



CREATIVE INTERPRETATION OF VARIABILITY

**40th South Asian Languages
Analysis Roundtable
(SALA-40)**

9--11th September, 2026

ABSTRACT BOOKLET

For more information visit our website:



40th South Asian Languages Analysis Roundtable (SALA-40)

9–11th September, 2026

Venue: CITEC Building, Closest tram stop: Wellensiek (Tram line 4)

Day 1: Room 1.204 on 1st floor

Time	Title
08:30–09:30	Registration
09:30–10:00	Welcome (Jutta M. Hartmann & Farhat Jabeen)
Session 1: Creativity	
Session chair: Indranil Dutta	
10:00–10:30	<i>When script changes but language remains: A computational study of Romanized Indic Texts</i> Chetana Ghosh
10:30–11:00	<i>Telugu or Thelagoo? Variability in civilian romanization among Telugu natives</i> Vikram Goudar
11:00–11:30	Coffee Break
11:30–12:30	Plenary talk: <i>Acoustic features of phonetic representations in early bilinguals and language learners are creatively similar.</i> Indranil Dutta
Session chair: Farhat Jabeen	
12:30–13:30	Lunch
Session 2: Comparative Linguistics	
Session chair: Shravani Patil	
13:30–14:00	<i>Pre-verbal and post-verbal negation in Odia: Distribution and typological patterns</i> Dibya Panda & Thapasya Jayaraj
14:00–14:30	<i>Dative “subjects” and “direct objects” – The dative case in Middle and Modern Kannada</i> John Peterson
14:30–15:00	<i>Typological universals and variation in objective case marking: Evidence from Eastern India</i> Shuvam Dutta
15:00–15:30	Coffee Break
Session 3: Morphology/Syntax	
Session chair: Sandhya Sundaresan	
15:30–16:00	<i>Morphological priming of Bangla prefixed and pseudo-prefixed words: A sentence-prime lexical decision study</i> Basita Biswas & Indranil Dutta
16:00–16:30	<i>On modally flavoured silent predicates in Vedic Sanskrit copular sentences</i> Davide Mocci
16:30–17:00	<i>Event integration in Hindi multi-verb predicates</i> Sakura Ishikawa
17:00–17:30	<i>Case and theta-driven processing biases in Malayalam non-subject antecedent preference</i> Anil Devaprabha

18:30 (Self-paid) Conference dinner at Brauhaus Joh. Albrecht, Hagenbruchstraße 8, 33602, Bielefeld

Day 2: Room 1.204 on 1st floor

Time	Title
Session 4: Dialect studies/ Descriptive / Sociolinguistics	
Session chair: Chinmay Dharurkar	
09:00–09:30	<i>Language shift in Jharkhand: A study of linguistic identities and ideologies</i> Ariba Khan
09:30–10:00	<i>Ergative variation in the Konkani varieties spoken in Karnataka</i> Ruta Paradkar
10:00–10:30	<i>Arranging converbs in Domaaki</i> Noboru Yoshioka
10:30–11:00	Coffee Break
11:00–12:00	Plenary Talk: <i>Some syntax puzzles from Tamil, and why we should care.</i> Sandhya Sundaresan Session chair: Jutta Maria Hartmann
12:00–13:00	Lunch
Session 5: Historical Linguistics	
Session chair: Désirée Kleineberg	
13:00–13:30	<i>Ingressive ↔ progressive: Bidirectional shifts in Indo-Aryan grammaticalisation</i> Chinmay Dharurkar
13:30–14:00	<i>Sources of polar question particles in some Indian languages</i> Aaditya Kulkarni & Chinmay Dharurkar
Session 6: Computational/Corpus Linguistics	
Session chair: Jutta Maria Hartmann	
14:00–14:30	<i>Reassessing relative constructions in 19th Century Urdu: A corpus based analysis of ‘jo’ correlatives and embedded relatives</i> Miki Nishioka & Muhammad Nawaz
14:30–15:00	<i>Exploring place names through new computational tools: Two case studies in India</i> Lennart Chevallier & Søren Wichmann
15:00–15:30	Coffee Break
Session 7: Phonetics & Phonology	
Session chair: Ákos Buza	
15:30–16:00	Diachronic and dialectal evidence for the presence of central vowels in Old Tamil Ramprashanth Venkatakrishnan & Chandan Narayan
16:00–16:30	<i>Acoustics of four-way laryngeal contrast in Kudmali stops</i> Baidyanath Mahato & Indranil Dutta
16:30–17:00	<i>Phonological transitions and contact phenomena in the Maithili–Angika continuum</i> Abhishek Kumar & Katyayani Singh
City tour: Meet up at 18:00. Jahnplatz clock	

Day 3: Room 1.204 on 1st floor

Time	Title
Session 8: Semantics / Pragmatics	
Session chair: Maryam Mohammadi	
09:00–09:30	<i>Subject obviation in directive sentence types in Tamil</i> Gopalakrishnan Ramamurthy
09:30–10:00	<i>On the clause-finality of Marathi polar question particles</i> Aaditya Kulkarni
10:00–10:30	Coffee Break
10:30–12:30	Tutorial: <i>Evaluating acceptability through statistical analysis</i> Farhat Jabeen
12:30–12:45	Closing Remarks

Acoustic features of phonetic representations in early bilinguals and language learners are creatively similar

Indranil Dutta

School of Languages and Linguistics, Jadavpur University, Kolkata, India

Shared phonological patterns between two or more languages, intuitively, should aid acquisition of L2 and/or Lx languages, when Lx learners need only analogically extend those patterns to newer paradigms. Japanese and Bengali both exhibit consonantal length contrasts, with the exception that Japanese doesn't natively exhibit length among voiced obstruents (Lahiri and Hankamer 1988; Kubozono et al. 2009). In the first part of this talk, we show that Japanese learners of Bengali, however, acquire a voiced singleton-geminate contrast akin to native level phonetic representations within a month of exposure to Bengali (Basak et al. 2025). Transfer from L1 by way of constraints such as Lyman's law are overcome by speakers as evidenced by acquisition of duration patterns mitigated by place of articulation and voicing. In addition to duration patterns, we also report on the acquisition of formant dynamics by Japanese L2/L3 learners of Bengali.

In the second part of the talk, we present results from two experiments; production and perception of code-switched Bengali-English word pairs. We show that balanced bilinguals do not exhibit switch cost patterns in the acoustics, however, mismatched coarticulatory information slows hearers in a concept monitoring task in both monolingual and bilingual speech. We argue that parallel activation in balanced bilinguals leads to creative phonetic representations that also encode acoustic variation due to coarticulation (Deb et al. 2026).

In the third part of the talk, we present results from measurements of anti-symmetric Fourier component coefficients in Bengali geminates. We show that Bengali voiced geminates exhibit active cavity expansion for facilitating voicing for a longer durations by shifting the location of the constriction to a more anterior position (Krishnaswamy et al. 2019).

Based on these disparate but linked studies, we claim that creative phonetic representations comprise of micro-phonetic acoustic variation among language learners and balanced bilinguals. These variations can be seen across durational and spectral acoustic cues. We argue that phonetic representations among language learners and bilinguals both provide a fecund site for exploring how parallel activation takes shape.

References

- Basak, Kounish, Sankhadeep Ghosh, Indranil Dutta, and Tariq Shekh. 2025. "Shared and Non-Shared Features in Interlanguage Phonology: A Study of Japanese Speakers of Bengali (L2/3)." *Proceedings of JSLS2025: 26th Annual International Conference of the Japanese Society for Language Sciences* (Matsuyama), July 12–13.
- Deb, S., I. Dutta, and N. S. Dash. 2026. "Phonetic Overlaps in Code-Switch: Evidence from Coarticulatory Cues in Early Bengali-English Bilinguals' Perception and Production of Mixed Speech." *6th International Symposium on Bilingual and L2 Processing in Adults and Children (ISBPAC 2026)* (Reading, United Kingdom).
- Krishnaswamy, Meghavarshini, Indranil Dutta, and Ushasi Banerjee. 2019. "Active Cavity Expansion Through Lingual Adjustments to Place of Constriction in Voiced Geminates." *Proceedings of Meetings on Acoustics* 33.
- Kubozono, Haruo, Junko Ito, and Armin Mester. 2009. "Consonant Gemination in Japanese Loanword

Phonology.” In *Current Issues in Unity and Diversity of Languages: Collection of Papers Selected from the 18th International Congress of Linguists [CIL 18]*, edited by The Linguistic Society of Korea. Dongam Publishing Co.

Lahiri, A., and J. Hankamer. 1988. “The Timing of Geminate Consonants.” *J. Phonetics* 16: 327–38.

Some syntax puzzles from Tamil, and why we should care

Sandhya Sundaesan

In this talk, I will focus on a pesky puzzle involving certain types of "shifty" agreement in my dialect of Tamil. These involve 1sg agreement being triggered by a 3sg and 2sg anaphor/logophor that is marked in the nominative case. In Sundaesan (2012), I analyzed this 1sg form as the result of agreement due to a silent 1st person shifted indexical in the clausal periphery. I am no longer convinced this is the right way to analyze these facts. The syntactic properties of agreement leave distinct morphological reflexes in Tamil; furthermore, agreement and case interact in telling ways. I propose that, together, these facts simultaneously restrict the possibility space of logical solutions while also paving the way for a distinct analysis of the complex agreement facts.

I will propose that shifty agreement results from the idea that the logophors in these constructions bear hybrid person features: 1st- and 2nd- in one case, and 1st- and 3rd-, in the other. I will also propose that person agreement may be implemented through a distinct probe-goal relation than gender and number agreement. Taken together, these entail that "shifty" person agreement is a kind of hybrid agreement, along the lines of what has been reported for hybrid gender agreement phenomena in languages like Russian and Bosnian-Serbo-Croatian.

There are two natural extensions to this idea. The first is that certain types of logophor (and, indeed, the kind of shifty agreement I report for Tamil has been reported for other languages, like Telugu (also Dravidian), and Aqusha Dargwa (Nakh-Dagestanian) and Mutki Zazaki (Indo-Iranian)) are defined by their unique blend of person features arranged in a particular syntactic hierarchy. The second is that the distinct properties of pro-forms -- locally-bound anaphors, unshifty logophors, shifty logophors, shifted indexicals, and unshifted indexicals & deictic pronouns -- can be made to fall out of incremental differences in the size and shape of their nominal projections.

Case and Theta-Driven Processing Biases in Malayalam Non-Subject Antecedent Preference

Devaprabha K Anil | devaprabhaphdlandp24@efluniversity.ac.in
The English and Foreign Languages University, Hyderabad

Abstract: This study investigates the morpho-syntactic primacy underlying antecedent selection for the Malayalam reflexive *taan*. As a critical test case for binding theory, *taan* exhibits unique characteristics: it is morphologically underspecified (number-marked but gender-neutral) and possesses exceptional domain flexibility. Unlike standard reflexives, its antecedent need not occupy the minimal clause, allowing for unbounded dependencies that can extend up to five clauses or more even though economically scarce, exceeding the range of similar markers like Icelandic *sig*.

Taan _{i/j/k}	paripaat _i -kkə	var-um	ennə	raanj _i y-ootə	Raadha _i
self	programme-DAT	come-FUT	COMP	rani-SOC	Radha.NOM
paraj _n -u	ennə	Siitaj _i -kkə	toon _n -i	ennə	Kamala _k
tell-PAST	COMP	Sita-DAT	feel-PAST	COMP	Kamala.NOM
paraj _n -u					
tell-PAST					

‘Kamala_j said Sita_j felt that Radha_i told Rani that self_{i/j/k} would go to the programme’

Theoretical Framework: We challenge conventional phase-based accounts of binding by proposing a model where case-marking and theta-theoretic hierarchies govern anaphora resolution. Our investigation demonstrates that these morpho-syntactic constraints systematically predict antecedent preferences in both local and non-local contexts.

Methodology: The study utilized a self-paced reading task involving fifteen native Malayalam speakers, aged 18-25, to examine the processing of the reflexive *taan* embedded in a case-sensitive theta-driven non-subject headed clause. The stimuli consisted of 30 sentences, comprising 24 target items with multiple embeddings and 6 unambiguous fillers, randomized to mitigate priming and pattern bias. By manipulating word order through scrambling and varying clausal segmentation across minimal pairs, the experiment evaluated the interaction between syntactic structure and pragmatic conditions in co-referential perception. Participants identified preferred antecedents through post-sentence comprehension checks, providing reaction time and accuracy data to clarify the processing hierarchy underlying dependency resolution.

Findings: The study's key findings establish that antecedent selection for the Malayalam reflexive *taan* is fundamentally governed by morpho-syntactic primacy rather than purely configurational rules. Central to these results is the discovery that while DPs embedded within a subject NP can serve as potential antecedents, their accessibility is strictly modulated by the verbal argument structure and the specific theta-role configuration of the subject. Notably, genitive DPs in these positions can only function as antecedents under a logophoric interpretation; any violation of this constraint induces strong ungrammaticality. Furthermore, the research highlights a syntax-semantics interface effect where experiencer DPs whether they function as dative subjects or oblique arguments, manifest subject-like binding properties in psych-verb and stative constructions. This suggests that theta-roles can override standard structural positions, supporting a relativized view of subjecthood within Dravidian syntax. These theoretical insights are empirically supported by self-paced reading experiments, which recorded significantly faster resolution times when antecedent selection adhered to case-theoretic and theta-theoretic constraints.

Experimental Validation: Controlled self-paced reading experiments validate these theoretical claims. Participants demonstrated significantly faster resolution times when antecedent selection adhered to case-theoretic and theta-theoretic constraints. These results corroborate Case Theory and the theta Criterion as robust predictive frameworks for Malayalam anaphora and clarify the processing hierarchy underlying dependency resolution.

Mean Preferential Bias Percentage in Target sentences		Mean Preferential bias percentage in Fillers	
Most likely antecedent	Most unlikely antecedent	Most likely antecedent	Most unlikely antecedent
79.13%	17.97%	97.77%	2.22%

Fig 1: Mean Preferential Bias

Keywords: Malayalam syntax, Processing, Co-referentiality, Case-marking, Reflexive binding, Logophoricity, theta-theory, Psych verbs, Dative subjects

References: Amritavalli, R. (1984). Anaphorization in Dravidian. CIEFL Working Papers in Linguistics, 1, 1–31.

Amritavalli, R. (1999). Lexical anaphors and pronouns in Kannada. In J. G. K. S. B. Lust, K. Wali (Ed.), Lexical anaphors and pronouns in selected South Asian languages: a principled typology (pp. 49–112). Berlin: Mouton de Gruyter.

Hariprasad, M. (1998). Anaphorization in Telugu and English. PhD thesis, Central Institute of English and Foreign Languages.

Heim, Irene. 1993. Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. Sfs-Report-07- 93. Tübingen: University of Tübingen.

