

# UC Merced

## Proceedings of the Annual Meeting of the Cognitive Science Society

### Title

Work Smarter...Not Harder: Efficient Minimization of Dependency Length in SOV Languages

### Permalink

<https://escholarship.org/uc/item/0120d8t1>

### Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

### Authors

Ranjan, Sidharth  
von der Malsburg, Titus

### Publication Date

2024

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Work Smarter...Not Harder!

## Efficient Minimization of Dependency Length in SOV Languages

Sidharth Ranjan (sidharth.ranjan@ling.uni-stuttgart.de)  
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

Dependency length minimization is a universally observed quantitative property of natural languages. However, the extent of dependency length minimization, and the cognitive mechanisms through which the language processor achieves this minimization remain unclear. This research offers mechanistic insights by postulating that moving a short preverbal constituent next to the main verb explains preverbal constituent ordering decisions better than global minimization of dependency length in SOV languages. This approach constitutes a least-effort strategy because it's just one operation but simultaneously reduces the length of all preverbal dependencies linked to the main verb. We corroborate this strategy using large-scale corpus evidence across all seven SOV languages that are prominently represented in the Universal Dependency Treebank. These findings align with the concept of bounded rationality, where decision-making is influenced by 'quick-yet-economical' heuristics rather than exhaustive searches for optimal solutions. Overall, this work sheds light on the role of bounded rationality in linguistic decision-making and language evolution.

**Keywords:** SOV languages; Word order; Syntactic choice; Locality; Production; Bounded rationality; Decision-making

### Introduction

The syntax of natural language enables humans to achieve what von Humboldt (1836) and Chomsky (1965) famously described as the 'infinite use of finite means.' This phenomenon allows for the creation of a diverse range of expressions, such as sentences and discourses, by combining a finite set of linguistic units, namely words and constituents or phrases. However, among the myriad sequences available to convey an idea, the speaker produces just one (refer to Figures 1 and 3 for an illustration). What factors influenced this decision, and what cognitive mechanisms underlie this decision-making process?

Dependency Locality Theory (DLT; Gibson, 2000, 1998) has been very influential in predicting such ordering decisions in natural languages, based on the idea that the human mind grapples with limited working memory capacity (Liu, 2008; Gildea & Temperley, 2010; Futrell, Mahowald, & Gibson, 2015; Futrell, Levy, & Gibson, 2020; Rajkumar, van Schijndel, White, & Schuler, 2016; Liu, Xu, & Liang, 2017; Temperley & Gildea, 2018; Ranjan, Rajkumar, & Agarwal, 2022). DLT predicts that language processing system strives to maintain syntactically related words (head-dependent pair) in close proximity within the sentence in order to minimize memory load. In SVO languages, the placement of SHORT-

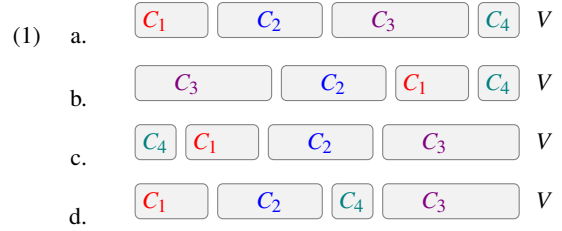


Figure 1: Preverbal constituent ordering problem in a language with SUBJECT-OBJECT-VERB (SOV) word order; Constituent ( $C_i$ ) is defined as a word or a group of words that functions as a single unit within a syntactic configuration

BEFORE-LONG constituents and in SOV languages, LONG-BEFORE-SHORT constituents in the sentence have consistently been identified as the most preferred syntactic ordering choices, as they minimize overall dependency length of the sentences (Hawkins, 1994, 2004; Yamashita & Chang, 2001; Temperley, 2007; Ranjan & von der Malsburg, 2023; Zafar & Husain, 2023). However, the extent to which dependency length minimization is employed in a given language is not well understood, and the cognitive mechanisms through which this minimization is accomplished also remain unclear.

This study lays the foundation for a mechanistic account that explains dependency length minimization (DLM) for preverbal constituent ordering preferences in SOV languages, including *Basque*, *Hindi*, *Japanese*, *Korean*, *Latin*, *Persian*, and *Turkish*. We test the hypothesis that language users minimize a sentence's dependency length by placing only a short preverbal constituent (the shortest if possible) next to the main verb, and we posit that this better explains preverbal constituent ordering in SOV languages compared to global minimization of dependency length. We deploy large-scale corpus analyses and computational simulations using data from the Universal Dependency Treebank (UD; Zeman et al., 2022) to investigate our hypothesis. We refer to this hypothesis as a 'least-effort' strategy because it concurrently shortens the length of all preverbal dependencies connected to the main-verb without the need to simulate through an entire search space of possible constituent orders for globally minimizing the dependency length within the sentence.

