South Asian Languages Analysis Roundtable -39

Adam Mickiewicz University, Poland

(29 September – 1 October)

Faculty of Modern Languages and Literatures, Institute of Oriental Studies

Venue: Collegium Novum, Al. Niepodległości 4, 61-713 Poznań

BOOK OF ABSTRACT

Programme

A SWITCH LANGUAGE PATTERN IN RAMPURHAT BANGLA

Abstract

This paper investigates the dialectal variations in the stress patterns of disyllabic words in Standard Colloquial Bangla (SCB) and Rampurhat Bangla (RB), spoken in Rampurhat region of Birbhum, West Bengal (India). While Hayes and Lahiri (1991) assert an inviolable rule of initial syllable stress in Bangla, Shaw (1984) suggests stress shifts to the second syllable whenever it is heavier. Bhattasali (2016) attributes these discrepancies to dialectal variations. SCB exhibits a predominantly quantity-insensitive stress pattern, whereas RB displays both quantity and quality sensitive stress pattern. RB employs phonological strategies like vowel lowering, vowel shortening in the first syllable, and vowel lengthening in the second syllable to facilitate stress shifting. This suggests that RB has either shifted from trochaic to iambic stress pattern entirely or functions as a 'switch language' Houghton (2013). In the case of RB, this can be a result of language contact with several indigenous languages spoken in the neighbouring regions. The primary data for the study was collected from 10 native speakers of RB using structured questionnaires and interviews. A sample data has been presented below.

CV.CV

SCB	RB	Gloss
/ˈburo/	[buˈrɔ]	‘old’
/ˈtʃʰele/	[tʃʰeˈlæ]	‘boy’
/ˈmulo/	[muˈlɔ]	‘radish’

CV.CVC

SCB	RB	Gloss
/ˈbiral/	[biˈræl]	‘cat’
/ˈɖokan/	[ɖuˈkan]	‘shop’
/ˈkoɖal/	[kuˈɖal]	‘an axe’

CVC.CV

SCB	RB	Gloss
/ˈdʒʰapʃa/	[ˈdʒʰapʃæ]	‘blurred’
/ˈʃondhe/	[ˈʃondhæ]	‘evening’
/ˈʃorʃe/	[ˈʃorʃæ]	‘mustard’

An Optimality Theoretic analysis (Prince and Smolensky, 1993) has been provided to explain the dialectal differences. Further, an acoustic analysis of the stress pattern of RB has also been conducted for the study. This was done by speech segmentation and sound analysis using the PRAAT software.

Alignment, Subjecthood and Transitivity Prominence in Indo-Iranian and Indo-European

Eystein Dahl & Tania Ahmad, Adam Mickiewicz University, Poznań

eystein.dahl@amu.edu.pl, tanya.ahmad@amu.edu.pl

This paper explores the relationship between alignment, subjecthood features and transitivity prominence in Indo-Iranian as reflected in Vedic Sanskrit compared with the other branches of the Indo-European family. In the present context, alignment refers to the morphosyntactic realization of core arguments. Subjecthood is a cover term for the morphosyntactic properties of the argument referred to as subject, which in this context is defined as the generalized syntactically privileged argument, a notion borrowed from RRG (cf. Van Valin and Lapolla 1997, Van Valin 2005). Transitivity prominence is understood as the degree to which predicates in a language select the pattern characteristic of core transitive verbs (cf. e,g, Haspelmath 2015, Creissels 2018). These three morphosyntactic dimensions are taken to constitute the argument realization system of a language. The empirical data are drawn from the oldest attested stages of representative languages of each of the branches of Indo-European by means of a three-tiered typological questionnaire, which is designed to capture fine-grained differences in each of the three dimensions. The subjecthood properties represent a subset of the ones identified by Falk (2006). Relative transitivity prominence is measured by means of the predicate list in Creissels (2018b), which comprises 30 verb meanings, each of which is assigned a score based on three variables (voice marking, case marking of first argument, case marking of second argument), the scores being summarized for each language, cf. Table 4 below. Taken together, this questionnaire enables a multivariate comparative analysis of the argument realization systems in the languages under scrutiny (e.g., Bickel 2015). It is unclear, however, whether the three morphosyntactic dimensions are diachronically interrelated, so that changes in one dimension may cause changes in another dimension, or not.

Assuming that each of the variables in each of the dimensions represent independent features, we aim to establish whether there are any observable correlations between features. To this aim, we submit the combined data to Principal Components Analysis (PCA), which allows for examining the correlations between variables, cf. Figures 1a and 1b (Le et al. 2008). These figures indicate that certain subjecthood properties tend to cluster, e.g., control and raising constructions, but are weakly associated with certain alignment features, e.g., ergative agreement. Figure 2 plots the resultant clustering of the languages in our sample, indicating that there is a fundamental distinction between languages where control and raising represent subject properties, notably Latin, Greek and Lithuanian (in Hittite, only raising) and languages where they don't (Le et al. 2008). Interestingly, neutral

agreement, which is found in Greek, Lithuanian and Hittite, seems to play an almost equally important role in the distribution of the languages. Moreover, raising and control constructions are not subject properties in the languages showing tense/aspect-based split ergative alignment, i.e., Vedic Sanskrit and Classical Armenian. These brief observations are indicative of considerable variation within the languages under consideration and may seem to suggest that language-specific developments within one of the three abovementioned dimensions may condition or restrict developments in the other dimensions.

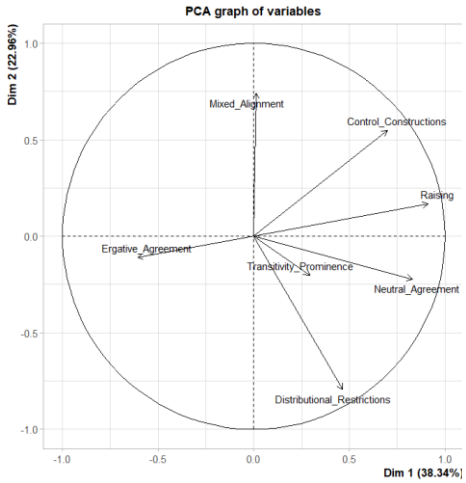


Figure 1a.

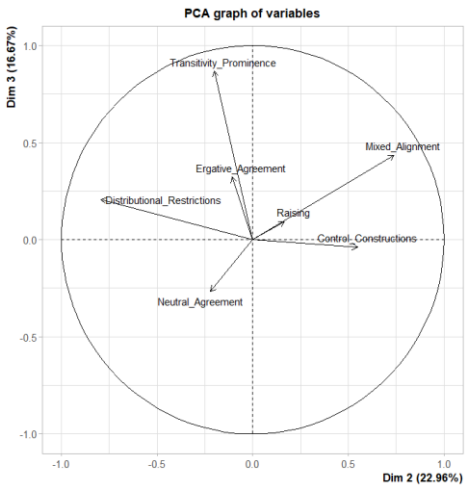


Figure 1b.

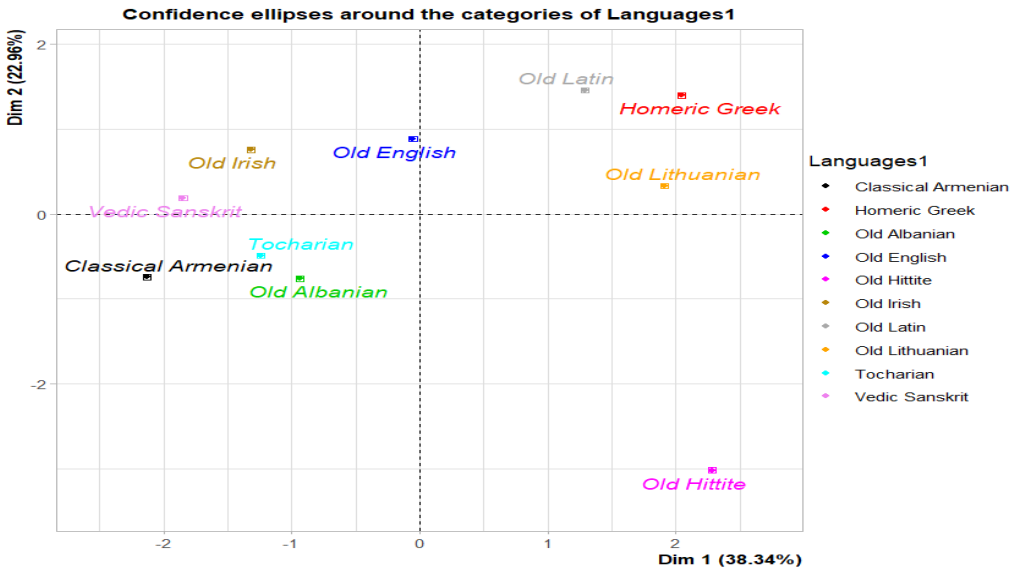


Figure 2.

References:

Bickel, Balthasar. 2015. Distributional Typology: Statistical Inquiries into the Dynamics of Linguistic Diversity. In Bernd Heine and Heiko Narrog (eds.) *The Oxford Handbook of Linguistic Analysis (2nd Edition.* Oxford: OUP, 901-924.

Cotticelli, Paola and Eystein Dahl. 2022. 'Argument alternation and non-canonical argument marking in some archaic Indo-European languages.' In Eystein Dahl (ed.) *Alignment and Alignment Change in the Indo-European Family*. Oxford: OUP.

Creissels, Denis. 2018a. The Obligatory Coding Principle in diachronic perspective. In Sonia Cristofaro and Fernando Zúñiga (Eds.). 2018. *Typological Hierarchies in Synchrony and Diachrony*. Amsterdam: Benjamins. 59–110.

Creissels, Denis. 2018b. Transitivity prominence in typological perspective: the case of Basque. In Joseba A. Lakarra and Blanca Urgell (eds.) *Studia Philologica et Diachronica in honorem Joakin Gorrotxategi Vasconica et Aquitanica*. Special issue of *Anuario del Seminario de Filología Vasca «Julio de Urquijo» International Journal of Basque Linguistics and Philology LII*: 1-2 (2018), 175-187.