Jayaseelan, K. A. (1997). Anaphors as pronouns. *Studia Linguistica*, 51, 186–234.

Jayaseelan, K. A. (1998). Blocking effects and the syntax of Malayalam taan. In R. Singh (Ed.), The yearbook of south asian languages and linguistics (pp. 11–27). New Delhi: Sage Publications.

Jayaseelan, K. A. (1999). Lexical anaphors and pronouns in Malayalam. In J. G. K. S. B. Lust, K. Wali (Ed.), Lexical anaphors and pronouns in selected South Asian languages: a principled typology (pp. 113–168). Berlin: Mouton de Gruyter.

Jayaseelan, K. A. (2001). IP-internal topic and focus phrases. *Studia Linguistica*.

Jayaseelan, K. A. (2017). Anaphora in Dravidian. *Oxford Research Encyclopedia of Linguistics*.

Jayaseelan, K. A. (2004). The possessor-experiencer dative in Malayalam. In P. Bhaskararao & K. Subbarao (Eds.), Non-nominative subjects, volume 1 (pp. 227–244).

Morphological priming of Bangla prefixed and pseudo-prefixed words: a sentence-prime lexical decision study

Basita Biswas and Indranil Dutta
Jadavpur University

Previous research using the primed lexical decision paradigm shows that complex words are morphologically decomposed in the early stages of processing—when primed by affixed words, affixed targets are facilitated, whereas pseudo-affixed targets are delayed (Taft & Forster, 1975; Marslen-Wilson et al., 1994; Rastle et al., 2004; Biswas & Mukherjee, 2024). However, most of the work has been done in isolated word paradigms, whereas sentential context has largely been used to determine the effects of semantic relatedness in natural language settings (Colombo & Williams, 1990). Therefore, research on morphological processing in natural sentence contexts remains comparatively limited. In this study, we employ a sentence-priming lexical paradigm to examine how morphological relatedness influences recognition of target words containing prefixation or pseudo-prefixation, going beyond isolated lexical activation.

Methodology: In this experiment, each prime sentence has an embedded critical word, either prefixed or pseudo-prefixed. Targets are words—prefixed, pseudo-prefixed, or non-words. All non-words are phonotactically legal in Bangla. Sentence length is controlled.

The experimental conditions are illustrated in Table 1 (word conditions) and Table 2 (non-word conditions).

To ensure that semantic relatedness is not playing a role in priming, the prime–target word pairs do not have semantically related stems and are designed to test for morphological relatedness only through the prefix. Word stems are semantically transparent free morphemes and carry a clear meaning in the whole word. No phonemic change is caused to the stems as a result of prefixation through vowel harmony.

Sixty percent filler conditions were used (consisting of both words and non-words) to minimise practice effects.

Both the primes and the targets were presented visually using OpenSesame, after which the participant was instructed to decide whether the target was a valid Bangla word or a non-word by pressing the corresponding keys on a laptop.

Forty literate native Bangla speakers participated in the study.

Discussion: If Bangla prefixed words are represented through morphological decomposition in a natural language comprehension context, targets that share prefixes with the word embedded in the prime sentence should show faster RTs. If, however, words are accessed holistically in context, the RTs should not be higher than the baseline. The experimental design also enables the detection of inhibitory effects, which would suggest lexical competition or context-driven prediction mismatches. The predicted priming result is graded facilitation, with the greatest priming effects in prefixed targets, followed by pseudo-prefixed targets in a prefixed prime condition, and no facilitation for non-words. This result would indicate prefixes to be a functional unit of representation in the mental lexicon of Bangla.

The collected data are undergoing analysis for response time (RT) and accuracy.

Table 1. The prime words are embedded in sentences, such as in condition 1: সবাই সুদিন ফেরার আশায় আছে (“everyone is awaiting the good days”)

Conditions	PRIME (word)	TARGET (word)
1	(prefixed)	(prefixed)

	সুদিন /ʃudin/ ‘good+days’	সুনাম /ʃunam/ ‘good+name’, ‘fame’
2	(prefixed) সুদিন /ʃudin/ ‘good+days’	(pseudo-prefixed) সুজি /ʃudʒi/ ‘semolina’
3	(pseudo-prefixed) সুতো /ʃuto/ ‘thread’	(pseudo-prefixed) সুধা /ʃudʰa/ ‘nectar’
4	(pseudo-prefixed) সুতো /ʃuto/ ‘thread’	(prefixed) সুপথ /ʃupɔtʰ/ ‘good+path’

Table 2. The prime words are embedded in sentences, such as in condition i: ভুমি কুপথ এড়িয়ে চলো (“avoid the bad path”)

Conditions	PRIME (word)	TARGET (non-word)
i	(prefixed) কুপথ /kupɔtʰ/ ‘bad+path’	(same segment) কুজাস /kudʒaʃ/
ii	(prefixed) বিদেশ /bideʃ/ ‘outside+country’	(different segment) আগুপ /agup/
iii	(pseudo-prefixed) কুমার /kumar/ ‘unmarried man’	(same segment) কুটাস /kutaʃ/
iv	(pseudo-prefixed) বিচার /bitʃar/ ‘judgment’	(different segment) লাফুর /lapʰur/

References

- Biswas, B., & Mukherjee, M. (2024). Processing of pseudo-affixes in words and non-words: A study on Bangla prefixes. In *Console xxxii: Proceedings of the 32nd conference of the student organization of linguistics in Europe* (V ol. 32, p. 261-272).
- Colombo, L., & Williams, J. (1990). Effects of word- and sentence-level contexts upon word recognition. *Memory & Cognition*, 18(2), 153–163
- Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3-33.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother’s brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11(6), 1090–1098.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638-647.

Exploring place names through new computational tools: Two case studies in India

Lennart Chevallier & Søren Wichmann

In our talk we will demonstrate the functionalities, scientific utility and limitations of our R package ‘toponym’ (Chevallier & Wichmann 2026). We will do this by discussing distributions of select generic elements in place names and their possible origins in two regions of India for which we are currently collecting and analyzing data. The two regions are the Konkan region of western India, where Indo-Aryan comes into contact with Dravidian, and the northeastern states of Assam and Nagaland, where Indo-Aryan is in contact among others with Tibeto-Burman.

Toponomastics or the study of place names is a well-established branch of humanities which has a history of more than two centuries of practice based on sound, scientific principles and which is of interdisciplinary importance (Eichler 1982: 8). Toponyms “permit historical inferences about languages and the people who spoke them” (Campbell 2013: 436). Sometimes they are unique sources of information about settlement histories, and, given that they are often treated conservatively rather than replaced, they may reveal processes of cultural contacts and population replacements.

In toponomastics, it is useful to distinguish between what we may call particularistic and pattern-finding approaches: The former include a focus on the etymologies of *individual* place names and what they can tell us about histories of ethnic groups. The latter are mainly concerned with the identification of morphemes recurring across numerous place names and what the *distributions* found may reveal. Since around 1990, the personal computer began to play a decisive role in some toponymic studies of the pattern-finding kind (e.g., Campbell 1991), and the availability of GIS software has made such studies more common. Nevertheless, the particularistic approaches are still vastly dominant in the literature on toponyms (cf. Tent 2015), and it is symptomatic that the recent textbook by Cacciafoco and Cavallaro (2023) is almost devoid of distributional maps. We believe, however, that the possibility to process and map large amounts of place names using dedicated software has a great potential to boost toponomastics. For this reason, we have developed the package within the computational environment of R, an environment that enjoys a wide popularity.

The package provides an interface to the data at <https://www.geonames.org/>, which currently contains around 13,5 million entries for toponyms from 254 countries with 9 feature classes (populated places, water bodies, mountains, etc.). This software facilitates the search for specific strings (sequences of letters) in the entire database and produces maps of toponyms filtered in a number of different possible ways.

The case studies targeting different regions of India serve not only to illustrate the utility of our package, but also pose challenges in terms of how to improve on the package so that it becomes better tuned to incorporate new information in addition to that of the single, agglomerative source that it currently accesses. We aim to show how it can be extended so as to allow for the inclusion of data sets in a flexible, reusable format with new kinds of information, such as historical attestations, etymologies, references, etc.

References

- Campbell, Jon C. 1991. Stream generic terms as indicators of historical settlement patterns. *Names* 39(4): 333–366.
- Campbell, Lyle. 2013. *Historical Linguistics: An Introduction*, 3rd edn. Edinburgh: Edinburgh University Press.
- Chevallier, Lennart & Wichmann, Søren. 2026. *toponym: Analyze and Visualize Toponyms*. Version 2.0.1. <https://cran.r-project.org/package=toponym>
- Eichler, Ernst. 1982. *Ergebnisse der Namenforschung im deutsch-slawischen Berührungsgebiet*. Berlin: Akademie-Verlag.
- Tent, Jan. 2015. Approaches to research in toponymy. *Names* 63(2): 65–74.
- Perono Cacciafoco, Francesco and Francesco Cavallaro. 2023. *Place Names: Approaches and Perspectives in Toponymy and Toponomastics*. Cambridge: Cambridge University Press.

Ingressive ↔ Progressive: Bidirectional Shifts in Indo-Aryan Grammaticalisation

Chinmay Dharurkar,

Assistant Professor, Dept of Humanities and Social Sciences, IIT Kanpur

chinmayd@iitk.ac.in, chinmayshubhada@gmail.com

This paper explores a grammatical aspectual shift (if not a clear case of a cycle) observed in Indo-Aryan languages (e.g., Marathi, Hindi, Punjabi, Haryanvi) and some Dravidian languages (e.g., Kannada), where the ingressive aspect evolves into a progressive aspect, and in certain languages (e.g., Gujarati, Sanskrit), progressive participles assume an inchoative sense. This pattern—*inchoative* → *progressive and back*—captures the interplay between aspectual meanings and their diachronic grammaticalisation pathways.

In Standard Marathi verb like *lagṇe* (to attach) and in a few varieties of Kannada *hattu* (to attach)¹ mark the ingressive aspect. However, in their contact varieties (Southern Marathi, Northern Kannada), these constructions acquire a progressive sense. (Briefly noted in Dharurkar, 2020).

Standard Marathi

Kaj	kər-u	lag-l-a	ahe-s?
What	do-INF	ATTACH-PRF-SG.M	be.PRS-2SG
What have you started doing?			

(This aspect in second person interrogative sounds iffy but is given here as a morphosyntactic parallel to the progressive version in the southern variety)

Southern Marathi

ka	kər	la-l-a-s
What	do-INF	ATTACH.PRF-SG.M-2SG
What are you doing?		

Standard Kannada

e:n	maḍa-toḍaḡ-iddi?
What	do-BEGIN-be.PRS-2SG
What have you started doing?	

Northern Kannada

e:n	ma:ḍa-li-k-(h)atti-ri?
What	Do-INF-DAT-ATTACH.pres-2PL
What are you doing?	

Parallel developments are found in other Indic languages, such as Haryanvi (and several other Pahadi languages), in addition to the ingressive use of *lagṇa*, it also combines with *reṇa* (to stay) and *hoṇa* (to be) to form progressive constructions.

vo	ga:ṇa:	ga:ṇ	la:g	rəja:	sɛ
he	song	sing.INF	ATTACH	stay.SG.M	be.PRES.SG
He is singing.					

The ingressive construction in Haryanvi is simpler:

vo	ga:ṇa:	ga:ṇ	la:g-j-a:
he	song	sing.INF	ATTACH-PAST-3.SG.M
He started singing.			

Conversely, in Gujarati (Cardona 1965) and Classical Sanskrit, present participles—typically progressive—convey ingressive meanings (e.g., Sanskrit *śyāmāyamānāni* ‘becoming dark’).

¹*hattu* (to attach) is dialectal and colloquial in several varieties of Kannada. In majority of the varieties wherein it is used, [h] is dropped. In Standard Kannada, the ingressive is by *toḍaḡu* (to engage), see Sridhar (2007: 234).

Gujarati [based on Cardona (1965: 108), glossing and translation mine]

brahməṇo	həve	indəṇ	khata	thəja	tʰe
Brahmins	now	eggs	eat-PRS.PTCP	leave-PRS.PTCP	be.PRS.3
Brahmins have now started eating eggs.					

Classical Sanskrit seems to convey the ingressive sense using its (indicative or progressive) participles. For instance, a verse from Raghuvamśa of Kālidāsa in Classical Sanskrit uses present participle to convey the ingressive sense.

śja:ma:ja-ma:na:ni	vāna:ni	pəcjan
darken-PRS.MID.PTCP.N.NOM.PL	forest.N.NOM.PL	see-PRS.ACT.PTCP-M.NOM.SG
(He) seeing forests that were becoming dark.		
From Raghuvamśa of Kalidāsa [See Kale (1922: 32)] ²		

I also discuss instances from Classical and Vedic Sanskrit where the ingressive interpretation is achieved by using participial forms and aorist respectively. [For aorist see Dahl (2011: 276)].

The paper argues that this bidirectional shift—where markers of ingressive aspect evolve into progressive and, in some cases, revert to ingressive—is driven by semantic and functional reanalysis, facilitated by the repurposing of aspectual markers. Drawing on Heine and Kuteva's (2005) framework, the study highlights the role of areal contact and typological convergence in shaping these pathways, while acknowledging that the process may not constitute a true grammaticalisation cycle in the strictest sense (Narrog & Heine 2021). The paper reflects upon the openness of the participles (as in the date, from Sanskrit and Gujarati), to be interpreted as ingressive, and of course upon the potential proximity of the inchoative and ingressive interpretations, in general. The findings contribute to broader discussions on recurrent grammaticalisation patterns (Bybee et al. 1994) and the dynamics of aspectual semantics in South Asian languages.

References

- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Cardona, G. (1965) A Gujarati Reference Grammar. Philadelphia: The University of Pennsylvania Press.
- Dahl, E. (2011) 'Tense and Aspect in Indo-Iranian Part 1: The Present and Aorist'. *Language and Linguistics Compass* 5/5 (2011): 265–281.
- Dharurkar, Chinmay. (2022) The Uncanny Sisterhood of Deccan's Languages'. Wire. 30 July 2022. <https://thewire.in/society/the-uncanny-sisterhood-of-deccans-languages>
- Heine, B; Kutewa, T. (2005) *Language Contact and Grammatical Change*. Cambridge: CUP.
- Kale, M. R. (1922) *The Raghuvamśa of Kālidāsa*. Bombay: Gopal Narayen & Co.
- Sridhar, S. N. (2007) *Modern Kannada Grammar*. New Delhi: Manohar.

² The full verse in IAST:

sa palvalottirṇavarāhayūthānyāvāsavr̥kṣonmukhabarhiṇāni.

yayau mṛgādhyāsitaśādvalāni śyāmāyamānāni vanāni paśyan. 2.17

Translation: He went on, beholding forests that were turning dark, with grassy meadows frequented by deer, where flocks of wild boars had emerged from marshes, and peacocks stood facing their roosting trees.

