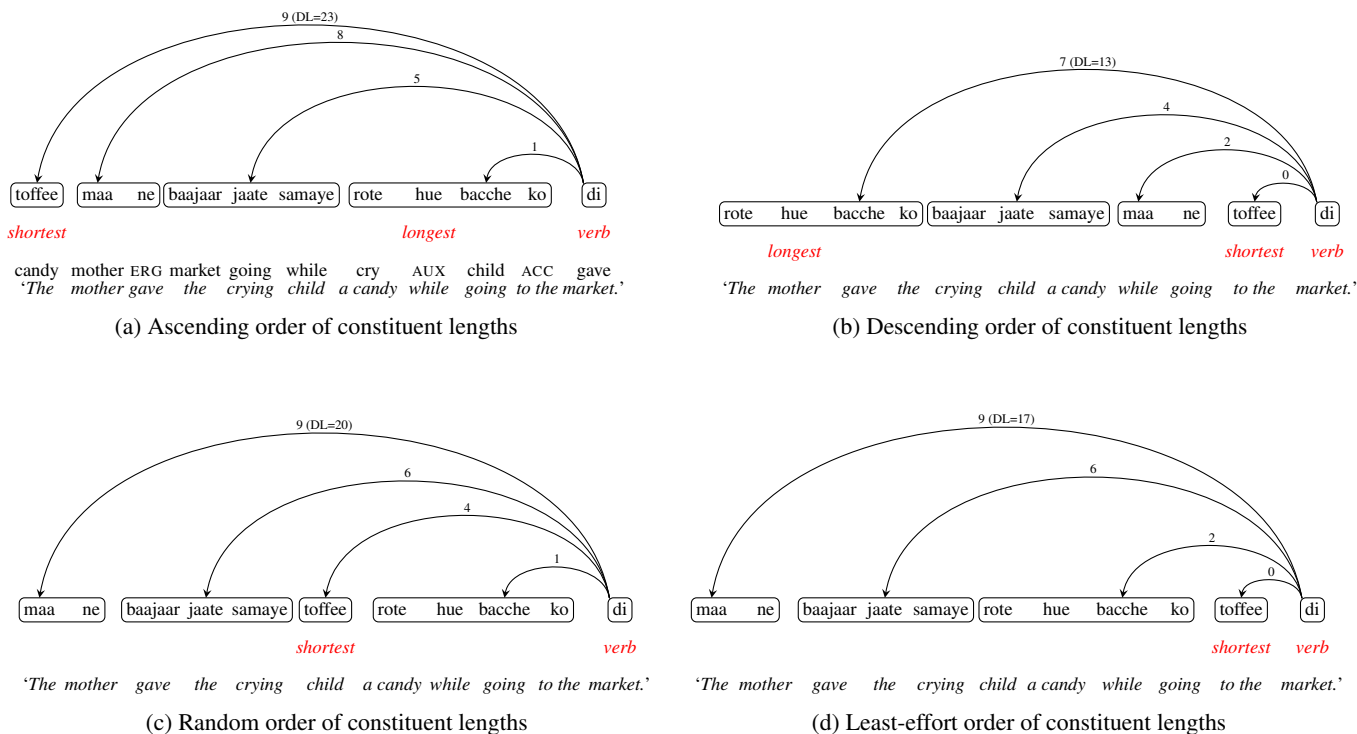(d) Least-effort order of constituent lengths

Figure 3: Dependency length and constituent ordering for Hindi head-final structure; Only main verb dependencies are depicted in the figures; Total dependency length (DL) of the structure indicated above each sub-figure; Constituent dependency length is mentioned above each dependency arc

product of w and the difference between the feature vectors must be greater than zero. The dataset was structured such that each variant sentence was paired with its corresponding reference sentence, with the order being balanced across these pairs. For instance, Example 1 would give (1a-1b) and (1c-1a). Feature vectors were subtracted and binary labels were assigned to each transformed data point. The "Reference-Variant" pairs were labeled as "1" and "Variant-Reference" pairs were labeled as "0", thus balancing the previously imbalanced classification task. Post Joachims's transformation, we had in total 184818 data points for our classification task. See Ranjan et al. (2022) for more details.