Dahl, Eystein. 2021. 'Pathways to split ergativity: The rise of ergative alignment in Anatolian and Indo-Aryan.' In *Diachronica*, 38:3 (2021), 413-456.

Dahl, Eystein. 2022. 'Alignment in Proto-Indo-European'. In Eystein Dahl (ed.) *Alignment and Alignment Change in the Indo-European Family*. Oxford: OUP.

Falk, Yehuda N. 2006. *Subjects and Universal Grammar. An Explanatory Theory.* Cambridge: CUP.

Haspelmath, Martin. 2015. 'Transitivity Prominence.' In Andrej L. Malchukov & Bernard Comrie (eds.), Valency classes in the world's languages: A comparative handbook, vol. 1. Berlin: De Gruyter Mouton, 131–147.

Le, Sebastien, Julie Josse and Francois Husson. 2008. FactoMineR: An R Package for Multivariate Analysis. In *Journal of Statistical Software*, 25(1), 1-18. 10.18637/jss.v025.i01

Van Valin, Robert D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: CUP.

Van Valin, Robert D. and Randy Lapolla. 1997. *Syntax. Structure, meaning and function*. Cambridge: CUP.

Diachronic stability and borrowability across Shinaic lects in High Asia

Henrik Liljegren (Stockholm University)

Shina is the geographically most widely dispersed phylogenetic cluster among the many Indo-Aryan languages represented in the mountainous Hindu Kush-Karakorum region. At least in the linguistic sense, these are—to borrow the words of Strand (2001: 253)—the descendant of the more adventurous Āryas who in ancient times left the plains, following the course of the Indus into the rugged terrain of Kohistan, to establish a new presence in the Chilas area, from where they subsequently expanded in all directions (Jettmar 1975: 20–21). Most of those varieties are in fact languages in their own right as of today, spoken in non-adjacent pockets, from Afghanistan's Kunar Valley in the west, across several watersheds running parallel with the Indus in northern Pakistan, to Indian Ladakh in the east. In the present study, nine Shinaic lects evenly distributed across this region, are investigated (see Figure 1). Largely based on recently collected primary materials, a comprehensive comparison of lexical as well as structural data (including basic vocabulary, pronouns, morphological paradigms, alignment patterns, segmental phonology, grammatical gender, numerals and kinship) has been conducted, revealing a fascinating adaptability—and resilience—of Shina firmly planted in their different areal-typological and sociolinguistic environments (Liljegren 2023).

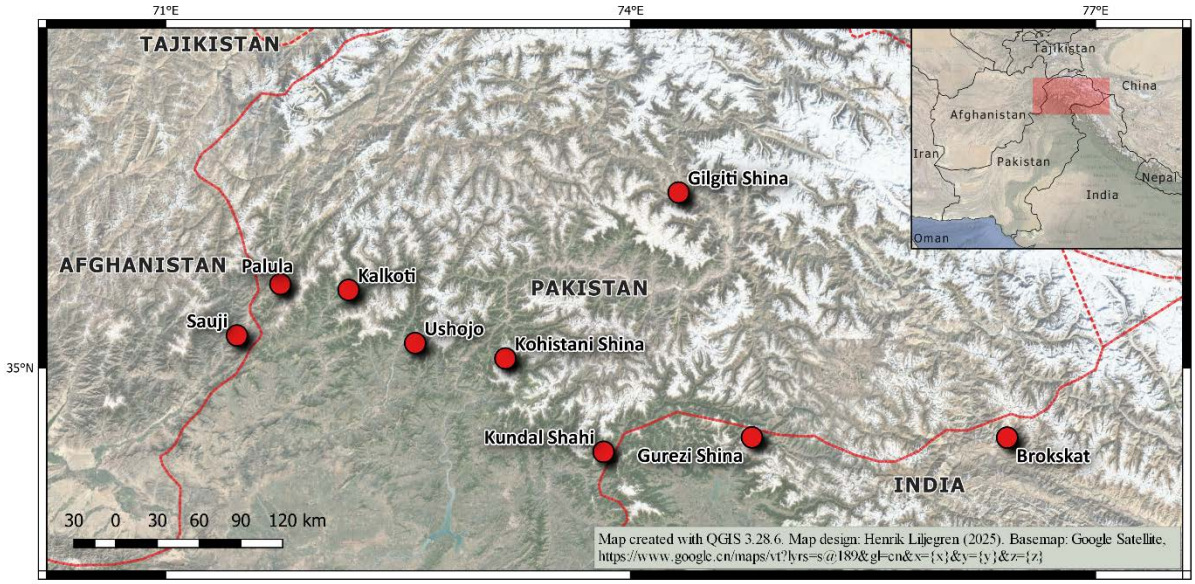


Figure 1: Map displaying the locations of the nine sample Shina lects in the Hindu Kush-Karakorum region.

Borrowability and convergence with surrounding and locally influential non-Shinaic languages is particularly tangible in the lexicon, affecting content words more than function words, nouns rather than verbs, and concepts related to certain semantic fields, such as religion and belief, to a much higher degree than more universal conceptualizations, thus confirming many previous claims (Tadmor 2009: 59–65; Matras 2007: 61–62). The two examples, (1) and (2), representing the two geographical extremes of the east-west Shinaic districution (Sauji spoken in Kunar, Afghanistan, and Brokskat spoken in Ladakh, India), illustrate the rather dramatic drifting apart, to the extent that probably only a single inherited concept—*ǰandráw; žun* < *jantú* (Turner 1966: 5110)—has survived in both of the two parallel sentences. Other categories giving evidence of high borrowability and contact-induced change across the Shinaic cluster are e.g., syllable structures, demonstratives, discourse particles and agreement patterns.

(1) Sauji [SDG; savi1242]
pi-eé *ǰandraw-eé* *rateé* *šaank-eé* *di* *marg-ãã̆* *ṣaṭ* *dití*
boy-OBL snake-OBL on stick-OBL with death-GEN hit(F) give.PST.F
'The boy beat the snake with a stick.' (SDG-Val-FR:027)

(2) Brokskat [BKK; brok1247]
byo *gulu-wé* *žun-uží* *khuṭé-t*
boy stick-INS snake-LOC beat-PST
'The boy beat the snake with a stick.' (BKK-Val-TL:027)

Properties that on the other hand appear to be particularly stable and resilient are certain types of core vocabulary (e.g., nuclear verbs, such as 'do' with a preserved *th*-root in all of the lects, vis-à-vis the ubiquitous *k(ar)*-root in all non-Shinaic Indo-Aryan lects of the region), personal pronouns, an inherent sex-based two-gender system and word/constituent order, in all essential aspects confirming Nichol's large-scale findings (2003: 291–305) regarding the relative propensity for inheritance vs. acquisition.

Word count: 494 words (excluding references)

Jettmar, Karl. 1975. *Die Religionen des Hindukusch* (Die Religionen Der Menschheit; Bd. 4, 1). Stuttgart: W. Kohlhammer.

Liljegren, Henrik. 2023. The languages of Peristan through the lens of areal typology. In Alberto Cacopardo & Augusto Cacopardo (eds.), *Roots of Peristan: The pre-Islamic cultures of the Hindukush/Karakorum* (Serie Orientale Roma, Nuova Serie 37), 391–431. Rome: ISMEO - Associazione Internazionale di Studi sul Mediterraneo e l'Oriente.

Matras, Yaron. 2007. The borrowability of structural categories. In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective* (Empirical Approaches to Language Typology 38), 31–73. New York: Mouton de Gruyter.

Nichols, Johanna. 2003. Diversity and stability in language. In Brian D Joseph & Richard D Janda (eds.), *The handbook of historical linguistics*, 283–310. Malden, MA: Blackwell Pub.

Strand, Richard F. 2001. The tongues of Peristân. Appendix 1. In Alberto Cacopardo & Augusto Cacopardo (eds.), *Gates of Peristan: History, Religion and Society in the Hindu Kush* (Reports and Memoirs 5), 251–257. Rome: Istituto Italiano per l'Africa e l'Oriente (IsIAO).

Tadmor, Uri. 2009. Loanwords in the world's languages: findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the World's Languages, A Comparative Handbook*, 55–75. Berlin, Boston: De Gruyter Mouton.

Turner, Ralph Lilley. 1966. *A comparative dictionary of the Indo-Aryan languages*. Oxford: Oxford University Press.

Aiding linguistic analysis with the help of Artificial Intelligence

Introduction

Artificial Intelligence is widely used in the task of Natural Language Processing. Its computational capabilities allow it to model the language on various levels: structural, conceptual or even pragmatic. However, due to occasional inaccuracies and hallucinations, the output of AI cannot always be trusted.

As opposed to scenarios where an entire task is carried by AI, this paper presents the use of AI to aid the work of linguists. Humans and AI work side by side to optimize the work and achieve deeper research results.

Computer-aided analysis of Braj

The paper [Braj, 2017] presents results of the research on early New Indo-Aryan languages, involving traditional linguistic analysis aided by computational methods. For the purpose of annotation of early NIA corpus the IA tagger system was designed. The system allows for manual word-level text annotation on several levels, which is performed by linguists. Collected data serves both for direct linguistic analysis and for training and testing automatic classifiers.

As the recent research on early Braj is focused on specific verb forms – converbs – automatic converb detector for Braj was developed. The performance of the classifier allows for its use in a scenario, where converbs are automatically detected in a large unannotated text collection and then verified by linguists. This method facilitates the process of a multilayered linguistic analysis of any type of converbal constructions.

Text embeddings for official languages of India

Text embeddings are mathematical representations of words. These representations encode the meaning of individual words and are used to predict if two given words are semantically similar.

The paper [LLM, 2023] presents the research on using text embeddings in the task of aligning English texts with the texts in selected official languages of India. The goal of the task was to perform analysis of texts translated from English into the Indian languages. Embeddings allowed for automatic detection corresponding words between English and the target language. See below examples of alignment for English-Marathi and English-Kannada:

	मागील	महिन्यात	खेळलेल्या	खेळाची	आकडेवारी
Statistics					
of					
the					
games					
played					
during					
last					
month					

	ಸ್ಥಿತಿ	ಪಟ್ಟಿಯನ್ನು	ತೋರಿಸಬೇಕೆ	ಅಥವ	ಬೇಡವೆ
Whether					
or					
not					
to					
show					
the					
status					
bar					

The information was used by translators in the process of proofreading.