Typological Universals and Variation in Objective Case Marking: Evidence from Eastern India
Shuvam Dutta

Indian Institute of Technology Kharagpur, India

The study investigates the typology of objective case marking across four genetically diverse languages of Eastern India: Bangla (Indo-Aryan), Dhimal (Tibeto-Burman), Kurux (Dravidian), and Santali (Austroasiatic), and situates their patterns within cross-linguistic typological universals. The study demonstrates how these languages collectively illustrate the interplay between universal principles of case alignment and areal convergence in contact situations. Except for Bangla, the other three languages are less studied. Therefore, this study primarily relies on primary data while also incorporating available secondary sources.

A typological universal observed cross-linguistically is that case marking serves to distinguish core arguments, particularly the agent (A) from the patient (P) in transitive clauses (Comrie 1989; Croft 1991). Another widely attested pattern is that case marking correlates with semantic prominence, specifically animacy, definiteness, and affectedness (Aissen 2003). These universals are strongly reflected in the four languages examined here, but with distinct morphosyntactic realizations.

Bangla and Dhimal exhibit syncretic objective systems, in which a single morpheme marks both accusative and dative roles (*/-ke/* and */-eŋ/* respectively). This reflects the universal tendency for dative extensions to cover accusative functions, a phenomenon of case syncretism also reported in other Indo-Aryan languages. Both languages display Differential Object Marking (DOM), governed by animacy and definiteness hierarchies (human > animate > inanimate), aligning with the typological prediction that higher-prominence objects are more likely to be overtly marked (See example 1).

Kurux, in contrast, retains distinct morphological markers for accusative and dative cases (*/-n/* or */-in/* vs. */ge/*), reflecting the structural case alignment typical of Dravidian languages. Yet its restriction of accusative marking in ditransitives parallels DOM-like behavior in neighboring Indo-Aryan systems, evidencing contact-induced typological convergence (See example 2).

Santali, a Munda language belonging to the Austroasiatic language family, diverges entirely from the pattern of dependent marking. Instead of overt case suffixes, object roles are encoded by head marking through verbal agreement, exemplifying the universal trade-off between flagging and indexing (Haspelmath 2005). Even so, animacy continues to govern object agreement, revealing how universal prominence hierarchies manifest across distinct morphosyntactic architectures (See example 3).

The comparative analysis highlights several typological universals: (1) **Semantic alignment universals**: case marking reflects animacy and affectedness; (2) **Structural universals**: languages employ either dependent or head marking to distinguish grammatical relations; (3) **Diachronic universals**—accusative-dative syncretism and extension of dative marking to experiencer and obligation constructions follow well-attested grammaticalization paths.

The case-marking systems of these languages reveal that, while Eastern Indian languages belong to distinct genealogical families, they have converged toward a shared areal typology. This convergence exemplifies a broader universal: languages in close contact tend to align morphosyntactic strategies to maximize functional transparency. Thus, this study contributes to the typology of case systems by demonstrating how universal principles of argument marking interact with regional contact dynamics, resulting in diverse yet predictable patterns of objective case marking across languages.

References

- Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21, 435-483.
- Comrie, B. (1989). *Language Universals and Linguistic Typology*. Oxford Blackwell.
- Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago Press.
- Haspelmath, M. (2005). Argument Marking in Ditransitive Alignment Types. *Linguistic Discovery* 3(1):1-21.

Examples 1

Bangla

- (1) *c^hele-ṭa montri-ke p^hele di-l-o*
boy-CLF minister-ACC/DAT push give-PST-3
The boy pushed the minister.
- (2) *ṣiṇho c^hagol-ṭa- (ke) k^he-l-o*
lion goat-CLF-ACC/DAT eat-PST-3
Lion ate the goat.
- (3) *kukur har-ṭa k^he-l-o*
dog bonet-CLF eat-PST-3
Dog ate the bone.

Dhimal

- (4) *ojaṇ-ṭ^hoi bejaṇ-eṇ sipi-hi*
boy-TOP girl-ACC/DAT push-PST
The boy pushed the girl.
- (5) *k^hōa mec^ha-eṇ sa-hi*
lion goat-ACC/DAT eat-PST
Lion ate the goat.
- (6) *k^hia mai sa-hi*
dog bone eat-PST
Dog ate the bone

Examples 2

Kurux

- (7) *jōq^həs kukoi ge pũp ciccəs*
boy girl DAT flower gave
The boy gave a flower to the girl.

Examples 3

Santali

- (8) *kora gidrə-do kuri gidrə q^haka-ked-e-ay*
boy-TOP girl push-PST--3SG.SUBJ-3SG.OBJ
The boy pushed the girl.
- (9) *sita-do jaṇ-e-ṣom-ked-a*
dog-TOP bone-3SG.SUBJ-eat-PST-FIN
Dog ate the bone.

When Script changes but Language remains : A computational study of Romanized Indic Texts

Chetana Ghosh

School of Languages and Linguistics, Jadavpur University

Abstract

With the rise in use of digital communication systems, non-standard orthographical practices have become the norm with the most common being, Romanized forms of native scripts. The variation in script is becoming a form of linguistic variation. These transliterated versions are user-driven as there are no standard norms for Romanization. This study examines Romanization as a computational and linguistic problem by analyzing how script variation affects subword tokenization. Previous research investigated the performance of twelve tokenizers on the twenty-two official languages of India. Building on this work, I investigate the performance of three tokenizers - GPT-4o (an English centric model), mBERT (a multilingual model) and Google MuRIL (a model trained specifically on Indic languages) on the native and transliterated versions of two Indian languages - Bengali and Hindi. 200 data points containing native sentences and their transliterated counterparts were randomly selected from the Bhasha Abhijanam dataset and they were tokenized using the above mentioned tokenizers. I quantified the performances of these tokenizers on the basis of two metrics:

- a) Mean Script Sensitivity Ratio (Mean SSR), defined as the ratio between the number of roman and native tokens within the same tokenizer,
- b) Normalized Sequence Length (NSL), defined as the length of an encoded sequence from a tokenizer in comparison to a baseline tokenizer.

The results reveal clear, model-dependent biases in how scripts are tokenized. mBERT shows relatively balanced behavior across native and Romanized inputs, whereas MuRIL fragments Romanized text more heavily, reflecting its Indic-focused pretraining. In contrast, GPT exhibits the opposite trend, disproportionately fragmenting native scripts (e.g., $SSR \approx 0.31$ for Bengali). Linguistically, these patterns suggest that tokenizers fail to treat transliterated forms as coherent lexical units, instead encoding script as a proxy for structure. NSL values further capture differences in tokenization efficiency, with disparities between Romanized and native forms indicating varying orthographic processing costs across models. As illustrated in Figures 1.1 and 1.2, Bengali is consistently fragmented into more subwords than Hindi, pointing to its relative under-representation in multilingual pretraining environments. Beyond identifying effective tokenization strategies for Romanized Indian text, this study proposes a quantitative framework for analyzing orthographic variation within a language. More broadly, the findings raise fundamental questions about how subword tokenization methods such as WordPiece and Byte Pair Encoding (BPE) define lexical units, and whether neural language models encode script as an aspect of linguistic identity. Future work will extend this analysis to downstream NLP tasks involving Romanized scripts. encode script as an aspect of linguistic identity. Future work should extend this analysis to downstream NLP tasks involving Romanized scripts.

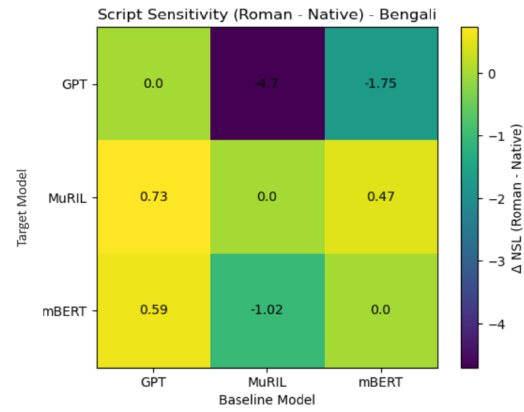


Fig. 1.1 : Difference in NSL values ($NSL_{\text{Roman}} - NSL_{\text{Native}}$) for Bengali. The NSL values were calculated using rotating baselines.

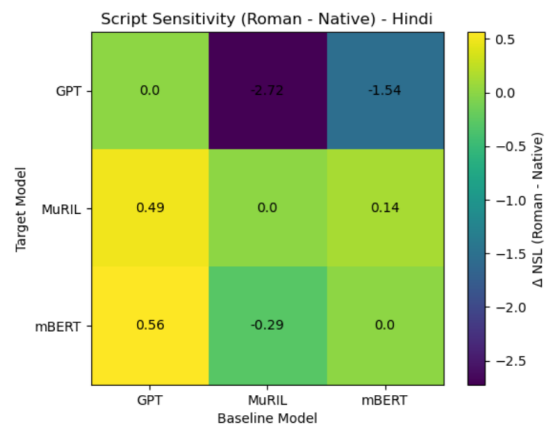


Fig. 1.2 : Difference in NSL values ($NSL_{\text{Roman}} - NSL_{\text{Native}}$) for Hindi. The NSL values were calculated using rotating baselines.

Keywords : orthographic variation, tokenizer performance, Romanized Indian text

Telugu or Thelagoo? Variability in Civilian Romanization Among Telugu Natives

Vikram Goudar

Background: Telugu, the Dravidian language of Andhra Pradesh and Telangana, is written with an abugida of 36 consonants and 16 vowels, but romanization is becoming increasingly popular (Sowmya, Varma, 2009). One romanization scheme with one-to-one correspondence is the NLKR (Nat'l Library at Kolkata Romanization), which makes heavy use of diacritics. However, the general population often omits diacritics (Demirsahin et al., 2022). Instead, they use what can be called ‘civilian romanizations,’ accepting ambiguity in exchange for speed and ease.

The ambiguity of these romanizations can be attributed to a few key disparities. Telugu employs aspiration, retroflex-dental distinctions, vowel length, and consonant sandhi, while English does not. Furthermore, in Telugu, aspiration displays inconsistently: take unaspirated-aspirated pairs పప, చచ, and దద, but జరు, తద, and టర. Such visual discrepancies often carry over in transcription. Because of English’s orthography, only about 48% of Telugu characters have relatively unique English graphemes.

Particularly of note in this study are Telugu’s eight retroflex and dental plosives, which can be arranged on a cube. This is similar to Turkish’s eight-vowel system, arranged on the vertices of a cube with each of the 3 axes representing a feature (frontedness, openness, and roundedness) (Pierce, 2016). Telugu’s plosive cube simply has different axes: voicing, retroflex-dental, and aspiration.

Methods: To test variability within civilian romanizations, a novel and self-designed method was used: ē-PADAM (most Efficient Phonetic Arrangements Designed Across Matrices), which revolves around testing phonological (in this case, transcriptional) features in as few survey questions as possible. Guidelines for this system include: a) using non-real words to avoid preexisting romanizations, b) keeping each word about 4 syllables to mimic Telugu, c) mirroring phonetic structure across words to avoid confounding variables, and d) assigning each consonant to a unique vowel to prioritize efficiency. The result was 26 different letters tested across only 5 words (see Table 1). A questionnaire was sent to 30 native speakers and their responses were compiled.

To analyze the results, frequencies of English graphemes for each Telugu grapheme were tabulated. An index of coincidence metric (IndCo) was calculated for each Telugu grapheme, measuring the degree of consensus among the respondents (IndCo is the probability that any two random respondents have the same response; it is shown in Table 2 out of 870, with 870 indicating unanimity). Additionally, for each of the three axes of the plosive cube (voicing, retroflex-dental, and aspiration), a calculation was done to indicate the consistency of that feature’s marking (e.g. for voicing, does the change in romanization from /t/ to /d/ exactly mirror that of /tʰ/ to /dʰ/, /t/ to /d/, and /tʰ/ to /dʰ/?). This corresponds to traversing an edge of the cube along a certain axis.

Results and Conclusions: Tables 1 and 2 show the words tested, transcriptions, and IndCo values.

10 of the 30 respondents used an “i” or “y” alone to represent /ai/, 15 used “ow” for /au/, 19 used “oo” for /u:/ and 9 of 30 used “f” or “fh” for /pʰ/, indicating English phonological influence on Telugu romanization, as these grapheme-phoneme pairs are fairly unique to English.

For the plosive cube, the average respondent displayed 38% consistency when aspirating an unaspirated plosive, 44% when dentalizing a retroflex plosive, and 75% consistency when voicing an unvoiced plosive. This ambiguity further resembles English phonology; English can consistently mark voicing (t → d), approximate retroflex-dental distinctions (t → th), but not cleanly mark aspiration.¹

All 30 used some form of “t” to mark unvoiced plosives and “d” to mark voiced plosives, with one key exception: 13 of the 30 used d, dh, or dhh to transcribe ఢ, the *unvoiced* dental plosive (tha in NLKR). English phonology alone cannot explain this. However, consider that ఢ was most commonly “da” and ఢ was “dha.” ఢ (/tʰ/) romanizing as a “d” indicates potential visual bias because ఢ looks more similar to voiced plosives (దదడద) than to its unvoiced counterparts (తటర). In other words, Telugu letters which look similar may have similar romanizations, absent a phonological basis. Another example of this visual bias can be found in Table 3.

Future Directions: Following this will be the testing of if other Telugu consonants romanize differently depending on the surrounding vowels. This study’s ē-PADAMs lend well to comparative analysis with Malayalam and Kannada. Findings could be used to design romanization systems that balance consistency and clarity with simplicity.

¹ Although some opted to use “h” to indicate aspiration, this was often accompanied by a doubling of the consonant or the use of a different consonant altogether, resulting in less overall consistency.

Original Telugu	NLKR	IPA	Composite Civilian Romanization
తోడంగః	tōḍhaṅgaḥ	/t̪oːd̪aŋgaḥa/	thodhangaha
ధేడచైక్ష	thēḍacaikṣa	/t̪h̥eːd̪at̪ʃaikaṣa/	thedachaiksha
ధీటచౌ	dhīṭachau	/d̪h̥iːt̪at̪ʃau/	dheetachow
దూఠచ్చం	dūṭhaccam	/d̪uːt̪at̪ʃam/	dootacham
ఫాక్రుచ్చశ	phākṛcchaśa	/p̪h̥aːkrut̪ʃaśa/	phakruchasa

Table 1: The set of five words in Telugu, NLKR, IPA, and Composite Civilian Romanization (the combination of the most common representation of each phoneme among the 30 respondents).