Our hypothesis is grounded in a substantial body of evidence from behavioral economics, specifically the princi-

ples of bounded rationality (Simon, 1955, 1982, 1990, 1991; Gigerenzer & Selten, 2002; Gigerenzer & Gaissmaier, 2011; Gigerenzer, 2015). These principles suggest that human decision-making, constrained by cognitive resource limitations, has evolved to prioritize choices that are satisfactory rather than strictly optimal. In the current work, the proposed least-effort strategy serves as a satisfactory solution for the constituent ordering task, as it streamlines the search space of possible constituent orders for efficient communication.

To test ordering preferences, we conducted a simulation in which we created various counterfactual variants by randomly permuting the preverbal constituents of sentences sourced from the UD Treebank. The preverbal domain serves as the primary locus of constituent order variation in SOV languages, with fewer post-verbal constituents (Ranjan et al., 2022). Conceptually, Fig. 1 illustrates this generation process for an SOV sentence containing four preverbal constituents ( $C_i$ ) directly dependent on the main verb ( $V$ ). The reference sentence in Example 1a, as it originally appeared in the corpus, served as the basis for generating variant sentences in Examples 1b to 1d, among others ( $4! = 24$  such possibilities). Sentences that originally appeared in the corpus are considered a human preferred syntactic choice over those that are counterfactually generated (Ranjan & van Schijndel, 2024). Next, we conducted corpus analysis, where we examined the distributions of length of preverbal constituents and total dependency lengths within reference and variant sentences. We then deployed these two features in a logistic regression classifier for distinguishing corpus reference sentences from the counterfactual variant sentences.

Our results indicate that naturally produced corpus sentences largely conform to the constituent orders predicted by the hypothesized least-effort strategy (*i.e.* placing only a short preverbal constituent next to the main verb) across all tested SOV languages. Furthermore, the pressure to employ this strategy increases as the number of preverbal constituents grows, potentially suggesting that speakers aim to balance their production effort with their available computational capacity. Notably, for the task of classifying corpus reference sentences amidst counterfactually generated variants, this strategy significantly improves the accuracy of the regression model above and beyond the total dependency length across SOV languages, suggesting that it represents a ‘satisficing’ (satisfactory and sufficient) solution that humans tend to employ when making constituent ordering decisions.

Our primary contribution is that we demonstrate the role of bounded rationality in constituent ordering decisions across SOV languages and offer cross-linguistic evidence imperative for the development of the theories of language and cognitive science (Norcliffe, Harris, & Jaeger, 2015).

## Data and Methods

Our dataset comprises sentences from all head-final (SOV) languages that have a significant amount of data available in the publicly accessible Universal Dependency Treebank

Type	Hindi	Japanese	Korean	Turkish	Basque	Persian	Latin
Reference	10,223	24,198	18,868	25,928	4,628	23,969	17,294
Variant	308,571	325,958	262,696	204,783	35,988	618,070	191,053

Table 1: Total number of reference sentences that originally appeared in the corpus and corresponding artificially generated variant sentences in our dataset

(Zeman et al., 2022) corpus.<sup>1</sup> We specifically selected SOV languages that met the criteria of having at least 2000 sentences and projective dependency trees featuring a minimum of two preverbal constituents. This selection criterion resulted in 7 languages in our dataset: *Basque, Hindi, Japanese, Korean, Latin, Persian, and Turkish*. We implemented counterfactual variant generation for each reference sentence in our dataset. This process involved random permutation of preverbal constituents within the dependency tree, specifically those directly dependent on the root verb (see Fig. 1 and Fig. 3). We restricted the variant generation to a maximum of 120 variants per corpus reference sentence, an arbitrary cutoff to keep our computation tractable.<sup>2</sup> Table 1 presents the total number of generated variants for each reference sentence in the corpus across the languages in our dataset.