Conclusion

Properly designed AI solutions can improve the work of linguists not by replacing them but by working side by side an aiding their work.

References

[Braj, 2017] Jaworski Rafał Tomasz, Stroński Krzysztof: Experiments with Automatic and Semi-automatic Detection of Sparse Word Forms in Old Braj, W: Human Language Technology. Challenges for Computer Science and Linguistics 8th Language and Technology Conference, LTC 2017, Poznań, Poland, November 17–19, 2017, Revised Selected Papers / Vetulani Zygmunt, Paroubek Patrick, Kubis Marek (red.), Lecture Notes In Computer Science, vol. 12598, 2020, Springer, ISBN 978-3-030-66527-2, s. 119-130, DOI:10.1007/978-3-030-66527-2_9

[LLM, 2023] Zydroń, Andrzej, Rafał Tomasz Jaworski i Szymon Kaczmarek. 2023. „Large Language Models and the Future of the Localization Industry”. W Human Language Technologies as a Challenge for Computer Science and Linguistics – 2023, edited by Zygmunt Vetulani i Patrick Paroubek, 353–356.

Vowel Harmony and its Licensing: A phonological profile of Bangla spoken in Assam

Anindita Choudhury

Abstract

Bangla, also spelt as Bengali, is spoken by 7 million people in Assam (Census of India, 2011), a region linguistically situated at the northeastern periphery of the broader Bengali-speaking landscape. While a majority of the speakers identify themselves as Sylheti Bengalis, the region exhibits a variety of Bangla, referred to by some as *Purvanchal Bangla*. It is a blend of various dominant and smaller languages spoken in the region due to language contact. Bangla, both the standard form (SCB) in formal domains and Purvanchal Bangla in informal domains, are used.

This study investigates the phonological phenomenon of vowel harmony with respect to vowel height in the Standard Colloquial Bengali (SCB) spoken in Assam. Previous linguistic research on Standard Bangla, has documented patterns of vowel harmony, especially involving height features among adjacent syllables (Thompson, 2010; Bhattacharya, 2017). These processes often manifest through assimilation, wherein mid vowels tend to harmonize with high vowels in specific morphophonemic contexts.

However, field data collected from Bengali speakers of Assam suggest a significant deviation from these previously established patterns. In this variety, vowel harmony based on height is either absent or extremely restricted, even in environments where harmony would be expected based on previous research and comparative dialectological evidence. As shown in the example below, the general phenomenon of height raising of the first mid vowel /e/ to assimilate with the second high vowel /i/ is not observed. Instead, the phonological structure shows a greater degree of vowel stability and preservation of contrast, which may be attributed to contact dynamics with neighbouring languages that do not exhibit vowel harmony.

Bangla: lekʰa “to write”:

1	ami	likʰi (in both SCB and Assam variety)
	I	write-1SG.PRS
	I write.	
2	t̪umi	lekʰo (SCB)
	t̪umi	likʰo (Assam variety)
	you	write-2SG.IMP
	You write.	
3	o	lekʰe (SCB)
	o	likʰe (Assam variety)
	he/she	write-3SG.PRS
	He/She writes.	

The study draws on a corpus of recorded speech from speakers between 18 and 74 years old and acoustic analysis to support the observations. The findings challenge assumptions about the uniformity of vowel harmony processes across varieties of Bangla and highlight the importance of considering regional variations and contact-induced changes in phonological analysis. This work contributes to a more nuanced understanding of Eastern Indo-Aryan phonology and underscores the linguistic uniqueness of language varieties.

Keywords: Bangla; Vowel harmony; Vowel height; Language contact; Variations.

Title: The Three-Language Problem in Multilingual India: Corpus-Assisted Discourse Analysis of Digital Language Wars on X

Abstract:
The Three-Language Formula for language education, proposed by the Kothari Commission (1964–66), remains one of the most enduring pillars of India's language policy and planning. Despite recurrent disputes and center–state frictions, its influence continues through successive National Education Policies (1968, 1986, 2020) and National Curricular Frameworks (2005, 2023). In February 2025, renewed opposition emerged in Tamil Nadu against the three-language provision of NEP 2020, followed by Karnataka and Maharashtra, framing it as a threat to linguistic autonomy and federalism. Allegations of financial coercion and the misuse of Samagra Shiksha funds amplified debates at the intersection of language, education, and India's federal structure, fueling digital discourses around 'language wars.'

This study critically examines media and citizen discourse on X (formerly Twitter) as a site of ideological struggle and linguistic negotiation. Using Corpus-Assisted Discourse Studies (CADS), it identifies how micro-level linguistic features, collocation clusters, sentiment trajectories, and patterns of multilingual code-switching, map onto broader socio-political narratives of federal tensions, linguistic hierarchies, and regional identity. The analysis addresses two central questions: How do media and netizens on X construct, contest, and circulate discourses on the Three-Language Formula and language imposition? How do processes of iconization, erasure, and fractal recursivity (Irvine & Gal, 2000) organize the negotiation of linguistic authority and cultural authenticity?

Data were collected through Selenium WebDriver with Chrome from 11 national media handles, preprocessed in Python, and analyzed using GPT-4o for multilingual context-sensitive insights. Beyond computational modeling, the study draws on Fairclough's (1995) Critical Discourse Analysis and Irvine & Gal's (2000) semiotic processes of language ideology, adapting both frameworks to digital discourse. Findings demonstrate how micro-level discourse patterns instantiate macro-level socio-political narratives while also refining theoretical discussions of language ideology in digitally mediated contexts.

By combining computational rigor with critical qualitative interpretation, the research sheds light on how digital discourse shapes public understandings of language policy, revealing emerging ideological fault lines in Indian federalism and the complex dynamics of

multilingualism. The approach offers a transferable methodological model for studying political discourse at scale through hybrid local–cloud architectures.

Keywords: National Education Policy (NEP) 2020, Language Ideologies, Corpus-Assisted Discourse Studies (CADS), Sentiment Analysis, Social Media Discourse, Multilingual Analysis, Critical Discourse Analysis

References

Fairclough, N. (1995). *Critical Discourse Analysis: The Critical Study of Language*. Longman.

Flowerdew, L. (2013). Corpus-based discourse analysis. In *The Routledge handbook of discourse analysis* (pp. 174-188). Routledge.

Irvine, J. T., & Gal, S. (2000). Language ideology and linguistic differentiation. In P. V. Kroskrity (Ed.), *Regimes of Language: Ideologies, Polities, and Identities* (pp. 35–83). School of American Research Press.

Kothari Commission. (1966). *Report of the Education Commission 1964–66*. Government of India.

Ministry of Education. (1968, 1986, 2020). *National Education Policy*. Government of India.

NCERT. (2005, 2023). *National Curricular Framework*. Government of India.

Sloan, L., & Quan-Haase, A. (2016). The SAGE Handbook of social media research methods. SAGE Publications Ltd, https://doi.org/10.4135/9781473983847

Torregrosa J, D'Antonio-Maceiras S, Villar-Rodríguez G, Hussain A, Cambria E, Camacho D. A Mixed Approach for Aggressive Political Discourse Analysis on Twitter. Cognit Comput. 2023;15(2):440-465. doi: 10.1007/s12559-022-10048-w. Epub 2022 Aug 17. PMID: 35996741; PMCID: PMC9385242.

Banter in Bambaiyā: Solidarity, Identity and Urban Speech

This paper is part of a doctoral project on Bambaiyā, an urban speech that has developed in the metropolitan region of Mumbai, India. It attempts to investigate how Bambaiyā speakers understand and/or value a general aspect of disrespect or lack of politeness in their speech, and how they use this impolite speech to position themselves as part of a local, collective, and often highly sociolinguistically-defined speech community. Such 'banter' or mock politeness, that signals solidarity and camaraderie (Leech, *The pragmatics of politeness* 238) , especially in an urban speech such as Bambaiyā, can reveal complex communicative practices, identity formations, and the position that speakers assign to themselves and others in a society.

Bambaiyā, alternatively termed Bambaiyā Hindi or Tapori Hindi, draws from the linguistic reservoirs of Marathi, Hindi/Urdu, Gujarati, Dakkhani, English and other nearby forms of bazaar speech. It is inherently hybrid in its formulation, created out of the city's socio-economically diverse demographic, where low and middle classes, castes, and national as well as international immigrants are ghettoised and sustainably coexist.

The analyses presented in this paper are based on data collected primarily in the form of field interviews conducted in 2023 and 2024 with Bambaiyā speakers in various public places in Mumbai, as well as interviews conducted with writers and filmmakers who have studied or participated in the re-production and representation of Bambaiyā in films etc. The body of data also includes digital self-representations of Bambaiyā speakers, published online as short-form video content. The discussion in this paper is informed by theories of politeness, solidarity, camaraderie, and slang (Leech*;* Lakoff; Lodge; Snell; Slotta; Adams; Androutsopoulos; Bradac and Wisegarver; Kisielewska-Krysiuk and Wałaszewska)

In the data that I have collected so far, Bambaiyā speakers display, firstly, conscious language awareness and maintenance that is attached to a local, collective sociolinguistic identity. Secondly, they display an awareness (and at times enjoyment) of socially-shared violations of the norms of politeness. This mock impoliteness has a flavor of humour to it, which can add to this effect of violation. Consequently, much of the conversations in this speech take the form of banter. The purposeful usage of such a speech may thus foster solidarity. The linguistic features of this speech, such as a lack of T-V distinction, also display the solidarity semantic. The paper further discusses how settings where this banter is found are also culturally seen as masculine and that such urban speech is thus mediated by norms and differences in the way men and women speak and are perceived.

Given the many pressures of standard regional Indian languages and English, the usage and maintenance of Bambaiyā becomes an exercise in expressing personal and group identities, as well as the sociolinguistic politics associated with this identity. In other words, banter in an urban speech such as Bambaiyā serves a multi-function of solidarity and camaraderie in specific speech communities as well as an expression of insubordination against larger linguistic-political pressures.

