# Do speakers minimize dependency length during naturalistic dialogue?

**Meghna Hooda** **Mudafia Zafar** **Samar Husain**
Indian Institute of Technology Delhi
meghnahooda1706@gmail.com mudafia@hss.iitd.ac.in samar@hss.iitd.ac.in

## Abstract

Dependency Length Minimization (DLM) is considered to be a linguistic universal governing word order variation cross-linguistically. However, evidence for DLM from large-scale corpus work is typically based on written (news) corpus and its effect on sentence production during naturalistic dialogue is largely unknown. Furthermore, Subject-Object-Verb languages are known to show a weaker preference for DLM. In this work, we test the validity of DLM using a dialogue corpus of Hindi, an SOV language. We also undertake a quantitative analysis of various syntactic phenomena that lead to DLM and compare the effect of DLM on both spoken and written modalities. Results provide novel evidence supporting a robust effect of DLM in spoken corpus. At the same time, compared to the written data, DLM was found to be weaker in dialogue. We discuss the implications of these findings on sentence production and on methodological issues with regards to the use of corpus data to investigate DLM.

## 1 Introduction

Understanding the structural complexity of natural language has been a key goal in psycholinguistics (e.g., Miller, 1962; Kimball, 1973; Hawkins, 1990; Levelt, 1972). This is because the formal properties of natural language can help us to uncover the underlying processes that subserve the generation and comprehension of such structures (Frazier, 1987; Levelt, 1989). These proposals are informed by our understanding of the severe resource constraint under which a dynamic system such as language comprehension/production operates (e.g., Just and Carpenter, 1992). An influential way to formalize complexity has been in terms of the arrangement of words in a sentence (Hudson, 1995; Wasow, 2002). On this account, called Dependency Length Minimization (DLM), two words that are syntactically related to each other would tend to appear in close proximity rather than away from each other (Gibson, 1998, 2000). DLM can be understood in terms of optimizing limited memory resources – establishing a dependency relation between two words will typically require memory retrieval (of the head or the dependent), and these retrievals are known to be subject to locality considerations (Gibson, 1998; Lewis and Vasishth, 2005). This implies that sentences with shorter dependencies will, on average, be easier to process. Indeed, there is experimental evidence that an increase in dependency length leads to difficulty during comprehension as well as production (Grodner and Gibson, 2005; Bartek et al., 2011; Scontras et al., 2015).

Recent corpus-based work has provided strong cross-linguistic validation for DLM (Liu, 2008; Gildea and Temperley, 2010; Futrell et al., 2015; Temperley, 2007). These studies clearly demonstrate that DLM can be deemed as a linguistic universal across languages. If true, this has implications for the design properties of natural language and its architectural underpinnings (Futrell et al., 2020). However, a key issue with this claim is that corpus-based evidence for DLM mostly comes from written data (e.g., news genre). While both speaking and writing involve the same production apparatus, it is easy to see that they may not be operating under similar constraints. For instance, the production system can be assumed to be under more time pressure when speaking than writing, where it is typical to make many edits to a sentence (Biber, 2009; Hayes and Flower, 1986; Chafe, 1985). One reason for this is that the visual feedback during writing is more stable while the acoustic feedback during speech is momentary. Thus, it is reasonable to assume that the DLM constraint might be more evident in written text where the writer tries to achieve high readability for the reader. Meanwhile, in speech, other speaker-centric pressures related to incrementality, accessibility, etc., could supersede the DLM constraint (cf.



*Proceedings of the Society for Computation in Linguistics (SCiL) 2024*, pages 139-149.
Irvine, California, June 27-29, 2024

Levelt, 1989; Gleitman et al., 2007; Wheeldon and Konopka, 2023).

Another interesting finding in the literature is that while DLM operates cross-linguistically, it does not appear to be as strong across all languages. In particular, research suggests that the effect of DLM in Subject-Object-Verb (SOV) languages is less strong (Futrell et al., 2020; Dyer, 2023; Liu, 2020). Indeed, recent work using news data suggests that DLM has a marginal role in determining word order variation in an SOV language like Hindi (Ranjan et al., 2022). Thus, it is unclear if the DLM constraint would also hold in naturalistic spoken data in an SOV language, Hindi.

To summarize, there are two reasons to doubt the cross-linguistic generalizability of DLM: (a) large-scale validation of DLM has primarily been observed with written data, and (b) the effect of DLM has been observed to be weaker in SOV languages. In this work, we investigate if DLM is indeed operational in an SOV language Hindi during naturalistic dialogue. Further, if we do find evidence for DLM in the spoken modality, we are interested in probing the source of this effect. In particular, we investigate two well-known word order related phenomena that are known to be triggered by DLM, these are, the long-before-short pattern (Hawkins, 2014) and right-extraposition (Wasow, 1997b). Finally, we compare the strength of DLM in the spoken vs written modality.

The paper is arranged as follows: in Section 2, we present our key experiment on investigating DLM using random baselines. In Section 3, we probe the results regarding DLM in dialogue corpus using two word order related phenomena. Following this, in Section 4, we compare the findings of the DLM experiment on dialogue corpus with written corpus. We consolidate all the findings and discuss their implications in Section 5. Section 6 concludes the paper.

## 2 DLM during dialogue

This section presents the key investigation of our work, i.e., can DLM be observed during naturalistic dialogue in an SOV language, Hindi? In order to test this question, we conduct a corpus-based study using the methodology proposed in Liu (2008); Futrell et al. (2015); Liu et al. (2017); Yadav et al. (2019). In particular, we compare real trees in a Hindi dialogue corpus with random baseline trees that match the real trees in certain formal properties. The DLM distribution between these pairs of trees is compared to investigate the question at hand.

