What determines the order of verbal dependents in Hindi? Effects of efficiency in comprehension and production

Kartik Sharma[†], **Richard Futrell**[‡*] **and Samar Husain**[†*]
[†]Indian Institute of Technology Delhi, [‡]University of California Irvine
`kartik.sharma.cs117@cse.iitd.ac.in,rfutrell@uci.edu,`
`samar@iitd.ac.in`

Abstract

Word order flexibility is one of the distinctive features of SOV languages. In this work we investigate whether the order and relative distance of preverbal dependents in Hindi, an SOV language, is affected by factors motivated by efficiency considerations during language comprehension and production. We investigate the influence of Head–Dependent Mutual Information (HDMI), similarity-based interference, accessibility and case-marking. Results show that preverbal dependents remain close to the verbal head when the HDMI between the verb and its dependent is high. This demonstrates the influence of locality constraints on dependency distance and word order in an SOV language. Additionally, dependency distance were found to be longer when the dependent was animate, when it was case-marked and when it was semantically similar to other preverbal dependents. Together the results highlight the cross-linguistic generalizability of these factors and provide evidence for a functionally motivated account of word order in SOV languages such as Hindi.

1 Introduction

Natural languages are known to be influenced by a pressure for efficiency, that is, languages should enable speakers to communicate as well as possible, subject to constraints on the complexity of production and comprehension (Zipf, 1949; Jaeger and Tily, 2011; Gibson et al., 2019; Hawkins, 2014). One area where efficiency pressures appear to have a large effect is in word order: for example, across languages, the distribution of word orders can be well predicted using the principle of dependency length minimization, the idea that words in syntactic dependencies are under a pressure to be close to each other (Liu et al., 2017; Temperley and Gildea,

2018). Dependency length minimization is motivated by efficiency because it results in lower working memory requirements for language production and comprehension.

Here, we take up the question of whether word orders can be predicted using efficiency in an area that goes beyond dependency length minimization. In particular, we examine how efficiency pressures may influence the order of pre-verbal dependencies in a verb-final language, Hindi. We formalize a number of measures of the complexity of comprehension and production, drawing from the psycholinguistic literature, and we test what effect these measures have on word order as observed in a large dependency treebank.

The paper is arranged as follows: in Section 2, we describe the various psycholinguistic factors used to investigate preverbal ordering and review related work. Section 3 describes the data and methods used to undertake the various analyses. In Section 4 we present the results. Section 5 discusses the implications of our results and concludes.

2 Psycholinguistic Factors influencing Word Order

We consider four psycholinguistically-motivated factors as predictors of the order of verbal dependents in Hindi.

Head–dependent mutual information (HDMI) While the theory of dependency length minimization holds that all words in dependencies should be close to each other, there is no theoretical consensus on *which* words in dependencies should be especially close to each other. In contrast, the more general theory of **information locality** holds that any two words w_1 and w_2 should be close to each other in proportion to their **pointwise mutual in-**

---

*Equal contribution by RF and SH.

1

formation (pmi):

$$\text{pmi}(w_1; w_2) \equiv \log \frac{p(w_1, w_2)}{p(w_1)p(w_2)}. \qquad (1)$$

In support of this idea, in previous work, Futrell (2019) found that dependencies in which the head and dependent have high pmi are under a stronger pressure to be close than dependencies with less head–dependent mutual information.

The idea of information locality is motivated by efficiency in language comprehension, based on information-theoretic models of incremental sentence processing. For example, the model of Futrell et al. (2020b) holds that the difficulty of processing a word in context is given by the surprisal (negative log probability) of the word given a lossy memory representation of the context. Information locality can be derived as a consequence of this model.

Following previous work, we operationalize the head–dependent mutual information as the pointwise mutual information between part-of-speech tags. Our main reason for using the pmi between part-of-speech tags, rather than the pmi between wordforms, is in order to avoid data sparsity issues in the estimation of mutual information (Paninski, 2003; Futrell et al., 2019). However, this choice also changes the interpretation of the HDMI measure. Instead of reflecting the full predictive information contained in one word about another word, the measure now only reflects something like the *syntactic* predictive information contained in one word about another.

For representing words for this measure, we use an augmented part-of-speech tagset described in Section 3.2.1. This tagset captures not only part-of-speech information, but also verb argument structure and nominal case marking.

Accessibility   A persistent generalization in the typological literature is that there is a bias for words which are more **accessible** to go earlier in a sentence. Words are more accessible when they are animate, definite, imageability, and/or salient in discourse (Ariel, 1990; Jaeger and Tily, 2011). The preference for more accessible words to go earlier in sentences appears to be motivated by ease of language production (Ferreira and Dell, 2000; Kurumada and Jaeger, 2015), under the theory that language producers will tend to produce the words which are most accessible as quickly as possible. Evidence for production ease comes from, amongst others, the observation that lemma selection during grammatical encoding is influenced by accessibility of the lemma (e.g., Bock and Warren, 1985; Bock et al., 1992; Prat-Sala and Branigan, 2000; Branigan et al., 2008).