References:

Adams, Michael. *Slang: The People's Poetry*. Oxford University Press, 2009, pdfcoffee.com/slang-the-peoplex27s-poetry-michael-adams-pdf-free.html.

Androutsopoulos, Jannis. "Extending the Concept of the (Socio)Linguistic Variable to Slang∗." *Mi a Szleng? ['what Is Slang?']*, edited by Tamás Kis, Kossuth Lajos University Press, 2000.

Bradac, James and Randall Wisegarver. "Ascribed Status, Lexical Diversity, and Accent: Determinants of Perceived Status, Solidarity, and Control of Speech Style." *Journal of Language and Social Psychology*, 3(4), 1984, 239-255. https://doi.org/10.1177/0261927X8400300401.

Kisielewska-Krysiuk, M., and E. Wałaszewska. *In the Mind and Across Minds: A Relevance-Theoretic Perspective on Communication and Translation*. Cambridge Scholars, 2010. EBSCO Ebook Academic Collection, books.google.de/books?id=olEaBwAAQBAJ.

Lakoff, Robin. *Context Counts: Papers on Language, Gender, and Power*. Edited by Laurel Sutton. Oxford University Press, 2017.

Leech, Geoffrey. *The Pragmatics of Politeness*. Oxford University Press, 2014.

---. *Principles of Pragmatics*. Routledge, 2016, www.taylorfrancis.com/books/mono/10.4324/9781315835976/principles-pragmatics-geoffrey-leech.

Lodge, Anthony. "The Pragmatics of Slang." *Web Journal of Modern Language Linguistics* 2, 1997, wjmll.ncl.ac.uk/issue02/lodge.htm.

Slotta, James. "Slang and the Semantic Sense of Identity." *Texas Linguistics Forum* 59, 2016, 119-128, repositories.lib.utexas.edu.

Snell, Julia. "Solidarity, Stance, and Class Identities." *Language in Society*, vol. 47, no. 5, 2018, pp. 665–91. https://doi.org/10.1017/S0047404518000970.

Cross-Linguistic perspectives on Indian English and Bangla Tag Questions

Atanu Saha & Sudipta Saha

Jadavpur University

This research focuses on the canonical and the invariant tag questions in Indian English and Bangla. Tag questions consist of an antecedent clause and a tag clause that is either the same or of reversed polarity. This paper intends to address two specific research questions.

i) Do canonical and invariant structures in Indian English and Bangla follow anaphoric dependency and elliptical structures?

ii) Do Indian language tag constructions influence Indian English tags?

To obtain data on tag questions in Indian English and Bangla, a separate and detailed questionnaire for both languages in the form of Google Forms was outlined to validate respondents' comprehension of tag questions. Apart from these questionnaires, Tatoeba, an open online, free, collaborative corpus, was also deployed to gather the data for tag questions.

Canonical tag questions in Indian English exhibit dependency between the antecedent and the tag clause for encoding the morpho-syntactic features like polarity, subject, tense, lexical operator/auxiliary, gender, and person, (1). They involve three syntactic operations: binding of long-distance anaphors, VP ellipsis, and morpho-syntactic feature dependency to generate the tag clause. The tag clause comprises a lexical operator and a pronominal.

1) John$_i$ is/ isnot playing, isn't he$_i$ / is he$_i$?

Canonical tag questions in Bangla adhere to a grammatical dependency between the stem and tag clauses. The tag clause is represented by the lexical verb and discourse particles like '*na*', '*t̪o*' and '*ki*'. However, they do not show an anaphoric dependency or long-distance binding since Bangla doesn't involve a pronominal in the tag clause (2). In (2), the lexical verb in the tag clause is dependent on the antecedent verb for phi-features. Bangla involves verb-stranding VP ellipsis, where the verb moves out of the ellipsis domain before ellipsis.

2) dipa boi-ta por-tʃʰ-e, por-tʃʰ-e t̪o / na / ki
Dipa book-CLS read-PRS-3P read-PRS-3P PRT
'Dipa had read the book, hadn't she?'

Invariant tag questions are called 'grammatically independent tag questions' since they do not show any grammatical dependency between the stem and tag clauses. In Indian English(IE),

they are represented by discourse particles like ‘right,’ ‘okay,’ and ‘na,’ among others, or by clauses like ‘am I right,’ etc, (3).

3) Everyone doesn’t keep their promise, right/ na/ nehi?

The invariant tag ‘*na* /no’ in (4) also expressed in (3) is considered an ‘indigenous tag’.

4) ɾuma kʰel-tʃʰ-e, ʈai ʈo/ʈai na/na
   ruma play-PROG-3P PRT PRT PRT
   Ruma is playing, isn’t it?

Following Lange (2012) we assume the invariant type na/nehi in IE has evolved due to the linguistic convergence of Indian languages like Bangla and Hindi. This phenomenon shows a prevalent contact situation of Indian English and Bangla.

References

Columbus, G. (2010). A comparative analysis of invariant tags in three varieties of English. *English World-Wide, 31*(3), 288–310.

Culicover, W. P. (1992). English tag questions in Universal Grammar. *Lingua 88*, 193-226.

Haldar, S. (2020). Bengali Verb-stranding VP Ellipsis and Ellipsis Identity Conditions.

Hoffman, S., Blass, A.-K., & Mukherjee, J. (2017). *Canonical Tag Questions in Asian Englishes.* New York: Oxford University Press.

Hsin, A.-L. C. (2016). An Analysis of Chinese Tag Questions with a cross-linguistic comparison to English Tags. *Taiwan Journal of Linguistics, 14*(1), 67-118.

Lange, C. (2012). *The Syntax of Spoken Indian English.* John Benjamins

Sailor, C. (2009). *Tagged for Deletion: A Typological Approach to VP Ellipsis in Tag Questions.* UCLA Master’s thesis.

Tottie, G., & Hoffman, S. (2009). Tag Questions in English: The First Century. *Journal of English Linguistics, 37*(2).

Distributional restrictions on Haryanvi retroflexes

Paroma Sanyal and Alokesh Majumder
Indian Institute of Technology Delhi

Abstract
This paper describes the distributional restrictions on the retroflex phonemes in Haryanvi a western Indo-Aryan language. Haryanvi has five retroflexes with four distinct manners of articulations. These are shown using Element-based characterization (Backley 2011):

	ʈ	ʈʰ	ɖ	ɖʰ	ɽ	ɽʰ	ɳ	ɭ
Place of articulation	|A|	|A|	|A|	|A|	|A|	|A|	|A|	|AU|
Manner of articulation	|ʔ|	|ʔH|	|ʔL|	|ʔLH|		|H|	|ʔL|	

Based on their distribution in five phonological environments, we propose that there are a total of six phonemes involving retroflex sounds. Of these four of them are in allophonic distribution with other sounds. This distributional pattern is summarized below.

	Word initial	Geminate	Word final	Intervocalic	Post nasal
	#**C**	**CC**	**C**#	V**C**V	N**C**
ʈ/ʈʰ	ʈoːɭ people ʈʰeːɭ interference	ɟoːʈʈa: buffalo kəʈʈʰe: together	baːʈ weight paːʈʰ chant	biʈoːɽa: storage hut uʈʰəɳ getting up	bənʈa: water pot
ɖ/ɖʰ	ɖaːk letter ɖʰuːŋga: waist	siɖɖi stairs ʈiɖɖaːɳa: nag			saːnɖ bull gaːnɖe sugarcane
ɽ/ɽʰ			baːɽ fence naːɽ neck	biʈoːɽa: storage hut saːɽʰu brother-in-law	
n/l	laːɭ red nuːɳ salt	ɟinna: ladder			
ɳ/ɭ			aːɭ wrestling gʰuːɳ termite	ɖoːɭa: water channel paːɳi: water	

The voiceless stops have no distributional restriction and can surface in all phonological contexts. These are also the consonants with maximum structural complexity as per GP2.0 (Pöchtrager 2006, 2024). The configuration of structural complexity in this theory is based on the idea that the more prototypical a segment is for onset or nucleus, larger would be the structural space it would occupy. Coronal voiceless stops being the most prototypical of consonants have the largest possible structure for onsets.

The voiced stops are in complementary distribution with the retroflex rhotic. Similarly, the retroflex nasal and lateral are also in allophonic configuration with their alveolar counterparts. Using the onset-rhyme constituent sequencing of Government Phonology (KLV1990) we observe that the rhotic, nasal and lateral retroflexes always need to be preceded by a nucleus. Further, there is a very strong preference for this preceding nucleus to have the element |A| in its constituency.

Since the place of articulation (PoA) for both vowels and consonants are described using the same factors, Element theory lends itself to establish theoretical correlations between vowels and consonants. The three elements |A|, |I| and |U|, embody the defining properties of the corner vowels in the cardinal space: /a/, /i/ and /u/. The PoA of Haryanvi consonants are depicted below using the same Elements.

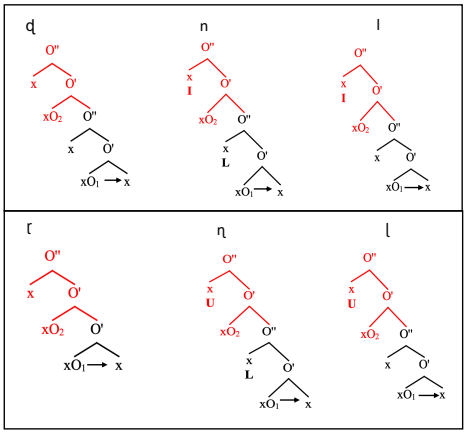
	Labial	Dental/ Alveolar	Retroflex	Palatal	Dorsal
	|U|	|I| or |A|	|A|	|I|	|U|
Stop |ʔ|	|U.ʔ| b	|I.ʔ| d̪	|A.ʔ| ɖ	|I.ʔ| ɟ	|U.ʔ| g
Rhotic |A|		|A.| r	|A| ɽ		
Nasal |ʔ.L|	|U.ʔ.L| m	|A.ʔ.L| n	|A.ʔ.L| ɳ		|U.ʔ.L| ŋ
Lateral |A.I|		|A.I| l	|A.I| ɭ		

In contrast to the other elements marking place of articulation, |I| and |U|, Pöchtrager 2024 and preceding work in GP2.0 argue that the element |A| is a structure rather than a melody. This would mean the retroflex consonants have larger structure compared to the palatal, labial or velar counterparts. The preceding vowel with |A| would also have more structure than those without |A|.