### 2.1 Data

The IIT Delhi Hindi Dialogue Corpus (Pareek et al., 2023) was used for the study. The dialogue data comprises of the Hindi segment of the CallFriend project (Canavan and George, 1996), which consists of 60 unscripted telephone conversations between Hindi native speakers. The spoken data was manually transcribed and later was (semi-)automatically annotated for part-of-speech and syntactic dependency relations. All annotations were finally validated manually. The current study is based on data comprising 31,020 sentences (mean sentence length = 6.13). For the purpose of this study, this dataset underwent a filtering process involving the exclusion of sentences containing code-switching, quotations, and incomprehensible content (i.e., words that were transcribed as *incomprehensible* because the audio was not clear). Non-lexical tokens such as laughter, pauses, and noise were also removed from the sentence. Finally, tokens representing disfluencies were also removed. This left us with a dataset comprising 28,953 sentences (mean sentence length = 5.68).

For generating the random baselines, we further subset this data to exclude sentences with lengths less than 3 and more than 19.[1] This gave us the final data comprising 22414 sentences that were used to generate the random baselines. The average sentence length in this data was 7.67.

### 2.2 Random Baselines

Following Yadav et al. (2022a,b), we generate a random baseline called random linear arrangement baselines (RLAs) for the real dependency trees obtained from the Hindi Dialogue Corpus. The algorithm chooses a random baseline tree from a uniform distribution of random linearization of a real tree through a rejection sampling method. The random tree is controlled for sentence length, the number of crossing dependencies and all topological properties (e.g., node arity, tree depth, etc.). Critically, the baseline preserves the dependency relations of a real tree. This makes RLAs a relatively strict baseline compared to simple random

[1] This was necessitated because the compute time to generate the conservative baselines for sentences more than 19 was very high. Sentences with less than 3 words were removed because the random baselines generated for such sentences remain invariant. Note that the sentence length was computed by excluding the punctuation.

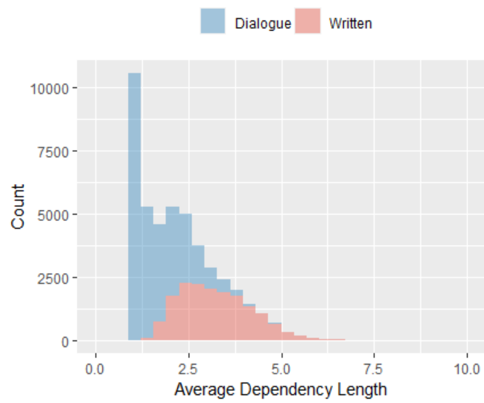


Figure 1: Distribution of Average Dependency Length in Dialogue and Text Corpus

structure baselines where the tree topology is not controlled (cf. Futrell et al., 2015). A random tree is generated for each real tree. We then compare the two trees to test if the average dependency length between the two trees differs (for more details see Yadav et al., 2022b).

### 2.3 Statistical Method

In order to investigate if DLM is indeed observed during a naturalistic speech in Hindi, we test whether the distribution of dependency length is significantly different between real trees and the baseline trees. Dependency length was computed as the number of words intervening between the head and its dependent. We compare the growth of average dependency length with sentence length in real vs random trees. If DLM holds in speech data, then compared to random baselines, this growth in real trees should be slower. We fit a linear mixed-effects model using the *lme4* (Bates et al., 2015) package in R (R Core Team, 2022) with dependency length as the dependent variable, sentence length, and tree type (real or random) as fixed effects, and Interlocutor-pair as the random effect (see Eq 1). The random effect captures the variation across different speakers in the dataset. The key coefficient of interest in the model is the interaction between sentence length and tree type. Note that in treating the interaction as the coefficient of interest we follow Ferrer-i Cancho and Liu (2013); Futrell et al. (2015); Gildea and Temperley (2010) who show that dependency length should be considered as a function of sentence length. This is because the effect of capturing the average difference of dependency length between real and random trees could be inaccurate, as dependencies could come from varying sentence lengths.

$$DL \sim Sentence.length * Tree.Type + (Sentence.length * Tree.Type | Interlocutor - pair) \quad (1)$$

Maximal models were fit, subject to model convergence (Barr et al., 2013).

### 2.4 Results

Table 1 shows the results of the linear mixed model analysis. Results show that the average dependency length grows slower with sentence length in real trees compared to random trees (p-value$< 0.001$). This can also be visually observed in Figure 2. These results show that DLM is observed in the dialogue corpus.

Table 1: Results from the linear mixed models. Tree.Type (Random vs Real) was coded as treatment contrast with the random tree as the baseline. Sentence length (SL) was scaled.

| | estimate | SE | t-value | Pr(>ltl) |
|---|---|---|---|---|
| Intercept | 2.34 | 0.005 | 397.79 | <0.001 |
| SL | 0.54 | 0.005 | 104.97 | <0.001 |
| Real | -0.36 | 0.01 | -31.99 | <0.001 |
| SL:Real | -0.17 | 0.008 | -21.08 | <0.001 |

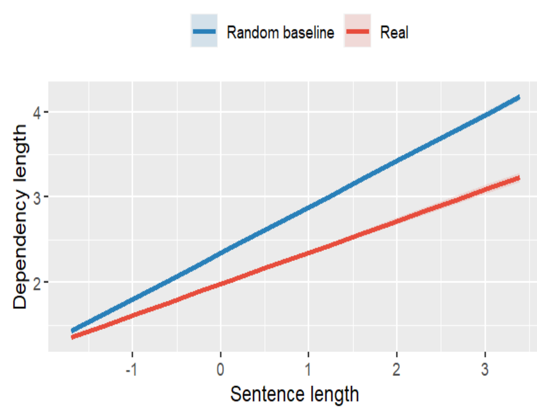


Figure 2: Fitted models showing the growth of dependency length with respect to sentence length in real language trees compared to Random Linear Arrangement (RLA) baselines.