Subsequently, we conducted quantitative analyses by examining the distributions of constituent lengths and dependency lengths within both reference and variant sentences containing varying preverbal constituents.<sup>3</sup> We define CONSTITUENT LENGTH as the total number of words in a constituent as indicated by boxes around the words in Fig. 3. We define DEPENDENCY LENGTH as the count of intervening words between head and dependent units within a dependency graph (Temperley, 2007). Total dependency length is then calculated by summing the dependency length of each word within the sentence. Afterwards, we set up a classification task to investigate whether the aforementioned features are effective in identifying the corpus reference sentences amidst competing counterfactual variants (see Table 1). For corpus analyses reported in this paper, we focus on sentences with five or fewer preverbal constituents due to data sparsity issues with those containing six or more. However, we use the entire dataset, including sentences with 6 or more preverbal constituents, to compute our prediction results.

## Results

### Constituent Length Analysis

According to our least-effort strategy, the preverbal constituent adjacent to the main verb should be amongst the short preverbal constituents in the sentence. In contrast, the global DLM would predict a gradual decrease in the lengths of preverbal constituents as they approach the main-verb. To test these predictions, we computed the average length of constituents at various preverbal positions in the sentences from

<sup>1</sup>Version 2.11; <http://hdl.handle.net/11234/1-4923>

<sup>2</sup>Our findings are consistent regardless of chosen cutoff.

<sup>3</sup>We conducted a similar analysis using only grammatical variants and found consistent results.

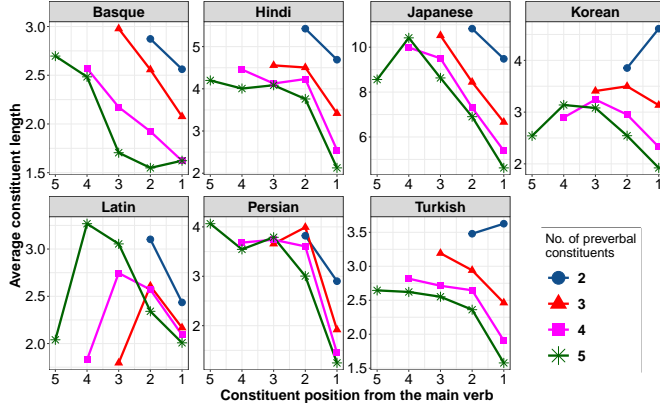


Figure 2: Average constituent length of preverbal constituents for corpus sentences with 2 to 5 constituents separately

the original corpus. Fig. 2 displays the average constituent length distribution for corpus sentences with 2 to 5 preverbal constituents in SOV languages.

In Persian, the average constituent length remains relatively consistent across positions until the position next to the main verb, where the length suddenly decreases. This suggests that the constituent immediately adjacent to the main verb is generally shorter than constituents in any other preverbal position in the sentence. This trend is consistent across all sentences types with varying numbers of preverbal constituents, confirming the least-effort strategy in Persian. A similar pattern is observed in Hindi. Turkish and Korean also follow a similar pattern for the most part, but with a deviation observed in sentences with two preverbal constituents. This deviation may indicate that these languages do not prioritize optimizing dependency length due to the low memory pressure associated with these sentences. Additionally, in Korean, while the shortest constituent on average is positioned next to the main verb, corpus sentences also often begin with a short constituent. These short constituents generally include left dislocated elements (29%) akin to clefting (Fernández-Sánchez & Ott, 2020), conjunctions (14%), clausal complements (13%), *inter alia*. This suggest that certain grammatical constraints, in addition to factors like information structure, discourse, and style, may take precedence over processing considerations. The plots for Basque and Japanese exhibit a more gradual decrease in average constituent length as one moves towards the main verb, with the on-average shortest constituent positioned next to the main verb.

Interestingly, in Latin (akin to patterns in Korean), while the shortest preverbal constituent is commonly positioned in the first position, the constituent at the last preverbal position is also short. Further linguistic analysis of Latin sentences revealed a specific stylistic preference, where a lot of sentences begin with the single-word conjunctions (23%) like “et”,<sup>4</sup> in addition to other categories (22%) involving short

phrases such as adverbial modifiers and negations, thereby considerably reducing the size of the first constituent in the sentence. Despite this stylistic preference in the corpus, the pressure to maintain a short preverbal constituent adjacent to the main verb compellingly illustrates the existence of expected pattern in Latin, consistent with other SOV languages under study. Additionally, Fig. 2 further illustrates that the pressure to place a short preverbal constituent adjacent to the main verb across SOV languages increases as the number of preverbal constituents with sentences grows, possibly due to increased memory load.

