Identifying Causal–Noncausal Light Verb alternations in Hindi

HSL748
Kanishka Jain (21HUZ8481)

Introduction (LVC)

Light Verb Constructions (LVC) consists two parts: Noun + Verb. The noun in an LVC gives the central meaning about the event and the light verb gives an additional semantics of agentivity, completeness, or permission (Vaidya et al., 2016).

1. Sam has taken a shower. (*Showering*)

   Verb Noun

Light Verb Constructions are quite frequent in languages like in Hindi. For instance, according to Vaidya et al. (2014), the Hindi Treebank (Palmer et al., 2009) has 37% of the predicates annotated as LVC.

2.

raam	ne	samajhne	ki	koshish	ki
Ram.3.sg.m	erg	understand.inf	gen	try	do.pst.f

   ‘Ram tried to understand.’ (*Trying*)

   Noun Verb

Introduction (Causality)

Causatives are common, however, complex lexical items as they increase the valency of the verb and thus demand investigation with more precision. For instance in Hindi:

3. /jamnaa/ ‘freeze’ → /jamaanaa/ ‘to (make) freeze’ (anticausative to causative)

4. /toDnaa/ ‘to break’ → /TuTnaa/ ‘break’ (causative to anticausative)

Image Source: Google Images

Related Work

Haspelmath (1993) investigates how causatives are derived from anticausatives and vice versa. He proposes that such verbs can be put on a scale of spontaneity (see 3 from Samardzic and Merlo, 2012) where spontaneity refers to the presence of external argument or simply causer.

5. freeze > dry > melt > ::::: > gather > open > break > close
   (spontaneous) (non-spontaneous)

Verbs with a low anticausative to causative (A/C) ratio describe events that are likely to happen with no external causer involved.

Samardžić and Merlo (2012) suggested calculating the ratio between the corpus frequencies of causative and anticausative. However they have calculated C/A ratio for their study.

Related Work (contd.)

Vaidya et al. (2014) has proposed that some nouns allow an agentive subject when they occur with the light verb like /kar/ ‘do’ however it does not have an agentive subject with /ho/ ‘be’

5.	Raam	ne	chori	ki	
	Ram.3.sg.m	erg	theft/steal	do.pst.perf.f	
	‘Raam has stolen.’				
6.	Raam	ka	saamaan	chori	hua
	Ram.3.sg.m	gen	belongings	theft/steal	be.pst.pfv.m
	‘Raam’s belongings were stolen.’				

Further certain nouns like /chori/ ‘theft’, /koshish/ ‘try’ are inherently agenitive in nature i.e. the presence of an agent or causer of an event is ‘default’ or presupposed, irrespective of the light verb.

7. ? apneaap koshish huyi
   ‘Trying itself’

8. apneaap kami huyi
   decreased itself’

Objective

This project investigates predicative nouns as part of LVCs in Hindi occurring as either causative or non-causative. Nominal predicates occuring with the light verb /kar/ “do” are causal constructions whereas nominals occurring with /ho/ “be” are considered as non-causal constructions.

- The main objective is to identify and classify LVCs occurring with ‘do’ and ‘be’ as causative and non-causative in Hindi corpus generating the scale of spontaneity (ranking).
- Second, to find the probability of a noun to occur as Causal/Causative Noun
- Third, to find the probability of a noun to occur as Causal/Causative Noun when not occurring with highly frequent verbs like ‘do’ and ‘give’.

Motivation

Why care about causality or LVs?

1. They are highly frequent and challenging.
2. Common strategy to deal with foreign terms. Consider examples like googling
3. Knowing that a predicate requires an external causer will help us to classify events better
4. It also helps in predicting the behaviour of lexical items.