## 3 What drives DLM?

Large-scale corpus based investigations are important because they help us in testing the ecological validity and generalizability of a theory like DLM.

At the same time, we also need to uncover the underlying cause for the observed results in terms of various syntactic configurations. One such configuration concerns the length of a constituent – the length of a constituent has been argued to guide word order changes that minimize the overall dependency length in that utterance (Hawkins, 1994, 2004, 2014). With regard to SOV languages, a long-before-short order of constituents can be deemed to lessen the overall dependency length compared to a short-before-long order. This can schematically be seen in Figure 3.

Evidence for a long-before-short preference has been found from production experiments in SOV languages (Yamashita and Chang, 2001; Ros et al., 2015; Faghiri and Samvelian, 2020). For example, in Japanese sentences 1a-1b, consisting of a short subject and a long object, Yamashita and Chang (2001) found that speakers produced more non-canonical (OSV) order sentences like 1b than the canonical (SOV) like 1a.

(1) a. [$_S$ keezi-ga] [$_O$ se-ga takakute
detective-NOM height-NOM tall
gassiri sita hannin-o]
and big-boned suspect-ACC
[$_V$ oikaketa]
chased

'The detective chased the suspect who is tall and big-boned.'

b. [$_O$ se-ga takakute gassiri sita hannin-o] [$_S$ keezi-ga] [$_V$ oikaketa]

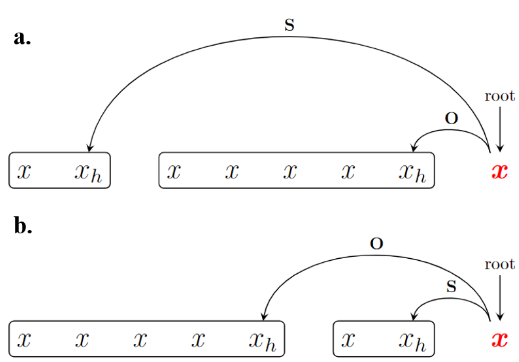


Figure 3: Two possible ordering patterns for a short subject (S) and a long object (O) in SOV languages. In (a) the short S is placed first while in (b), the long O is placed first. Order (b) has a shorter dependency length than (a).

The other configuration that has been argued to be driven by DLM is right-extraposition (Hawkins, 2014; Wasow, 1997b; Arnold et al., 2000; Wasow and Arnold, 2003; Szmrecsányi, 2004; Yngve, 1960; Gibson, 1998).[2] In the context of SOV languages, this would imply placing the phrase in question *after* the clause final verb. Here we explore two such configurations, one where the noun is modified by a relative clause (Kothari, 2010; Zafar and Husain, 2023) (see, Example 2a, 2b), and the other where the noun is modified by a non-verbal phrase (e.g., another noun phrase; Example 3a, 3b).

### 3.1 Long-before-short order

In the previous section we demonstrated DLM in the Hindi Dialogue corpus. In this section we ask if the DLM effect is driven by a long-before-short pattern. We do this by examining the word order patterns for core arguments such as subjects and objects;[3] in addition, we also investigate this effect for adjuncts. In particular, we investigate if increase in object/adjunct length will increase a shift from canonical to non-canonical OSV/AdjSV order to align with a long-before-short pattern and whether this shift leads to reduced average dependency length of the utterance. The key prediction is that preverbally, a long argument/adjunct should be placed before a short argument/adjunct when this leads to reduced dependency length.

#### 3.1.1 Method

We extracted SOV and OSV utterances from the dialogue corpus. While doing so, we ensured that the sentences had (a) only S and O as the two core arguments, (b) both arguments were dependent on the same verbal head, (c) both the arguments preceded the verb, and (d) did not involve any crossings. For computing phrasal length, case-markers were considered as part of their respective noun. We obtain 1152 SOV instances and 274 OSV instances for the analysis.

Generalized linear models with the logit function (Nelder and Wedderburn, 1972) were fit to the data. The order of arguments (SOV or OSV) was the dependent variable, and phrasal length was the independent variable consisting of 3 levels: Equal; Subject Long and Object Long. The Equal condition served as the baseline because DLM would

[2] Explanations for right-extraposition have traditionally been cast in terms of phrasal length or complexity (e.g., Yngve, 1960; Wasow, 2002) However, this point is not relevant for the current discussion as right-extraposition due to increased phrasal length leads to DLM (see, Yadav et al., 2022a).