Here we operationalize accessibility as the animacy of a referent. We leave investigations of other factors affecting accessibility (definiteness, imageability, etc.) to future work.

Semantic similarity   A great deal of work in psycholinguistics has focused on the effect of **similarity-based interference** on processing: the idea that difficulty results when a comprehender must retrieve a target item from working memory, but there is another distractor item in working memory which interferes with the retrieval of the target item (Jäger et al., 2017). From a production perspective, presence of similar phrases in an utterance can lead to competition between the phrases and thereby difficulty in planning. Successful articulation of a phrase therefore requires inhibition of one of the phrases, thereby resulting in a delayed articulation of the inhibited phrase. The magnitude of this interference increases as the target and the distractor become more similar (Gordon et al., 2006; Gennari et al., 2012). For comprehension, this effect has been modeled within frameworks based on cue-based retrieval such as ACT-R (Lewis and Vasishth, 2005).

In terms of word order, it has been shown that when syntactically/semantically similar nominals appear adjacent (or close) to each other, processing suffers (e.g., Lewis and Nakayama, 2002; Vasishth, 2003; Gordon et al., 2006; Apurva and Husain, 2020). Relatedly, there has been some work on the influence of semantic interference during production (e.g., Ferreira and Firato, 2002; Humphreys et al., 2016; Gennari et al., 2012; MacDonald, 2013). For example, Gennari et al. (2012) found that an animate head noun leads to a higher chance for producing passive relative clause construction in English, compared to an active relative clause, thus keeping the animate head noun distant from the animate passive subject. On this account, a pre-verbal dependent that is similar to other phrases should appear earlier in the sentence because increased dependency distance between similar nouns should lead to production ease.

Therefore, the prediction based on psycholinguistics is that a word should be pushed out towards the beginning of the sentence when it is semantically similar with other words. This creates dis-

tance between the two similar words and facilitates language production and comprehension.

Morphological marking A feature of a majority of head-final languages is the presence of case marking on nouns that signifies the syntactic relation between a nouns and its verbal head.[2] Cross-linguistically, the presence of case-markers has been shown to increase the dependency distance between a nominal and its verbal head (e.g., Yadav et al., 2020), perhaps because nominal case-markers help in making robust predictions about upcoming verbs, as shown by Husain et al. (2014). Therefore, we predict that nominals with case marking should be farther out from the verb than those without.

2.1 Related Work

There have been some previous corpus-based investigations on word order variation in Hindi (e.g., Husain et al., 2013; Ranjan et al., 2019); (also see, Vasishth, 2004). For example, Ranjan et al. (2019) investigated the role of case-marking on word order choices in Hindi and found evidence for the Easy-First and Reduce Inference principles of the Production Distribution and Comprehension (PDC) hypothesis (MacDonald, 2013). Production ease was operationalized in Ranjan et al. (2019) as low n-gram/dependency surprisal value (Hale, 2001) and interference was operationalized as case-marker similarity between preverbal nominals. Jain et al. (2018) investigated the Uniform Information Density hypothesis (Jaeger, 2010) with regard to predicting Hindi word order in corpus sentences vs random sentences and found no support for UID in capturing such a distinction. They also did not find UID to predict non-canonical word order in Hindi.

With regard to ordering of co-siblings in dependency trees, Dyer (2018) has argued that the relative predictability of the head at the dependent (which can be operationalized as the entropy of heads given dependents) determines which sibling is closer to the head. In particular, he shows that co-siblings that induce lower entropy tend to be closer to the head compared to co-siblings that induce higher entropy. The current work will complement the above investigations by exploring the role of some novel factors such as HDMI, semantic interference, and animacy. The effect of these factors on preverbal ordering and dependency distance in SOV languages is largely unknown.

3 Methodology

3.1 Data and Tools

We use the monolingual Hindi corpus developed by Kunchukuttan et al. (2017). This corpus includes raw sentences of Hindi, collected from various sources (HindMonoCorp (Bojar et al., 2014), BBC, Wikipedia etc.). We restricted our analysis[3] to only the first 5 million sentences of this data, resulting in a dataset of 14 million verbal dependencies. Since we needed to extract noun–verb dependency relations, we parsed this data using the ISC dependency parser for Hindi.[4] The parser is trained on the Hyderabad Dependency Treebank (Bhatt et al., 2009) that is based on the Computational Paninian Grammar (CPG, henceforth) (Bharati et al., 1995).

3.2 Model

In this section, we provide details regarding the computation of the factors discussed in Section 2.

3.2.1 HDMI

As mentioned in Section 2, we calculate the point-wise mutual information between words using augmented part-of-speech tags. Here, we describe the augmentation in detail. A verb is identified with its verb class. Verbal classes are defined based on the different argument structures that a verb can have. A class of a verb is characterized by the set of core argument relations (Subject, Noun complement for Copula, Direct Object, Indirect Object). For example, two verbs will belong to the same class if the set of the core argument relations they have is identical. Thus, we classify a verb into one of the 16 classes according to this scheme. A 16 way classification is based on Sharma et al. (2019). This constitutes an exhaustive set of possible argument structures a verb can have, considering only the core-arguments.