Like /a:/ the retroflexes in Haryanvi also readily cooccur with the short open vowel that is usually denoted as a schwa.

Consonantal			#C		C#	
Stop	|A.ʔ|	/ɖ/	/ɖəs/	string		
Rhotic	|A|	/ɽ/			/gaɖəɽ/	jackal
Nasal	|A.ʔ.L|	/ɳ/			/kirəɳ/	light ray
Lateral	|A.I|	/ɭ/			/bəɭ/	fire

Since schwa is an empty xN head without further structure or melody, we note that the environment that the retroflex prefers is not |A|, but rather it avoid being preceded by xN head annotated with melody (|I| or |U|) without any structure. Further, when the retroflex sounds are preceded by high vowels, there is a tendency to lower the vowel in the phonetic representation. For example, *ɖʰi:ɽ ~ ɖʰe:ɽ* 'stubborn'. Alternatively, when preceded by a short high vowel, they vowel loses its melody to become a schwa. For example, Hindi *nagin* becomes *nagəɳ* 'female cobra' in Haryanvi.



In conclusion, we have argued for the following structural configuration for the retroflex sounds that are in complementary distribution. Further, based on these structure we have shown why they appear in these particular phonological environments.

References

Cited References

Backley, Phillip. 2011. *An introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Kaye, Jonathan, Jean Lowenstamm and Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7(2). 193–231.

Pöchtrager, Markus A. 2006. *The structure of length*. Doctoral dissertation. University of Vienna, Vienna.

Pöchtrager, Markus A. 2024. Vowel reduction in European Portuguese and the removal of structure. *Isogloss* 10(3)/2. 1–22.

Domains Do Not Matter: Analysis of Borrowing Pattern Across Semantic Domains in Shabar Kharia

Raj Adhikary (University of Florida)

Overview

This study aims to investigate contact-induced lexical borrowing in Shabar Kharia (henceforth SK), a mixed tribal language spoken in Purulia, West Bengal. Based on fieldwork conducted in four villages, this study presents a compiled lexicon of over 1,000 words to analyze borrowing patterns from Bengali and Hindi.

Background

SK is an undocumented, endangered tribal language spoken in Purulia, India. Historically, SK was classified as a lower variety of the Bangla Manbhumi dialect (Grierson, 1927). However, SK also shows Austroasiatic features like locative variation based on animacy, alongside borrowing from Bangla and Hindi. Due to the variation in the grammar and lexicon, SK is considered a mixed language for this study. SK shows contact-induced changes across several semantic domains. For example, domains like kinship and environment remain relatively 'original', but domains like emotions, body parts, and household items show high borrowing rates. Variation is observed across speaker age, education, and gender. Here are some of the data-

1. **Semantic Field- Body Parts**

   [rɑdʒɑ] 'eyeball' (SK: older male speaker)

   Vs.

   [moni] 'eyeball' (used by younger male speakers; Bangla loan)

2. **Semantic Field- House and Home**

   [l̥oɽa] 'beam' (SK: middle-aged female speaker)

   Vs.

   [ʈ̪ʰamba] 'beam' (younger female speaker, Hindi loan)

Puzzle and Hypothesis

Now the question arises, why are some semantic fields more vulnerable to borrowing than others in SK, and what does this reveal about the contact situation? Domains with high social prestige or institutional exposure (e.g., education, administration) correlate with Indo-Aryan lexical borrowings. I hypothesize that not only domains associated with modernity, schooling, and institutional life, domains that usually resist borrowing are affected by the contact situation due to several reasons, including attitude and motivation.

Methodology

The data for this study were collected through fieldwork conducted in four villages of Purulia, West Bengal, involving five speakers from diverse demographic backgrounds (age, gender, and education). A wordlist of over 1,000 lexical items was compiled through elicitation sessions in free and controlled environments. The lexicon was categorized by semantic field and compared across speaker variables. In this study, 'SK native' refers to items historically retained from the

non-Indo-Aryan substrate, differentiated from recent Bengali/Hindi loans by etymological and phonological criteria. For each word, the lexical origin (SK native vs. borrowings) was identified through etymological and phonological comparison. Domains showing high interspeaker variation were analysed for correlation with external factors such as institutional exposure, literacy level, and language attitudes. Additional observations were made on phonological adaptation where present. The results were also compared with the LWT project. (Haspelmath, Tadmor, 2009)

Conclusion

This study demonstrates that lexical borrowing in Shabar Kharia (SK) is not uniformly distributed across semantic domains. Notably, domains such as 'The body' and 'The House', which the LWT project identifies as having low average borrowability scores (0.17 and 0.40, respectively), exhibit significant borrowing in SK. These patterns suggest that sociolinguistic factors, such as age, education, and exposure to dominant languages, play a crucial role in facilitating borrowing, even in typically resistant domains.

Reference

- Grierson, G. A. (1927). *Linguistic Survey of India*
- Haspelmath, Martin & Tadmor, Uri (eds.) 2009. *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wold.clld.org, Accessed on 2024-05-31.)

Many Ways to Be - Copular Strategies in Kodava

Current literature treats the copula as a 'dummy' element, a purely formal device that serves as a hat-rack for verbal morphology and links a subject to its predicate without contributing semantic content of its own. Kodava, a little-studied South Dravidian language, has no single element that fulfils this role. Instead, it recruits different lexical items which, when pressed into service as copulas, undergo contextual semantic bleaching. Crucially, this bleaching happens only in certain contexts: the same items retain their full lexical meaning when used in non-copular clauses. The choice of item depends on both the tense of the clause and the syntactic category of the predicate.

Two verbs are central to an analysis of Kodava copular constructions. The first is the stative verb *ʊŋɖ*, an intransitive verb that takes an entity or a property and asserts its existence (1). The second is the dynamic or eventive verb *aa*. Used transitively, *aa* denotes a change from one state to another (2); used intransitively, it denotes a state coming into existence from nothing (3).

(1) ɪlli sʊmar sɔl̪ɛ ʊŋɖ
here many mosquitoes exist.PRS.3
'There are many mosquitoes here.'

(2) munɖ̪ ʈ̪ɪŋgəʈ̪-lɨ nanɨ ʈ̪əllənɛ api
three months.OBL-LOC 1Sg thin become.NPST.1Sg
'In three months I will become thin.'

(3) na-kɨ gaja a͡tʃi
1Sg-DAT injury become.PST.3
'I got injured.' (or I went from a state of having no injury to a state in which I have one)

With *ʊŋɖ* meaning 'to exist' and *aa* 'to become', Kodava cannot use either verb alone to express 'to be'. Instead, it creates a periphrastic construction from both, by marking *aa* for perfective aspect (*aɪʈ̪*), which presents the event of ‘becoming’ as a completed whole, and marking tense on *ʊŋɖ,* the default tense auxiliary in the language. This combination of perfective morphology on the main verb along with a tense auxiliary presents the resulting event (of 'having become') as a state; *aɪʈ̪ ʊŋɖ* thus integrates eventive and existential semantics to realize the copular function (4).

(4) əʊə̃ dɔkʈ̪ər aɪʈ̪ ɪnd͡ʒəʈ̪ʈ̪
3MSg doctor become.PFV exist.PST.3 (AUX VERB)
'He was a doctor.'

For nominal predication in the present tense, Kodava adopts a different copular strategy. Changing the tense on the auxiliary in (4) to the present does not yield ‘He is a doctor.’ Instead, the unmarked strategy to express nominal predication in the present tense is a zero copula (5). An overt copula appears in the present tense only in contexts where an additional nuance is intended (6).

(5) əʊə̃ dɔkʈər
3MSg doctor
'He is a doctor.'

(6) əʊə̃ a hɔspiʈəl-lɨ dɔkʈər aɪ̯̱ ʊɳɖ
3MSg that hospital-LOC doctor become.PFV exist.PRS.3 (AUX)
'He has become/is functioning as a doctor in that hospital.'

But Kodava doesn't use the zero copula for nominal predication in the present tense alone. The language has highly productive strategies for deriving nouns from verbs; ex. the verb *kamba* 'see' can be nominalized into: *kəɳɖəʊə̃* 'the one I saw' (using the past stem), *kambə̃ʊə̃* 'the one I will see/the one I habitually see' (non-past stem), and *kəɳɖəɳɖ ʊllə̃ʊə̃* 'the one I am seeing right now' (progressive form). Because these nominalized verbs encode tense information, Kodava copular clauses may receive a past tense interpretation even when no copula is present (7).

(7) lɪft-lɨ kəɳɖ-əʊə̃ suruʈ̱-ra kɛlsa ʈ̱ənɖ̱-əʊə̃
lift-LOC see.PST-3MSg first.OBL-GEN job give.PST-3MSg
'The man i met in the lift was the man who gave me my first job.'

We can in fact argue that since copular clauses encode states, tense only becomes semantically relevant when the speaker intends to indicate that a past state no longer holds or that a future state is not yet in effect.

The story develops further when we turn to adjectives. Unlike nominal predicates, adjectival predicates in Kodava cannot stand without a copula. Since the unmarked strategy for nominal predication is a zero copula, the appearance of an overt copula in the present tense—namely *aɪ̯̱ ʊɳɖ*—carries with it a change-of-state interpretation, contributed by the main verb *aa* 'become.' As a result, the purely relational function of linking (with no independent semantic contribution) is now assumed instead by the verb *ʊɳɖ* when the predicate is adjectival (cf. (8) vs. (9)).

(8) əʊə̃ ʈ̱əllənɛ ʊɳɖ
3MSg thin exist.PRS.3
'He is thin.'