[3] We did not include ditransitive sentences as part of the analysis because they were less in number.

not prefer one order over the other in this condition. Scaled average dependency length was also added as an additional predictor (see Equation 2). The effect of Object Long, as well as its interaction with average dependency length, formed the coefficients of interest – this is because while the effect of Object long captures the shift due to length, the interaction tells us if this shift leads to DLM. Specifically, if the long-before-short effect exists in dialogue then we expect the likelihood of shifts from the SOV to the OSV order to increase when the Object is long compared to when both the Subject and the Object are of equal length. This would mean that the coefficient of Object Long should have a positive sign. Additionally, the interaction of dependency length with Object Long should have a negative sign. This would tell us that the likelihood of shifts to the OSV order in the O Long condition decreases as the dependency length of the sentence increases. This implies that higher OSV shifts when the Object is long correspond to lower DLM.

$$Order(SOV|OSV) \sim (Subject.Long + Object.Long) * Avg.Dependency.Length \quad (2)$$

We additionally investigated shifts from the Subject Adjunct Verb (SAdjV) patterns to the Adjunct Subject Verb (AdjSV) pattern as a function of length. This analysis helps us investigate if the long-before-long order exists irrespective of the nature of the verbal modification, i.e., argument or adjunct. Using the criterion mentioned for SOV sentences previously, 699 SAdjV instances and 644 AdjSV instances were extracted for analysis. All sentences had only one core argument – the subject, and one adjunct. The *glm* model for this analysis is shown in Equation 3 and is similar to the analysis we ran for argument shifts.

$$Order(SAdjV|AdjSV) \sim (Subject.Long + Adjunct.Long) * Avg.Dependency.Length \quad (3)$$

#### 3.1.2 Results

Table 2 shows the results. With regard to the SOV/OSV analysis, we find that compared to the Equal condition, in sentences with long objects, the tendency to place the object initially (leading to a long-before-short OSV order) increases (p=0.01) (also see Figure 4). In addition, we also observe a significant interaction between object length and average dependency length (p=0.03) such that compared to the equal length subject/object, the proportion of OSV order decreases when the average dependency length increases. An opposite trend was observed for utterances with long subjects suggesting that the tendency to shift decreased when the subject was longer than the object.[4]

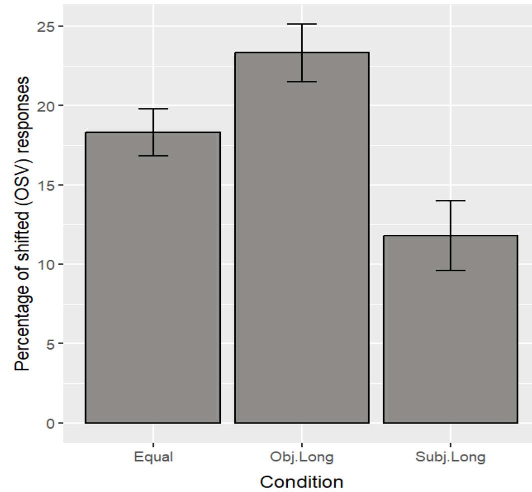


Figure 4: Percentage of Object-fronted responses in the Equal, Subject-Long and Object-Long conditions in the Dialogue Corpus.

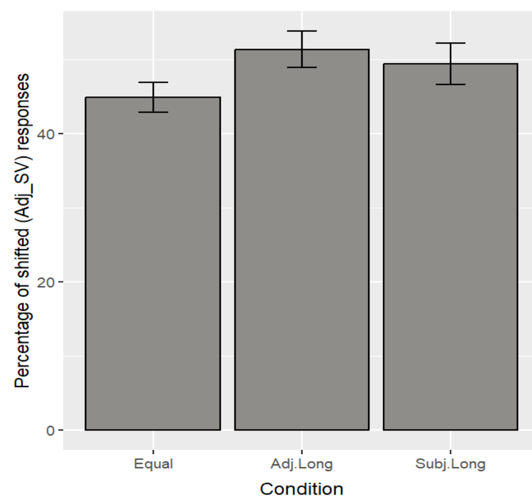


Figure 5: Percentage of Adj-fronted responses in the Equal, Subject-Long and Adj-Long conditions in the Dialogue Corpus.

[4] A careful reader will observe that the effect of Avg.DL in the models is not really meaningful for the discussion at hand; it represent the effect of dependency length in the utterances

A similar long-before-short pattern was found in the SAdjV/AdjSV utterances. Long adjuncts were fronted to form an AdjSV more when the adjunct was longer than the subject compared to when they were of equal length (p=0.009) (see Figure 5). In addition, there was a significant interaction between the length of the adjunct and average dependency length (p<0.001) such that the tendency to form an AdjSV utterance reduced with increased dependency length. Broadly, the above results show that during dialogue, speakers follow a long-before-short pattern for DLM.

### 3.2 Right-extraposition

The long-before-short word order configuration discussed above leads to reduced dependency length through changes in word order preverbally. However, as discussed earlier, in certain configurations dependency length in SOV languages can also be minimized by placing a long phrase *after* the matrix verb via right extraposition. Below we report the analysis for two types of right extrapositions, one where the noun is modified by a relative clause (Kothari, 2010; Zafar and Husain, 2023) (see, Example 2b), and the other where the noun is modified by a non-verbal phrase (e.g., another noun phrase; Example 3b).

#### 3.2.1 Method

We began by extracting all instances of nominal modifications from the dialogue corpus. These (clausal or non-clausal) modifiers could either appear *in-situ* or could be right-extraposed postverbally.[5] This gave us 199 right-extraposed utterances and 1710 in-situ utterances. See examples 2a and 3a for clausal and non-clausal *in-situ* modifications, respectively; Examples 2b and 3b show their right-extraposed counterparts.