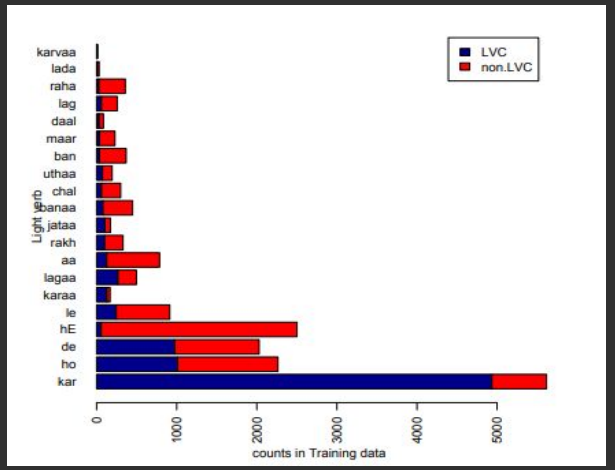


Figure 1: Light verb distribution in the Training and Development section of the Treebank (Vaidya et al. 2016)

Resources and Tools

Corpus: the Hindi Dependency Treebank (HDTB) - has 16647 sentences and 351704 tokens.

Python Libraries: UDAPI - an open-source framework providing an application programming interface (API) for processing Universal Dependencies data.

```
# sent_id = dev-s1
# text = रामायण काल में भगवान राम के पुत्र कुश की राजधानी कुशावती को 483 ईसा पूर्व बुद्ध ने अपने अंतिम विश्राम के लिए चुना ।
1	रामायण	रामायण	PROPN	NNPC	Case=Nom|Gender=Masc|Number=Sing|Person=3	2	compound	_	Vib=0|Tam=0|ChunkId=NP|ChunkType=child|Translit=rāmāyaṇa
2	काल	काल	PROPN	NNP	Case=Acc|Gender=Masc|Number=Sing|Person=3	23	obl	_	Vib=0_में|Tam=0|ChunkId=NP|ChunkType=head|Translit=kāla
3	में	में	ADP	PSP	AdpType=Post	2	case	_	ChunkId=NP|ChunkType=child|Translit=meṁ
4	भगवान	भगवान	NOUN	NNC	Case=Nom|Gender=Masc|Number=Sing|Person=3	5	compound	_	Vib=0|Tam=0|ChunkId=NP2|ChunkType=child|Translit=bhagavāna
5	राम	राम	PROPN	NNP	Case=Acc|Gender=Masc|Number=Sing|Person=3	7	nmod	_	Vib=0_का|Tam=0|ChunkId=NP2|ChunkType=head|Translit=rāma
6	के	का	ADP	PSP	AdpType=Post|Case=Acc|Gender=Masc|Number=Sing	5	case	_	ChunkId=NP2|ChunkType=child|Translit=ke
7	पुत्र	पुत्र	NOUN	NN	Case=Acc|Gender=Masc|Number=Sing|Person=3	8	nmod	_	Vib=0|Tam=0|ChunkId=NP3|ChunkType=head|Translit=putra
8	कुश	कुश	PROPN	NNP	Case=Acc|Gender=Masc|Number=Sing|Person=3	10	nmod	_	Vib=0_का|Tam=0|ChunkId=NP4|ChunkType=head|Translit=kuśa
9	की	का	ADP	PSP	AdpType=Post|Case=Acc|Gender=Fem|Number=Sing	8	case	_	ChunkId=NP4|ChunkType=child|Translit=kī