Taken together, these plots offer crucial evidence that naturally occurring sentences consistently exhibit a preference for optimizing only or at least primarily the length of the constituent next to the main verb in SOV languages. This approach may be employed to achieve a balance between production effort and the cognitive constraints experienced by listeners consistent with core assumptions of the framework of bounded rationality.

### Counterfactual Analysis

In this analysis, our objective is to simulate human preferences for constituent orders across SOV languages. We aim to identify which constituent orders closely align with the dependency lengths observed in the corpus data, a true representative of human behavior. For this, we compared the total dependency length of corpus reference sentences to four different types of variants, as shown in Fig. 3.

1. **RANDOM ORDER:** Randomly arrange the preverbal constituents in the sentence (average case, Fig. 3a).
2. **ASCENDING ORDER:** Arrange the preverbal constituents in increasing order of constituent lengths, resulting in maximal dependency length of the sentence (worst case, Fig. 3b).
3. **DESCENDING ORDER:** Arrange the preverbal constituents in decreasing order of constituent lengths, globally minimizing the dependency length of the sentence (best case, Fig. 3c).
4. **LEAST-EFFORT ORDER:** Start with any random order of preverbal constituents and then move shortest constituent next to main verb (efficient solution, Fig. 3d).

Fig. 4 illustrates the trends in average dependency lengths across reference and variant sentences with varying preverbal constituents and ordering types. The grey shaded region in the plot represents the entire spectrum of dependency length of sentences normalized by their sentence lengths, ranging from the maximal dependency length (achieved via ascending order of preverbal constituent lengths) to the minimal dependency length (via descending order). This shaded region serves as the strict boundary within which the language processor can adjust its dependency length by reordering preverbal constituents. Interestingly, we observe that the dependency length of corpus sentences, across languages, aligns with the dependency length of sentences predicted by least-effort. In essence, the dependency length obtained by the

<sup>4</sup>Example (Thomas Aquinas): *Et ita deum esse per se notum erit.* (‘And so it will be known that god exists by himself.’) In terms

of dependency length, *et* is an ideal candidate for placement next to the verb, but grammar prevents it.

