Cognitive Factors Influencing Word Order Variation in Hindi Actives and Passives

Sidharth Ranjan[1] Rajakrishnan Rajkumar[2] Sumeet Agarwal[1]
IIT Delhi[1], IISER Bhopal[2]
sidharth.ranjan03@gmail.com, rajak@iiserb.ac.in, sumeet@iitd.ac.in

Choices in language production are governed by both *production ease* and *communicative considerations* (Jaeger and Buz, 2017). In this work, we studied the impact of processing costs defined by Dependency Locality Theory (Gibson, 2000, DLT) and Surprisal Theory (Hale, 2001; Levy, 2008) on word order variation in Hindi active and passive sentences. A long line of work attests that languages tend to prefer sentences which minimize dependency length and surprisal (Hawkins, 2004; Gildea and Jaeger, 2015). Though both measures were originally defined to model comprehension load, recent works have extended them to model language production difficulty. Following the insights from previous works described above, our study used 7586 declarative sentences (6733 active and 853 passive) from the Hindi-Urdu Treebank (HUTB) corpus of written Hindi to create grammatical variants by permuting preverbal constituents as described in Page 4 (143721 active; 15170 passive). We computed the dependency length of each reference sentence and variant by summing the head-dependent distance of each dependency relation. Per-word surprisal estimated using trigram as well as a dependency parsing model was summed across all words to give sentence-level values of those measures.

First we trained separate regression models on a transformed version of our dataset (transformation procedure discussed in final page) consisting of reference-variant sentence pairs belonging to active and passive constructions. Based on the literature cited above (inter-alia), we predicted that the log odds of predicting the reference sentence increased with decreasing processing costs (*i.e.* dependency length and both kinds of surprisal). As shown in Table 2, all three predictors have significant negative coefficients in the case of both active and passive voice constructions, validating the original hypothesis in the literature stated above.

In order to understand the success of dependency length in predicting active voice sentences and failure in the case of passives, we examined the relative distribution of active and passive sentences in 7 bins of absolute dependency length difference between referent and variant sentences. As discussed in work under review (Ranjan et al., 2020), dependency length is most effective in the final bin (absolute dependency length difference of 32). As shown in Table1, in all the bins except the last one, passives constituted roughly 9-10% of the total number of data points (active voice pairs constituting the remaining cases). In the final bin, passive reference-variant pairs constituted only 5.76% of the cases in that bin *i.e.*, around 4% less than the passive voice cases in the previous bins. So fewer passives in this bin resulted in lower effectiveness of dependency length for this construction compared to active voice pairs. Passive reference sentences in the HUTB have lower average dependency length (29.46 words) compared to their active counterparts (38.98 words). This difference is because 86.04% of passives do not have an overt agent-NP. In spontaneous production experiments, Perera and Srivastava (2016) noted the tendency of Hindi speakers to drop the agent NP in passive sentences and interpreted it as a drive to avoid interference between animate nouns. Conceptual similarity betweeen two nouns can reduce their accessibility and lead to planning difficulties in speech production as attested by Smith and Wheeldon (2004). Thus minimization of dependency length is not powerful enough in the case of the passive voice construction, where surprisal effects dominate. In contrast to passives, subjects are realized overtly more frequently in active voice sentences leading to greater number of active cases in the final bin, where dependency length is very effective. A potential explanation for the efficacy of surprisal in passives is the primacy of accessibility based considerations. The psychological reality of our findings need to be validated using spoken datasets.