(2) a. vo vali job **jo** **kar rahey**
DEM PART job REL do PROG
**thein** chor diye
be.PST.PRF leave give.PST.PRF
'I left that job which I was doing'

b. vo vali job chor diye **jo kar rahey thein**

(3) a. aur **tumhare dushman ki** shadi
and your enemy POSS marriage
ho gayi
be.PRES go.PST.PRF
'And your enemy got married?'

b. aur shadi hogayi **tumhare dushman ki**

Similar to section 3.1, we ran a generalized linear model with the logit link function where the order (Right-extraposed or In-situ) was the dependent variable, and scaled phrasal length was the independent variable. Scaled average dependency length was added as an additional predictor (see Equation 4). If right-extraposition is driven by dependency length minimization, then we should observe an effect of phrasal length such that as phrasal length increases, right-extraposition should increase, i.e., we should observe a positive sign on the coefficient. In addition, we ought to also observe a negative coefficient for the interaction between phrasal length and dependency length – this will suggest that shifts to the right-extraposed order are less likely when such shifts increase the total dependency length of the sentence. Together the two effects would imply that right-extraposition for long phrases correspond to lower DLM.

$$Order(RightExtraposed|InSitu) \sim \\ PhrasalLength * DependencyLength \quad (4)$$

#### 3.2.2 Results

Table 3 shows the results for the glm analysis. The key finding was that the shift to the right-extraposed order increased with an increase in phrasal length ($p<0.001$). However, the interaction between phrasal length and dependency length was also positive. This means that the shift from in-situ to right-extraposition in fact increased with an increase in dependency length for higher values of phrasal length( p=0.01). Thus, the results suggest that the increase in right-extraposition for long phrases in dialogue is not driven by DLM.

## 4 DLM in speech vs written text

As discussed in Section 1, a key difference between dialogue and written text concerns the time window under which the final utterance is produced. Typically, the time available to produce an utterance (such as the ones found in the current dialogue corpus) during naturalistic dialogue will be much less than the time taken to produce an edited sentence in written corpus. Indeed, it is well known that turn-taking during dialogue is very fast (Clark, 2014). This suggests that DLM, which will require considerable resources due to planning, could be more

where the length of both subject and object is equal.

[5] For clausal modifications *in-situ* implies a post-nominal modification; while for non-clausal modifications *in-situ* means a pre-nominal modification.

Table 2: Word Order analysis for the long-before-short experiment in Section 3.1. Treatment contrast was used in both models (Equal length argument/adjunct formed the baseline); ARG = Arguments, ADJ = Adjuncts, Avg.DL = Average Dependency Length, Subj.Long = Subject longer than Object, Obj.Long = Object longer than Subject, Adj.Long = Adjunct longer than Subject. Avg.DL was scaled before fitting the model. Significant effects where p-value<0.05 have been highlighted.

| | | Dialogue | | | | Written | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | estimate | SE | z-value | p-value | estimate | SE | z-value | p-value |
| ARG | Intercept | -1.59 | 0.09 | -17.2 | <**0.001** | -2.6 | 0.06 | -38.1 | **<0.001** |
| | Subj.Long | -0.61 | 0.26 | -2.36 | **0.001** | -0.27 | 0.18 | -1.51 | 0.12 |
| | Obj.Long | 0.36 | 0.14 | 2.49 | **0.01** | -0.11 | 0.15 | -0.71 | 0.47 |
| | Avg.DL | 0.06 | 0.08 | 0.78 | 0.43 | -0.26 | 0.07 | -3.8 | **0.0001** |
| | Subj.Long:Avg.DL | 0.80 | 0.21 | 3.69 | <**0.001** | 0.13 | 0.18 | 0.74 | 0.45 |
| | Obj.Long:Avg.DL | -0.32 | 0.15 | -2.07 | **0.03** | -0.42 | 0.16 | -2.53 | **<0.01** |
| | | Dialogue | | | | Written | | | |
| | | estimate | SE | z-value | p-value | estimate | SE | z-value | p-value |
| ADJ | Intercept | -0.02 | 0.57 | -0.33 | 0.73 | 0.05 | 0.03 | 1.43 | 0.15 |
| | Subj.Long | 0.18 | 0.13 | 1.32 | 0.18 | 0.3 | 0.09 | 3.3 | <**0.001** |
| | Adj.Long | 0.34 | 0.13 | 2.58 | **0.009** | -0.03 | 0.09 | -0.4 | 0.68 |
| | Avg.DL | -0.05 | 0.05 | -0.93 | 0.35 | -0.01 | 0.03 | -0.38 | 0.69 |
| | Subj.Long:Avg.DL | 0.08 | 0.14 | 0.62 | 0.53 | 0.28 | 0.09 | 3 | **0.002** |
| | Adj.Long:Avg.DL | -0.54 | 0.13 | -4.02 | <**0.001** | -0.32 | 0.0.9 | -3.49 | <**0.001** |

Table 3: Results from the glm models for right extraposition in dialogue and written. Treatment contrast was used in the model. DL: Dependency Length and PL: Phrasal Length. Significant effects where p-value<0.05 have been highlighted.

| | Dialogue | | | | Written | | | |
|---|---|---|---|---|---|---|---|---|
| | estimate | SE | t-value | p-value | estimate | SE | t-value | p-value |
| Intercept | -2.27 | 0.08 | -28.09 | **<0.001** | -4.29 | 0.05 | -74.136 | **<0.001** |
| PL | 0.5 | 0.06 | 7.53 | **<0.001** | 0.97 | 0.02 | 37.13 | **<0.001** |
| DL | -0.03 | 0.08 | -0.41 | 0.67 | -0.02 | 0.056 | -0.52 | 0.6 |
| PL:DL | 0.16 | 0.06 | 2.39 | **0.01** | -0.1 | 0.01 | -5.67 | **0.01** |

visible in written corpus while speaker-centric factors such as accessibility, etc., could be more dominant in speech production (cf. Arnold et al., 2000; MacDonald, 2013; Ferreira and Dell, 2000). Below we investigate this possibility.

