

Investigating the Probability of External Causation in Hindi Light Verb Constructions

Kanishka Jain Ashwini Vaidya
Indian Institute of Technology Delhi



Goal of the Study

- Extend the notion of causativity to **nouns** in Light Verb Constructions (LVCs)
- In Hindi, predicating nouns affect the causalness of an LVC.
- Frequency can help us to identify the preferred causal or noncausal alternations for nouns
- Our preliminary results show that nouns can encode degrees of causalness based on corpus frequency

Introduction

- **Form-Frequency correspondence:** Frequent forms = less coded [Haspelmath 2008]
- causal alternation: noncausal $>_{\text{more frequent}}$ causal, noncausal alternation: causal $>_{\text{more frequent}}$ noncausal

alternation	gloss	noncausative		causative	
		form	freq	form	freq
causal	<i>freeze</i>	ganda	82	<i>gandisha</i>	20
noncausal	<i>break</i>	<i>vunjika</i>	336	vunja	883

Table 1. Examples of verb alternations in Swahili [Haspelmath 2008]

- Previous work has been limited to lexical and morphological causative alternations

Causal Properties of Nouns in LVCs

- In Hindi, nouns alternating with LVs **/kərna/** *do* and **/hona/** *be* signal causal and noncausal meaning.
- In Table 2, noun */cori/* *theft* + light verb */hui/* *be* is noncausal form whereas *cori/* + light verb */ki/* *do* is the causal form

Arg 1	Arg 2	NV	
1. -	gehnō-ki jewellery-GEN.F	cori	hui
<i>There was theft of jewellery.</i>			
2. lərke-ne boy.3.SG.M-ERG	gehnō-ki jewellery-GEN.F	cori	ki
<i>The boy stole the jewellery.</i>			

Table 2. Argument structure of noun */cori/* with */hona/* *be* and */kərna/* *do*

- Properties like type of arguments and their semantic roles (like agent and patient) are tied to nouns in LVCs. For instance,
 - */cori/* *theft* is an agentive noun and needs an agent, but
 - nouns like */izafa/* *increase* and */pəsənd/* *like* are stative nouns
- Nouns also posit selection restrictions on the light verbs [Butt 2010]
- In Table 3, */yad/* *memory* can combine with 3 LVs but */pəresʰani/* *trouble* doesn't
- Considering how nouns shape the structure and meaning of an LVC, it is interesting to ask if nouns can be ranked based on their degree of causalness

noun	light verb	meaning
yad	kərna (do)	intentionally remembering something/someone
	hona (be)	having a memory of someone/something
	ana (come)	unintentionally remembering something/someone
pəresʰani	hona (be)	having trouble

Table 3. Selectional restrictions on light verbs

- However, LVCs have *symmetric* coding¹, hence form-frequency correspondence cannot apply directly

¹ According to Haspelmath (2021), constructions like Hindi LVCs are examples of a 'uniformly explicit' coding system where efficiency is less important than the explicit coding of meaning

Methodology

- We compute the frequencies */kərna/-/hona/* (do-be) alternation from an annotated corpus
- **Corpus:** Hindi-Urdu Dependency Treebank (HUTB) [Bhatt et al. 2009]
- Calculated their *degree of causalness* [Haspelmath, 2008; Samardžić and Merlo, 2012]

$$\text{causalness} = \frac{\text{no. of do alternations} * 100}{(\text{no. of do} + \text{no. of be})}$$

- To evaluate for the validity and variability of our findings we test for:
- **Agency:** if nouns with high degree of causalness also have higher probability of occurring with the ergative marker (-ne) than other nouns
- **Reproducibility:** if similar pattern is observable in another corpus of Hindi
- For this, we have used a part of Hindi TimeBank (HTB) [Goel et al. 2020]

Results

- Found 25 common alternating pairs in HUTB and HTB
- Nouns do show a *general tendency* to occur either as a causative or noncausative across different corpus
- Degree of causalness is also reflected in the percentage of ergatives
- To verify the relationship between ergativity and casualness we calculated Spearman's rank correlation coefficient (ρ)
- ρ : 0.606 (level of significance = .01 (one-sided)) in HUTB and 0.323 in HTB

Noun	gloss	Caus HUTB	Caus HTB	%E HUTB	%E HTB
gʰoʃŋa	announce-ment	97.7	83.3	82.9	62.5
fəsla	decision	93.9	60	74.7	40
palən	comp-liance	87.5	90	37.5	0.0
ʃadi	marriage	57.1	61.1	42.8	55.6
bətʰək	meeting	37.5	66.7	15.6	41.7
prarəmbʰ	start	25	33.3	25	16.7
izafa	increase	16.7	28.6	8.3	0.0

Table 4. Sample of common LVC pairs from HUTB and HTB corpus. Caus=Causalness, %E= percentage of ergatives

Discussion

- The results suggest that nouns with high degree causalness tend to occur frequently with causal verb */kərna/* than with */hona/*.
- This is further verified by the correlation between causalness and ergativity for HUTB.
- This may be useful in the creation of lexical resources for predicating nouns

References

- Goel, P. et al. (May 2020). "Hindi TimeBank: An ISO-TimeML Annotated Reference Corpus". English. In: *16th Joint ACL-ISO Workshop on Interoperable Semantic Annotation Proceedings*, pp. 13–21. ISBN: 979-10-95546-48-1. URL: <https://aclanthology.org/2020.isa-1.2>.
- Samardžić, T. and P. Merlo (2012). "The meaning of lexical causatives in cross-linguistic variation". In: *Linguistic Issues in Language Technology* 7.
- Butt, M. (2010). "The light verb jungle: Still hacking away". In: *Complex predicates in cross-linguistic perspective*, pp. 48–78.
- Bhatt, R. et al. (2009). "A multi-representational and multi-layered treebank for Hindi/Urdu". In: *Proceedings of LAW III*, pp. 186–189.
- Haspelmath, M. (2008). "Frequency vs. iconicity in explaining grammatical asymmetries". In: *Cognitive Linguistics* 19.1, pp. 1–33. DOI: 10.1515/COG.2008.001.