A verb POS tag (*VM*) is augmented with the verb class that it belongs to. Nominal POS tags (*NN*—Common noun, *NNP*—Proper noun) are augmented by case-marker information (if any).

---

[2]Lexically, these case-markers can either appear as a suffix or as an independent token.

[4]The parser is an implementation of the incremental transition-based arc-eager parsing algorithm (Nivre, 2008). The parser is reported to have a UAS of 93.52% and a LAS of 87.77% (Bhat, 2017). `https://bitbucket.org/ account/user/iscnlp/projects/ISCNLP`.

3.2.2 Semantic Similarity

Semantic similarity of a noun in its context is modeled as the maximum cosine similarity (Salton, 1972) of the noun with other pre-verbal dependents of the corresponding verb. The cosine similarity is calculated from the word vectors taken from a pre-trained model for Hindi (Grave et al., 2018). Semantic similarity is

$$\text{sim}(d) = \max_{\substack{d':(h,d')\in\text{Dep} \\ id(d')<id(h)}} \frac{wv(d) \cdot wv(d')}{\|wv(d)\| \, \|wv(d')\|}, \quad (2)$$

where $wv(x)$ denotes the word vector of a word X, $id(x)$ denotes the index of X in the sentence, and $(h, d) \in \text{Dep}$. Note that Dep is the set of all dependencies in the sentence.

We choose to use word vectors to generate similarity scores because collecting human judgments for all words in our large dataset is impractical. There is mixed evidence about whether semantic similarity as defined using word vectors truly captures psycholinguistically relevant aspects of similarity. Despite arguments that word vectors do not capture certain properties of human similarity judgments (Griffiths et al., 2007; De Deyne et al., 2016) and certain semantic interference effects (Merlo and Ackermann, 2018), they have been used successfully in psycholinguistic models of interference (Smith and Vasishth, 2020).

3.2.3 Accessibility

In the current study, the accessibility of a noun is determined based on the notion of humanness. Nouns that are +Human as assumed to have conceptual prominence while nouns with a -Human feature are assumed to have low conceptual prominence. Thus, this is a categorical variable in our study. Animacy information was obtained from hand annotations in a version of the Hyderabad Dependency Treebank augmented with nominal semantic features (Jena et al., 2013). In this text, animate will be used interchangeably for +Human, unless otherwise specified.

3.2.4 Case Marking

The information regarding nominal case-marking is extracted from the parsed data using the DEPREL tag. A noun is case-marked if it has a dependent with DEPREL equal to *lwg__psp* in CPG. Thus, we mark a noun-verbal dependency as case-marked if the corresponding noun is case-marked, otherwise it is deemed unmarked.

3.3 Granularity of Analysis

In order to get a fine-grained view of the data, we will separately analyze the four different relation types: subjects, direct objects, indirect objects, and adjuncts. We label a relation as a subject if its DEPREL tag is either *k1* or *k1s*, as a direct object if it is *k2*, as an indirect object if it is *k4*, and as a verbal adjunct otherwise.

For all analyses, we consider only dependencies with length greater than 1. This is necessary because in Hindi there cannot be any noun-verb dependency at distance 1 when the noun is case-marked—a case-marker appears post-nominally and is adjacent to the noun. Since case-marker is a factor in our analysis, any real effect may be confounded by this constraint of the grammar.

4 Results

In this section we discuss the length and ordering of preverbal nominal dependencies. We first discuss how the length is affected by the factors discussed above. We then discuss how they affect the word-order.
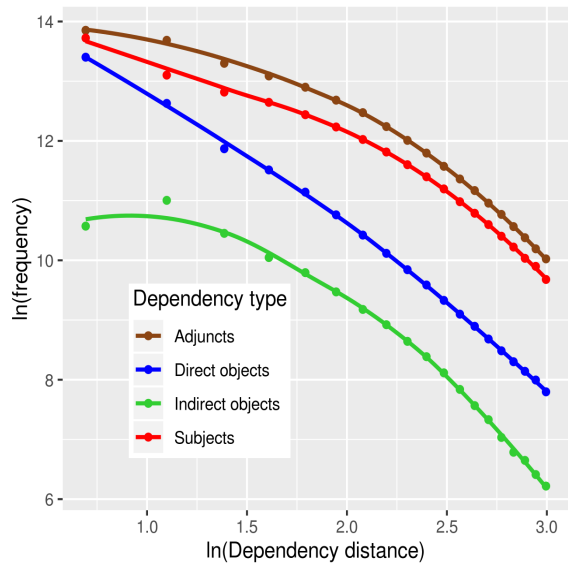

Figure 1: Frequency of various arguments at different dependency distance in logarithmic scale

4.1 Length of Preverbal Dependencies

Figure 1 shows the distribution of dependency distance (the number of words from the head to the dependent) for various arguments in a log–log plot. We find that verbal adjuncts are more frequent than all the arguments at all distances. Among the arguments, subjects are the most frequent at almost all

distances followed by direct objects and then indirect objects. With increasing distances, the frequencies of all dependencies decrease but the decay is strikingly slower in subjects and adjuncts compared to the objects. Direct objects decrease steadily following a power law while the others show a nonlinear trend until large distances (> 12).