T	N	I	T	N	I	T	N	I	T	N	I
డ	ḍ	870	ఛ	ch	572	ఌ	ā	438	ౌ	au	268
ఱ	ṛ	812	ఢ	dh	558	ఔ	ū	432	ఢ	ḍh	266
ట	ṭ	812	చ	c	554	శ	ś	416	ఠ	ṭh	264
ం	m	812	త	t	518	ః	aḥ	394	చ్చ	cc	222
క్ష	kṣ	758	ి	ī	504	ై	ai	384	చ్చ	cch	188
్	ṛ	602	ంగ	ṅg	476	థ	th	320			
ీ	ē	600	ద	d	438	ఫ	ph	272			

Table 2: The 26 consonants and vowels ‘tested’ in their original Telugu (Column T) and NLKR (Column N), as well as their indices of coincidence (Column I, out of a maximum possible score of 870, which indicates perfect consensus).

Voiced	Unasp.		Asp.	
Dent.	ద	d	ఢ	dh
Retr.	డ	ḍ	ఢ	dh
Unv.	Unasp.		Asp.	
Dent.	త	th	ఢ	th
Retr	ట	t	ఠ	t

Table 3: Telugu’s 8 plosives and their most common representations in the survey. The 4 Telugu consonants on top mark aspiration visually, and their romanizations do as well. However, the 4 on the bottom do not mark aspiration in a consistent way, and neither do their romanizations (which omit markings of aspiration, instead using “h” to mark retroflex-dental distinctions), indicating visual bias.

Sowmya, V. B., & Varma, V. (2009). Transliteration based text input methods for Telugu. In W. Li & D. Mollá-Aliod (Eds.), *Computer processing of oriental languages: Language technology for the knowledge-based economy (ICCPOL 2009)* (Lecture Notes in Computer Science, Vol. 5459). Springer.
https://doi.org/10.1007/978-3-642-00831-3_12

Demirsahin, I., Johny, C., Gutkin, A., & Roark, B. (2022). Criteria for useful automatic romanization in South Asian languages. In N. Calzolari et al. (Eds.), *Proceedings of the thirteenth language resources and evaluation conference (LREC 2022)* (pp. 6662–6673). European Language Resources Association.

Pierce, D. (2016). Turkish vowels (David Pierce) (David Pierce, Matematik, M.S.G.S.Ü.). Msgsu.edu.tr; Mimar Sinan Güzel Sanatlar Üniversitesi.
<http://mat.msgsu.edu.tr/~dpierce/Language/Turkish/turkish-vowels.shtml>

Event integration in Hindi multi-verb predicates

Sakura Ishikawa (sakura.ishikawa@tufs.ac.jp), Tokyo University of Foreign Studies

Keywords: associated motion, directional, co-expression, complex predicate, iconicity

Introduction Associated motion (AM) and deictic directionals (DIR) have been considered distinct grammatical categories. However, they are often co-expressed by the same means crosslinguistically (Belkadi 2015, 2021; Dryer 2021). Belkadi (2021) argues that they constitute a single category with a difference in the degree of event integration: AM is lower while DIR is higher. This difference reflects a verbal semantics hierarchy, as given in Figure 1, which shows the preference for an AM or DIR reading when verbs occur in formally identical constructions (Belkadi 2016: 64).

In Hindi, Subsequent AM (DO&COME/GO type of AM) and DIR are co-expressed through two strategies of multi-verb predicates: a linker strategy (using a converbal suffix) and a non-linker strategy (using a bare verb stem), as shown in (1)–(3). This raises a question: Does the choice between these strategies reflect the degree of event integration in AM and DIR events?

Aim of this paper This study analyzes the morphosyntactic properties of the two strategies of AM and DIR Constructions to examine the iconic relationship between form and function. It is hypothesized that the linker strategy represents lower event integration, while the non-linker strategy represents higher. To examine this, the present study focuses on (i) the degree of coherence, measured by occurrences of intervening elements between the two verbs, (ii) the lexical preference of V1, based on Belkadi's (2016) verbal semantics hierarchy.

Findings The results show differences between the two strategies, as summarized in Table 1: (i) with a negation marker, the linker strategy maintains adjacency of verbs in 10% of cases, while the non-linker strategy exhibits stricter adjacency, maintaining it in 90%, (ii) 90% of path verbs occur in the non-linker strategy, while approx. 68% of non-motion action verbs and stative verbs occur in the linker strategy. Almost equal numbers of manner, causative motion, and perception verbs occur in both strategies. This indicates that the linker strategy tends to occur with the verbs preferring an AM reading in Belkadi's hierarchy, while the non-linker strategy tends to occur with the verbs preferring a DIR reading.

Discussions This study shows that the strategy choice reflects event integration: the linker strategy encodes a lower degree (functioning primarily as AM), while the non-linker strategy encodes a higher degree (functioning primarily as DIR). While previous accounts attributed this strategic variation to stylistics (Montaut 2004: 93), I argue that form-function iconicity plays a decisive role. This study provides morphosyntactic evidence that AM and DIR are distinguished by their degree of event integration, reflecting the presence/absence of a linker.

- (1) AM:
ye **pahan(-kar)** **jā-o**
these wear-CVB go
'Put on them and go.'
(based on blogspot.com)
- (2) DIR (path verb):
ek laṛkā bāhar **nikal(-kar)** **ā-y-ā**
one boy outside exit-CVB come-PRF-M.SG
'A boy came outside.'
(based on neerajmusafir.com)
- (3) DIR (manner verb):
vo **douṛ(-kar)** **ā-y-ā**
3SG.NOM run-CVB come-PRF-M.SG
'He came running.'
(blogspot.com)

Figure 1 Verbal semantics hierarchy (adopted from Belkadi 2016: 64)

Strategy	Cohesion	Lexical preferences		
		Path	Manner,Causative,Perception	Action,State
Linker Strategy	10%	10%	Approx. 53%	Approx. 68%
Non-linker Strategy	90%	90%	Approx. 47%	Approx. 32%

- Belkadi, Aïcha. 2015. Associated motion with deictic directionals: A comparative overview. *SOAS Working Papers in Linguistics* 17. 49–76.
- Belkadi, Aïcha. 2016. Associated Motion Constructions in African Languages. *Africana Linguistica* 22. 43–70.
- Belkadi, Aïcha. 2021. Deictic directionality as associated motion: motion, complex events and event integration in African languages. In Antoine Guillaume & Harold Koch (eds.), *Associated Motion*, 163–200. Berlin: De Gruyter Mouton.
- Dryer, Matthew S. 2021. Associated motion and directionals: where they overlap. In Antoine Guillaume & Harold Koch (eds.), *Associated Motion*, 129–162. Berlin: De Gruyter Mouton.
- Montaut, Annie. 2004. *A grammar of Hindi*. Muenchen: LINCOM Europa.

Language shift in Jharkhand: A study of linguistic identities and ideologies

This paper investigates the sociolinguistic situation of indigenous communities who live in proximity with one another in central and western Jharkhand. Jharkhand, a central-eastern state in India, is home to a mixture of populations from different indigenous and non-indigenous communities. Members of the indigenous communities use the term Adivasi for self-identification. A little under 30% of the state's population is Adivasi. These communities are economically poor, have low literacy rates, and belong to the bottom of the social hierarchies present in the country.

The Oraon, Kharia, Munda(ri), and Ho indigenous communities, traditionally speaking Kurukh, Kharia, Munda(ri), and Ho, respectively, were chosen for this study. The paper explores the language-use pattern of all four communities. This pattern is used for tracing potential intergenerational language shifts. Language-use preference stemming from various reasons, both personal and practical, contributes to forming individuals' and groups' language identities and ideologies.

Previous studies on Jharkhand have explored language-use patterns that are domain-specific (Mohan 2006-2007) or community-specific (Al-Huri & Ranjan 2016). Moreover, situations of language contact and convergence (Peterson 2010; Paudyal & Peterson 2021) have been studied in addition to the compilation of grammars of several languages such as the lingua franca (Peterson & Baraik 2022), Khortha (Paudyal 2025), etc. These studies neither focused on intergenerational language change across multiple communities nor examined major indigenous communities from the western and southern parts of Jharkhand.

The data was collected through semi-structured interviews. Intergenerational language shift was analyzed across three generations, using content analysis after a categorization of the data according to communities and relevant sociocultural factors such as education, residence region, contact with urban and rural regions, religion, proficiency in one's traditional language, etc.

The results show that the indigenous communities in the central and western parts of Jharkhand are currently undergoing a language shift, with the younger generation (Gen III) moving towards the regionally dominant languages like the lingua franca and Hindi, while the older generation (Gen I) continue using their traditional languages. The middle generation (Gen II) presented a balanced use of the traditional languages as well as the lingua franca. This shift, however, is not uniform among the four indigenous communities. While the Oraon and Munda(ri) communities are well underway this shift, the Kharia community, though, on the same path shows slow progress. In contrast, the Ho community, does not show any signs of such a shift. This paper presents the varying results of the four indigenous communities, discussing the reasons behind their language-use pattern and highlighting the linguistic identities and ideologies attached to these languages.

References:

1. Al-Huri, Ibrahim & Ranjan, Mayank. 2016. The consequences of language contact on the sociolinguistic situation of Kurukh in Jharkhand: A case study of Mandar Region. *English Literature and Language Review* 2. 121-130.
2. Mohan, Shailendra. 2006-2007. Patterns of language use among the tribal communities of Jharkhand. *Bulletin of the Deccan College Post-graduate and Research Institute* 66/67. 335-347.
3. Paudyal, Netra & Peterson, John. 2021. How one language became four: the impact of different contact-scenarios between “Sadani” and the tribal languages of Jharkhand. *Journal of South Asian Languages and Linguistics* 7. 327-358.
4. Paudyal, Netra. A grammar of Khortha. Leiden: Koninklijke Brill.
5. Peterson, John & Baraik, Sunil. 2022. A grammar of Chotanagpuri Sadri: An Indo-Aryan lingua franca of eastern central India. Mysuru: Central Institute of Indian Languages.
6. Peterson, John. 2010. Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9. 56-86.

On the clause-finality of Marathi polar question particles

Aaditya Kulkarni

All India Institute of Speech & Hearing

Claim: In this paper, I analyze polar questions in Marathi to show that despite their surface complexity, Marathi polar questions are strictly a clause-edge phenomenon as apparent clause-medial instances of polar question particles do not lead to polar questions.

Background: Polar questions in Marathi are characterized by a final rising intonation and an optional polar question particle (PQP) – either *kā* or *kāy*. While the PQPs canonically appear clause-finally, they also seem to appear at various positions within the clause (see 1).

(1) Apparent distribution of Marathi PQPs in a clause

(a) (* <i>kāy</i>)	<i>rām-ne</i>	(<i>kāy</i>)	<i>kāl</i>	(<i>kāy</i>)	<i>āmbā</i>	(* <i>kāy</i>)	<i>khāllā</i>	(<i>kāy</i>)
(b) (* <i>kā</i>)	<i>rām-ne</i>	(<i>kā</i>)	<i>kāl</i>	(<i>kā</i>)	<i>āmbā</i>	(<i>kā</i>)	<i>khāllā</i>	(<i>kā</i>)
PQP	Ram-ERG	PQP	yesterday	PQP	mango	PQP	ate	PQP
‘Did Ram eat a/the mango yesterday?’								

The apparent distribution of Marathi PQPs thus parallels that of its better known counterparts in languages such as Bangla, Hindi-Urdu, and Odia. Such clause-medial instances of PQPs are often associated with focus effects in these languages. The optionality of PQPs, along with their varied semantic properties resulting from their syntactic positions has also led to a set of relatively complex (and contrasting) accounts of polar questions in these languages (see Syed & Dash 2017, Bhatt & Dayal 2020, Biezma et. al. 2025, a.o.). This raises the question of whether a similar explanation is necessary to account for Marathi polar questions and the distribution of Marathi PQPs.

Analysis: Following a closer inspection of semantic-pragmatic properties of the apparent clause-medial PQPs in Marathi, I show that the apparent clause-medial instances of polar questions are not polar questions at all, and that Marathi polar questions are strictly a clause-edge phenomenon; and thus, a much ‘simplified’ version of the analyses of Indo-Aryan polar questions might be needed to account for them.

The evidence in support of this for Marathi PQP *kāy* comes from Deo’s (2023) analysis of hybrid interrogatives in Hindi-Urdu. Under this analysis, the apparent clause-medial instances of Hindi-Urdu PQP *kyā* are instances of an underspecified wh-expression – namely KYĀ – which acts as a placeholder for other semantically richer wh-expressions (see also Munaro & Obenauer 1999 for more on underspecified wh-expressions). In addition, a polar question particle can (co-)occur with this underspecified wh-expression. Thus, Deo analyzes the apparent clause-medial occurrences of PQPs as instances of ‘hybrid interrogatives’, where the speaker asks a backgrounded wh-question and a foregrounded polar question in a single discourse move. This explanation straightforwardly extends to clause-medial *kāy* in Marathi: the apparent clause-medial instance of *kāy* from (2a) can be replaced with a relevant wh-expression in (2b) without any change in meaning. An optional polar question particle (either of *kā* and *kāy*) is felicitous in both instances.

(2) Hybrid interrogatives in Marathi

(a) <i>rām-ne</i>	<i>kāy</i>	<i>kāl</i>	<i>āmbā</i>	<i>khāllā</i>	(<i>kā</i> / <i>kāy</i>)
Ram-ERG	KĀY	yesterday	mango	ate	PQP
(b) <i>rām-ne</i>	<i>kadhi</i>	<i>kāl</i>	<i>āmbā</i>	<i>khāllā</i>	(<i>kā</i> / <i>kāy</i>)
Ram-ERG	when	yesterday	mango	ate	PQP
‘When did Ram eat a/the mango? Yesterday?’					

On the clause-finality of Marathi polar question particles

As to PQP *kā*, I argue that clause-medial instances of *kā* are not questions in a pragmatic sense, even though their syntax/semantics might resemble that of questions. Following Littell et. al.'s (2010) three-way typology of question-types based on expectations of speaker/addressee knowledge of the answer, an interrogative expression qualifies as an ordinary question (wh or polar) when the speaker does not know the answer to the question but expects that the addressee does. However, examining the felicity conditions of the clause-medial occurrences of *kā*, it becomes clear that they do not lead to ordinary (i.e. information seeking) questions but rather serve to acknowledge or express surprise at some newfound information (see 3).

- (3) Context: Sita had her birthday yesterday. Among those who attended the party was A who knows that Sita received many gifts including a book, from many people including Ram. However, A does not know what Ram's gift to Sita was. The next day, A makes an inquiry to B about what Sita is doing.

A: What is Sita up to?

B: She is reading the book Ram gave her yesterday.

A: *acchā rām-ne kā sitā-lā pustak dila*
Right Ram-ERG KĀ Sita-DAT book gave

'Right, it was Ram who gave Sita a/the book!'