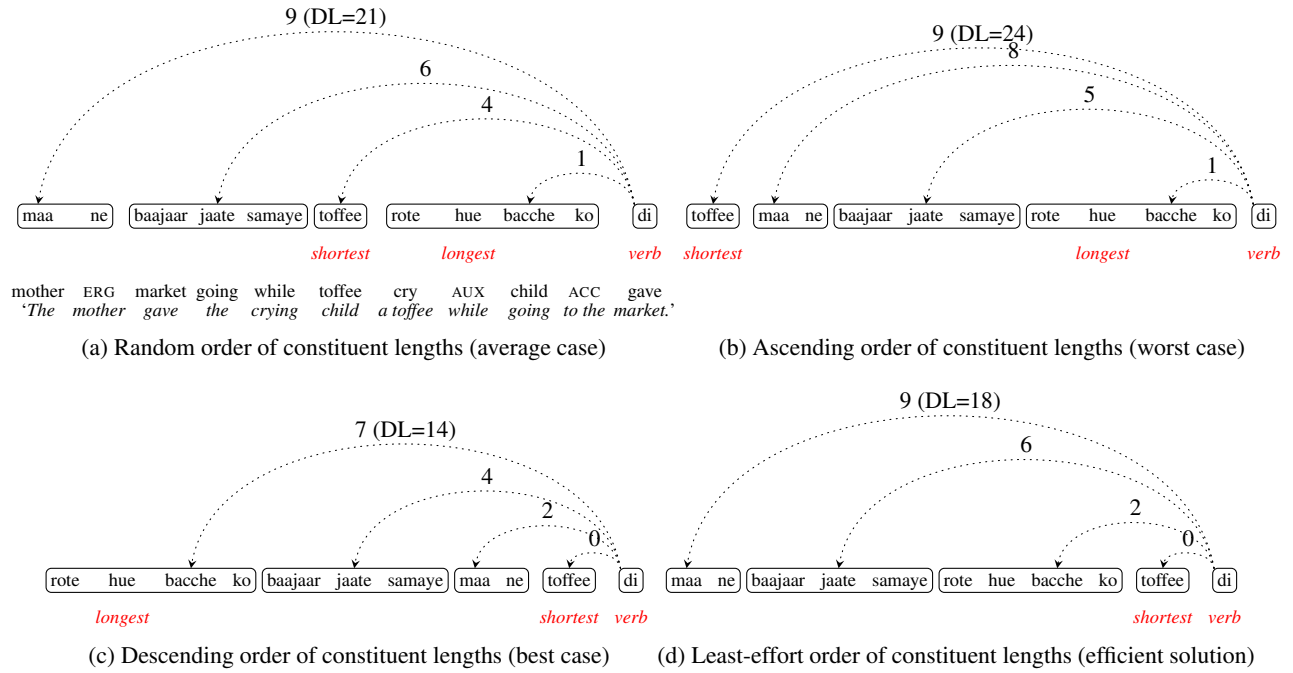


Figure 3: Preverbal constituent ordering in Hindi (SOV); Only main verb dependencies are depicted; Total dependency length (DL) and Constituent’s DL are indicated above each arc; Image reproduced from Ranjan and von der Malsburg (2023)

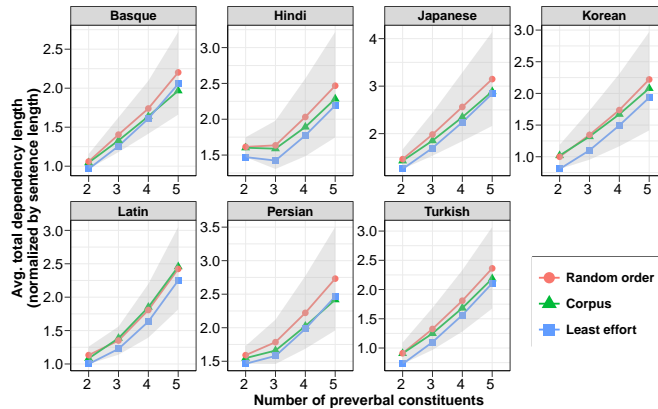


Figure 4: Average total dependency length (*i.e.*, total dependency length normalized by the number of words in a sentence) for different constituent orderings and for different numbers of preverbal constituents

least-effort strategy estimates the lower bound to the corpus sentences, indicating that the language system typically does not minimize dependency length beyond this point.

Through this simulation we enforced a strict constraint on the hypothesized least-effort strategy: first, identify the shortest preverbal constituent, and then position it next to the main verb. A boundedly rational language system would efficiently operate within this lower bound by placing only a short preverbal constituent next to the main verb, without concerning

itself with other preverbal constituents. This strategic optimization adopted by the speakers aims to minimize the sentence’s dependency length, reflecting considerations of memory load. Consistent with these insights, Fig. 4 also suggests that the tendency for corpus sentences to align with the least-effort solution strengthens as the number of preverbal constituents—and therefore memory load—goes up.

Additionally, based on these plots, we also find that corpus sentences, on average, exhibit lower dependency length than sentences with random preverbal constituent orderings. This trend aligns with previous studies (Liu, 2008; Gildea & Temperley, 2010; Futrell et al., 2015), except for Latin and, perhaps, Korean. As previously discussed, a significant number of corpus sentences in Latin begin with a single-word categories directly linked to the main verb. In case of Korean, although the shortest constituent is often next to the main verb, the sentences generally start with a short constituent. This results in a substantially longer dependency length within sentences for both the languages. Nonetheless, we contend that our least-effort strategy still holds in both Latin and Korean (see Fig. 2), as there is a consistent effort to reduce the preverbal constituent length next to the main verb to the extent possible. Therefore, we conjecture that a greater average dependency length for corpus sentences compared to sentences with random orders does not necessarily imply the absence of dependency length minimization in a language (Ferrer-i-Cancho & Liu, 2014). In summary, these observations reinforce the efficacy of the least-effort strategy in predicting constituent ordering choices in SOV languages.

## Corpus Prediction Task

If speakers employ the hypothesized least-effort strategy, we should be able to predict whether a sentence is a corpus reference sentence or a counterfactual variant by only examining the length of the preverbal constituent next to the main verb. Further, these predictions should be *better* than those obtained when total dependency length is used as the predictor. For this, we deployed logistic regression model to identify reference sentences amidst the generated variants using features such as preverbal constituent length next to the main verb (CL Last) and/or total dependency length (Total DL). Table 1 presents the count of corpus reference sentences and alternative variants for each language in our dataset. The presence of a greater number of variants per reference sentence introduces a substantial class imbalance issue (REFERENCE / VARIANT) for our binary classification task. The subsequent section addresses this class imbalance problem by transforming the dataset (explained below) with balanced class labels and then reports computational modeling analyses in Table 2.