Bin	Len=0	Len=1	Len=2	2<len<=4	4<len<=8	8<len<=16	16<len<32	32<len<Inf
%Passive	10.23	10.07	9.53	9.86	9.84	9.72	9.13	5.76
#points	9469	13561	15423	25096	33782	33864	20230	7466

Table 1: Bin-wise percentage of passive sentences

Predictor /Intercept	Estimate	Std. Error	z-value
Intercept	0.002	0.011	-0.189
n-gram surprisal	-1.017	0.006	-155.43
parser surprisal	-0.781	0.015	-51.65
dependency length	-0.019	0.001	-22.29

(a) Active sentences (143721 points)

Predictor /Intercept	Estimate	Std. Error	z-value
Intercept	-0.008	0.031	-0.266
n-gram surprisal	-0.949	0.018	-52.45
parser surprisal	-1.019	0.051	-20.10
dependency length	-0.019	0.001	-22.29

(b) Passive sentences (15170 points)

Table 2: Regression results for construction-wise models containing all three predictors

References

Davison, A. (1982). On the form and meaning of hindi passive sentences. *Lingua*, 58(1):149 – 179.

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In Marantz, A., Miyashita, Y., and O'Neil, W., editors, *Image, Language, brain: Papers from the First Mind Articulation Project Symposium*. MIT Press, Cambridge, MA.

Gildea, D. and Jaeger, T. F. (2015). Human languages order information efficiently. *CoRR*, abs/1510.02823.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *NAACL Proceedings*, NAACL '01, pages 1–8, Pittsburgh, Pennsylvania. Association for Computational Linguistics.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press.

Jaeger, T. F. and Buz, E. (2017). Signal reduction and linguistic encoding. In Fernandez, E. M. and Cairns, H. S., editors, *Handbook of Psycholinguistics*, page To appear. Wiley-Blackwell.

Joachims, T. (2002). Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD*, KDD '02, pages 133–142, New York, USA. ACM.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126 – 1177.

Perera, C. K. and Srivastava, A. K. (2016). Animacy-based accessibility and competition in relative clause production in hindi and malayalam. *Journal of Psycholinguistic Research*, 45(4):915–930.

Ranjan, S., Rajkumar, R., and Agarwal, S. (2020). Locality and Expectation Effects in Hindi Preverbal Constituent Ordering. *Cognition*, Under revision.

Smith, M. and Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3):675–686.

Hindi

Hindi (Indo-Aryan language; Indo-European language family) has SOV canonical order along with a rich case-marking system and relatively free word order. The following examples illustrate word order flexibility in Hindi.[1]

(1) a.

aaj	maa=ne	bacce=se	kitaab	padh-ne=ko	kah-aa
today	mother=ERG	child=ACC	book	read-INF=ACC	say-PFV.M.SG

Today the mother told the child to read the book.

b. aaj kitaab maa ne bacce se padhne ko kahaa

c. aaj bacce se kitaab maa ne padhne ko kahaa

d. bacce se kitaab maa ne padhne ko aaj kahaa

e. maa ne kitaab aaj bacce se padhne ko kahaa

In the above examples, permuting the preverbal constituents of the first sentence results in the remaining variant sentences, which express the same propositional content. As noted in pioneering work, sentences having canonical order (like Example 1a above) can be considered to be neutral with respect to the preceding discourse context. Sentences with other orders (like the remaining sentences above), in contrast, are marked structures signifying various kinds of alternate emphases which require context for complete interpretation.

As mentioned in pioneering work by Davison (1982), Hindi passive sentences are typically used to signify one of three meanings, *viz.*, ability, prohibition and neutral. She also notes that the passive verbal morphology (VERB-perfective marker-*jaa*-'go'+ tense and aspect information) makes a passive sentences marked structures in comparison to their active counterparts. As mentioned in Davison's paper, the agent is optional in Hindi passive sentences. She further illustrated that passive sentences are more restricted semantically compared to the corresponding active sentence since passives cannot have inanimate agents and verbs denoting non-volitional acts. So researchers have argued against devices like transformations stating the equivalence of active and passive forms because of the shades of meaning conveyed by passives. The choice of the passive is linked to register and style as well as special choices made by the speaker. Typically, passives are preferred when the speaker wishes to adopt an elliptical strategy to convey a particular idea, instead of concise and direct communication. In the examples below, the writer does not want to explicitly mention the person or organization which removed them from an official position. In fact, there is a need for a model which factors in contextual information to facilitate conversational inferences related to passive sentences. In the following passive examples taken from our data, the surprisal scores for the reference sentence, Example 2a, are lower than those of the variant sentence, Example 2b (dependency length, n-gram surprisal, and parser surprisal shown alongside):

(2) a.