(Felicitous only when speaker knew that Sita had been gifted a book yesterday and has just learnt or inferred that it was Ram who gifted it)

Outlook: Arguments based on (2) and (3) affirm that the apparent clause-medial instances of PQPs *kā* and *kāy* do not lead to polar questions, and that Marathi polar questions are strictly a clause-edge phenomenon. This necessitates that the account of Marathi polar questions to be much 'simpler' than the ones proposed for other Indo-Aryan languages for it need not account for the clause-medial position of PQPs or the focus effects that come therewith. Secondly, it underlines the need to revisit the semantic-pragmatic properties of polar questions containing clause-medial PQPs in Bangla, Odia, and other languages in the light of these findings.

References:

- Biezma, M., M. Butt, F. Jabeen, & B. Mumtaz. 2025. Urdu/Hindi polar kya as an expression of uncertainty. *Natural Language & Linguistic Theory*, 43:1417–1468.
- Bhatt, R. & V. Dayal. 2020. Polar question particles: Hindi-Urdu kya:. *Natural Language & Linguistic Theory*, 38:1115–1144.
- Deo, A. 2023. Invoking a wh-question by asking a polar question. Paper presented at the 13th (formal) Approaches to South Asian Languages, University of Michigan.
- Littell, P., L. Matthewson, & T. Peterson. 2010. On the Semantics of Conjectural Questions. In *Evidence from Evidentials*. Tyler Peterson and Uli Sauerland (eds), Vancouver: UBCWPL.
- Metslang, H., K. Habicht, & K. Pajusalu. 2017. Where do polar question markers come from? *Sprachtypologie und Universalienforschung* 70. 489–521.
- Munaro, N. & H. G. Obenauer. 1999. On underspecified wh-elements in pseudo-interrogatives. *University of Venice Working Papers in Linguistics* 9, 181–253.
- Syed, S. & B. Dash. 2017. A unified account of the yes/no particle in Hindi, Bangla and Odia. *Proceedings of GLOW in Asia XI*, 1:201–212.

Sources of polar question particles in some Indian languages

[Aaditya Kulkarni](#) (All India Institute of Speech & Hearing)

[Chinmay Dharurkar](#) (Indian Institute of Technology Kanpur)

In this paper, we explore strategies of polar question marking in Indian languages with a focus on polar question particles to a. look at their genealogical and areal distribution, and b. to account for the possible diachronic pathways that might have led to their emergence.

A cursory look at polar questions in some major Indo-Aryan and Dravidian languages reveals that polar question markers in these languages are often homophonous with thematic question words (see 1 for Hindi-Urdu *kya*), disjunction markers (see 2 for Malayalam *-oo*), or both (see 3 for Marathi *ka*). We show that this is consistent with the documented pathways of emergence of question (Metslang et. al. 2017, Walkden et. al. 2025, a. o.) where polar question markers often result from the grammaticalization of thematic question particles and disjunctors, among other sources.

1 (Hindi)	<i>kya ram-ne k^hana k^haya</i> PQP Ram-ERG dinner ate 'Did Ram have dinner?'	<i>ram-ne kya k^haya</i> Ram-ERG what ate 'What did Ram eat?'
2 ¹ (Malayalam)	<i>Mary wannu-oo</i> Mary came-Q 'Did Mary come?'	<i>John-oo Bill-oo Peter-oo wannu</i> John-DISJ Bill-DISJ Peter-DISJ came 'John or Bill or Peter came.'
3 (Marathi)	<i>ram gela ka</i> Ram went PQP 'Did Ram go?'	<i>ram ka gela</i> Ram why went 'Why did Ram go?'
		<i>ram ka sham gela</i> Ram or Sham went 'Ram or Sham went.'

Further, we grapple with the question of Marathi polar questions, and why two different polar question particles (*ka* and *kay*) with differing properties are attested in the language. By looking at their synchronic properties and their diachronic trajectory, we propose that the two PQPs in the language have likely resulted from two different sources. The same, we argue, applies to Kannada polar questions where the connection between polar question markers and their sources is more transparent: both disjunctive *-o* and wh-question particle *e:nu* 'what' are homophonous with polar question markers (Sridhar 2007).

Lastly, we take a broader perspective to look at polar question marking in the history of Indo-Aryan languages. An interesting insight in connection with this is that while Old Indo-Aryan has a PQP in '*api*', we barely find any reflexes of that in the New Indo-Aryan languages. However, cognates of other Old Indo-Aryan PQPs in OIA such as '*kim*', '*wa*', and '*kim wa*' do seem to make their way in marking polar questions in New Indo-Aryan languages. In the light of this observation, we explore whether there exists any continuity in the polar question marking across different stages of Indo-Aryan languages.

¹ Malayalam examples are from Jayaseelan & Amritavalli (2017, 208-209)

Sources of polar question particles in some Indian languages

References:

- Jayaseelan**, K. A., & R. Amritavalli. 2017. *Dravidian Syntax and Universal Grammar*. Oxford. Pp. 208-221.
- Metslang**, H., K. U. Habicht, & K. Pajusalu, 2017. ‘Where Do Polar Question Markers Come From?’. *STUF- Language Typology and Universals* 70(3). 489–521.
- Sridhar**, S. N. 2007. *Kannada*. Delhi: Manohar publications.
- Walkden**, G., T. Li, & L. Weisshaar. 2025. Sources for question particles. In A. Cournane & G. Thoms (eds.), *Proceedings of the 23rd Diachronic Generative Syntax (DiGS) Conference* (special issue in *Journal of Historical Syntax*) 9(10), 1-39.

Phonological Transitions and Contact Phenomena in the Maithili-Angika Continuum

Abhishek Kumar¹, M.A. Linguistics, Jawaharlal Nehru University,
abhijha2221@gmail.com

Katyayani Singh², PhD Scholar, University of Hyderabad,
Katyayani.vns01@gmail.com

Abstract

Maithili and Angika are Indo-Aryan languages, mainly spoken in Bihar. At the same time, Maithili is officially recognised under the Eighth Schedule of the Indian Constitution, it has a huge speaker base of over 13 million in Bihar (Linguistic Survey of Bihar, 2020) a total of around 20.9 million people worldwide speak Maithili as their mother tongue, while Angika remains less institutionally acknowledged despite a comparable speaker population of around 15 million (Ninan, 2007; Masica, 1993). The contact zone between these languages, spanning from Maithili-dominant districts like Darbhanga to Angika-speaking regions such as Banka, forms a complex linguistic landscape characterised by gradual rather than abrupt phonological transitions.

The transition zone surrounding districts like Samastipur, Begusarai, Khagariya, and Bhagalpur represents a dialect continuum (Chambers & Trudgill, 1998) in which intermediate varieties, such as *Thethi*, emerge through contact. The number of Maithili speakers notably declines in this region, while Angika influence becomes increasingly prominent (Linguistic Survey of Bihar, 2020). This demographic gradient aligns with a corresponding shift in phonological features, including segmental substitutions and suprasegmental markers, indicating processes of phonological assimilation and divergence. The study aims to systematically map and analyse these phonological features both in isolation (within Maithili and Angika) and in contact (within transitional varieties).

Methodologically, the study draws from dialect geography, contact linguistics and socio-phonetics to trace the modification of specific phonemes and prosodic structures across the continuum (Labov, 1994; Hock & Joseph, 2009). The elicitation wordlist consisted of high-frequency lexical items and morphologically comparable verb forms, selected to capture potential variation in aspiration, vowel quality, nasalisation, and future-tense morphology across the Maithili-Angika continuum. The list was pilot tested with native speakers from selected districts and refined through expert consultation to ensure regional familiarity and phonological relevance.

Gloss	Darbhangā	Samastipur	Begusarai	Banka
Will do	kərbəh	kərbho	kərb ^h o	kərb ^h ̃
Will go	jəybəh	jəybho	jəyb ^h o	jəyb ^h ̃
Will eat	k ^h əybəh	k ^h əybho	k ^h əyb ^h o	k ^h əyb ^h ̃
Will beat	marbəs	marbho	marb ^h o	marb ^h ̃

These examples demonstrate a clear case of sound change as we move geographically across dialects likely influenced by assimilation processes. This shift reflects regional variation in the assimilation of phonetic features over a spatial continuum, a common phenomenon in dialect continua. Such changes often occur gradually, with adjacent dialects sharing more phonetic features while distant dialects diverge more significantly. (Labov, 2001; Chambers, 1995).

In addition to linguistic factors, the study also considers socio-cultural variables, such as rural settlement patterns, language attitudes, and domains of language use, which mediate the phonological outcomes in contact zones (Weinreich, 1953; Thomason & Kaufman, 1988).

This research contributes both empirically and theoretically: it helps preserve transitional speech varieties threatened by erosion, and it advances our understanding of phonological change in geographically contiguous contexts. It also invites further exploration into the interface between linguistic structure and social practice in the Indo-Aryan context.

Keywords- Language contact, Phonological transition, Dialect continuum, Transition zone, Socio-phonetics, Dialect geography, Language shift

References-

- Abbi, A. (1994). *Languages of the Eastern Indo-Aryan Group*. Lincom Europa.
- Chambers, J. K., & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge University Press.
- <https://doi.org/10.1017/CBO9780511805103>
- Grierson, G. A. (1880). *An introduction to the Maithili dialect of the Bihari language as spoken in North Bihar, Part I* (2nd ed., 1909). Calcutta: Asiatic Society.
- Grierson, G. A. (1903). *Linguistic Survey of India, Vol. V, Part II, Bihari & Oriya Languages* (Reprinted in 1968). Delhi: Munshi Ram Motilal Banarsidass.
- Hock, H.H., & Joseph, B.D. (2009). *Language History, Language Change, and Language Relationship*. Mouton de Gruyter.
- Labov, W. (1994). *Principles of Linguistic Change: Internal Factors*. Blackwell.
- Linguistic Survey of Bihar. (2020). Department of Languages, Government of Bihar.
- Masica, C.P. (1993). *The Indo-Aryan Languages*. Cambridge University Press.
- Ninan, P. (2007). "Angika Language and Identity." *Economic and Political Weekly*, 42(36),

3666–3670.

Thomason, S.G., & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. University of California Press.

Trudgill, P. (2008). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, 37(5), 1–28. Cambridge University Press. <https://doi.org/10.1017/S0047404508080299>

Weinreich, U. (1953). *Languages in Contact: Findings and Problems*. Mouton.

Acoustics of Four-way Laryngeal Contrast in Kudmali Stops

Baidyanath Mahato¹ Indranil Dutta²

School of Languages and Linguistics, Jadavpur University^{1,2}
bmahatos114@gmail.com¹, indranildutta.lnl@jadavpuruniversity.in²

Abstract

Four-way laryngeal contrast in stop consonants is an important typological feature of Indo-Aryan Languages. Several acoustic studies in major Indo-Aryan Languages such as Hindi (Dutta, 2009) (Jessamyn Schertz, 2020), Marathi (Dmitrieva, 2019), Nepali (Martha Schwarz, 2019), etc., provide an account of the four-way contrast system in stop consonants. Among them, Hindi (Maddieson, 1984) and Marathi (Dmitrieva, 2019) have accounted for a stable laryngeal contrast system.

Kudmali, also known as Kurmali, spoken in parts of West Bengal, Jharkhand, Odisha, and Assam, is a lesser-known and under-resourced language. This language is sometimes confused with the Manbhum dialect, although in some recent studies it has been recognized as a separate language belonging to the eastern Maghi branch of the Indo-Aryan language family. Beyond its status as an under-documented language, Kudmali provides an important testing ground for theories of laryngeal contrast and cue integration in speech. Additionally, the findings have practical implications for speech technology applications, including automatic speech recognition and text-to-speech systems for low-resource languages.

This work aims to document and study the stability of the four-way contrasts (voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced aspirated) in bilabial, dental/alveolar, and velar stop consonants of Kudmali.

The dataset consists of controlled elicitation of lexical items designed to represent the full range of stop contrasts in Kudmali. Tokens were recorded in a consistent phonological environment, primarily in word-initial position before low vowel /a/, in order to minimize contextual variability in f₀ realization. The dataset includes four repetitions of each item across 20 speakers differing in age (20-50) and gender, all of whom are native speakers of Kudmali. For recording audio Zoom H6 recorder was used with a sample rate of 44100 Hz.

Acoustic parameters such as VOT, fundamental frequency at onset of voicing, closure duration, Spectral tilt index (H1-H2), Spectral tilt (H1-A3, H1-A2) are measured.

Primary results indicate that VOT and f₀ onset measure clearly differentiates across four categories of Kudmali, and it also preserves the four-way contrast system like other Indo-Aryan Languages of India, though it is largely in contact with Austroasiatic languages. Moreover, spectral and phonation-related acoustic measures provide similar evidence as stated above.

Keywords: Kudmali, Acoustic analysis, Laryngeal contrast, Stop consonants

References

- Abramson, L. L. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, 20(3), 384-422. doi:{10.1080/00437956.1964.11659830
- Anvita, A. (2001). *A manual of linguistic field work and structures of Indian languages*. Lincom Europa.
- Cho, T. W. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72(4), 52-65. doi:10.1016/j.wocn.2018.11.002
- Dmitrieva, O. &. (2019). Acoustic Correlates of the Four-Way Laryngeal Contrast in Marathi. *Phonetica*, 77(3). doi:10.1159/000501673
- Dutta, I. (2009). *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*. VDM Publishing.
- Ethnologue*. (2025). Retrieved from <https://www.ethnologue.com/language/kyw/>
- Jessamyn Schertz, S. K. (2020). Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics*, 81, 100979.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
- Martha Schwarz, M. S. (2019). *Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism* (Vol. 73). doi:10.1016/j.wocn.2018.12.007

On modally flavoured silent predicates in Vedic Sanskrit copular sentences

1. The puzzle: An unnoticed generalization

In the *R̥gveda* and the *Atharvaveda*, namely in the earliest strata of Vedic Sanskrit, there is a class of copular sentences where: the predicate is a dative-marked (*-tave/-tavaī*) deverbal noun functionally identical to an infinitive; the subject is a nominative-marked nominal; and the copula (*as-* ‘to be’; *bhū-* ‘to become’) is generally left silent when it is conjugated in the indicative present (see next page, (1)-(2)), and overtly realized otherwise, as in (3) (Keydana 2013:60). In this class of copular sentences, a patientive subject correlates with deontic modality, while an agentive subject correlates with dynamic modality. This generalization is numerically illustrated in (4a-b): the figures to the right of the slash indicate the total number of *-tave/-tavaī* infinitives occurring in the texts under examination, including non-predicative infinitives. This generalization has gone unnoticed in modern scholarship.

Thus, Gippert (1984:218) correlates patientive and agentive subjects of the infinitive with the stronger versus weaker periphrastic relation between the copula and the infinitive. Lühr (1997) distinguishes between authority-related necessity and circumstance-related possibility, thereby preventing the identification of dynamic modality as an independent modal category. On the other hand, Keydana (2013:72) assumes that the modality is always epistemic in (1)-(3). Adopting a synchronic, formal approach couched within the framework of generative grammar, we show that the generalization in (4a-b) naturally follows from Moro’s (1997) analysis of copular sentences combined with Manzini’s (2017) analysis of the passive.

