# Degrees of animacy and telicity in Hindi intransitives: A corpus study





## Background

- ► Two syntactic classes of intransitives: **unergatives** with only an external argument and unaccusatives with only an internal argument (Burzio, 1986; Perlmutter, 1978)
- Different roots have different preferences for one syntax over another (e.g., break vs. work)  $\rightarrow$  What determines these preferences?
- Claim in previous research (Dowty, 1991; Sorace, 2000, 2004, 2011; a.m.o.):
  - Highly agentive semantics favors an unergative syntax Telic events tend to be realized with an unaccusative syntax
- Limitations of previous research:
  - Effect of agentivity and telicity is largely reported based on intuitive judgments; little quantitative evidence
  - Studies often presuppose that a root's behavior is categorical and/or that agentivity and telicity are binary features (see also Kim et al., 2024)
  - Much research focuses on well-studied European languages

### Our goal

We investigate whether agentivity is correlated with unergativity, and telicity with unaccusativity, using corpus data from Hindi. We assume that roots might possess agentivity and telicity to varying degrees, and that this might result in gradient preferences for unergative/unaccusative structures.

#### Methods

- Data are taken from the Hindi Dependency Treebank

- All code is available via the QR code
- ▶ For each root, we compute its **animacy score** as its likelihood to take an animate subject
  - Animacy is an (imperfect) proxy for agentivity
  - In Hindi, nouns in subject position bearing ergative case are highly likely to be animate, nouns in direct object position not bearing direct object marking are highly likely to be inanimate
- We use case marking to compile two lists of animate and inanimate nouns, respectively, then cleaning the lists manually
- We also compute for each root its telicity score as the number of times it occurs with a telic light verb (a: 'come' or ja: 'go')
  - To normalize the telicity score, we divide by the total number of basic intransitive occurrences of the root

#### Results

Root	Translation	Frequency	Animacy	<b>Telicity score</b>	Telicity score
			score	(absolute)	(normalized)
bik	be sold	5	0	0	0
sulajh	get untangled	5	0	2	0.4
chha	cover	6	0	0	0
bigaṭ	be spoiled	6	0	0	0
bixt	elapse	6	0	0	0
sudhar	improve	8	0	1	0.125
ţuuţ	break	15	0	12	0.8
khul	open	25	0.04	0	0
baṛh	increase	66	0.08	33	0.5
paṛ	fall	58	0.1	6	0.1
ghaṭ	reduce	9	0.11	0	0
phail	spread	9	0.11	0	0
ruk	stop	9	0.11	4	0.4
ţik	stay	8	0.125	0	0
bah	flow	7	0.14	8	1.14
jam	freeze	5	0.2	6	1.2
bandh	be tied	5	0.2	0	0
nikal	emerge	33	0.21	13	0.39
gir	fall	9	0.22	10	1.11
ho	be	12221	0.23	755	0.06
chal	walk, function	134	0.28	9	0.07
juṛ	be connected	18	0.33	4	0.22
ban	become	136	0.36	70	0.51
bach	be saved	24	0.375	11	0.46
ha:r	lose	5	0.4	4	0.8
ghir	be surrounded	5	0.4	0	0
baiṭh	sit	15	0.47	0	0
ubhar	emerge	6	0.5	0	0
gujar	elapse	8	0.5	2	0.25
mar	die	8	0.5	8	1
pahunc	arrive	107	0.5	61	0.57
jax	go	111	0.55	0	0
lauṭ	return	21	0.57	16	0.76
dauṛ	run	5	0.6	4	0.8
bha:g	run, escape	8	0.75	8	1
haṭ	remove oneself	6	1	1	0.17

Results for roots with a total number of occurrences > 4

## Results (cont.)

- For agentivity, the numbers are largely as expected:
- High agentivity scores for prototypical unergatives such as go and run
- Low scores for prototypical unaccusatives such as break, open and fall
- For telicity, the results are more complex:
  - High telicity scores for prototypical unaccusatives such as break, freeze and fall, but low scores for open and improve
  - → Are the latter more open-ended processes?
  - Low score for the prototypical unergative *go*, but high score for *run* → The latter can switch to an unaccusative use in telic environments, then licensing reduced relatives (1):
- run-pfv boy \*'the run boy'
- a. # daur-a: larka: b. sku:l=tak daur-a: larka: school=to run-pfv boy 'the boy who ran to school'

#### Discussion

#### **Limitations**:

- Data set is too small to draw reliable conclusions: total number of occurrences is low for most verbs; few clear unergatives in the sample
- We do not have a quantitative measure of the syntactic behavior of roots we could compare the results to
- Animacy is only one aspect of agentivity (besides intentionality, volition, causal power...)

#### ► Takeaways and open questions:

- Two strategies for investigating root meaning: counting contexts (e.g., our work) and collecting speakers' intuitions (e.g., Kim et al., 2024)
  - → Which of them is superior? What if they diverge?
- Are agentivity/telicity properties of the root or of the sentence? How strong a predictor is the root for sentence-level agentivity/telicity?

#### Summary

Our data tentatively support an effect of animacy and telicity on verbal syntax, suggesting that investigating contextual cues is a promising strategy for understanding the semantic correlates of the unergative/unaccusative distinction.

# Selected references and acknowledgments

Burzio (1996). Italian syntax: A Government and Binding approach. Dowty (1991). Thematic proto-roles and argument selection. Kim, Binder, Humphries & Conant (2024). Decomposing unaccusativity: a statistical modelling approach. Perlmutter (1978). Impersonal passives and the Unaccusative Hypothesis. Sorace (2000). Gradients in auxiliary selection with intransitive verbs. **Thanks** to Kyle Johnson, Brian Dillon and Katrin Erk.

eva-neu.github.io eneu@umass.edu