10	राजधानी	राजधानी	NOUN	NN	Case=Acc|Gender=Fem|Number=Sing|Person=3	11	nmod	_	Vib=0|Tam=0|ChunkId=NP5|ChunkType=head|Translit=rājadhānī
11	कुशावती	कुशावती	PROPN	NNP	Case=Acc|Gender=Fem|Number=Sing|Person=3	23	obj	_	Vib=0_को|Tam=0|ChunkId=NP6|ChunkType=head|Translit=kuśāvatī
12	को	को	ADP	PSP	AdpType=Post	11	case	_	ChunkId=NP6|ChunkType=child|Translit=ko
13	483	483	PROPN	NNPC	Case=Nom|Gender=Masc|Number=Sing|Person=3	15	compound	_	Vib=0|Tam=0|ChunkId=NP7|ChunkType=child|Translit=483
14	ईसा	ईसा	PROPN	NNPC	Case=Nom|Gender=Masc|Number=Sing|Person=3	15	compound	_	Vib=0|Tam=0|ChunkId=NP7|ChunkType=child|Translit=īsā
15	पूर्व	पूर्व	PROPN	NNP	Case=Nom|Gender=Masc|Number=Sing|Person=3	23	obl	_	Vib=0|Tam=0|ChunkId=NP7|ChunkType=head|Translit=pūrva
16	बुद्ध	बुद्ध	PROPN	NNP	Case=Acc|Gender=Masc|Number=Sing|Person=3	23	nsubj	_	Vib=0_ने|Tam=0|ChunkId=NP8|ChunkType=head|Translit=buddha
17	ने	ने	ADP	PSP	AdpType=Post	16	case	_	ChunkId=NP8|ChunkType=child|Translit=ne
18	अपने	अपना	PRON	PRP	Case=Acc|Gender=Masc|PronType=Prs	20	nmod	_	Vib=0|Tam=0|ChunkId=NP9|ChunkType=head|Translit=apane
19	अंतिम	अंतिम	ADJ	JJ	Case=Acc	20	amod	_	ChunkId=NP10|ChunkType=child|Translit=aṁtima
20	विश्राम	विश्राम	NOUN	NN	Case=Acc|Gender=Masc|Number=Sing|Person=3	23	obl	_	Vib=0_के_लिए|Tam=0|ChunkId=NP10|ChunkType=head|Translit=viśrāma
21	के	के	ADP	PSP	AdpType=Post	20	case	_	ChunkId=NP10|ChunkType=child|Translit=ke
22	लिए	लिए	ADP	PSP	AdpType=Post	20	case	_	ChunkId=NP10|ChunkType=child|Translit=lie
23	चुना	चुन	VERB	VM	Aspect=Perf|Gender=Masc|Number=Sing|VerbForm=Part|Voice=Act	0	root	_
ChunkId=VGF|ChunkType=head|Stype=declarative|Tam=yA|Translit=cunā|Vib=ा
24	।	।	PUNCT	SYM	_	23	punct	_	ChunkId=BLK|ChunkType=head|Translit=.
```