2. The analysis: A silent predicative expletive

According to Moro (1997:195-196), the pronominal *it* of sentences like *it seems that John is sad* and *it’s that John is sad* is not a subject expletive, but a predicative expletive that serves as a sentential predicate, semantically equivalent to the adjective *true* (see next page, (5)). Let us now assume that Vedic Sanskrit has a pronominal expletive *pro*^x that only differs from the *it* of (5) in that *pro*^x is systematically silent and is not restricted to be semantically equivalent to *true* (i.e., *pro*^{TRUE}, bearing an unmarked modal value), but ranges over different modal flavours: *pro*^{PERMITTED}, *pro*^{REQUIRED}, and *pro*^{(UN)ABLE}. This yields the underlying structures (6a)-(7a) for (1)-(2), respectively. Let us now see how generalization (4a) can be deduced from (6).

According to Manzini (2017:239), the agent starts out in the aspectual layer *Asp*, ultimately resulting in nominative alignment; however, when *Asp* is missing, such as in the past passive participle, the agent is introduced as an oblique argument in Spec-vP: see (8). By hypothesis, Vedic infinitives, qua reduced verbal categories, also lack *Asp*. Accordingly, in (6) the agent is introduced as an oblique argument (instrumental-marked: *anena* ‘by him’), while the patient takes the nominative (*havīs* ‘oblation’) by raising to the Spec of the silent copula. The different featural makeup of the (non-oblique) patient and the (oblique) agent ensures that Relativized Minimality is not violated (Manzini 2017:241). This derives the facts of (4a).

Let us now turn to the deduction of generalization (4b) from (7). Crucially, unlike deontic *pro*^{PERMITTED}, dynamic *pro*^{(UN)ABLE} is no sentential predicate. Thus, *pro*^{UNABLE} selects its nominal subject (Wurmbrand 2001:201): *háriḥ* ‘tawny’ (i.e., Fire). This permits *Fire* to be directly introduced into the Spec of the silent copula in (7), where it is nominative-marked. To this one should add that the subject of dynamic predicates like (*un*)able typically controls the agent, as opposed to the patient, of the infinitive (Nuyts 2016: 34 n. 4). Therefore, *Fire* (the subject of *pro*^{UNABLE}) controls *PRO*, interpreted as the agent of ‘to eat’ (*áttave*). This accounts for (4b).

All in all, the assumption that Vedic Sanskrit copular sentences of the type in (1)-(3) involve a silent predicative expletive endowed with different modal flavours permits the derivation of the generalization in (4a-b) from independently assumed principles of grammar.

(1) *Śaunakīya Atharvaveda* 12.2.37ab (Patient + **DEONTIC** modality)

<i>a-yajñiyó</i>	<i>hatávarcā</i>	<i>bhavati/</i>	<i>nái=nena</i>
NEG-sacrificeble.NOM	smitten.splendour.NOM	becomes	NEG=he.INS
<i>havír</i>	<i>áttave</i>		
oblation.NOM	eat.INF		

‘He (= Fire) becomes unfit for sacrifice, of smitten splendor. **The oblation must/ may not be eaten** by him.’

(2) *R̥gveda* 10.79.6c (Agent + **DYNAMIC** modality)

[...]	<i>á-kṛīlan</i>	<i>kṛīlan</i>	<i>hárir</i>	<i>áttave</i>	<i>’dán</i>
	NEG-playing.NOM	playing.NOM	tawny.NOM	eat.INF	NEG-tooth.NOM

‘(What sin of omission or commission have you done before the gods? I now ask you, Fire, since I do not know.) Whether or not he likes it, **tawny (Fire) has to eat without teeth/ is unable to eat with his own teeth.**’

(3) *Paippalāda Atharvaveda* 19.4.11 (Agent + **DYNAMIC** modality)

<i>oktā</i>	<i>āpah</i>	<i>karmaṇyā</i>	<i>muñcantv</i>
invoked.ACC	waters.ACC	connected.to.action.ACC	free.IMPV.3PL
<i>itah</i>	<i>prañitaye/</i>	<i>sadyo</i>	<i>bhavantv</i>
from.here	guidance.DAT	instantly	become.IMPV.3PL
			<i>etave </i>
			go.INF

‘May they release from here the invoked waters connected to the action from this place to guide [us]; let [the waters] instantly become able to flow.’

(4) a. **Patient** + (copula) + infinitive = **deontic** modality (‘X must not be eaten’)

‘-tave: 5/77 (*R̥gveda*); 3/44 (*Śaunakīya Atharvaveda*); 3/54 (*Paippalāda Atharvaveda*).

‘-tavaí: 3/26 (*R̥gveda*); 1/10 (*Śaunakīya Atharvaveda*); 2/14 (*Paippalāda Atharvaveda*).

b. **Agent** + (copula) + infinitive = **dynamic** modality (‘X is unable to eat’)

‘-tave: 1/77 (*R̥gveda*); 0/44 (*Śaunakīya Atharvaveda*); 1/54 (*Paippalāda Atharvaveda*).

‘-tavaí: 0/26 (*R̥gveda*); 0/10 (*Śaunakīya Atharvaveda*); 0/14 (*Paippalāda Atharvaveda*).

(5) a. It seems [that John is sad]. c. ___ seems **pro**^{TRUE} [John to be sad].

b. [That John is sad] seems true. d. John_i seems **pro**^{TRUE} [_i to be sad].

(6) a. *Underlying structure*: [_{Asp} **he** [to eat the oblation]] is not **pro**^{PERMITTED}.

b. *Oblique marking of Agent*: [_{Asp} ___ [_{vP} by him [to eat the oblation]]] is not **pro**^{PERMITTED}.

c. *Extraposition*: ___ is not **pro**^{PERMITTED} [_{Asp} ___ [_{vP} by him [to eat the oblation]]].

d. *Raising of Patient*: the oblation_i is not **pro**^{PERMITTED} [_{Asp} ___ [_{vP} by him [to eat _{t_i}]]].

e. *Fronting of oblique Agent*: by him_j the oblation_i is not **pro**^{PERMITTED} [_{Asp} ___ [_{vP} _{t_j} [to eat _{t_i}]]].

f. *Surface structure*: náinena_j havír_i **pro**^{PERMITTED} [_{Asp} ___ [_{vP} _{t_j} [áttave _{t_i}]]].

(7) a. *Underlying structure*: Fire is **pro**^{UNABLE} [_{Asp} ___ [_{vP} PRO [to eat with teeth]]].

b. *Control*: Fire_i is **pro**^{UNABLE} [_{Asp} ___ [_{vP} PRO_i [to eat with teeth]]].

c. *Surface structure*: hárir_i **pro**^{UNABLE} [_{Asp} ___ [_{vP} PRO_i [áttave ’dán]]]].

(8) ‘He was killed by John’: ... was [_{vP} [_{vP} v [_{vP} killed he]]] [(_⊆)_P by John]] (Manzini 2017:239)

Gippert 1984, Zum “prädikativen” Infinitiv, in *Zeitschrift für Vergleichende Sprachforschung* 97. **Keydana 2013**, *Infinitive im R̥gveda*, Brill. **Lühr 1997**, Zur ‘Kasusattraktion’ in *altindischen dativischen Infinitivkonstruktionen*, in *Sound law and analogy*, Rodopi. **Manzini 2017**, *Passive, smuggling and the by-phrase*, in *A Schrift to Fest Kyle Johnson*, Linguistics Open Access Publications. **Moro 1997**, *The Raising of Predicates*, Cambridge University Press. **Nuyts 2016**, *Analyses of the Modal Meanings*, in *The Oxford Handbook of Modality and Mood*, Oxford University Press. **Wurmbrand 2001**, *Infinitives*, De Gruyter Mouton.

Reassessing Relative Constructions in 19th-Century Urdu: A Corpus-Based Analysis of *jo* Correlatives and Embedded Relatives

Miki Nishioka
University of Osaka

Muhammad Nawaz
COMSATS University Islamabad

This study reassesses relative clause patterns in nineteenth-century Urdu in order to determine whether any central-embedded relatives—sensu Puri (2011)—coexisted with the traditional *jo ... vo/us* correlative system. While descriptive grammars emphasize the productivity of classical correlatives, large-scale quantitative analyses of historical Urdu prose remain scarce. To address this gap, a stratified sample of approximately one hundred *jo*-tokens was extracted from two nineteenth-century Urdu narrative works (as preserved in mid-twentieth-century printed editions). The sample was drawn from a corpus comprising 1,227 occurrences of *jo* (including *jis/jin*) and subjected to detailed structural analysis.

Each token was categorized into four types following established Indo-Aryan distinctions: (i) post-nominal Head N + *jo* (type 1a), (ii) correlative *jo/jis/jin ... vo/us* (type 2), (iii) topic-fronted *vo* + Head N ... *jo* (type 1b), and (iv) potential embedded relatives. Because *jo ... vo/us* correlatives in Urdu function as discourse-based topic–comment constructions rather than movement-derived relatives (Butt & King 1996), their structural behavior differs fundamentally from NP-internal post-nominal relatives. Following Puri (2011), the present study defines “embedded” strictly as central-embedded relatives—e.g., *vah baccā [jo nahā rahā hai] sundar hai*—in which the relative clause interrupts the matrix predicate. Consequently, post-nominal N + *jo/jis/jin* sequences were not treated as embedded, even when no correlative occurred, unless they exhibited genuine matrix-internal embedding.

The analysis found no instances of central-embedded relatives among the approximately one hundred sampled tokens; all examples were either post-nominal relatives or members of the *jo ... vo/us* correlative system. This suggests that nineteenth-century Urdu narrative prose maintained the classical correlative system and did not develop the central-embedded relativization documented in English-influenced translated Hindi texts (Puri 2011). This outcome is consistent with the view that central-embedded relativization is not an inherited Indo-Aryan structure but rather a contact-induced innovation arising primarily in translation contexts. Further examination of all 1,227 tokens will clarify whether the absence of central embedding reflects a broader characteristic of Urdu or a genre-specific distribution in the texts analyzed.

Table 1. Representative Examples by Type

Type	Urdu Example	IPA-style Romanization
1a: N + <i>jo</i> ... <i>vo/us</i>	وہ کتاب جو تم نے دی تھی وہ بہت اچھی ہے۔	<i>vo kitāb jo tum ne dī thī</i> <i>vo bahut acchī hai.</i>
1b: <i>vo/us</i> + N ... <i>jo/jis/jin</i>	وہ آدمی آیا ہے جو کل تم سے ملا تھا۔	<i>vo ādmī āyā hai jo kal</i> <i>tum se milā thā.</i>
2: <i>jo/jis/jin</i> ... <i>vo/us</i>	جو کچھ اُس نے کہا تھا وہ درست تھا۔	<i>jo kuch us ne kahā thā vo</i> <i>durust thā.</i>
Central- embedded	وہ بچہ جو نہا رہا ہے بہت خوبصورت ہے۔	<i>vo baccā jo nahā rahā</i> <i>hai bahut khūbsūrat hai.</i>

Table 2. Classification of Relative Constructions in Urdu

Type	Structure	Correlative?	NP-internal?	Embedded?
1a	N + <i>jo</i> ... <i>vo/us</i>	optional	Yes (N+ <i>jo</i>)	No
1b	<i>vo/us</i> + N ... <i>jo</i>	Yes	No	No
2	<i>jo/jis/jin</i> ... <i>vo/us</i>	Yes	Not relevant	No
Central- embedded	vo N [<i>jo/jis/jin</i> ...] X-Predicate	No	Yes	Yes

Main References

- Butt, M., & King, T. H. (1996). Structural Topic and Focus without Movement. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG Conference*.
- Masica, C. (1991). *The Indo-Aryan Languages*. Cambridge University Press.
- Puri, V. (2011). The Influence of English on the History of Hindi. In R. Singh (Ed.), *The Yearbook of South Asian Languages and Linguistics*.
- Schmidt, R. L. (1999). *Urdu: An Essential Grammar*. Routledge.
- Urdu Corpus on COSH CQL Search (2024), available at <https://cql.cosh.site/>, accessed January 22, 2026.

Pre-verbal and Post-verbal Negation in Odia: Distribution and Typological Patterns

Abstract

Negation in Indo-Aryan languages is widely recognized as structurally complex and resistant to uniform characterization (Masica 1991: 389). Building on this assessment, the present study offers a comprehensive analysis of the negation system in Odia. This research argues that negative markers in Odia are not structurally homogeneous; rather, their distribution and functional properties vary according to the type of negation involved. A central pattern is the alternation between pre-verbal and post-verbal negation, which is conditioned by tense, aspect, and modality distinctions, as well as by finiteness.

Within clausal negation, pre-verbal and post-verbal negators exhibit systematic distributional tendencies. Post-verbal negators typically occur in standard negation (example 1) and non-declarative contexts (example 2), whereas pre-verbal negators are represented with particular moods (example 3), non-main clauses (example 4), and stative predications (example 5). The paper typologically examines the positional distribution of negative markers across different domains of clausal negation, including standard negation, modal negation, imperatives, and subordinate clauses. The analysis is further situated within Miestamo's (2005) typology of symmetric and asymmetric negation, which compares affirmative and negative constructions to determine whether the introduction of a negative marker results in structural variations. Thus, this study enables a deeper understanding of the typological and structural patterns of negation in Indo-Aryan languages by examining Miestamo's framework on Odia.

Keywords: Odia, Clausal negation, pre-verbal and post-verbal, symmetric and asymmetric

Examples

1. mǔ Germany dʒa-u-n-i
 1SG Germany go-PROG-NEG-1SG
 I am not going to Germany.
2. dʒʌtʌmi kɔr-ni
 mischief do-NEG
 Don't do any mischief.
3. c^hua-tʃa kuɔ pak^hɔku nɔ-dʒa-u
 kid.3SG-CL well near NEG-go-3SG
 Do not let the child go near the well.
4. tɔme nɔ-k^ha-i cali-go-l-ɔ
 2SG NEG-EAT-CP walk-go-PST-2SG
 You went away without eating.
5. mɔsiŋa tɔle sapɔ nɔ-t^hi-l-a
 mat under snake NEG-be-PST-3SG
 There was no snake under the mat.

References

- Miestamo, Matti. 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. Berlin: Mouton de Gruyter.
- Masica, Colin P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Title: Ergative variation in the Konkani varieties spoken in Karnataka

Author: Ruta Paradkar

Academic affiliation: Institute for Scandinavian Studies, Frisian Studies and General Linguistics, Kiel University (CAU), Germany.