We now investigate the effect of our psycholinguistically-motivated factors on dependency distance. Figure 2a shows a plot for HDMI at various dependency distances. The trend for preverbal subjects is quite clear: as dependency distance increases, HDMI decreases.

The figure unexpectedly shows that the HDMI for adjuncts and indirect object tends to increase starting around dependency distance 7. One possible reason for this increase is estimation error in the pmi values: given small amounts of data from a relatively large number of classes, pmi is likely to be overestimated (Futrell et al., 2019), and the frequency of these dependents is less at higher dependency distance. Therefore, the HDMI – dependency distance relation for these dependents seems to be more stable at shorter dependency distance. Another explanation for the apparent HDMI increase at large dependency distances could have to do with certain discourse-related dependencies between verbs and adjuncts, where the adjunct appears preferentially at the beginning of a clause or sentence (Butt and King, 1996). Such discourse dependencies might have high HDMI while also having large dependency distance due do the constraint of appearing early.

Figure 2b shows a plot for mean semantic similarity at various dependency distance. We find that for all the arguments, an increase in semantic similarity leads to increase in dependency distance. Figure 2c shows a plot for proportion of animate arguments at various dependency distance. The most clear trend is for the subjects – the proportion of +Human subjects tend to increase with increase in dependency distance. This positive trend is also seen for the other dependents, but for short distances.

Finally, Figure 2d shows a plot for proportion of case-marked arguments at various dependency distance. Here again, the expected positive trend, i.e., an increase in case-marked arguments with increasing length, is seen only for subjects. For direct objects and adjuncts, this trend is reversed. In fact, the negative trend of fewer case-marked

nominals at longer dependency distance is quite strong for direct objects. Note that indirect objects are always case-marked.

In order to quantify the effects of our predictors, we fit a linear regression predicting dependency distance as a function of our predictors:

$$\text{distance} \approx \beta_0 + \beta_1 \times \text{HDMI} + \beta_2 \times \text{Similarity} + \beta_3 \times \text{Animacy} + \beta_4 \times \text{Case} \quad (3)$$

The above model was fit separately for all 4 relations (subject, direct object, indirect object, and adjunct). The variables HDMI and similarity are continuous. On the other hand, the variables animacy and case are categorical in nature. Based on the discussion in Section 2, we expect β_1 to be negative, i.e., dependency distance is expected to decrease with increase in HDMI. The coefficients β_2, β_3 and β_4 are expected to have positive values. In other words, dependency distance is expected to increase when there is increased semantic similarity between nouns, when the argument is +Human and when it is case-marked.

Coefficient	S	DO	IO	Adjuncts
Intercept	-0.19	1.18	0.13	-0.60
HDMI	-0.43	3.50	-0.88	-1.69
Similarity	0.01	0.06	0.02	0.07
Animacy	0.07	0.003	0.03	0.03
Case	1.50	-0.35	—	0.66

Table 1: Linear regression coefficients predicting distance for different dependency types; see Eq. 3. S = Subject; DO = Direct Object; IO = Indirect Object. Variables are centered. Note that case is excluded as a factor for indirect objects.

The values of the fitted coefficients β are shown in Table 1. All coefficients are significant at $p < .05$. For the subjects, adjuncts and indirect objects, we find that all the coefficients are in the expected direction. On the other hand, mirroring the unexpected effects visible in Figure 2, the coefficients of HDMI and case-marker for direct objects[5] are not in the expected direction.