We used the `glm` function in R to test the stated hypothesis at the outset (dependent variable $\sim$ independent variables): *choice* $\sim \delta$ *dependency length* and *choice* $\sim$ δ *constituent length*. Here, choice is a binary choice dependent variable (1 stands for reference sentence preference, and 0 denotes the variant sentence preference). The delta (δ) refers to the difference between the feature vectors of reference sentence and its paired variant. The classifier predicts label "1" if corpus reference sentence outranks a variant paired with it and "0" if variant outranks its paired reference sentence. All the independent variables were normalised to z-scores, *i.e.*, the predictor's value (centered around its mean) was divided by its standard deviation. We estimated indepen-

dent variables as below (also see Figure 3 for illustration):

1. TOTAL DEPENDENCY LENGTH: Summation of head-dependent distances in a dependency tree of a sentence. And the head-dependent distance was estimated by counting the number of intervening words between them.

2. CONSTITUENT DEPENDENCY LENGTH: Number of words spanned between the head of the preverbal constituent and the main root on which the constituent is dependent.

3. CONSTITUENT LENGTH: Total word counts in the constituent.

Regression analysis

In this section, we test our main hypothesis that speakers optimize word order choices in Hindi by adopting least-effort strategy (i.e., shifting the shortest constituent closest to the main verb) under the constraints of dependency length minimization. Figure 1b depicts the distribution of variants with varying preverbal constituents. Since our classification dataset is dominated by reference-variant pairs with 5 preverbal constituents, we test the least-effort hypothesis on the transformed version of our dataset with 5 preverbal constituents only (87143 data points) using a logistic regression model. We estimated the dependency length and constituent length of all 5 preverbal constituents and introduced them as predictors in our regression models (*choice* $\sim \delta const1 + \delta const2 + \delta const3 + \delta const4 + \delta const5$). However, before
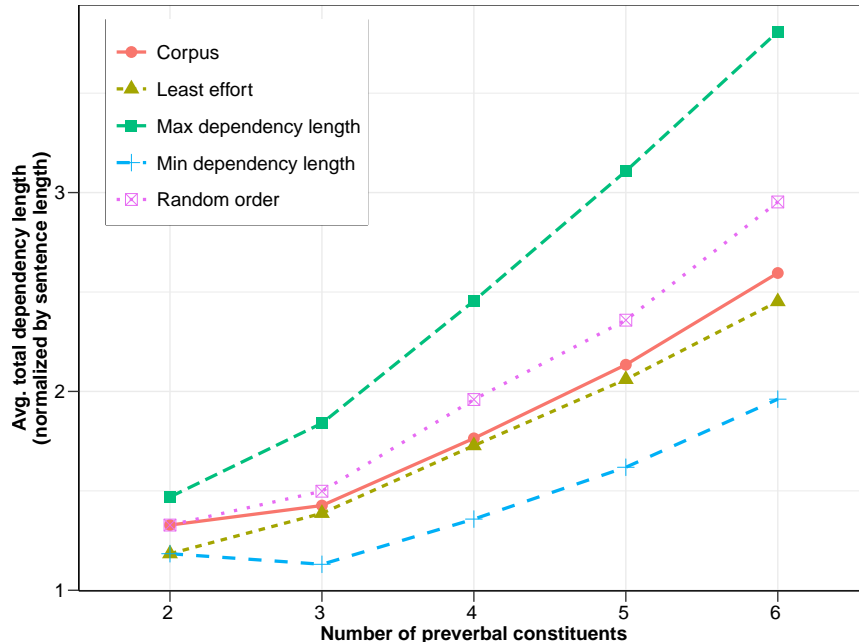
Figure 4: Average total dependency length (*i.e.,* total dependency length normalized by the number of words in a sentence) for different constituent orderings and for different numbers of preverbal constituents (7586 data points for each ordering)

doing so, we first deployed the Recursive Feature Elimination Cross-validation algorithm (Guyon, Weston, Barnhill, & Vapnik, 2002, RFECV) to find the best predictors among five preverbal constituents and then reported the regression results using the obtained optimal features.