### 4.1 Method

To investigate the strength of DLM in dialogue and written data, we follow the method discussed in Section 2. Similar to the experiment for the dialogue corpus, RLA random baselines were used for comparison. To do this comparison, we needed a baseline that can be compared with both dialogue as well as written text trees. So, we select those sentences in the dialogue and written data that match in three critical topological features – sentence length, max arity, and max tree depth. This enables us to generate an RLA baseline, which can be compared with real trees from two modalities. Using this criterion, we got 869 triplets of baseline and dialogue/written trees.

$$Dependency.Length \sim Sentence.length * Tree.Type \tag{5}$$

A linear model was used to investigate the increase in dependency length with respect to sentence length in random vs real trees of dialogue and written text (Equation 5). As before, the key coefficient (which should be negative) will be the 2-way interactions between real trees in dialogue/written text and sentence length.

### 4.2 Results

Table 4 shows the results. The results show a significant interaction between sentence length with

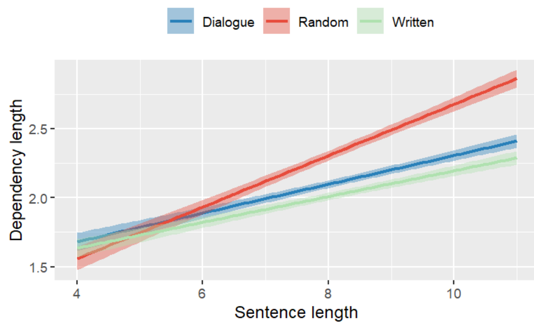


Figure 6: Fitted models showing the growth of dependency length with respect to sentence length in real trees from dialogue and written data compared to RLAs.

Table 4: Estimates from the fitted linear models comparing DL in real and RLAs baselines tree for the two modalities for the effect of dependency length(scaled). SL = sentence length.

| | Estimate | SE | t-value | p-value |
|---|---|---|---|---|
| Intercept | 2.32 | 0.01 | 151.23 | <**0.001** |
| SL | 0.35 | 0.01 | 23.07 | <**0.001** |
| $\text{Real}_D$ | -0.21 | 0.02 | -9.99 | <**0.001** |
| $\text{Real}_W$ | -0.3 | 0.02 | -14.15 | <**0.001** |
| $\text{SL:Real}_D$ | -0.15 | 0.02 | -7.18 | <**0.001** |
| $\text{SL:Real}_W$ | -0.17 | 0.02 | -8.14 | **<0.001** |

both dialogue trees (p<0.001) and written trees (p<0.001). This shows that DLM is being minimized in both written and dialogue data when compared to the common baseline.[6] Interestingly, the effect-size of the coefficient suggests that DLM is comparatively stronger in written data compared to the dialogue data. This can be clearly seen in Figure 6.

In additional analyses, we also investigated if the effects discussed in Section 3.1 and 3.2 can be observed in written data using the HTDB corpus. Results show that, unlike in dialogue, we do not find evidence for a long-before-short pattern in the written data (no significant increase in fronting when the object or the adjunct was long) in both SOV ($p = 0.47$) as well as in the SAdjV utterances ($p = 0.68$). Interestingly, the data showed evidence for DLM in the case of right-extraposition ($p < 0.001$). The details of these analyses can be found in Tables 2, 3.

[6]Similar to the dialogue data analysis, we also tested for DLM in written data independently using the entire written corpus (Bhatt et al., 2009). For analysis, sentences with length more than 2 and less than 12 were used. As expected, DLM is minimized in Hindi written data (Futrell et al., 2015; Dyer, 2023).

## 5 Discussion

The current paper provides novel evidence in support of DLM in a dialogue data for an SOV language, Hindi. We investigated two phenomenona that are implicated in DLM, namely, long-before-short and right-extraposition. Our results show that while long-before-short leads to DLM in the dialogue data, right-extraposition does not.

With regard to the comparison between dialogue and written data, while we find evidence for DLM induced long-before-short pattern in the dialogue data, we did not find any evidence for this in the written data (see Table 2). On the other hand, while we find evidence for DLM induced right-extraposition in the written data, we did not find such evidence in the dialogue data (see Table 3). To probe this further, we investigated the modifier type in right-extraposed situations in both dialogue and written data. We find that right-extraposition is dominated by clausal modifiers in written data – in the dialogue data, only 21% of right-extraposed modifiers are clausal; while this was 85% in the written data. It is known that such clausal extrapositions lead to DLM in Hindi (see, Zafar and Husain, 2023). In addition, in such configurations, the average length for clausal modifiers in dialogue was 6.65 words, while it was 10.8 words in the written data. Similarly, the average length for right-extraposed modifiers (clausal and non-clausal) was longer in written (11.1 words) than in dialogue (3.7 words) (cf. Biber, 2009). This shows that right-extraposition is an important DLM strategy in written data but not in dialogue. One might ask, if right-extraposition is not motivated by DLM, why do we find increased shifting with an increase in phrasal length in dialogue? This trend for right-extraposition of long phrases after the verb could be due to other reasons such as information structure (Butt and King, 1996; Huck and Na, 1990), ease of planning (Wasow, 1997a) or syntactic expectation (Levy et al., 2012).