---

[5]At first glance this might seem inconsistent with the distribution of Object and case-marker in Husain et al. (2013) who report that objects closer to the verb are more likely to be bare compared to when they are further away. However, note that Husain et al. (2013) report the overall count, while here we show the proportion at each dependency distance. The trend reported in Husain et al. (2013) can be driven by short dependencies (which form the bulk of all instances); the pattern can, in fact, be seen in Figure 2d for dependency lengths 2-5.
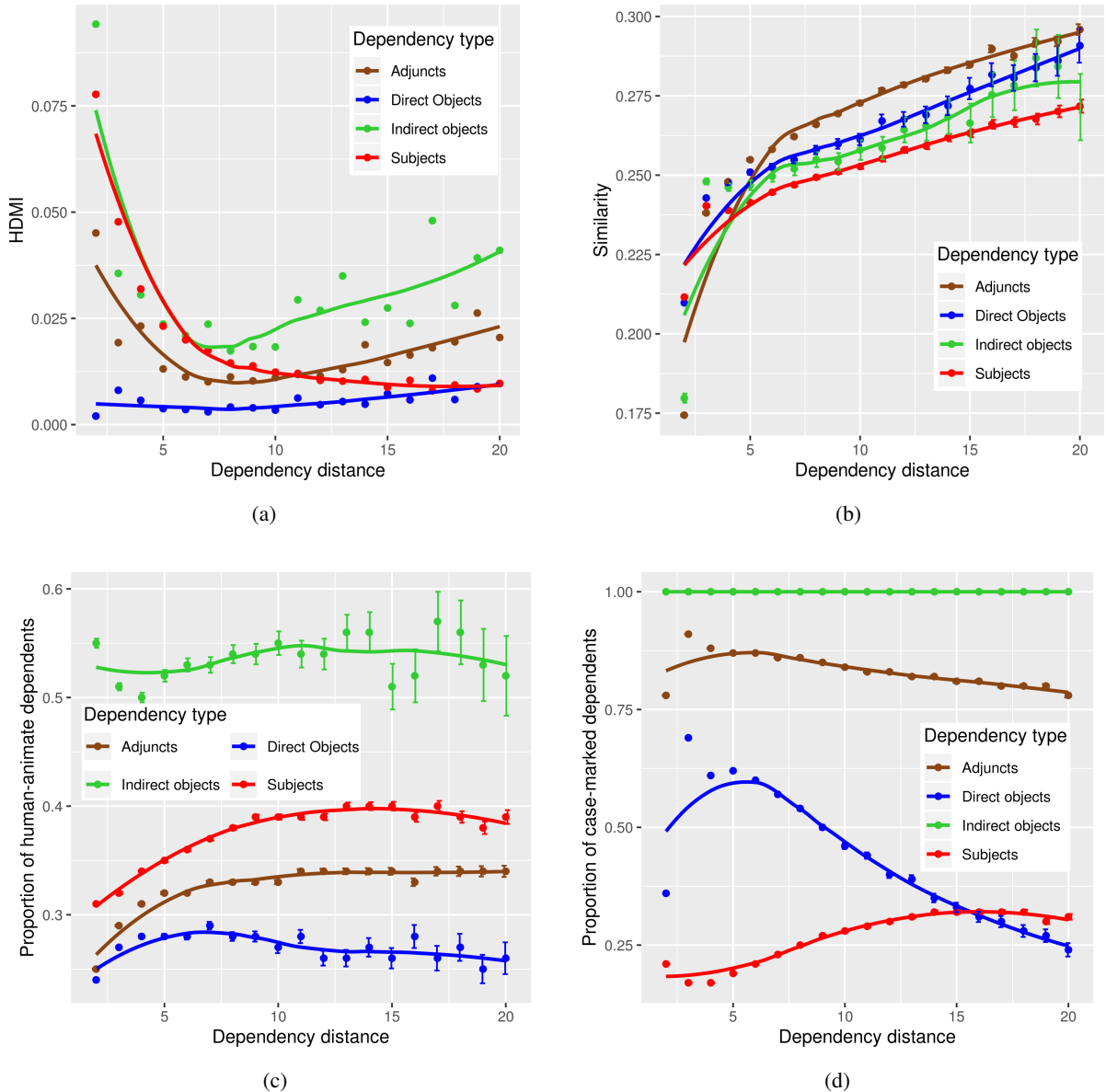
(a)

(b)

(c)

(d)

Figure 2: Distribution of factors (mean and 95% CI) at different distances greater than 1.

4.2 Order of Preverbal Dependencies

We now consider how different dependencies of a verb are ordered relative to each other. We consider order between 3 pairs of constituents: (i) arguments and adjuncts, (ii) subjects and objects, and (iii) direct objects and indirect objects. Note that objects constitute both direct objects and indirect objects, and arguments constitute both subjects and objects.

All nominal dependencies attached to a common verb were collected and grouped into the above-mentioned pairs according to their dependency relations. Then, the order between the constituents is calculated by considering the sign of the difference between their dependency distances. Henceforth, we denote an order between constituents X and Y

as X–Y if X precedes Y and Y–X otherwise.

We fit a logistic regression predicting the order of two dependents given the *difference* in the values of the factors for the two dependents (as in Morgan and Levy, 2016; Futrell et al., 2020a). The factors are same as used in the previous section. In particular, we find coefficients β to fit

$$
\begin{aligned}
\log \frac{P(X\text{–}Y \text{ order})}{P(Y\text{–}X \text{ order})} &\approx \beta_0 + \beta_1 \times \Delta\mathrm{HDMI} \\
&+ \beta_2 \times \Delta\mathrm{Similarity} + \beta_3 \times \Delta\mathrm{Animacy} \\
&+ \beta_4 \times \Delta\mathrm{Case},
\end{aligned} \tag{4}
$$

where $\Delta\mathrm{HDMI}$ indicates the HDMI of X minus the HDMI of Y, etc. This model is evaluated separately for argument (X) and adjunct (Y), subject

6

(X) and object (Y), and direct object (X) and indirect object (Y), along with random intercepts by sentence.