(9) əʊə̃ ʈ̱əllənɛ aɪ̯̱ ʊɳɖ
3MSg thin become.PFV exist.PRS.3
'He has become thin.'

Kodava thus illustrates that copular function can be distributed rather than realized through a single semantically empty item. The language uses a cluster of strategies to link subject with predicate - zero copula, *ʊɳɖ*, and *aɪ̯̱ ʊɳɖ*.

References

Arche, Maria J & Fábregas, Antonio & Marín, Rafael. (2019). Main questions in the study of copulas Categories, structures, and operations. 10.1093/oso/9780198829850.003.0001.

Balakrishnan, R. (1977). *A grammar of Kodagu* (1st ed.). Annamalai nagar: Annamalai University.

Burrow, T., and M. B. Emeneau. *A Dravidian etymological dictionary.* 2nd ed. Oxford [Oxfordshire]: Clarendon Press, 1984.

den Dikken, M., & O'Neill, T. (2017, March 29). Copular Constructions in Syntax. *Oxford Research Encyclopedia of Linguistics.* Retrieved 16 Jun. 2024, from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-137.

Ebert, Karen. 1996. Kodava. (Languages of the world/Materials 104). München/Newcastle: Lincom Europa.

Garman,M.A.G.(1973). *A grammar of the Coorg language.* University of Edinburgh doctoral dissertation.

Government of India. (2011). *Census of India 2011: Provisional Population Totals.* Office of the Registrar General & Census Commissioner, India.

Krishnamurti, Bhadriraju. (2003). *The Dravidian Languages (Cambridge Language Surveys).* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486876

Mohanan, T., & Mohanan, K. P. (1999, May). Two forms of BE in Malayalam. In *Proceedings of the LFG99 Conference. Stanford, CA: CSLI Publications. University of Manchester, Manchester, UK*.

Moro, A. (2000). *Dynamic antisymmetry*. Cambridge, MA: MIT.

Pustet, R. (06 2003). *Copulas: Universals in the Categorization of the Lexicon*. doi:10.1093/acprof:oso/9780199258505.001.0001

Stassen, L. (1997). Intransitive Predication. Oxford: Clarendon.

Zvelebil, K. (2017). *A Sketch of Comparative Dravidian Morphology: Part One* (Vol. 180). Walter de Gruyter GmbH & Co KG

Multilingual policy ideals, monolingual tendency and bilingual practice in education: Policy actors agency paradox in Nepal

Prem Prasad Poudel, PhD
Department of English Education,
Mahendra Ratna Campus, Tribhuvan University, Kathmandu, Nepal.

Multilingual education has reemerged as a new educational reform agenda in a global scale as a part of social inclusion, especially in the multilingual educational contexts (Spolsky, 2004; Hult, 2012). This agenda has been driven by diverse aspirations relating to linguistic social justice. The exponential growth of English medium instruction has further exacerbated the issue of multilingual education, demonstrating emergence of monolingual tendency (Choi & Poudel, 2024; Poudel & Costley, 2023). This paper draws on the case of Nepal, a multilingual and multicultural country in South Asia. In Nepal, although the policy structure that evolved as an outcome of a democratic political process provides sufficient space for multilingualism to be adopted at all levels of the governance and education, in practice primarily bilingualism continues. What is yet to be known is why the multilingual education policy has not been realized in schools. To address this issue, the study gathered data from a census survey with local level education officials, in-depth interviews with head teachers and teachers from 30 basic schools across seven provinces of Nepal. Drawing on the theory of policy enactment, this paper analyzed the policy actors (education officials, head teachers and teachers) agency; constraints and enablements in realizing the long-standing commitment of multilingual education in Nepal. The findings indicate that monolingual tendency (towards English or Nepali) continues to reduce the multilingual ideals to bilingual practice. Further to this, enactment of multilingual policies resulted in bilingual practice, further strengthening the dominant languages rather than local ethnic/indigenous languages. The findings also link to critical language policy research appealing to seriously consider the context in theorising equitable language policy. The paper further presents some implications for future work in both policy and practice in relation to bi/multilingual education, which could be of interest to the bi/multilingual polities globally.

Keywords

Multilingualism, monolingualism, policy enactment, policy actors, Nepal

References

Choi, T. H., & Poudel, P. P. (2024). Enactment of English medium instruction in under-resourced educational contexts: A case of multilingual public secondary schools in Nepal. *System*, *121*, 103223.

Hult, F. M. (2012). English as a transcultural language in Swedish policy and practice. *Tesol Quarterly*, *46*(2), 230-257.

Poudel, P. P., & Costley, T. (2023). *Understanding the impact of languages and language policies on children's learning outcomes in Nepal*. UNICEF.

Spolsky, B. (2004). *Language policy*. Cambridge university press.

Null Elements in Mewati

Miray Gökkaya
University of Konstanz

This study investigates the distribution of argument omission in Mewati by adopting the discourse functions proposed for Urdu by Butt and King (1997) within Lexical Functional Grammar (LFG) Framework. Before turning our attention to discourse functional notions, we first explore syntactic notions such as word order, agreement, and case following from traditional pro-drop theories (Huang, 1984; Rizzi, 1982; 1986; Taraldsen, 1978).

Mewati belongs to the Indo-Aryan language family and it is closely related to Northeastern Rajasthani (Fatihi, 2005). It is spoken in the Mewat region in India and in Pakistan. The data presented here is collected from a native speaker of Mewati from Pakistan.

In Mewati, both the subject (1) and the object (2) can be omitted in the right discourse configurations.

(1) Speaker A: mariam=ne kal k^hanoo k^ha-yo ho?
mariam=ERG yesterday food.M eat-PST.3SG.M be.PST.SG
'had Mariam eaten food yesterday?'
Speaker B: hã, (∅) kal k^hanoo k^ha-yo ho.
yes (∅) yesterday food.M eat-PST.3SG.M be.PST.SG
'yes, (∅) had eaten food yesterday.'

(2) Speaker A: seb kis=ne k^ha-yo?
apple.M who=ERG eat-PST.3SG.M
'who ate the apple?'
Speaker B: mariam=ne k^ha-yo (∅).
mariam=ERG eat-PST.3SG.M (∅)
'mariam ate (∅).'

Butt and King (1997) propose four discourse functions: topic (T), focus (F), backgrounded (BI) and completive information (CI). Topic is [-new] and [+prominent], focus is [+new] and [+prominent], backgrounded information is [-new] and [-prominent], completive information is [+new] and [-prominent]. According to this analysis, only continuing topic and backgrounded information can be omitted. The implementation of Butt and King (1997)'s analysis to Mewati is presented in (3).

(3) Speaker A: [mariam=ne]$_T$ [seb]$_F$ k^ha li-o ho?
mariam=ERG apple.M eat take-PST be.PST.SG.M
'did [mariam]$_T$ eat [the apple]$_F$?'

Speaker B: hã, [(Ø)]$_{Cont.Topic}$ k^{h}a li-o ho [(Ø)]$_{BI}$.
yes (Ø) eat take-PST be.PST.SG.M (Ø)
'yes, [(Ø)]$_{Cont.Topic}$ ate [(Ø)]$_{BI}$.'

As *mariam* 'Mariam' is in the topic position in the question, it is interpreted as the continuing topic in the answer, s*eb* 'apple' is interpreted as the backgrounded information. This observation for Mewati is in alignment with Butt and King (1997)'s analysis for Urdu. However, it cannot account for the data in (4) where the Continuing Topic cannot be omitted unless the question is a polar question as in (5).

(4) Speaker A: What happened to [Mariam]$_{T}$?
Speaker B: *[(Ø)] $_{Cont.Topic}$ ulti ayi.
(Ø) vomit come.PST.3SG
'[(Ø)]$_{Cont.Topic}$ vomited.'

(5) Speaker A: [mariam-e]$_{T}$ ulti ayi?
mariam-DAT vomit come.PST.3SG
'did mariam vomit?
Speaker B: hã, [(Ø)]$_{Cont.Topic}$ ulti ayi.
yes (Ø) vomit come.PST.3SG
'yes, [(Ø)]$_{Cont.Topic}$ vomited.'

The observations in (4) and (5) show that there are different factors that affect argument omission in Mewati in comparison to Urdu. Hence, these factors require further exploration in Mewati.

References

Butt, Miriam. & King, Tracy. (1997). Null Elements in Discourse Structure. In *Papers from the NULLS Seminar*, ed. K.V. Subbarao. Delhi: Moti Lal Banarsi Das.

Fatihi, A.R. (2005). The economy of articulation in Mewati Phonology. In *Language in India*. Volume 5

Huang, C.-T. James. (1984). On the Distribution and Reference of Empty Pronouns. Linguistic Inquiry 15(4):531–574

Rizzi, Luigi. (1982) Issues in Italian Syntax. Foris, Dordrecht

Rizzi, Luigi. 1986. Null Objects in Italian and Theory of pro. Linguistic Inquiry 17(3):501–557

Taraldsen, K.T. (1978). On the Nominative Island Condition, vacuous application and the that-t filter. Bloomington: Indiana University Linguistics Club.

Realization of glides in Bangla: An Element Theory analysis

Paroma Sanyal
Indian Institute of Technology Delhi

Abstract
The lexicon of Indo-Aryan languages is replete with glides /j/ and /w/ which fail to appear in the phonetic uttarance of words in Bangla. However, these missing glides are ubiquitous in the orthography and can surface as optional epenthetic consonants in contexts of vowel hiatus. Further, based on the phonological environment, there are five distinct strategies to modify the structure of an underlying glide in the surface representation. In this paper I build on the theoretical insights from Government Phonology (KLV 1990), Element Theory (Backlet 2011) and GP2.0 (Pöchtrager 2006, 2021, 2024, 2025) to explore the structural restrictions that underlie these five glide modification strategies in Bangla.
Building on the idea that the amount of phonological structure associated with a segment is directly proportional to the degree to which it contains the prototypical property of an onset or nucleus, I propose that Bangla lacks glides for the same reason it lacks minimal vowels like schwa in its phonetic realization. Glides being least consonantal and schwa being least sonorant have the smallest possible phonological structure for an onset and nucleus. I show that the minimal structure of both an onset and nucleus in Bangla must have at least one unannotated x and call it the Minimal Structure Parameter. The glides, although present in the lexical representation, are unable to surface, being sub-minimal in phonological structure. The five phonological strategies to transform underlying sub-minimal glides into utterable surface representation have been summarized below.