Tables 1 and 2 display the results of our regression analysis. Negative regression coefficients in Tables 1 and 2 suggest that reference sentences tend to have lower constituent-level dependency length and the length of constituent, respectively, than their paired variant. The larger magnitude of the regression coefficient of the last two preverbal constituents closest to the verb (*const*4 & *const*5) as opposed to other preverbal constituents in the model suggests their greater role in minimizing overall dependency length of reference sentence than the paired variants. Notably, our results were consistent across reference-variant sentences with 2 to 6 preverbal constituents, indicating that our findings are not limited to rare constructions, but are applicable to frequent ones as well in the Hindi natural corpus. In sum, these findings suggest that speakers indeed, apply the least-effort strategy to optimize for word orders such that the preverbal constituents closest to the verb are shorter and also have lower dependency length.

Classification analysis

Based on the insights from the previous section, we deployed the last two preverbal constituent's dependency length and constituent length (closest to the main verb: *2nd last* & *last constituents*) as predictors into the classification model aimed at predicting reference sentences amidst the counterfactual variants. We make use of our entire dataset (184818 reference-variant pairs) and conduct 10-fold cross-

Predictor	Estimate	Std-Error	z-value
Intercept	**0.018**	0.008	1.98
1st constituent's deplen	-0.004	0.012	-0.35
2nd constituent's deplen	**0.088**	0.014	6.20
3rd constituent's deplen	**-0.147**	0.015	-10.12
4th constituent's deplen	**-0.529**	0.016	-32.90
5th constituent's deplen	**-2.720**	0.023	-118.03

Table 1: Regression model containing dependency lengths (deplen) of preverbal constituents as predictors (87143 data points); significant predictors denoted in bold with $p < 0.001$

validation to evaluate model's classification accuracy, *i.e.,* the percentage of data points where a model correctly predicted the referent sentence over a paired variant. Tables 3 and 4 present the prediction performance of our models. As illustrated in Table 3, in terms of individual classification accuracy, the dependency length of the last constituent turned out to be the best predictor (72.70%), thus validating the efficacy of the proposed least-effort principle. Interestingly, this measure supersedes the performance of the overall dependency length measure (62.69%) of the sentence as well.

Over a baseline model containing the last constituent's dependency length, adding the 2nd last constituent's dependency length induced a significant increase of 4.47% in the classification accuracy (p < 0.001 using McNemar's two-tailed test). We find similar results when the constituent lengths of the last two constituents were deployed as predictors in the classifier, as shown in Table 4. These findings lend credence to the idea that speakers adopt least-effort strategy,

Predictor	Estimate	Std-Error	z-value
Intercept	-0.003	0.008	-0.41
1st constituent's length	**-0.083**	0.009	-8.44
3rd constituent's length	**0.058**	0.010	5.72
4th constituent's length	**-0.148**	0.009	-15.26
5th constituent's length	**-1.549**	0.016	-97.82

Table 2: Regression model containing preverbal constituent lengths as predictors (87143 data points); significant predictors denoted in bold with $p < 0.001$; constituent length = word counts in the constituent

Predictor(s)	Accuracy
total dependency length	62.69
2nd-last preverbal constituent's deplen	68.48***
last preverbal constituent's deplen	72.70***
last + 2nd last preverbal constituent's deplen	77.17***

Table 3: Prediction accuracy of distinct models with dependency length as predictor on full dataset (184818 data points; deplen = dependency length; McNemar's two-tailed significance test compared to previous row: *** $p < 0.001$)

as measured by dependency length minimization, in determining Hindi word-order preferences.

Discussion and conclusion

Our main findings demonstrate that speakers in Hindi choose optimal word orders by placing the shortest phrase closest to the main verb, consistent with the central ideas of the bounded rationality view of decision making. This least effort approach not only reduces the overall dependency length but also enables a selective search procedure in a vast search space where finding the optimal solution is difficult. Therefore the sentence planning system uses a simple yet powerful heuristic. Additionally, as depicted in Figure 2, the pressure to shorten the constituent closest to the main verb increases as the length of the sentence increases. We also observed that sentences that originally appeared in the corpus had dependency lengths distribution similar to the proposed least-effort solution evincing the efficacy of bounded rationality in choosing the appropriate word order in Hindi (See Figure 4). In addition, the HUTB corpus sentences had lower overall dependency length than the random word orders consistent with the earlier studies (Liu, 2008; Futrell et al., 2015, 2020).