Based on the discussion in Section 2, we expect β_1 to be negative, since a higher value of HDMI of X than Y should imply that X is closer to the verb than Y, i.e. the order should be $Y - X$. The similarity coefficient β_2 is expected to be positive based on the effect of interference on distance seen in Section 4.1, and β_3 and β_4 are also expected to have positive values. In other words, a dependent is expected to be farther away from the verb than another dependent when it is more +Human (and case-marked) than the other dependent.

The values of the fitted regression coefficients β are shown in Table 2. Significance is according to the criterion $p < .05$. All effects except ΔSimilarity are in the theoretically expected directions. Figure 3 shows the effect of each factor on the word-order. One can verify the trends with the sign of the coefficients in the regression.

Coefficient	Arg–Adj	S–O	DO–IO
Intercept	0.03	2.32	-1.96
Δ HDMI	-0.99	-1.26	-1.95
Δ Similarity	-0.11	-0.09	0.01[a]
Δ Animacy	0.21	0.16	0.11
Δ Case	1.25	1.50	0.57

[a]Not significant

Table 2: Logistic regression coefficients predicting X–Y vs. Y–X order based on the difference in predictor values for X and Y (see Eq. 4). S = Subject; O = Object, DO = Direct Object; IO = Indirect Object.

5 Discussion

The results show that comprehension/production efficiency based factors affect dependency distance and ordering of preverbal dependents in Hindi.

Previous work investigating dependency distance has demonstrated that SOV languages allow for longer dependency distance between the verb and its prior dependents (Futrell et al., 2020c; Yadav et al., 2020; Konieczny, 2000). The current work makes an important contribution by highlighting that compared to preverbal adjuncts, the core arguments (subject, indirect object and direct object) tend to be closer to the verb in an SOV language like Hindi. Additionally, when the HDMI between the preverbal dependent and the verb is high, they tend to be close to each other. The HDMI difference between a pair of dependencies also correctly predicts their order. This provides compelling evidence that, all else being equal, the preverbal dependents in an SOV language such as Hindi are under information-theoretic locality constraints. Indeed, recent behavioral studies (e.g., Apurva and Husain, 2020) suggest that clause final verb prediction in Hindi suffers with increased distance between prior arguments and the upcoming verb.

The current findings also highlight certain conditions under which locality constraints can be overridden, leading to increased distance between preverbal arguments and the verb. These factors relate to well established processing constraints such as animacy-driven accessibility and similarity-based interference. Language-specific characteristics also influence syntactic configurations. In the current study, the presence of nominal case-markers tends to increase the dependency distance between the verb and its prior dependents. This could highlight the predictive strength of case-markers vis-à-vis a verb type (Husain et al., 2014; Vasishth and Husain, 2014). Under such a setting, the locality constraint can be violated, leading to longer dependency distance.

The results also raise some inconsistencies. One such inconsistency was that the dependency distance prediction for factors such as HDMI and case-marking in the case of direct object is not borne out. It is possible that compared to other arguments, the relationship between a direct object and verb is distinct (Momma and Ferreira, 2019). If this is true, then we should be able to replicate the DO pattern found in the current study for other SOV languages. The other inconsistency in the result was that the prediction for word order with respect to similarity-based interference turned out to be incorrect in the order-based analyses.