Methodology

To find C/A ratio:

1. Filtered all the occurrences of nouns with /karnaa/ ‘do’ and /honaa/ ‘be’ and created a list of all their immediate dependents : 4183 (nouns + kar tokens), 792 (nouns + ho tokens) → 616 (common tokens) —> 133 (types)
2. Filtered the data for instances where the frequency of both causal and non-causal event 1 or less. Now we got list of 54 nouns. Calculated the C/A ratio for these 54 nouns

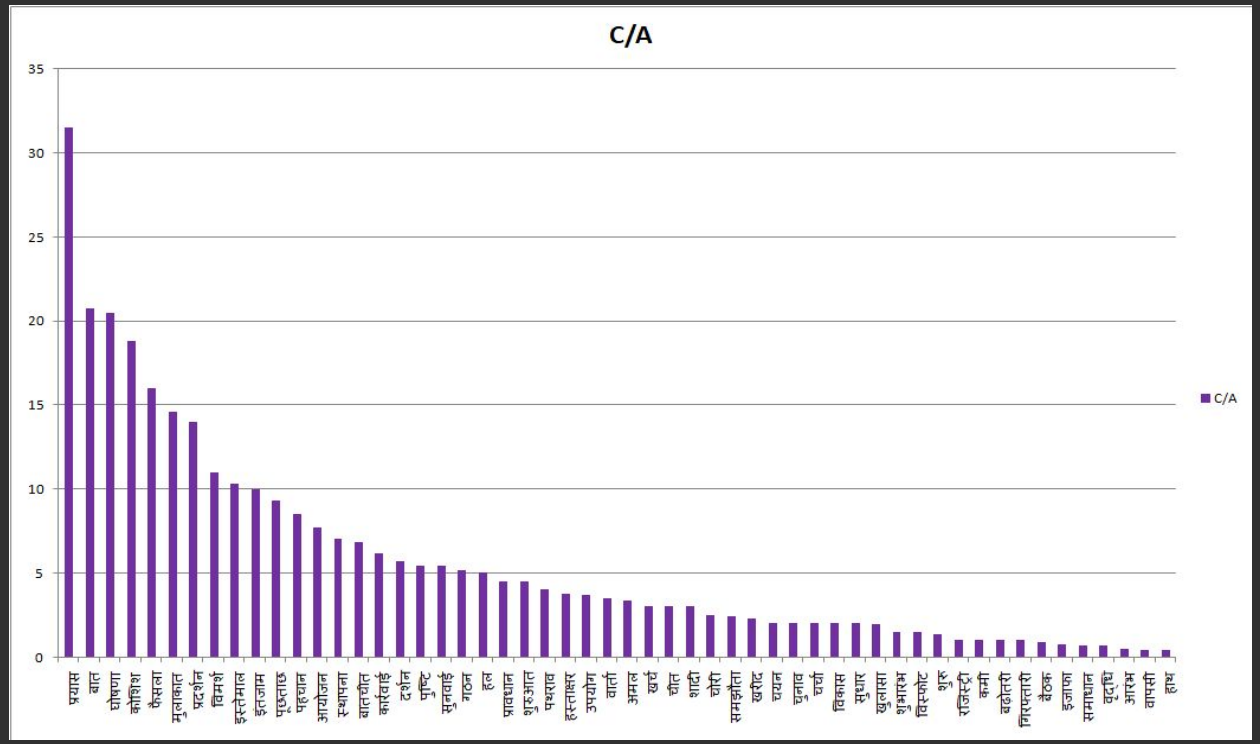
To find the probability of a noun to occur as a Causal Noun:

3. Calculated the frequency of all predicative nouns in LVC sentences by filtering those that have agentive subject (noun + /ne/)
4. Divided it by the frequency of all predicative nouns from the corpus for all verbs

To find the probability of a noun to occur as a Causal Noun when not occuring with common light verbs:

5. Calculated the frequency of all predicative nouns in LVC sentences by filtering those that have agentive subject (noun + /ne/)
6. Divided it by the frequency of all predicative nouns from the corpus for all verbs except (/kar/, /de/ and /le/)

Results (Part-1)



Causative <— Spontenity Scale —> Non-causative

Figure 2: Spontaneity Scale (C/A ratio) for 54 predicative nouns found in HDTB corpus

Noun	Causative	Anticausative	C/A
खरीद	9	4.0	2.250000
समझौता	12	5.0	2.400000
चोरी	5	2.0	2.500000
चोट	9	3.0	3.000000
शादी	9	3.0	3.000000
खर्च	24	8.0	3.000000
अमल	10	3.0	3.333333
वार्ता	7	2.0	3.500000
उपयोग	11	3.0	3.666667
हस्ताक्षर	15	4.0	3.750000
पथराव	8	2.0	4.000000
प्रावधान	9	2.0	4.500000
शुरुआत	27	6.0	4.500000
हल	10	2.0	5.000000
गठन	31	6.0	5.166667
पुष्टि	27	5.0	5.400000
सुनवाई	27	5.0	5.400000
दर्शन	17	3.0	5.666667
कार्रवाई	37	6.0	6.166667
बातचीत	75	11.0	6.818182
स्थापना	14	2.0	7.000000
आयोजन	23	3.0	7.666667
पहचान	17	2.0	8.500000
पूछताछ	28	3.0	9.333333

Figure 3: Table showing a subset of predicative nouns, their frequency as causative and anticausative and their C/A ratio

Results (Part-2)