**Ranking Model** In order to address the class imbalance issue for appropriate classification, we adopt an approach proposed by Joachims (2002), originally designed for ranking web pages. Previous work dealing with the syntactic choice prediction task (Ranjan et al., 2022), similar to our current work, has also employed this method. This approach converts a binary classification task into a pairwise ranking task involving feature vectors of a reference sentence and each of its variants. In this case, the feature vectors consist of the input features used for each reference and variant sentence in the model. Subsequently, we train a logistic regression model on the difference between feature vectors of reference and variant pairs, as illustrated in the equations below:

$$\mathbf{w} \cdot \phi(\text{Reference}) > \mathbf{w} \cdot \phi(\text{Variant}) \quad (1)$$

$$\mathbf{w} \cdot (\phi(\text{Reference}) - \phi(\text{Variant})) > 0 \quad (2)$$

Equation 1 represents a standard classification model that evaluates whether the reference sentence ranks higher than one of its variants. This decision is made by comparing the dot product of the feature vector associated with the reference sentence and the learned feature weights  $\mathbf{w}$  with the corresponding dot product of the variant sentence. This relationship can also be expressed in the form of Equation 2, where feature values of the first member are subtracted from the second member (Joachims, 2002). Now, the model’s decision for a particular referent-variant pair can be made by evaluating the sign of the dot product involving learned feature weight and difference between the feature vectors (see Equation 2).

We created ordered pairs consisting of feature vectors of reference (REF) and variant (VAR) sentences, ensuring that the counts of orders of each type (REF-VAR, VAR-REF) are balanced. For instance, the reference sentence 1a in Fig. 1 would generate three such pairs: (1a-1b), (1c-1a), and (1a-1d), where sentence (1a) is the corpus reference sentence since it happens to be originally present in the corpus, and

rest all are counterfactual variants (1b-1d). The pairs that alternate between ‘REF-VAR’ were assigned the label ‘1,’ while the pair ‘VAR-REF’ received the label ‘0.’ This alternate coding results in a balanced dataset and contains an equal number of labels of each type when the total number of variants is even, and a difference of one label when the total is odd.

The transformed feature values were incorporated into a logistic regression model using the `glm` function in R. We used the following `glm` equation<sup>5</sup> to investigate our hypothesis:  $\text{choice} \sim \delta \text{ constituent length}$  and  $\text{choice} \sim \delta \text{ dependency length}$ . Here, choice is a binary choice dependent variable (1 denotes reference sentence preference, and 0 denotes variant sentence preference). The delta ( $\delta$ ) refers to the difference between the feature vectors of reference sentence and its paired variant. All the independent variables (CL Last and Total DL) were normalised to z-scores, *i.e.*, the predictor’s value (centered around its mean) was divided by its standard deviation. We evaluate our model’s performance using 10-fold cross-validation and report the classification accuracy in Table 2b. However, for estimating regression coefficients shown in Table 2a, we used the entire transformed test data for our experiments. The Pearson’s correlation coefficient between ‘ $\delta$  CL Last’ and ‘ $\delta$  Total DL’ was consistently around 0.70 for all languages.

**Regression Analysis** Table 2a presents the regression coefficients (all significant with  $p < 0.001$ ) of different models on the entire test dataset for each language under investigation. In the models with individual feature (second column in Table 2a), the negative regression coefficient for ‘CL Last’ suggests that sentences naturally occurring in the corpus tend to have shorter constituents adjacent to the main verb than alternative variants across all SOV languages, thus supporting the presence of the least-effort strategy. Additionally, consistent with previous studies, the negative regression coefficient for ‘Total DL’ in the model with individual feature (third column in Table 2a) suggests that sentences in SOV languages consistently minimize their dependency length owing to memory constraints. When both these features are included in the model (see the last column in Table 2a), we consistently observe negative regression coefficients for both features across SOV languages, except for Basque where the coefficient sign for ‘CL Last’ flips. The variance inflation factor (VIF) for each feature in the aforementioned combined model consistently remained below 1.75 across all analyzed languages. We discuss Basque in the next subsection.

**Prediction Accuracy** Table 2b presents the classification performance of our various models on held-out dataset via 10-fold cross-validation to determine how many sentences are affected by each predictor. In terms of individual performance, the model with feature ‘CL Last’, corresponding to the least-effort strategy (second column in Table 2b), ac-

<sup>5</sup>We followed the R GLM format: the dependent variable is on the left of ‘ $\sim$ ,’ and the independent variables are on the right, with  $\delta$  indicating feature differences.