The current work does not probe the effect of discourse/information structure on word order. Butt and King (1996) have proposed a discourse-centric mapping of word position vis-à-vis the verb and discourse function – Topic maps to the sentence initial position, Focus maps to the immediate preverbal position, Background Information maps to the postverbal position and Completive Information maps to the non-immediate preverbal position. Such information structure roles can be difficult to ascertain automatically, but future work can attempt to investigate the role of these factors in the
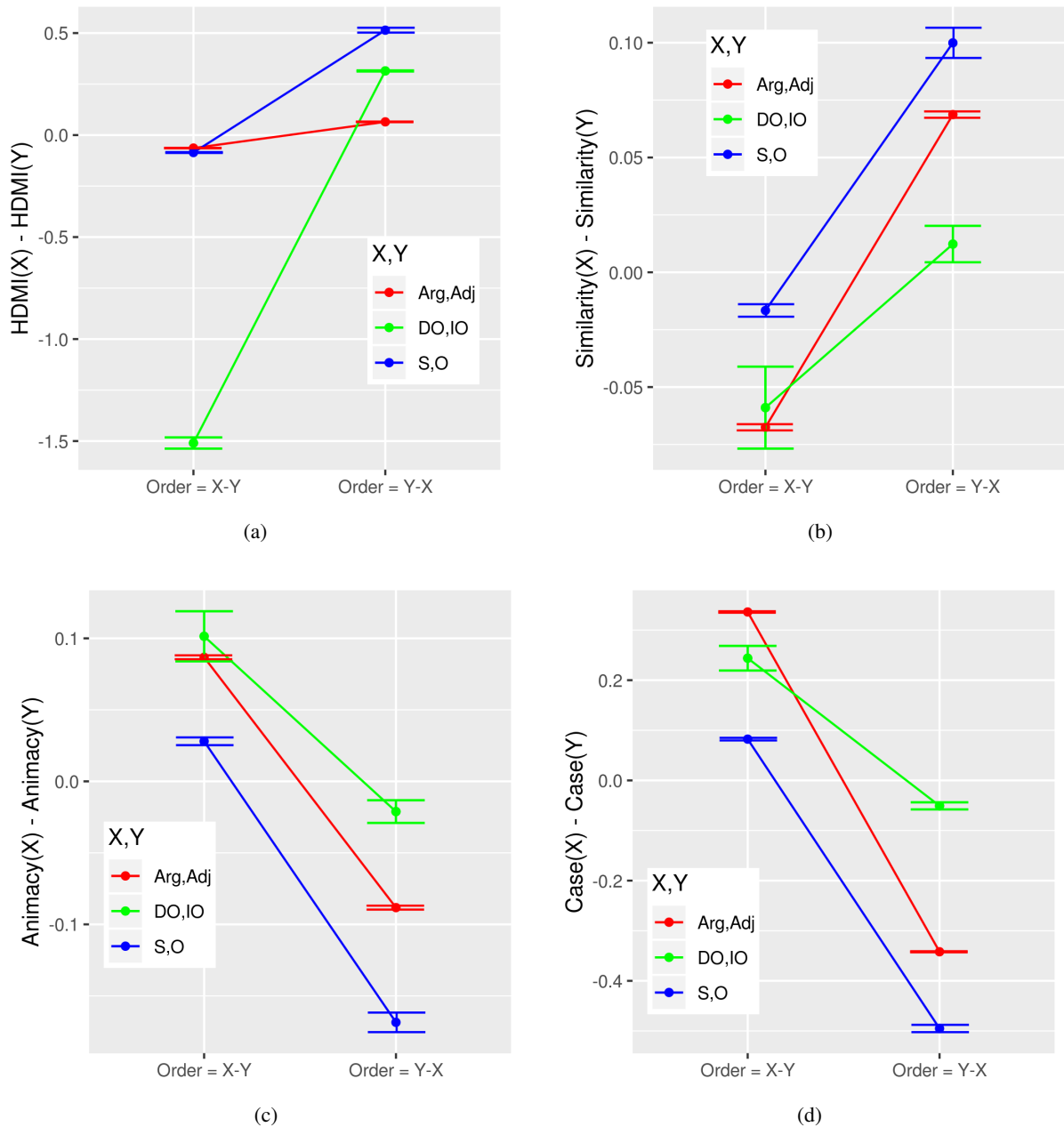
Figure 3: Effect of factors (mean and 95% CI) on the order between constituents X and Y. Arg=Argument; Adj=Adjunct; S=Subject; O=Object, DO=Direct Object; IO=Indirect Object.

context of the current work.

6 Conclusion

Our work provides evidence in support of the information locality hypothesis which states that head-dependent are under pressure to be close to each other when they are strongly associated with each other. It also supports other efficiency-driven factors such as similarity-based interference, accessibility, and the presence of case-markers, which act as countervailing forces to keep the distance between the dependent and its verbal head long.

Together these factors influence the dependency distance and word order patterns in an SOV language like Hindi.

References

Apurva and Samar Husain. 2020. Parsing errors in hindi: Investigating limits to verbal prediction in an sov language. *In submission*.

Mira Ariel. 1990. *Accessing noun-phrase antecedents*. Routledge.

A. Bharati, V. Chaitanya, and R. Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. Prentice-Hall of India, New Delhi.

Riyaz Bhat. 2017. *Exploiting Linguistic Knowledge to Address Representation and Sparsity Issues in Dependency Parsing of Indian Languages*. Ph.D. thesis, IIIT Hyderabad India.

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*.

J. Kathryn Bock, Helga Loebell, and Randal Morey. 1992. From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99:150 – 171.

J. Kathryn Bock and R. K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21:47–67.

Ondřej Bojar, Vojtěch Diatka, Pavel Rychlý, Pavel Straňák, Vít Suchomel, Aleš Tamchyna, and Daniel Zeman. 2014. HindEnCorp - Hindi-English and Hindi-only Corpus for Machine Translation. In *Proceedings of LREC'14*.

Holly P. Branigan, Martin J. Pickering, and Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118(2):172 – 189.

M. Butt and T. C. King. 1996. Structural topic and focus without movement. In *M. Butt and T. H. King, eds., The First LFG Conference. CSLI Publications.*

Simon De Deyne, Amy Perfors, and Daniel J Navarro. 2016. Predicting human similarity judgments with distributional models: The value of word associations. In *Proceedings of COLING*.

William Dyer. 2018. Integration complexity and the order of cosisters. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 55–65.

Victor S. Ferreira and Carla E Firato. 2002. Proactive interference effects on sentence production. *Psychonomic bulletin review*, 9(2):795 – 800.

V.S. Ferreira and G.S. Dell. 2000. Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production* 1. *Cognitive Psychology*, 40(4):296–340.