The prediction performance of two distinct classification models aimed at distinguishing reference sentences amidst artificially generated variants using just the constituent lengths (70.28%) and just the dependency lengths of the last two preverbal constituents that are nearest to the main verb (77.17%) emphasizes the significance of the proposed least-effort strategy in shaping the constituent order (Liu et al., 2017; Temperley & Gildea, 2018; Futrell et al., 2020). It's also interesting to note that the overall dependency length (62.69%) performs poorly in predicting corpus reference sentences (amidst

Predictor(s)	Accuracy
2nd-last preverbal constituent length	54.35
last preverbal constituent length	69.62***
last + 2nd last preverbal constituent length	70.28***

Table 4: Prediction accuracy of distinct models with constituent length as predictor (184818 data points; McNemar's two-tailed test compared to previous row: *** $p < 0.001$)

variants) than the least-effort strategy (last preverbal constituent's dependency length = 72.70%; last preverbal constituent length = 69.62%). As a part of future work, we plan to examine the effectiveness of our proposed least-effort measure across various languages from different language families. Studying how languages may differ in the way they implement bounded rationality strategies can provide further insight into the relationship between language and cognition.

The broader implication of our work is that word order choices can be seen as a result of the pressures for efficient communication within the constraints of bounded rationality (Gigerenzer & Goldstein, 1996; Gigerenzer, 2015). The speaker's attempt to minimize dependency length is one such frugal strategy to deal with limited cognitive resources and the complexity of the task. Given that we often lack the cognitive and computational resources to solve problems exactly, we tend to use approximation methods as a more viable strategy, leading to bounded rationality (Simon, 1990; Gigerenzer & Selten, 2002). According to Simon, the following mechanisms are used by people to deal with complex problems and make decisions under the constraints of bounded rationality – *pattern recognition, heuristic search,* and *extrapolation.* Along similar lines, in the case of language, communicative efficiency can be achieved by using most *common word-order patterns*, doing *selective search* among possible word orders that place the syntactically related words close together or salient/predictable words early in the sentence to overcome the memory load. In addition, speakers can make use to *certain properties of languages*, such as word-order flexibility, which provides helpful cues to the listener for improved understanding. From the listener's perspective, flexibility in word order can be challenging, and relying on various language-specific cues could be a rational strategy for them to navigate this challenge within the constraints of bounded rationality. Therefore, our results and interpretations of bounded rationality hold from both speaker as well as hearer perspectives. A thorough investigation from these perspectives will pave the way for better understanding. To make more tangible claims about language production in general, it will also be necessary to investigate the efficacy of the least-effort principle on spoken data.

Although bounded rationality may not fully account for all aspects of language, it may provide an explanation for certain aspects, such as the extent of dependency length minimization, word-order patterns, individual differences in decision-making processes, and the cognitive mechanisms involved in

language use. Future work needs to investigate how different factors, such as context, task demands, and individual cognitive differences, may affect how speakers and listeners use boundedly rational strategies.

Overall, our results suggest that within the constraints of bounded rationality, dependency length minimization is a significant predictor of Hindi word order patterns.

Acknowledgments

References

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*, 28–55. Retrieved from `https://www.jstor.org/stable/417392`

Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., & Xia, F. (2009). A multi-representational and multi-layered treebank for Hindi/urdu. In *Proceedings of the third linguistic annotation workshop* (pp. 186–189). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from `http://dl.acm.org/citation.cfm?id=1698381.1698417`

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47–67. Retrieved from `https://www.sciencedirect.com/science/article/pii/001002778590023X`

Choi, H.-w. (2007). Length and Order: A Corpus Study of Korean Dative-Accusative Construction. *Discourse and Cognition*, *14*(3), 207–227. Retrieved from `https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE00986720`

Faghiri, P., & Samvelian, P. (2020). Word order preferences and the effect of phrasal length in sov languages: evidence from sentence production in persian. *Glossa: a journal of general linguistics (2016-2021)*, *5*(1), 86.

Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70*(5), 056135. Retrieved from `https://core.ac.uk/download/pdf/286456939.pdf`

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371–412. Retrieved from `https://sites.socsci.uci.edu/~rfutrell/papers/futrell2020dependency.pdf`

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336-10341. Retrieved from `http://www.pnas.org/content/112/33/10336.abstract` doi: 10.1073/pnas.1502134112

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Carnegie Mellon University. Retrieved from `http://tedlab.mit.edu/tedlab_website/researchpapers/Gibson_1991_PhDthesis.pdf`

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76. Retrieved from `https://doi.org/10.1016/S0010-0277(98)00034-1`

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: Papers from the first mind articulation project symposium.* Cambridge, MA: MIT Press. Retrieved from `http://www.ling.uni-potsdam.de/~vasishth/Papers/Gibson-Cognition2000.pdf`

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019, May 01). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389-407. Retrieved from `https://doi.org/10.1016/j.tics.2019.02.003` doi: 10.1016/j.tics.2019.02.003

Gigerenzer, G. (2015). *Simply rational: Decision making in the real world*. Evolution and Cognition.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451–482.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, *103*(4), 650.

Gigerenzer, G., Hertwig, R. E., & Pachur, T. E. (2011). *Heuristics: The foundations of adaptive behavior*. Oxford University Press.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, *46*, 389–422. Retrieved from `https://link.springer.com/article/10.1023/A:1012487302797`

Hawkins, J. A. (1994). *A Performance theory of order and constituency*. New York: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press.

Hudson, R. A. (1995). Measuring syntactic difficulty. *University College London*. Retrieved from `http://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf`

Jaeger, T. F., & Norcliffe, E. (2009). The cross-linguistic study of sentence production: State of the art and a call for action. *Language and Linguistics Compass*, *3*(4), 866–887. Retrieved from `http://dx.doi.org/10.1111/j.1749-818X.2009.00147.x`

Joachims, T. (2002). Optimizing search engines using clickthrough data. In *Proceedings of the eighth acm sigkdd international conference on knowledge discovery and data mining* (pp. 133–142). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/775047`

.775067 doi: 10.1145/775047.775067

Kachru, Y. (1982). Conjunct verbs in hindi-urdu and persian. *South Asian Review*, *6*(3), 117-126. Retrieved from `https://doi.org/10.1080/02759527.1982.11933096` doi: 10.1080/02759527.1982.11933096

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159-191. Retrieved from `http://www.lingviko .net/JCS.pdf`

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171 - 193. Retrieved from `http://www.sciencedirect.com/science/article/ pii/S1571064517300532` doi: https://doi.org/10.1016/ j.plrev.2017.03.002

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ranjan, S., Rajkumar, R., & Agarwal, S. (2022). Locality and expectation effects in hindi preverbal constituent ordering. *Cognition*, *223*, 104959. Retrieved from `https://www.sciencedirect.com/science/article/ pii/S0010027721003826` doi: https://doi.org/10.1016/ j.cognition.2021.104959

Sharma, K., Futrell, R., & Husain, S. (2020). What determines the order of verbal dependents in hindi? effects of efficiency in comprehension and production. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 1–10). Retrieved from `https:// aclanthology.org/2020.cmcl-1.1/`

Simon, H. A. (1982). Models of bounded rationality, vols. 1 and 2. *Economic Analysis and Public Policy, MIT Press, Cambridge, Mass*.

Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology*, *41*(1), 1–20. Retrieved from `https://www.annualreviews.org/doi/ 10.1146/annurev.ps.41.020190.000245`

Simon, H. A. (1991). Cognitive architectures and rational analysis: Comment. In K. VanLehn (Ed.), *Architectures for intelligence: The twenty-second carnegie mellon symposium on cognition* (pp. 25–39). Lawrence Erlbaum Associates, Inc. Retrieved from `https://apps.dtic.mil/ sti/citations/ADA219199`

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*(2), 300–333. Retrieved from `http://www.sciencedirect .com / science / article / B6T24 -4M7CDMS -2 / 2 / e095449f6439b30003822a5838e53786` doi: DOI:10.1016/j.cognition.2006.09.011

Temperley, D., & Gildea, D. (2018). Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Reviews of Linguistics*. Retrieved from `https:// www.cs.rochester.edu/u/gildea/pubs/temperley -gildea-ar18.pdf`

Yamashita, H., & Chang, F. (2001). "Long before short" preference in the production of a head-final language. *Cognition*, *81*. Retrieved from `https://www.sciencedirect .com/science/article/pii/S0010027701001214`