Strategy	Environment	Cognate comparison with Hindi					Analysis
		Hindi		Bangla		Gloss	
Fortition	Initial onset	युग	jug	যুগ	ɟu:g	age	Increase structure
		वायु	wa:ju	বায়ু	ba:ju:	wind	
Coalescence	Initial Onset cluster	त्याग	t̪ja:g	ত্যাগ	t̪ɛg	sacrifice	Delink glide onset. Move melody |I| to N
		ज्ञान	ɡja:n	জ্ঞান	gɛn	knowledge	
		व्यायाम	wja:.ja:m	ব্যায়াম	bɛm	exercise	
Gemination	Medial Onset Cluster	योग्य	jo.ɡjə	যোগ্য	ɟog.go	capable	Delink glide onset. Link the remnant onset to R.
		यज्ञ	jə.ɡjəh	যজ্ঞ	ɟog.go	yagna	
		विश्व	wi.ʂwə	বিশ্ব	biʂ.ʂo	world	
		गद्य	gə.d̪jə	গদ্য	god̪.d̪o	prose	
Lowering	Word finally	व्यय	wjəj	ব্যয়	bɛe	expenditure	Delink glide onset. Move |I| to N
		आय	a:j	আয়	a:e	earning	
Deletion	Initial onset cluster	स्वाद	swa:d̪	স্বাদ	ʂa:d̪	taste	Delink glide onset.
		स्वार्थ	swa:rt̪^{h}	স্বার্থ	ʂa:rt̪^{h}o	self interest	

In contrast to these five contexts where the underlying lexical glide fails to surface, they can optionally surface in intervocalic contexts.

বিয়ে	bije	bi.e	marriage
জোয়ার	ɟo.war	ɟo.ar	high tide

Further, the underlying lexical glide that appears in the morpheme initial position of derivational morphemes, for instance the nominalizer [-war] produces multiple pronunciation options.

Verb + [-war]	Glide	Stop	Ø	Geminate		In slow speech or contexts with syntactic focus, the morpheme initial glide always becomes a stop.
[ɟa] + [-war]	ɟa**w**ar	ɟa**b**ar			go	
[nam] + [-war]		nam**b**ar	namar	na**bb**ar	decend	
[d̪ɛk^{h}] + [-war]		d̪ɛk^{h}**b**ar	d̪ɛk^{h}ar		see	

Cited References

Backley, Phillip. 2011. *An introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Kaye, Jonathan, Jean Lowenstamm and Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology* 7(2). 193–231.

Pöchtrager, Markus A. 2006. *The structure of length*. Doctoral dissertation. University of Vienna, Vienna.

Pöchtrager, Markus A. 2021b. Towards a non-arbitrary account of affricates and affrication. *Glossa: A Journal of General Linguistics* 6(1). 1–31. https://doi.org/10.5334/gjgl.1116.

Pöchtrager, Markus A. 2024. Vowel reduction in European Portuguese and the removal of structure. *Isogloss* 10(3)/2. 1–22.

Pöchtrager, Markus A. 2025. A Neo-Jakobsonian merger of aperture, [−ATR], lowering, emphaticness, and retroflexion. *Acta Linguistica Academica* 72 (1), 20–45.

Accentual archaisms and innovations in the Indo-Aryan languages of the Hindukush–Karakoram

Samopriya Basu

Contrastive tones have been described for several New Indo-Aryan languages at the northwestern and eastern peripheries. They range from simple pitch accent systems as in Punjabi [Evans *et al.*, 2018], Chakma [Huziwara, 2019] and Shina [Degener, 2008] to more complex tonal inventories as described for languages like Gawri [Baart, 1997] and Khalkoti [Hultman, 2023]. However, even with better and more accurate descriptions of these and other varieties coming in over the past couple of decades, only the vague outlines of the origins of the tonal systems are currently known. In the planned talk, both using existing descriptions (e.g., Liljegren [2019]; Zoller [2005]; Huziwara [2019]) as well as newly gathered (or rechecked) data, the present author presents a diachronic picture of the varied developments that led to the modern systems.

As is well-known, in Punjabi and related dialects, tonal contrasts arise from loss of historical breathy voice, though with subtle differences across varieties. We have observed a similar (but hithertofore unreported) process in Kishtwari, where Punjabi-type tonogenesis (including initial devoicing of former breathy voiced stops) has occurred within the past century. This is also partly true in the far eastern Indo-Aryan languages (Chakma, Chittagonian-Rohingya, Sylheti), but aspiration and (historical) sibilants, even when now lost, are an additional source. Much more interesting is the situation in eastern Hindukush–Karakoram Indo-Aryan languages, viz., the Shinaic and "Kohistani" languages, to which most of the discussion will be dedicated.

In contrast to the languages mentioned above, in Shinaic languages like Shina [Degener, 2008] and Palula [Liljegren, 2016], and also in (non-Shinaic) Indus Kohistani [Zoller, 2005], the present-day phonemic contrast between a falling pitch contour (Gurez Shina *báal* 'hair' < Sanskrit *vā́laḥ*) and a rising contour (Gurez Shina *baál* '(male) child' < Sanskrit *bāláḥ*) on long vowels directly reflex the attested Vedic contrast between a barytone accentuation pattern against oxytone in inherited words that are not secondarily extended. This point is illustrated through several other examples from Shina, Palula and Indus Kohistani, proving in particular that loss of historical breathy voicing in Shina has nothing to do with tonogenesis unlike Punjabi or Chittagonian as has sometimes been presumed.

We also identify exact conditions under which the reflexed accent seemingly differs from a naïve historical expectation, e.g., originally unaccented long vowels attract accent from a preceding accented short vowel (Jijal Indus Kohistani *kuláal* 'potter' <

Sanskrit *kúlālaḥ*), and show that (at least in Shinaic languages) for substantival stems with short vowels, where rising and falling contours do not contrast, old accentual patterns are still faithfully represented by accentuation (or lack thereof) of inflectional endings. In thematically extended inherited lexemes, the old accent isn't preserved, but rather it is the stem that is almost always accented instead of the extension. Furthermore, if the stem is polysyllabic, then it is precisely the penultimate syllable of the word (i.e., the syllable preceding the extension, the final one of the stem) that gets accent regardless of what the accentuation was in the Sanskrit ancestor before extension (Atsharet Palula *kriṣíṇ-u/-i* 'black' < Sanskrit *kṛṣṇá-*). Other minor patterns are discussed (some not yet resolved) and using this new knowledge of accentuation, possible loanwords from (and to) other languages (like Kashmiri) identified.

More complex tonology to what exists in Shina, Palula and Indus Kohistani is found in Gawri, Torwali and Khalkoti. This talk only uses data from Gawri to explain the diachrony of its tonology. In particular, it is shown that Gawri builds on a base of the Shina-type inherited system described above, but adds to its archaism by introducing an innovative lowering of tone on the mora directly following aspiration, (historical) breathy voice or *h*. Different combinations of these two old and new tonal sources lead to one of the five tone contrastive contours described by Baart [1997]. Thus differ Gawri *láám*ˋ `H(L)` 'village' < Sanskrit *grā́maḥ*, Gawri *káál* `H` 'year' < Sanskrit *kāláḥ*, Gawri *tʰùún* `LH` 'pillar' < Sanskrit *stʰū́ṇā*, Gawri *ʦʰììr* `L` 'milk' < Sanskrit *kṣīrám*. Baart's `HL` pattern (as in *bóòr* `HL` 'lions' from *bóór* `H` 'lion') is usually either polysyllabic or bimorphemic and its genesis is also explained. In addition, vowel changes by *i*-umlaut induced by the now-lost *-ikā*-extension are described, and further contrasted with Kashmiri-Kishtwari where both *i*-umlaut from the feminine *-ikā*-extension and *u*-umlaut from the masculine *-akaḥ*-extension lead to more complex intra-paradigmatic vowel (and consonant) changes, but without tone.

This work builds toward a larger project of clarifying the details of historical phonology and morphology of the languages of the Hindukush–Karakoram. In particular, it is hoped that evidence from this study will help dispel the commonly held and often cited belief that Vedic accent was lost in post-Vedic Indo-Aryan. Indeed, clear accentual evidence of survival of Vedic accent down to the present day should invite more effort in trying to explain at least some of the unexplained apparent irregularities in vocalic and consonantal inheritance patterns in post-Sanskrit Indo-Aryan languages as had already been begun by such Indological pioneers as Pischel [1900] and Turner [1916], and recently continued by Tegethoff and Basu [forthcoming].

References

Joan L. G. Baart. *The sounds and tones of Kalam Kohistani*. National Institute of Pakistan Studies, Quaid-i-Azam University, and Summer Institute of Linguistics, 1997.

Almuth Degener. *Shina-Texte aus Gilgit (Nord-Pakistan)*. Harrassowitz-Verlag, 2008.

Jonathan Evans, Wen-Chi Yeh, and Rukmini Kulkarni. Acoustics of tone in Indian Punjabi. *Transactions of the Philological Society*, 116(3):509–528, 2018.

David Hultman. *Topics in the grammar of Kalkoti*. Stockholm University Bachelor's thesis, 2023.

Keisuke Huziwara. Tyakuma-go on'inron. *Gengokidyutu dyaanaru*, (11):51–102, 2019.

Henrik Liljegren. *A grammar of Palula*. Language Science Press, 2016.

Henrik Liljegren. Palula dictionary. *Dictionaria*, (3):1–2700, 2019.

Richard Pischel. *Grammatik der Prakrit-sprachen*. Verlag von Karl J. Trübner, 1900.

Robert Tegethoff and Samopriya Basu. The historical phonology of Hindukush–Karakoram Indic: New insights from Kalasha, Khowar, Shina and Grangali. *Münchener Studien zur Sprachwissenschaft*, forthcoming.