Richard Futrell. 2019. Information-theoretic locality properties of natural language. In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)*.

Richard Futrell, William Dyer, and Gregory Scontras. 2020a. What determines the order of adjectives in english? comparing efficiency-based theories using dependency corpora. In *Proceedings of ACL 2020*.

Richard Futrell, Edward Gibson, and Roger P. Levy. 2020b. Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44:e12814.

Richard Futrell, Roger P. Levy, and Edward Gibson. 2020c. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–413.

Richard Futrell, Peng Qian, Edward Gibson, Evelina Fedorenko, and Idan Blank. 2019. Syntactic dependencies correspond to word pairs with high mutual information. In *Proceedings of the Fifth International Conference on Dependency Linguistics*.

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, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences*.

Peter C. Gordon, Randall Hendrick, Marcus Johnson, and Yoonhyoung Lee. 2006. Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6):1304–1321.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the LREC 2018*.

Thomas L Griffiths, Mark Steyvers, and Joshua B Tenenbaum. 2007. Topics in semantic representation. *Psychological review*, 114(2):211.

John T. Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the NAACL*, pages 1–8.

John A. Hawkins. 2014. *Cross-linguistic variation and efficiency*. Oxford University Press, Oxford.

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.

Samar Husain, Rajesh Bhatt, and Shravan Vasishth. 2013. Towards a psycholinguistically motivated dependency grammar for Hindi. In *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*, pages 108–117.

Samar Husain, Shravan Vasishth, and Narayanan Srinivasan. 2014. Strong expectations cancel locality effects: Evidence from Hindi. *PLOS ONE*, 9(7):e100986.

T. Florian Jaeger. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

T. Florian Jaeger and Harry J. Tily. 2011. On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3):323–335.

Lena Jäger, Felix Engelmann, and Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94:316–339.

Ayush Jain, Vishal Singh, Sidharth Ranjan, Rajakrishnan Rajkumar, and Sumeet Agarwal. 2018. Uniform information density effects on syntactic choice in Hindi. In *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing*.

Itisree Jena, Riyaz Ahmad Bhat, Sambhav Jain, and Dipti Misra Sharma. 2013. Animacy annotation in the Hindi treebank. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*.

Lars Konieczny. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6):627–645.

Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya. 2017. The iit bombay english-hindi parallel corpus. *arXiv preprint arXiv:1710.02855*.

Chigusa Kurumada and T. Florian Jaeger. 2015. Communicative efficiency in language production: Optional case-marking in japanese. *Journal of Memory and Language*, 83:152 – 178.

Richard L Lewis and Mineharu Nakayama. 2002. Syntactic and positional similarity effects in the processing of japanese embeddings. *Sentence processing in East Asian languages*, pages 85–110.

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, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193.

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

Paola Merlo and Francesco Ackermann. 2018. Vectorial semantic spaces do not encode human judgments of intervention similarity. In *Proceedings of the 22nd CoNLL*, pages 392–401.

Shota Momma and Victor S. Ferreira. 2019. Beyond linear order: The role of argument structure in speaking. *Cognitive Psychology*, 114:101228.

Emily Morgan and Roger Levy. 2016. Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, 157:382–402.

Joakim Nivre. 2008. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics*, 34(4):513–553.

Liam Paninski. 2003. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253.

Mercè Prat-Sala and Holly P Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in english and spanish. *Journal of Memory and Language*, 42(2):168 – 182.

Sidharth Ranjan, Sumeet Agarwal, and Rajakrishnan Rajkumar. 2019. Surprisal and interference effects of case markers in Hindi word order. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 30–42.

Gerard Salton. 1972. A new comparison between conventional indexing (medlars) and automatic text processing (smart). *Journal of the American Society for Information Science*, 23(2):75–84.

Kartik Sharma, Kaivalya Swami, Aditya Shete, and Samar Husain. 2019. Can Greenbergian universals be induced from language networks? In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories*.

Garrett Smith and Shravan Vasishth. 2020. A principled approach to feature selection in models of sentence processing.

David Temperley and Daniel Gildea. 2018. Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, 4:1–15.

Shravan Vasishth. 2003. *Working memory in sentence comprehension: Processing Hindi center embeddings*. Routledge.

Shravan Vasishth. 2004. Discourse Context and Word Order Preferences in Hindi. In *R. Singh (ed.), The Yearbook of South Asian Languages and Linguistics*, pages 113–127.

Shravan Vasishth and Samar Husain. 2014. The interaction of working memory and expectation-based processes in sentence comprehension. In *Technical report of IEICE. Thought and language*, volume 114(176), pages 113–114.

Himanshu Yadav, Ashwini Vaidya, Vishakha Shukla, and Samar Husain. 2020. Word order typology interacts with linguistic complexity: a cross-linguistic corpus study. *Cognitive Science*, (44).

George Kingsley Zipf. 1949. *Human behavior and the principle of least effort*. Addison-Wesley Press, Oxford, UK.