Email address: rparadkar@isfas.uni-kiel.de

In this talk, I discuss variation in the ergative construction in varieties of Konkani, drawing on historical texts, dialect descriptions, and our ongoing work in the Konkani-Kannada border region. I report variation in the split-ergative pattern in Karwar located on Goa-Karnataka border in the Uttara Kannada district of Karnataka, and contrast it with other varieties of Konkani in more intense contact with Kannada that have not undergone this change.

Restructuring in the split-ergativity pattern in terms of overt case-marking and verbal agreement has been noted as a feature of convergence of IA towards Dravidian, an extensively-discussed phenomenon for varieties of Marathi in contact with Dravidian languages. (Gumperz and Wilson, 1975; Kulkarni-Joshi, 2023). In the neighbouring and closely related language Konkani, however, Miranda (2003, p. 760) reports a retention of the Indo-Aryan split-ergativity in Konkani spoken in Karnataka. In standard Konkani, in the transitive-perfect clauses, ergative case-marking appears on all subjects except the first and second person plural, and the verb variably shows object/default agreement. Miranda's claim of retention of this pattern holds up after examining various dialect documentations of Konkani, spoken in Goa and the Dakshina Kannada district for instance, which show a continuation of the same pattern from the 16th century texts shown in examples (3) and (4). In our recent fieldwork at Karwar located on the Goa-Karnataka border, however, variation in ergative-marking, and verbal agreement with the unmarked subjects in transitive perfect clauses were observed as demonstrated in examples (1) and (2) – resembling varieties of Marathi spoken in the Marathi-Kannada border region reported in Kulkarni-Joshi, 2023.

The previous accounts of modern Konkani-Kannada language contact assume that contact-induced changes are triggered by the functional dominance of Kannada in the minority Konkani settlements in Kannada-speaking regions formed as a result of migration since the 16th century (Nadkarni 1975; Miranda 2003). However, this explanation cannot account for this restructuring observed in the Konkani-dominant Karwar adjoining Goa, while it is retained in other varieties with a more intensive contact with Kannada.

Drawing on insights from sociolinguistic typology (Trudgill, 2011), I argue that examining kinship networks and sociopolitical history of the region may explain why other southern varieties have retained the older pattern in Kannada-dominant regions whereas the Karwar variety in the Konkani-dominant region shows signs of change. Karwar Konkani communities have stronger ties with the neighbouring Kannada-majority, and Kannada-Marathi bilingual areas than with Konkani-dominant Goa due to the sociopolitical history of the region. Since Konkani is widely spoken in Karwar as a lingua franca, changes are likely to spread more easily among all Konkani speakers there. On the other hand, Konkani spoken elsewhere in Karnataka is largely restricted to the home domain and has links to other Konkani-speaking regions through endogamous kinship networks. In this paper, I will interpret my observations in the context of these sociolinguistic settings of Konkani varieties in Karnataka.

Examples:

1. kaḷi bridʒ pəḷəy-l-ã n^{hi} tũ?
Kali bridge see-PST-SG.N no 2SG
You have seen the Kali bridge, right?
(In most Konkani varieties, a younger woman is referred to in the neuter gender, so verb shows neuter agreement here with the second person subject referring to a young female.)
2. te doni tæblætʃ g^{het}-l-ə
3PLM two tablets take-PST-SG.M
I (M) took those two tablets.

(Karwar fieldwork data, January-February 2026)

3. tãcẽ vadhapaṇa hãvẽ kãḍ-unu g^{heta}-l-ã
3.GEN.SG.N old age 1SG.ERG take-CVB take-PST-SG.N
I have taken away his old age.
4. tuvẽ apulə soməvamsu ud^{hər}-l-ə
2SG.ERG REFL.GEN.SG.M Soma clan elevate-PST.SG.M
You have elevated your Soma clan.

(16th century Konkani examples from Nayak, 1990)

References:

- Gumperz, J.J. and Wilson, R. (1971). Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In D. Hymes (ed.), *Pidginization and creolization of languages*. 151–168. Cambridge: CUP.
- Kulkarni-Joshi, S. (2023). Variation and Change in Dialects of Marathi: A Social-Dialectological Approach. In: Chandra, P. (eds) *Variation in South Asian Languages*. Springer, Singapore. https://doi.org/10.1007/978-981-99-1149-3_9
- Miranda, R. (2003). Konkani. The Indo-Aryan languages, George Cardona and Dhanesh Jain (Eds.), *The Indo-Aryan languages* (pp. 729-65.)
- Nadkarni, M. (1975). Bilingualism and syntactic change in Konkani. *Language*, 672-683.
- Nayak, P. (1990). *Solavya Shenkadyantle Konkani Mhabharat: Adi Parva*. Thomas Stephens Konkani Kendra.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press, USA.

**Dative “subjects” and “direct objects” –
The dative case in Middle and Modern Kannada**
John Peterson, Kiel University

In my talk I discuss the use of the dative case in Middle and Modern Kannada (glcode nucl1305, ISO 693-3 kan). In both stages of Kannada, the main function of the dative is to mark “indirect objects” and goals (e.g., Sridhar, 1990: 159ff.; Zydenbos, 2020: 87ff.), in addition to a relatively small number of other functions. In typological terms (and somewhat simplified here), Kannada has an “indirective” alignment pattern (Haspelmath, 2005), where P (the “direct object” of a monotransitive verb) and T (the “direct object” of a ditransitive verb) both appear in the accusative (if definite / human) or are unmarked (if indefinite / nonhuman), while R, the goal of a ditransitive verb or the “indirect object”, appears in the dative. P and T thus receive the same marking, which is different from that of R; i.e., $P=T \neq R$ (exs. (1a) and (1b) from Sridhar, 1990: 10 and 93, respectively).

As Sridhar (1990) notes for Modern Kannada, the dative is also found with the so-called dative-subject construction, which has a nominative-marked emotional or cognitive state as the grammatical subject (often with a copula or other light verb) and a dative-marked experiencer (ex. 2; from pp. 159-160). The dative also optionally marks P with certain verbs, instead of the accusative, e.g., *hode* ‘hit’, *pusala:yisu* ‘coax’, *re:gisu* ‘irritate’, *muttiḍu* ‘kiss’, *hogaḷu* ‘praise’, etc. (ex. 3; from p. 162). It is interesting to note that most of these verbs typically take a human / animate object.

While the functions of the dative are largely similar in the texts of the Middle Kannada *Vacana* literature of the 12th century CE, there are nevertheless important differences. Like Modern Kannada, Middle Kannada has the same dative-subject construction with a nominative-marked emotional / cognitive state as the grammatical subject and a dative-marked experiencer; cf. ex. (4) (the number refers to the respective verse number in Na:gabhu:ṣaṇasva:mi, 2004).

However, these dative-subject constructions are quite marginal in the Middle Kannada texts of my corpus, unlike in the modern language, where they are very common. Instead, Middle Kannada has a class of experiencer verbs with a nominative experiencer and a dative-marked stimulus, such as with *oḷi / oli* ‘be pleasing; like’, *aṃju* ‘fear’, *eḷasu* ‘take pleasure in; desire’, *na:cu* ‘be embarrassed / ashamed of; fear’ *a:saru* ‘become weary of’, and many others (exs. 5 and 6). Structurally, this class resembles that illustrated in example (3), where the “direct object” can be marked as dative, however there are clear semantic differences between the two constructions, despite these similarities, which I discuss in my talk.

A major shift in the marking of experience and cognition predicates has thus occurred in Kannada since the 12th century, from the predominantly NOM/ACC and NOM/DAT alignment of Middle Kannada to the now predominant dative-subject construction. In this respect it is noteworthy that the Modern Kannada alignment pattern is structurally identical to the marking of these predicates in Indo-Aryan, where “dative-subject constructions” abound.

I therefore end my talk with a discussion of the possibility that this dative-subject construction in Kannada is not an entirely internal development but derives at least in part from Indo-Aryan, which is spoken just to the north (Marathi, Konkani), as well as in Karnataka itself as an important and prestigious minority language (Konkani).

Such direct morphosyntactic restructuring from Indo-Aryan is also found elsewhere in the morpho-syntax of some Kannada varieties: E.g., in the speech of bilingual speakers of Marathi and Kannada directly along the Indo-Aryan / Dravidian border, where the historical dative in some North Kannada dialects has entirely ousted the accusative and is now used to mark all definite P’s, as in Marathi (Dhongde & Wali, 2009: 181; cf. examples 7a and 7b), resulting in a secundative alignment pattern in these dialects, as in Indo-Aryan, but unlike the indirective alignment found elsewhere in Kannada.

- (1)a. *sudha:ma ja:far-n-ige na:[e [...] ka:lcedmā-annu koḍ-al=idda:ne.*
 Sudhama Jafar-n-DAT tomorrow football-ACC give-INF=COP.PRS.3SG.M
 R=DAT T=ACC

‘Sudhama is going to give Jaffar the football [...] tomorrow.’

- b. *a:ta se:q-ig=a:gi u:r-ann=e: suṭṭ-a.*
 3SG revenge-DAT=ADV town-ACC=FOC burn.PST-3SG.M
 P=ACC

‘He burned the whole town out of vengeance.’ P = T = ACC ≠ R = DAT

- (2) *huḍugan-ige hedarike a:-yitu.*
 boy-DAT fright.NOM become-PST.3SG.N

‘The boy was frightened (= fright became to the boy).’

- (3) *avar-a aṇṇa avan-annu / avan-ige dinavu: hoga[-utta-r]=amte.*
 3SG.HON-GEN elder.brother 3SG-ACC / 3SG-DAT daily praise-PRS-3HON=HEARSAY

‘It’s said that his elder brother praises him everyday.’

- (4) *ī trividha-van ari-d-a-vamge [...] śamke-y illa [...] bhīti-y illa.*
 this Trividha-ACC know-PST-PCTP-3SG.M.DAT doubt-y NEG.COP fear-y NEG.COP

‘He who knows the trividha has no doubt [...] or fear [...]. [Amuge Rāyamma, No. 393]

- (5) *nānu cennamallikārjunan emba gamḍa-mge mige~mige*
 1SG Chennamallikarjuna say.NPST.PTCP husband-DAT intensely~REDUPL

oli-d-e ele avva-ga[irā.
 be.pleased-PST-1SG VOC mother-PL VOC.PL

[Akkamahadevi, No. 964]

‘Oh mothers, I have fallen intensely in love with my husband called Chennamallikarjuna.’

- (6) *anya-kk e[as-uv-ud=e enna mana?*
 other-DAT take.pleasure.in-IPFV-3SG.NHUM=INTER 1SG.GEN mind

‘Does my mind desire / take pleasure in anything else?’ [Akkamahadevi, No. 495]

(7a) North Kannada

na: giḍa-k no:q-əd-e.
 1SG tree-DAT see-PST-1SG

‘I saw the tree.’

(own data, speaker from
 Sankeshwar, Belagavi)

(7b) Standard Kannada

na:nu mara-vannu no:q-id-e.
 1SG tree-ACC see-PST-1SG

‘I saw the tree.’

References

- Dhonde, Ramesh Vaman & Kashi Wali. 2009. *Marathi*. [London Oriental and African Library 13]. Amsterdam / Philadelphia: Benjamins.
- Haspelmath, Martin. 2005. Argument marking in ditransitive alignment types. *Linguistic Discovery* 3/1: 1–21. DOI: 10.1349/PS1.1537-0852.A.280
- Na:gabhu:ṣaṇasva:mi, O.L. (ed.). 2004. *Vacana sa:vira* [A thousand *vacanas*]. Hampi: Prasa:raṅga Kannaḍa Viṣṭavidya:laya.
- Sridhar, S.N. 1990. *Kannada*. [Descriptive grammars series.] London / New York: Routledge.
- Zydenbos, Robert J. 2020. *A manual of modern Kannada*. Heidelberg University: CrossAsia-eBooks: DOI: <https://doi.org/10.11588/xabooks.736>

Subject obviation in directive sentence types in Tamil

Gopalakrishnan Ramamurthy, *Independent* (gopalkrishnan11251@gmail.com)

BACKGROUND. Tamil has two sentence types that conventionally give rise to directive speech acts (Stegovec 2019): the *Imperative* (1–2), and what we dub the *Jussive* (5–9). Neither can be used for descriptive speech acts; and following their utterance, ‘that’s true/false’ are infelicitous responses. They are thus distinguished from deontic modals used performatively (Kaufmann 2019a: §4). This study examines the subjects of Imperative and Jussive sentences in Tamil.

DATA. Imperatives (IMP) are restricted to 2nd person subjects (2.SBJ) (1–2). They cannot take the polar question particle (PQP) =*ā*, or *wh*-words (3–4) (except in echo questions, which we set aside). Basic Jussives (JUSS; verb suffix *-ttum*) can only have 3.SBJ (5) (no SBJ agreement). Jussives have the same “functional heterogeneity” as IMP: (5) with 3.SBJ may be a command, request, advice, etc. (Kaufmann 2019a). But, Jussives *can* take PQP =*ā* and *wh*-words; we dub these *Jussive Interrogatives* (JUSS.INT). JUSS.INT can have 1.SBJ or 3.SBJ, but not 2.SBJ (6–7).

MAIN CLAIM. Our goal is to explain the asymmetric pattern of acceptable subjects in JUSS sentences (see Table below). We focus here on singular subjects for brevity. Since IMP and JUSS have the same set of discourse functions (both may be used for commands, advice, etc.), our starting point is the assumption that they share the same semantics, differing only in that IMP is specified for 2.SBJ, while JUSS is unspecified for subject. We hence propose that 1.SG.SBJ is unavailable in basic JUSS, and 2.SG.SBJ in JUSS.INT, due to subject obviation (Stegovec 2019; Kaufmann 2019b), while basic JUSS with 2.SG.SBJ are blocked by competition with the IMP.

OBVIATION. Obviation refers to a cross-linguistic ban on co-reference of subjects of matrix attitude predicates with the subject of the clausal complement: **he_i wants that he_i run* (Stegovec 2019). Kaufmann (2019b) examines *directive obviation*: restrictions on available referents for subjects of directive sentences, which she analyses as arising from the semantics of directives.

ANALYSIS. Following Kaufmann (2019b), we pursue a semantic account of directive obviation, couched in her *modal operator theory*. Kaufmann assumes that directive speech acts involve two discourse roles: *director*, who “selects and promotes the course of events”; and *instigator*, who “sees to it that [they occur]”. Further, directive sentences have a covert modal operator, and trigger a set of propositions that are satisfied only by performative uses of modals. The relevant modal flavour is *decisive modality*: which provides “agreed-upon criteria” to solve a decision problem (DP). The two presuppositions relevant are: (i) director holds both *p* (prejacent) and $\neg p$ as possible; (ii) director has perfect knowledge of solutions to the DP. From this, the principle of *Director’s Anticipation* follows: if director *d* commits to the belief that instigator *i* believes *p* is necessary to solve the DP, then *d* commits to the belief that *p* will become true.