The fact that DLM is minimized more in written data than dialogue data (cf. Table 4) is consistent with not only the fact that written data is a product of an extensive editing process, but also that the writing process itself can be very different from speaking (Wengelin et al., 2009); also see, Roeser et al. (2019). The comparative analysis of written vs dialogue data is also quite instructive from a methodological perspective. Corpus-based investigations on DLM link their findings to the underlying cognitive processes (e.g., Futrell et al., 2015). The current work shows that while DLM can be observed in these different modalities, the underlying causes for the manifestation of DLM might be quite different. Therefore, any such generalizations should also be based on the syntactic configurations that lead to DLM. For example, there could be other causes to DLM in addition to the ones explored here, e.g., elision (Kramer, 2021).

The DLM constraint in the dialogue data has implications for production models that assume incrementality (Levelt, 1989). In particular, DLM minimization implies that speakers would have to structurally plan some components of the utterance before articulating them. This would mean that planning during language production is non-incremental to a certain degree (cf. Wheeldon and Konopka, 2023). At the same time, these results also highlight certain constraints on planning scope. The results suggest that, in dialogue, speakers do not plan very long post-nominal modifiers.

Together, these results highlight the over-arching influence of working-memory constraints on production process in an SOV language like Hindi (cf. Slevc, 2011; Gennari et al., 2012; Humphreys et al., 2016). Future work needs to investigate how such constraints interacts with other factors such as accessibility (cf. Ranjan et al., 2022).

## 6 Conclusion

In a corpus-based investigation, we test the generalizability of DLM as a cognitive principle for word order variation in a Hindi naturalistic spoken data. Our results show that the real trees attested in a dialogue corpus of Hindi have on average shorter dependencies when compared to random trees that match the real trees in topological features. Furthermore, to understand the sources of DLM in dialogue, we zoom into two phenomena known to minimize dependency length. We find that DLM in dialogue is primarily minimized by fronting longer arguments and adjuncts and not by right extraposing clausal or nominal modifiers. Finally, we compare the strength of DLM in spoken and written data. We posit that DLM is minimized in both the modalities, its effect being stronger for written than spoken. Overall, these results shed light on the overarching influence of working memory constraints in governing syntactic choices during both language comprehension and production.

## Acknowledgments

We would like to thank the anonymous reviewers whose insightful comments contributed to enhancing the quality of the paper. Sincere thanks to Himanshu Yadav for comments on the early draft of the paper and for offering valuable suggestions regarding the analysis presented in Section 4. We are also grateful to the team involved in the development of the IIT Delhi Hindi Dialogue Corpus for providing the data. This work was partially funded by a grant from the Department of Science and Technology's Cognitive Science Research Initiative (Grant no. DST/CSRI/2018/69) to Samar Husain. The project details and code used in this work is available at `https://osf.io/vxd6w/` and `https://github.com/MeghnaHooda/DLM-in-naturalistic-dialogue.git`.

## References

Jennifer E. Arnold, Anthony Losongco, Thomas Wasow, and Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1):28–55.

Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Brian Bartek, Richard L Lewis, Shravan Vasishth, and Mason R Smith. 2011. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5):1178.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, and Fei Xia. 2009. A multi-representational and multi-layered treebank for hindi/urdu. In *proceedings of the third linguistic annotation workshop (LAW III)*, pages 186–189.

Douglas Biber. 2009. Are there linguistic consequences of literacy? comparing the potentials of language use in speech and writing. *Cambridge handbook of literacy*, pages 75–91.

Miriam Butt and Tracy Holloway King. 1996. Structural topic and focus without movement. *Online Proceedings of LFG*.

Alexandra Canavan and Zipperlen George. 1996. CALLFRIEND Hindi LDC96S52. *Web Download. Philadelphia: Linguistic Data Consortium*.

Wallace Chafe. 1985. Linguistic differences produced by differences between speaking and writing. *Literacy, language, and learning: The nature and consequences of reading and writing*, 105:105–123.

Herbert H Clark. 2014. 18 spontaneous discourse. *The Oxford Handbook of Language Production*, page 292.

Andrew Thomas Dyer. 2023. Revisiting dependency length and intervener complexity minimisation on a parallel corpus in 35 languages. In *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 110–119.

Pegah Faghiri and Pollet Samvelian. 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*.

Victor S Ferreira and Gary S Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive psychology*, 40(4):296–340.

Ramon Ferrer-i Cancho and Haitao Liu. 2013. The risks of mixing dependency lengths from sequences of different length. *arXiv preprint arXiv:1304.3841*.

Lyn Frazier. 1987. Syntactic processing: Evidence from dutch. *Natural Language & Linguistic Theory*, 5(4):519–559.

Richard Futrell, Roger P Levy, and Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–412.

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341.

Silvia P. Gennari, Jelena Mirković, and Maryellen C. MacDonald. 2012. Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, 65(2):141 – 176.

Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, pages 95–126.

Daniel Gildea and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science*, 34(2):286–310.

Lila R Gleitman, David January, Rebecca Nappa, and John C Trueswell. 2007. On the give and take between event apprehension and utterance formulation. *Journal of memory and language*, 57(4):544–569.

Daniel Grodner and Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, 29(2):261–290.

John A Hawkins. 1990. A parsing theory of word order universals. *Linguistic inquiry*, 21(2):223–261.

John A Hawkins. 1994. *A performance theory of order and constituency*. 73. Cambridge University Press.

John A Hawkins. 2004. *Efficiency and complexity in grammars*. OUP Oxford.

John A Hawkins. 2014. *Cross-linguistic variation and efficiency*. OUP Oxford.