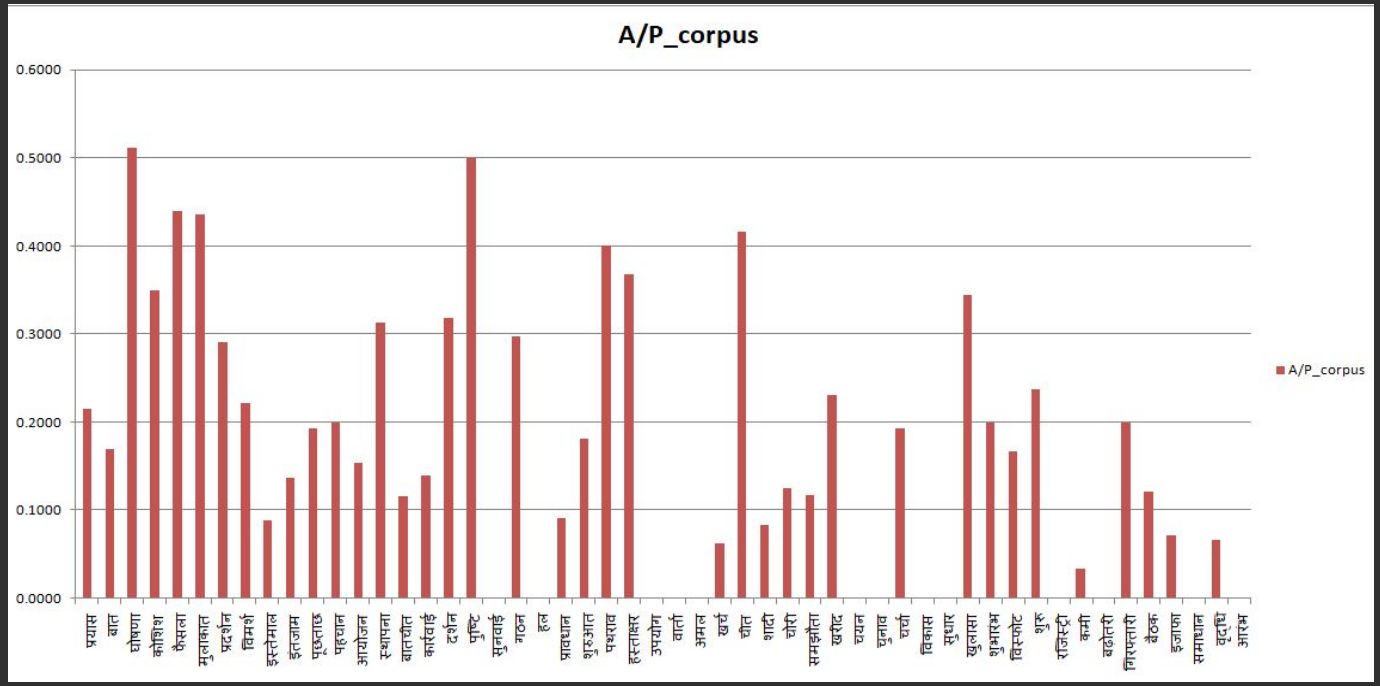


Figure 4: Ratio of the count of 54 predicative nouns occurring with an agentive noun in subject and the total count of 54 predicative nouns occurring as LVCs

Results (Part-3)