Language	CL Last	Total DL	Total DL + CL Last
Basque	-0.41	-0.83	-0.95, 0.17
Hindi	-1.15	-0.73	-0.18, -1.02
Japanese	-0.71	-0.68	-0.40, -0.44
Korean	-0.37	-0.31	-0.12, -0.29
Latin	-0.33	-0.26	-0.08, -0.28
Persian	-2.96	-1.22	-0.31, -2.76
Turkish	-0.61	-0.55	-0.27, -0.42

(a) Regression coefficients of distinct models with constituent length (CL) of last preverbal constituent and total dependency length (Total DL) as predictors; all regression coefficients are significant with  $p < 0.001$

Language	CL Last	Total DL	Total DL + CL Last
Basque	55.07	61.71	62.01
Hindi	69.49	63.39	69.23
Japanese	62.80	63.09	64.36
Korean	56.92	55.11	56.44
Latin	51.48	48.51	49.55
Persian	74.57	69.04	75.17
Turkish	61.72	60.00	62.02

(b) Classification accuracy (%) of distinct models (10-fold cross-validation) with constituent length (CL) of last preverbal constituent and total dependency length (Total DL) as predictors (Random accuracy = 50%)

Table 2: Logistic regression models determining corpus REFERENCE sentence and its paired counterfactual VARIANT

counted for the majority of our data for all SOV languages, except Basque and Japanese. In these two languages, ‘Total DL’ achieved slightly higher prediction accuracy than ‘CL Last’, with a small but statistically significant margin ( $p < 0.001$  using McNemar’s two-tailed significance test). Crucially, however, over a baseline model containing only ‘Total DL’ measure, adding the feature ‘CL Last’ (see the last column in Table 2b) induced a significant increase in classification accuracy across all SOV languages, including Basque ( $p < 0.001$  using McNemar’s two-tailed significance test compared to the previous column in Table 2b). Therefore, despite the sign flip in the regression coefficient of ‘CL Last’ in the combined model for Basque (last column in Table 2a), the significant increase in 10-fold cross-validation accuracy suggests that the preverbal constituent next to the main verb, typically shorter in reference sentences than in the paired variants, continues to play an important role in Basque as well. Thus these observations highlight the effectiveness of our least-effort strategy in predicting constituent ordering choices across SOV languages.

## Discussion

While the minimization of dependency length has been recognized as a universal quantitative property of natural languages, this work investigates the extent of minimization and the cognitive mechanisms driving the process. We found that speakers of SOV languages, consistent with the central ideas of bounded rationality in decision-making, choose constituent orders that minimize dependency length by employing a least-effort strategy as a fundamental mechanism. This strategy involves placing a short preverbal constituent, possibly the shortest one, next to the main verb, rather than aiming for the globally optimal solution, *i.e.* minimizing the dependency length entirely in the sentence. We also found that the pressure to move a short preverbal constituent next to the main verb increases as the number of preverbal constituents grows, suggesting a trade-off between production effort and limited cognitive resources across SOV languages. Finally, for the task of identifying corpus sentences amidst the alternative variants, the feature adhering to least-effort strategy significantly improved the prediction accuracy beyond total dependency length for all SOV languages under study.

Consistent with previous work (Liu, 2008; Gildea & Tem-

perley, 2010; Futrell et al., 2020), our experiments show that corpus sentences consistently exhibit lower dependency length than sentences with random preverbal constituent orderings in all SOV languages, except Latin. Despite the stylistic preference to place single-word conjunctions at the beginning of Latin sentences, the preverbal constituent adjacent to the main verb still indicates an attempt to keep dependency length small, confirming the least-effort strategy for Latin as well. Our fine-grained analysis therefore shows that a lack of total dependency length minimization (compared to random constituent orders) does *not* imply that producers of Latin forgo opportunities to optimize dependency length. Moreover, the example of Latin also illustrates how language users formulate sentences not only by optimizing dependency length; their choices are simultaneously influenced by many other factors, including grammar, discourse, communicative efficiency, and style. The proposed least-effort strategy necessarily operates in concert with these other constraints.