In basic JUSS, *speaker* = *director*. However, in JUSS.INT, *director* = *addressee*, due to interrogative flip. We also assume that when subject is 1p or 2p, the *instigator* is the subject by default. Obviation thus occurs when the *director* and *instigator* roles co-refer. In 1.SG.SBJ JUSS, *speaker* = *director* = *instigator*. Hence, the speaker commits to the belief that *p* will become true (*Director’s Anticipation*), but also that both *p* and $\neg p$ are possible: thus contradiction. This does not occur with 3.SG.SBJ. The same mechanism, *mutatis mutandis*, explains the fact that obviation blocks JUSS.INT with 2.SG.SBJ. Finally, JUSS with 2.SBJ is blocked by competition with the IMP.

PLURAL SBJ. Obviation also blocks 1.PL.EXCL.SBJ in JUSS, and 2.PL.SBJ in JUSS.INT. Further, JUSS with 2.PL.SBJ is blocked by competition with IMP.PL. We suggest that 1.PL.INCL.SBJ is blocked in JUSS *and* JUSS.INT due to competition with the performative use of the possibility modal *-lām*, which (conventionally?) yields exhortative interpretations (Asher 1989: 166).

EXCEPTIONS. This pattern has exceptions. E.g., with non-agentive predicates like ‘fall’, 1.SG & 1.PL.EXCL subjects are ameliorated (8–9). 1.SBJ JUSS are also acceptable when the speaker has

no control over the actualisation of *p* (10–11). These are predicted by the present account.

(Blue = acceptable; Red = unacceptable)

- (1) (**nān/nī*/**avan*) *ōḍḍ*
(**I/you.sg*/**he*) run.IMP.SG
'(**I/you*/**he*) run!'
- (2) *ellār=um* *ōḍḍ-nga*
all.SUBS.HUM=ADD run.IMP-PL
'Everyone (of **us/you*/**them*) run!'
- (3) **nī* *ōḍ=ā*
**you.sg* run.IMP.SG=POLQ
INTENDED: 'Should you run?'
- (4) **engē/yārū* *pō*
**where/who* go.IMP.SG
INTENDED: ???
- (5) **nān*/**nī/avan* *ōḍ-a-ttūm*
I*/you.sg/he* run-INF-JUSS
'Let **me*/**you/him* run!'
- (6) *nān*/**nī/avan* *engē* *pōg-a-ttūm*
I/**you.sg/he* where go-INF-JUSS
'Where shall *I*/**you/he* go?'
- (7) *nān*/**nī/avan* *ōḍ-a-ttūm=ā*
I/**you.sg/he* run-INF-JUSS=POLQ
'Shall *I*/**you/he* run?' / 'May *I*/**you/he* run?' / 'Is it okay with you if *I*/**you/he* run(s)?'
- (8) %*nān* *viḷ-a-ttūm*
%*I* fall-INF-JUSS
I should fall!/May *I* fall!
- (9) %*nānga* *viḷ-a-ttūm*
%*we.EXCL* fall-INF-JUSS
We.EXCL should fall!/May *we.EXCL* fall!

For *p* to be true, someone beside the speaker must take some action (Paramasivam 1981: 126).

- (10) *enn-ay viḍḍ*, *nān pōg-a-ttūm* (11) *pēs-ādē*, *nāngaḷ pōg-a-ttūm*
I-ACC leave.IMP.SG, *I* go-INF-JUSS speak-NEG.IMP.SG, *we.EXCL* go-INF-JUSS
Leave me, let me go. Don't say (anything), let us.EXCL go.

Subject	Basic jussives		Jussive interrogatives	
	Acceptable?	Unacceptable because:	Acceptable?	Unacceptable because:
1.SG	✗	SBJ obviation	✓	N/A
1.PL.EXCL	✗	SBJ obviation	✓	N/A
1.PL.INCL	✗	Competition w/ exhortative use of POSS.MOD - <i>lām</i>	✗	Competition w/ exhortative use of POSS.MOD - <i>lām</i>
2.SG	✗	Competition with IMP.SG	✗	SBJ obviation
2.PL	✗	Competition with IMP.PL	✗	SBJ obviation
3.SG	✓	N/A	✓	N/A
3.PL	✓	N/A	✓	N/A

REFERENCES.

- Asher, Ronald E. 1989. *Tamil*. 2nd edn. Routledge.
- Kaufmann, Magdalena. 2019a. Fine-tuning natural language imperatives. *Journal of Logic and Computation* 29(3). 321–348. DOI: [10.1093/logcom/exw009](https://doi.org/10.1093/logcom/exw009).
- Kaufmann, Magdalena. 2019b. Who controls who (or what). *Proceedings of Semantics and Linguistic Theory (SALT)* 29. 636–664. DOI: [10.3765/salt.v29i0.4643](https://doi.org/10.3765/salt.v29i0.4643).
- Paramasivam, K. 1981. Why the permissive is not the permissive. In M. Arunachalam (ed.), *Proceedings of the Fifth International Conference-Seminar of Tamil Studies*, 121–134. International Association of Tamil Research.
- Stegovec, Adrian. 2019. Perspectival control and obviation in directive clauses. *Natural Language Semantics* 27. 47–94. DOI: [10.1007/s11050-019-09150-x](https://doi.org/10.1007/s11050-019-09150-x).

Diachronic and dialectal evidence for the presence of central vowels in Old Tamil

Ramprashanth Venkatakrishnan¹ and Chandan Narayan²

¹*Madurai Kamaraj University, Tamil Nadu, India*

²*York University, Toronto, Canada*

Proto-South Dravidian is typically reconstructed with a five-vowel system (*i, e, u, o, a* and their longer counterparts). However, the well-known sound change of Proto-Dravidian **k > c* (palatalization) in Old Tamil suggests a more complex vocalic landscape. While it is widely accepted that palatalization occurs before front vowels (*i, ī, e, ē*), the lack of palatalization in certain environments, along with the presence of *central* vowels in related South Dravidian languages like Muduga, Toda, Irula, Kodava etc. (Krishnamurti, 2003), suggests that Old Tamil likely possessed central vowel allophones ([ɨ], [ə]) that blocked or mediated the sound change.

We present two types of evidence, the historical ordering of the palatalization rule and dialectal data. We argue that the failure of **k* to palatalize in specific contexts, despite the apparent presence of a front vowel, points toward an intermediate stage of vocalic centralization. This analysis aligns with Zvelebil's (1970) hypothesis that central vowels may represent a pre-Dravidian substratum influence. By mapping the conditions under which **k* remained velar, we can reconstruct the specific phonological environments where front vowels were retracted to central positions in the pre-literary period of Old Tamil. The environment in which front vowels centralize is before apical consonants (plosives and liquids), consistent with an assimilatory retraction of the tongue.

A significant challenge to this historical reconstruction is the near-total loss of central vowels in Modern Standard Tamil (MST) and Malayalam, except for the “enunciative” post-plosive release [ɨ] in surface forms. However, we identify two crucial *living fossils* in modern Tamil dialects: Jaffna Tamil, which maintains a robust inventory of central vowels, serving as a geographic and linguistic outlier that preserves the archaic status quo (Kuno, 1958); and Brahmin Tamil (BT), which exhibits sporadic and inconsistent use of central vowels before retroflex consonants. A few examples common to both JT and BT to illustrate the central allophones of short (a, c) and long (b, d) front vowels:

- | | | | | |
|-----|--------------------|---------------|-----------------------|-----------------------|
| (1) | a. /kiɨ/ > [kiɨ] | ‘parrot’ | b. /vi:ɨa/ > [vi:ɨa] | ‘stringed instrument’ |
| | c. /veɨi/ > [vəɨi] | ‘firecracker’ | d. /tɕe:t/ > [tɕə:dɨ] | ‘search’ |

This dialectal centralization is blocked when a non-apical consonant follows in (2):

- | | | | | |
|-----|---------------------|------------|-------------------------|---------------|
| (2) | a. /miɨɨ/ > [miɨɨ] | ‘to stamp’ | b. /vi:ɨi/ > [vi:ɨi] | ‘street’ |
| | c. /ceNp/ > [cəmbi] | ‘copper’ | d. /tɕe:Nk/ > [tɕe:ŋgi] | ‘to stagnate’ |

We suggest that this variation is a vestigial phonetic marker of a once-wider phonological distribution. By synthesizing historical sound laws (palatalization) with dialectal data (Jaffna and Brahmin Tamil), this paper argues that central vowels were not an anomaly but a core feature of the South Dravidian phonological history.

References

- Kuno, S. (1958). Phonemic structure of colloquial Tamil. *Gengo Kenkyu: Journal of the Linguistic Society of Japan*, 41-62.
- Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge University Press.
- Steever, S. B. (1998). *The Dravidian Languages*. Routledge.
- Zvelebil, K. V. (1970). *Comparative Dravidian Phonology*. Mouton: The Hague.

Arranging Converbs in Domaaki

Noboru YOSHIOKA (National Museum of Ethnology, Japan)

The Domaaki language [dmk (ISO 639-3)], an endangered Indo-Aryan language spoken in only three settlements in the Karakoram mountain range (see Figure 1), has several converbs, as in a neighbouring language isolate, Burushaski [bsk]. This presentation aims to sort converbs and the like in Domaaki, as mentioned in previous studies and observed in my fieldwork. What I call ‘converb’ here is an apparently non-finite form derived from a verbal stem mainly carrying both functions as a predicate of the clause where the form is and as a conjunction chaining the clause to a following clause.

Some converbal forms have been identified in previous studies (Lorimer 1939 and Вайнрайх 2011), while there seem to be mismatches between them, and I have found a few other forms that can be counted as converbs. Now I will sort them again and make the full list of converbs in Domaaki according to functional and morphological features, considering noun-modifying constructions and morphological restrictions.

As a result of (re)arranging, it becomes clear that there are eight converbs in the language (see Table 1 on the next page). These can formally be classified into two or three types based on their base: converbs based on the basic stem and those based on the perfective stem. Furthermore, the former has two subtypes, so-called conjunctive participle and infinitive-based forms. Converbs based on the basic stem can freely switch the subject reference between the joining clauses, while converbs based on the perfective stem can only be used when the subjects in the chained clauses are not the same, (1). Functionally, Domaaki converbs make several different relationships, including finalis, causal, simultaneous, manner, and just relative temporal (anterior) nexuses.

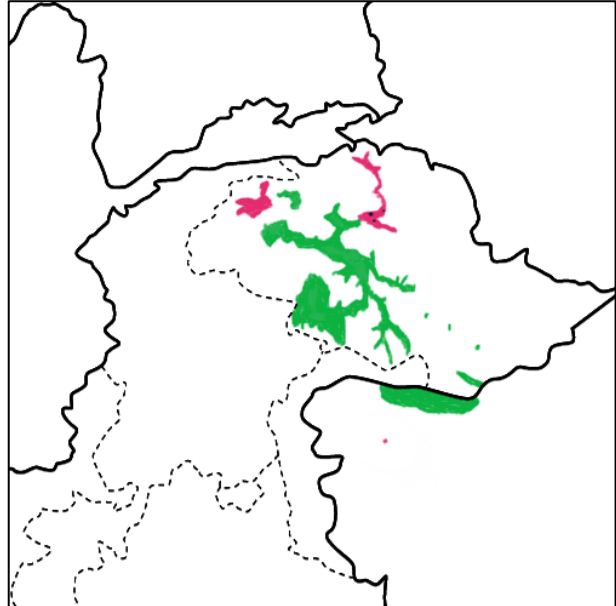


Figure 1: Distribution Map of Domaaki (navy dots), Burushaski (pink), and Shina (green) in and around northern Pakistan

(1)	<i>u</i>	<i>giásmo</i>	<i>(pačhi)</i>	<i>e</i>	<i>aayá</i>
	<i>ú</i>	<i>gi-as-mo</i>	<i>pačhi</i>	<i>é</i>	<i>aai-a</i>
	<i>I</i>	<i>go:PFV-OBL-ABL</i>	<i>behind</i>	<i>ANAPH.M</i>	<i>come:PFV-3SG.M:INTR</i>

‘(After) I left, and then he came’ [with an Anterior Sequential converb ₈]

While Burushaski has approximately twenty converbs, Domaaki has fewer. This is because Domaaki lacks verbal participles proper, thus limiting the potential for the regular production of many non-finite (converbal) forms compared to Burushaski. Separately, the neighbouring Indo-Aryan language, Shina [scl], has several converbs derived with morphologically unrelated and independent suffixes. I will present that Domaaki typologically exhibits a morphological orientation similar to Burushaski, not Shina.

Table 1: Eight converbal forms in Domaaki [S(witch).R(eference)., F(ree subjects), D(ifferent subjects)]

BASE	COMPOSITION	LOR. 1939	Вай. 2011	MY PRESENTATION	S.R.
Neutral stem	STEM- <i>ii</i>	YES (F)	YES	‘Conjunctive participle’ ₁	F
	STEM- <i>ina</i> -OBL		YES	Manner converb ₂	F
	STEM- <i>ina</i> -OBL-DAT	YES		Finalis converb ₃	F
	STEM- <i>ina</i> -OBL-ABL	YES		Causal converb ₄	F
	STEM- <i>ina</i> -ELA	YES	YES	Causal converb ₅	F
	STEM- <i>inee</i>		YES	Simultaneous converb ₆	F
	(STEM- <i>ineebel</i>)		YES (D)	[certainly refers to the following one]	
	(STEM- <i>inee</i> + <i>bulo</i>)			(Simultaneous converb ₆ + particle)	
	STEM-INS	?		- N/A -	
Perfective stem	STEM:PFV-OBL	?		Anterior Perfect converb ₇	D
	STEM:PFV-OBL-ABL			Anterior Sequential converb ₈	D
	(STEM:PFV- <i>in</i>)	YES		(finite verb)	

Abbreviations:

ABL	ablative	ELA	elative	OBL	oblique
ANAPH	anaphoric	INTR	intransitive	PFV	perfective
DAT	dative	M	masculine	SG	singular

References:

- Lorimer, D.L.R. 1939. *The Dumāki Language: Outlines of the Speech of the Doma, or Bēricho, of Hunza*. Nijmegen: Dekker & van de Vegt N.V.
- Вайнрайх, М. 2011. Домааки язык. In Т.И. Оранская, Ю.В. Мазурова, А.А. Кибрик, Л.И. Куликов, and А.Ю. Русаков (eds.) *Языки Мира: Новые индоарийские языки*. Москва: Academia. 165–94. (Weinreich, M. 2011. Domaaki language. In Tatiana I. Oranskaja, Yulia V. Mazurova, Andrej A. Kibrik, Leonid I. Kulikov, and Aleksandr Y. Rusakov (eds.) *Languages of the World: New Indo-Aryan languages*, 165–194. Moscow: Academia. [in Russian])