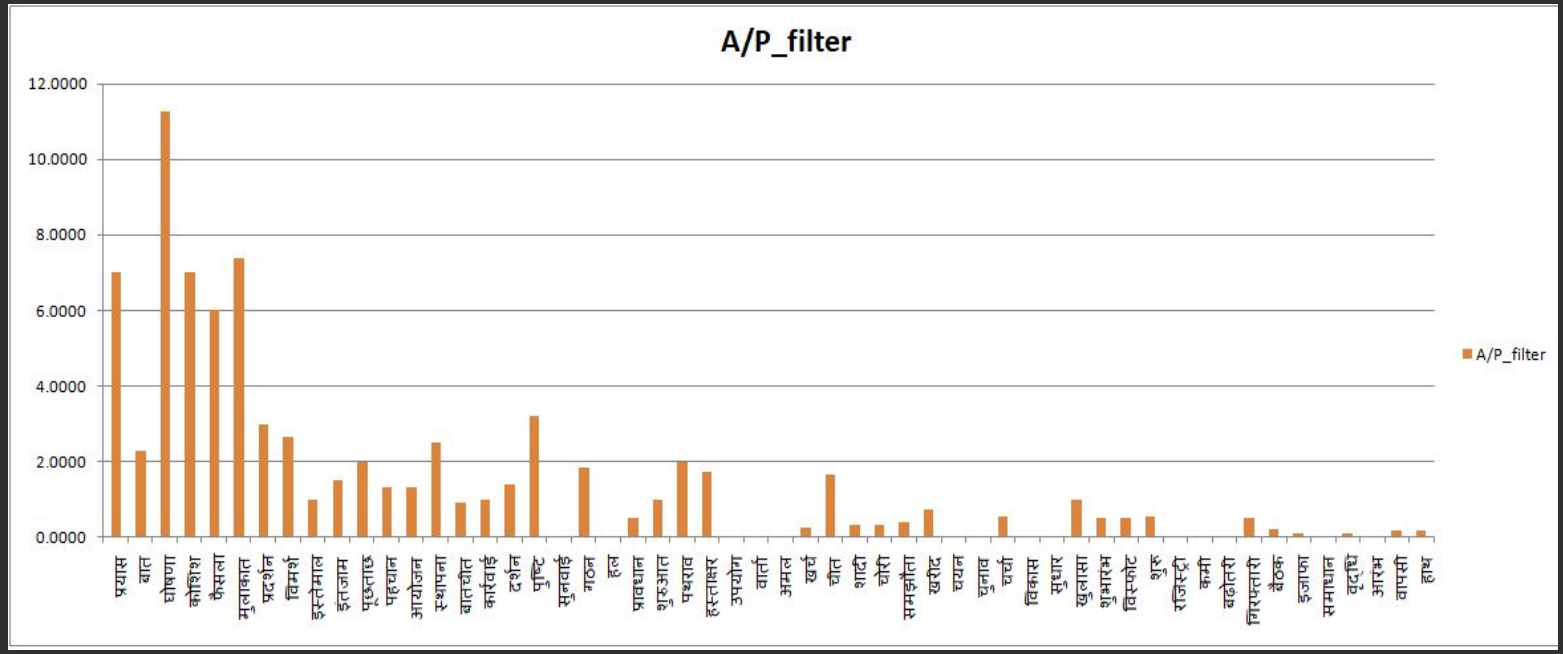


Figure 5: Ratio of the count of 54 predicative nouns occurring with an agentive noun in subject by filtering 3 most common light verbs (do, give, take) and the total count of 54 predicative nouns occurring as LVCs

Discussion

We have:

1. calculated the spontaneity scale for the 54 predicative nouns
2. seen that nouns at the causal end have higher probability of occurring with agentive subjects/external arguments (noun + /ne/) as compared to non-causal(s)
3. further shown that such nouns (causal) even if they occur with other less frequent light verbs they still occur more frequently in constructions with agentive nouns/external argument (noun + /ne/)

References

Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., & Xia, F. (2009, August). A multi-representational and multi-layered treebank for hindi/urdu. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)* (pp. 186-189).

Chomsky, N. (2014). *The minimalist program*. MIT press.

Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. *Causatives and transitivity*, 23, 87-121.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries.

Haspelmath, M., Calude, A., Spagnol, M., Narrog, H., & Bamyaci, E. (2014). Coding causal–noncausal verb alternations: A form–frequency correspondence explanation1. *Journal of Linguistics*, 50(3), 587-625.

Samardžić, T., & Merlo, P. (2012). The meaning of lexical causatives in cross-linguistic variation. *Linguistic Issues in Language Technology*, 7.

Popel, M., Žabokrtský, Z., & Vojtek, M. (2017, May). Udapi: Universal API for universal dependencies. In *Proceedings of the NoDaLiDa 2017 workshop on universal dependencies (UDW 2017)* (pp. 96-101).

Vaidya, A., Rambow, O., & Palmer, M. (2014, August). Light verb constructions with ‘do’and ‘be’in Hindi: A TAG analysis. In *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing* (pp. 127-136).

Vaidya, A., Agarwal, S., & Palmer, M. (2016, December). Linguistic features for Hindi light verb construction identification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 1320-1329).