Besides, our interpretation of bounded rationality may hold for both speakers and listeners, since both rely on shared linguistic cues to navigate the challenges associated with word order flexibility (Chater, McCauley, & Christiansen, 2016; Mamma & Phillips, 2018). Future work needs to test the proposed strategy for real-time production and comprehension, in addition to developing a full-fledged mechanistic account of the cognitive processes underlying DLM.

The broader implication of our findings is that flexibility in language and constituent order preferences can be understood through the framework of bounded rationality in human behavior. This behavior is limited by human’s cognitive abilities, information availability, and the time pressure under which individuals must respond (Simon, 1955, 1982). These perspectives emphasize that language, as a tool for communication, has evolved to align with the cognitive abilities of its users, making it both efficient and comprehensible. Simultaneously, these perspectives also offer valuable insights into how linguistic patterns and preferences emerge as a result of humans having to deal with their cognitive and informational limitations (Gigerenzer, Hertwig, & Pachur, 2011).

Overall, our results provide converging evidence that preverbal constituent ordering preferences in SOV languages are shaped by the minimization of dependency length within the constraints of bounded rationality.



## Acknowledgments

We thank Daniel Hole, Judith Tonhauser, Lisa Hofmann, Samar Husain, Shravan Vasishth, and the attendees at the Linguistic Evidence (LE-2024) conference, the Sentence Processing colloquium at the University of Potsdam, and the Linguistic Theory colloquium at the University of Stuttgart for their insightful comments and feedback on this work. We would also like to express our gratitude to the anonymous reviewers of SAFAL-2023, EACL-2024, LE-2024, and CogSci-2024 for their invaluable comments and suggestions.

## References

- Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, 89, 244 - 254. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0749596X15001412> (Speaking and Listening: Relationships Between Language Production and Comprehension) doi: <https://doi.org/10.1016/j.jml.2015.11.004>
- Chomsky, N. (1965). *Aspects of the theory of syntax* (Vol. 11). The MIT press.
- Fernández-Sánchez, J., & Ott, D. (2020). Dislocations. *Language and Linguistics Compass*, 14(9), e12391.
- Ferrer-i-Cancho, R., & Liu, H. (2014). The risks of mixing dependency lengths from sequences of different length. *Glottometry*, 5(2), 143–155.
- Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, 96(2), 371–412.
- 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. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 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: 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>
- 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., 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.
- 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.
- Joachims, T. (2002). Optimizing search engines using click-through 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
- 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>
- Momma, S., & Phillips, C. (2018). The relationship between parsing and generation. *Annual Review of Linguistics*, 4(1), 233–254. Retrieved from <https://doi.org/10.1146/annurev-linguistics-011817-045719> doi: 10.1146/annurev-linguistics-011817-045719
- Norcliffe, E., Harris, A. C., & Jaeger, T. F. (2015). *Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances* (Vol. 30) (No. 9). Taylor & Francis.
- Rajkumar, R., van Schijndel, M., White, M., & Schuler, W. (2016). Investigating locality effects and surprisal in written english syntactic choice phenomena. *Cognition*, 155, 204–232. Retrieved from <http://www.sciencedirect.com/science/article/pii/S001002771630155X/>
- 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>
- Ranjan, S., & van Schijndel, M. (2024). Does dependency locality predict non-canonical word order in Hindi? In L. K. Samuelson, S. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th annual meeting of the cognitive science society*. Rotterdam, Netherlands: Cognitive Science Society.
- Ranjan, S., & von der Malsburg, T. (2023). A bounded rationality account of dependency length minimization in hindi. In M. Goldwater, F. Anggoro, B. Hayes, & D. Ong (Eds.), *Proceedings of the 45th annual meeting of the cognitive science society*. Sidney, Australia: Cognitive Science Society. Retrieved from <https://escholarship.org/uc/item/85q7x1nd>



- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99–118.
- 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>
- von Humboldt, W. (1836). *Über die verschiedenheit des menschlichen sprachbaues*. Dr. d. Kgl. Akad. d. Wiss.
- Yamashita, H., & Chang, F. (2001). “Long before short” preference in the production of a head-final language. *Cognition*, 81.
- Zafar, M., & Husain, S. (2023). Dependency locality influences word order during production in SOV languages: Evidence. In *Proceedings of the annual meeting of the cognitive science society*, 45 (45).
- Zeman, D., Nivre, J., Abrams, M., Ackermann, E., Aepli, N., Aghaei, H., ... Ziane, R. (2022). *Universal dependencies 2.11*. Retrieved from <http://hdl.handle.net/11234/1-4923> (LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University)