[mujhe]	[is	pad=se]	[apmanjanak]	[tareeke=se]	hataya	ga-ya
1.SG.DAT	this	position=INS	humiliating	manner=INS	removed	go-PST

hai (12, 23.77, 0.14)
be.PRS.SG

I have been removed from this position in a humiliating manner.

b. is pad se mujhe apmanjanak tareeke se hataya gaya hai. (10, 24.23, 0.21)

Thus the model chooses the reference sentence over the variant.

[1] We follow the Leipzig Glossing Rules (`https://www.eva.mpg.de/lingua/resources/glossing-rules.php/`) in this paper: ACC: accusative; ERG: ergative; FUT: future; INF: infinitive; INS: instrumental; M: masculine; PFV: perfective; PL: plural; PROG: progressive; PRS: present; PST: past; SG: singular

Condition	Label	Dependency length	n-gram surprisal	Parser surprisal
Reference	1	16	16.34	0.18
Variant_1	0	18	20.00	0.15
Variant_2	0	17	17.36	0.14

(a) Original feature values

Condition	Label	δ Dependency length	δ n-gram surprisal	δ Parser surprisal
Variant_1 -Reference	0	2	3.66	-0.03
Reference -Variant_2	1	-1	-1.02	0.04

(b) Transformed feature values

Table 3: Joachims' transformation

Variant Generation and Data Transformation

We created variants for each HUTB reference sentence using a reordering algorithm which takes as input the dependency tree corresponding to that reference sentence. The reordering algorithm permuted the preverbal dependents of the root verb and linearized the resulting tree to obtain variant sentences. In order to automatically ensure that only grammatical variants were chosen, we filtered out variants which did not contain root-level dependency-relation sequences attested in the gold standard corpus of HUTB trees. In cases where the number of variants exceeded 100, we chose 99 variants randomly. After the aforementioned procedure, we obtained a dataset comprising of 7586 reference sentences and 158891 variants. Our dataset contained substantially more variants than reference sentences. In order to mitigate this imbalance, we transformed our data set using a technique originally proposed by Joachims (2002) for ranking web pages. The transformation converts a binary classification task (labelling a sentence as reference vs variant) into a pairwise ranking task involving the feature vectors of a reference sentence and each of its variants. We trained a machine learning model on the difference between the aforementioned feature vectors as per the equations below:

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

Equation 1 above shows a data point where the model predicts that the reference sentence outranks one of its variants when the dot product of the feature vector of the reference sentence and $\mathbf{w}$ (learned feature weights) is greater than the corresponding dot product of the variant sentence. This relationship can also be expressed in the form of Equation 2 , where the feature values of the first member of the pair were subtracted from the corresponding values of the second member. We created ordered pairs consisting of the feature vectors of reference-variant sentences. Examples 1a-1b and Examples 1c-1a constitute two such sentence pairs whose feature vectors were paired. Pairs alternate between *reference-variant* (coded as "1") and *variant-reference* (coded as "0"), resulting in a balanced data set that contained either equal number of classification labels of each kind (if the total number of variants is an even number) or a difference of one (if total number of variants is an odd number). Table 3 illustrates the original and transformed values of the independent variables calculated as follows:

1. **Dependency length**: We defined dependency length as the number of intervening words between each head and dependent.
2. n**-gram surprisal**: We estimated n-gram surprisal using a trigram model (n=3) over words trained on 1 million sentences from the EMILLE corpus with Good-Turing discounting using the SRILM toolkit (`http://www.speech.sri.com/projects/srilm/`)
3. **Dependency parser surprisal**: We used the incremental probabilistic dependency parser developed at IIT Delhi (`https://github.com/samarhusain/IncrementalParser/`) to estimate dependency parser surprisal using a corpus of 12,000 HUTB projective trees. We divided this corpus into 10 sections and models trained on 9 sections were used to get incremental surprisal scores for the remaining section, thus covering the entire corpus.