John R Hayes and Linda S Flower. 1986. Writing research and the writer. *American psychologist*, 41(10):1106.

Geoffrey J Huck and Younghee Na. 1990. Extraposition and focus. *Language*, pages 51–77.

Richard Hudson. 1995. Measuring syntactic difficulty. *Manuscript, University College, London*.

Gina F. Humphreys, Jelena Mirković, and Silvia P. Gennari. 2016. Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language*, 89:200 – 221. Speaking and Listening: Relationships Between Language Production and Comprehension.

Marcel A Just and Patricia A Carpenter. 1992. A capacity theory of comprehension: individual differences in working memory. *Psychological review*, 99(1):122.

John Kimball. 1973. Seven principles of surface structure parsing in natural language. *Cognition*, 2(1):15–47.

Anubha Kothari. 2010. *Processing constraints and word order variation in Hindi relative clauses*. Ph.D. thesis, Stanford University.

Alex Kramer. 2021. Dependency lengths in speech and writing: A cross-linguistic comparison via youdepp, a pipeline for scraping and parsing youtube captions. In *Proceedings of the Society for Computation in Linguistics 2021*, pages 359–365.

Willem JM Levelt. 1972. Some psychological aspects of linguistic data. *Linguistische Berichte*, 17:18–30.

Willem JM Levelt. 1989. *Speaking: From intention to articulation*. Cambridge: The MIT Press.

Roger Levy, Evelina Fedorenko, Mara Breen, and Edward Gibson. 2012. The processing of extraposed structures in english. *Cognition*, 122(1):12–36.

Richard L Lewis and Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3):375–419.

Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159–191.

Haitao Liu, Chunshan Xu, and Junying Liang. 2017. Dependency distance: a new perspective on syntactic patterns in natural languages. *Physics of life reviews*, 21:171–193.

Zoey Liu. 2020. Mixed evidence for crosslinguistic dependency length minimization. *STUF-Language Typology and Universals*, 73(4):605–633.

Maryellen C. MacDonald. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology*, 4:226.

George A Miller. 1962. Some psychological studies of grammar. *American psychologist*, 17(11):748.

John Ashworth Nelder and Robert WM Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384.

Benu Pareek, Mudafia Zafar, Karan Yadav, Meghna Hooda, Ashwini Vaidya, and Samar Husain. 2023. Introducing the hindi dialogue corpus. In *Poster presented at the fourth South Asian Forum on the Acquisition and Processing of Language*. https://osf.io/76akt.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sidharth Ranjan, Rajakrishnan Rajkumar, and Sumeet Agarwal. 2022. Linguistic complexity and planning effects on word duration in hindi read aloud speech. *Proceedings of the Society for Computation in Linguistics*, 5(1):119–132.

Jens Roeser, Mark Torrance, and Thom Baguley. 2019. Advance planning in written and spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(11):1983.

Idoia Ros, Mikel Santesteban, Kumiko Fukumura, and Itziar Laka. 2015. Aiming at shorter dependencies: the role of agreement morphology. *Language, Cognition and Neuroscience*, 30(9):1156–1174.

Gregory Scontras, William Badecker, Lisa Shank, Eunice Lim, and Evelina Fedorenko. 2015. Syntactic complexity effects in sentence production. *Cognitive science*, 39(3):559–583.

L Robert Slevc. 2011. Saying what's on your mind: working memory effects on sentence production. *Journal of experimental psychology: Learning, memory, and cognition*, 37(6):1503.

Benedikt Szmrecsányi. 2004. On operationalizing syntactic complexity. In *7th International Conference on Textual Data Statistical Analysis*, pages 1032–1039. Presses Universitaires de Louvain, Louvain-la-Neuve.

David Temperley. 2007. Minimization of dependency length in written english. *Cognition*, 105(2):300–333.

Thomas Wasow. 1997a. End-weight from the speaker's perspective. *Journal of Psycholinguistic research*, 26:347–361.

Thomas Wasow. 1997b. Remarks on grammatical weight. *Language Variation and Change*, 9(1):81–105.

Thomas Wasow. 2002. *Postverbal Behaviour*. CSLI Publications.

Thomas Wasow and Jennifer Arnold. 2003. Post-verbal constituent ordering in english. *Topics in English Linguistics*, 43:119–154.

Åsa Wengelin, Mark Torrance, Kenneth Holmqvist, Sol Simpson, David Galbraith, Victoria Johansson, and Roger Johansson. 2009. Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, 41(2):337–351.

Linda Ruth Wheeldon and Agnieszka Konopka. 2023. *Grammatical Encoding for Speech Production*. Elements in Psycholinguistics. Cambridge University Press.

Himanshu Yadav, Samar Husain, and Richard Futrell. 2019. Are formal restrictions on crossing dependencies epiphenominal? In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 2–12.

Himanshu Yadav, Samar Husain, and Richard Futrell. 2022a. Assessing Corpus Evidence for Formal and Psycholinguistic Constraints on Nonprojectivity. *Computational Linguistics*, 48(2):375–401.

Himanshu Yadav, Shubham Mittal, and Samar Husain. 2022b. A reappraisal of dependency length minimization as a linguistic universal. *Open Mind*, 6:147–168.

Hiroko Yamashita and Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition*, 81(2):B45–B55.

Victor H Yngve. 1960. A model and an hypothesis for language structure. *Proceedings of the American philosophical society*, 104(5):444–466.

Mudafia Zafar and Samar Husain. 2023. Dependency locality influences word order during production in sov languages: Evidence from hindi. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.