Ralph Lilley Turner. The Indo-Germanic accent in Marathi. *Journal of the Royal Asiatic Society of Great Britain and Ireland*, pages 203–251, 1916.

Claus Peter Zoller. *A grammar and dictionary of Indus Kohistani*, volume 1: Dictionary. De Gruyter Mouton, 2005.

Title of the paper: Sonorant Gemination in Old Tamil

Author: Ramprashanth Venkatakrishnan, Madurai Kamaraj University, India

Abstract:

Old Tamil exhibits two kinds of morphologically derived sonorant geminates (sonorants in Old Tamil include nasals, liquids and glides), as seen in the Sangam poetic texts and as described by the traditional Tamil Grammars like Tolkāppiyam and Naṉṉūl (Vijayavenugopal (1968)). These can be described by two linear phonological rules:

(1) Type A

$\#(C)V\mathbf{C_i}\# + \#V \rightarrow \#(C)V\mathbf{C_iC_i}$

(2) Type B

$\#(C)V\# + \#\mathbf{C_i} \rightarrow \#(C)V\mathbf{C_iC_i}$

$\mathbf{C_i}$ represents a sonorant consonant.

Examples of the two kinds of sonorant gemination:

(3) Type A /kal/ + /il/ > [kɐ**ll**ɪl]
'stone' LOC 'in the stone'

(4) Type B /mu/ + /maʈi/ > [mʊ**mm**ɐɖɪ]
three folds 'three folds'

In both these cases there is gemination at the vowel-consonant juncture between the monosyllabic root/word and the following suffix/word. The conditions for gemination are that the monosyllable contain only a short vowel with an optional onset, and that the geminating consonant be a sonorant. There is blocking of gemination when the vowel in the monosyllabic root/word is long:

(5) Type A /ka:l/ + /il/ > [ka:lɪl] *[ka:**ll**ɪl]
'leg' LOC 'in the leg'

(6) Type B /ma:/ + /n̪ilam/ > [ma:n̪ɪlɐm] *[ma:**n̪n̪**ɪlɐm]
big land 'earth'

Also, we find that obstruents do not geminate in the morpho-phonological context described in (1):

(7) /ʊiɻ/ + /aː/ > [ʊɪɖaː]
‘leave’ NEG ‘not leave’

However, obstruents seem to geminate in the context described in (2) as seen in example (8). We propose that this is not motivated by a constraint to achieve the optimal bimoraic initial syllable but to preserve the voicelessness of the initial obstruent of the conjoining word, since only word-initial and geminate obstruents are voiceless in Old Tamil. This motivation for obstruent gemination is clear as it happens even in the context of the monosyllabic root/word having a long vowel, as in example (9).

(8) /mu/ + /kulam/ > [mʊ**kk**ʊlɐm]
three clan ‘three clans’

(9) /ʈiː/ + /paɻakkam/ > [ʈiː**pp**ɐɻɐkkɐm]
‘bad’ ‘habit’ ‘bad habit’

We propose an optimality theoretic (Kager (1999)) treatment of this phenomenon where these seemingly ‘separate’ phonological rules are unified as a single Conspiracy (‘conspiracy’ is defined as per Kisseberth (1970)). A Prosodic Word constraint (ProsWd) is postulated, that requires prosodic words to begin with a heavy/bimoraic syllable (a similar constraint is proposed for Japanese by Ito et al. (2017)). This explains both kinds of sonorant geminates, and the case of blocking of sonorant gemination. Tableaux (10) and (11) analyse Type A and Type B geminates respectively:

(10) /kal/+ /il/ ‘in the stone’ (Onset >> ProsWd >> NoCoda >> *Gem)

/kal/ + /il/	Onset	ProsWd	NoCoda	*Gem
a. [kɐ.lɪl]		*!	*	
b. [kɐl.ɪl]	*!		*	
c. ☞ [kɐl.lɪl]			*	*

(11) /mu/ + /maʈi/ ‘three folds’ (Onset >> ProsWd >> NoCoda >> *Gem)

/mu/ + /maʈi/	Onset	ProsWd	NoCoda	*Gem
a. [mʊ.mɐ.ɖɪ]		*!		
b. [mʊm.ɐ.ɖɪ]	*!		*	
c. ☞[mʊm.mɐ.ɖɪ]			*	*

The introduction of ProsWd constraint is substantiated with evidences that Old Tamil is a moraic sensitive language (Steever (1987)).

We introduce another highly ranked markedness constraint *Gem(Obs) which prohibits obstruent geminates in a word.

(12) /ʋiɽ/ + /aː/ ‘not leave’ (Onset >> *Gem(Obs), ProsWd >> NoCoda >> *Gem)

/ʋiɽ/ + /aː/	Onset	*Gem(Obs)	ProsWd	NoCoda	*Gem
a. ☞[ʋɪ.ɖaː]			*		
b. [ʋɪɽ.ʈaː]		*		*!	*
c. [ʋɪɖ.aː]	*!				

References:

Ito, J., Kubozono, H., & Mester, A. (2017). A prosodic account of consonant gemination in Japanese loanwords. In H. Kubozono (Ed.), *The Phonetics and Phonology of Geminate Consonants* (Vol. 1). Oxford University Press.

Kager, R. (1999). *Optimality Theory*. Cambridge University Press.

Kisseberth, Charles. (1970). *On the Functional Unity of Phonological Rules*. Linguistic Inquiry. 1. 291-306.

Steever, S. B. (1987). Tamil and the Dravidian languages. In B. Comrie (Ed.), The world’s major languages (pp.725 – 746). London: Croom Helm.

Tolkāppiyam: Text (with non-metrical segmentation), transliteration and translations in English verse and prose (Edition first). (2021). Central Institute of Classical Tamil.

Vijayavenugopal, G. (1968). *A Modern Evaluation of Naṉṉūl, Eḻuttatikāram*. Annamalai University.

Negative Existential Cycle in Eastern Indo-Aryan Languages: A Typological Analysis

Abstract

The evolution of negative markers can be attested as a cyclical process, wherein new expressions emerge to encode pre-existing functional domains. Drawing on comparative data, Croft (1991) proposes that negative existential markers progressively extend their functional scope to encompass standard negation, eventually replacing earlier standard negators, a process conceptualized as the Negative Existential Cycle (NEC). For instance, the standard negation in Odia is expressed by a suffix *-nɔ* (example 2) and the existential predication is negated by a negative existential copula *nah-* (example 4). Gradually, the negative existential *nah-* enters the domain of standard negation and replaces the verbal negator in a certain TAM category (example 6).

Odia

1. Jagu gɔʈɔ kali kʰeɭ-u-ʈʰi-l-a
   Jagu.3SG.NOM yesterday play-PROG-AUX-PST-3SG
   Jagu was playing yesterday.

2. Jagu gɔʈɔ kali kʰeɭ-u-nɔ-ʈʰi-l-a
   Jagu.3SG.NOM yesterday play-PROG-NEG-AUX-PST-3SG
   Jagu was not playing yesterday.

3. puri-rɔ belabhuĩ-re ɖʰeu ɔcʰ-i
   puri-GEN seashore-LOC tide.3SG COP-3SG
   There is high tide on the seashore of Puri.

4. puri-rɔ belabhuĩ-re ɖʰeu nah-ĩ
   puri-GEN seashore-LOC tide.3SG NEG.EX-3SG
   There is no high tide on the seashore of Puri.

5. mũ a:mbɔ kʰa-u-cʰ-i
   1SG.NOM mango eat-PROG-AUX-1SG
   I am eating a mango.

6. mũ a:mbɔ kʰa-u-nah-ĩ
   1SG.NOM mango eat-PROG-NEG-1SG
   I am not eating mango.

In this framework, Croft identifies six distinct stages of the cycle: three *stable,* i.e., Type A, Type B, and Type C, with no variation between standard and existential negation, and three *transitional* stages, viz., Type A~B, Type B~C, and Type C~A, showing variation in either domain. Building on this typological foundation, Veselinova (2013, 2016, 2022) expands this framework cross-linguistically, tracing the interaction of existential negators with the scope of verbal negation. Aligning with these theoretical perspectives, the present study investigates the interaction between negative existentials and verbal negators across eight Eastern Indo-Aryan languages, namely, Assamese, Bengali, Odia, Angika, Bajjika, Bhojpuri, Magahi, and Maithili. Conversational data were elicited and recorded, and a questionnaire was also administered to speakers of the aforementioned languages. The collected data were subsequently transcribed and analyzed.

The analysis aims to examine the existential constructions in these languages, focusing on the forms and functional range of negative existential. Furthermore, this study tests Croft's NEC to classify these languages into specific stages of the existential cycle in the original model, thereby situating them within the broader framework of typological variation.

Keywords: Existential Negation, Standard Negation, Negative Existential Cycle, Eastern Indo-Aryan languages, Typology

Word count- 402

References

Croft, William. 1991. The evolution of negation. *Journal of Linguistics* 27: 1- 27.

Veselinova, Ljuba. 2013. Negative existentials: a cross-linguistic study. *Italian Journal of Linguistics: Special Issue on Existential Constructions,* 25(1):107–145.

Veselinova, Ljuba. 2016. The Negative Existential Cycle through the lens of comparative data. In van Gelderen, E (ed), *The Linguistic Cycle Continued,* pages 139–187. John Benjamins Publishing Co., Amsterdam/New York.

Veselinova, Ljuba. and Arja Hamari. (2022). Introducing the Negative Existential Cycle. In Veselinova, Ljuba. and Arja Hamari (eds.), *The Negative Existential Cycle,* pages 1–56. Language Science Press, Berlin.

Middle voice in Kannada: Markers in competition

Lennart Chevallier & John Peterson
Kiel University

The middle as a third voice category has been associated with various phenomena. A number of authors (e.g. Barber 1975: 18, 21; Klaiman 1991: 92; Kemmer 1993: 4) consider the middle voice to involve *subject-affectedness*. More recent approaches, however, such as Inglese (2021), do not seek a single core meaning. Instead, he considers the middle to encode a variety of valency-related functions but not limited to that. For Inglese, in a middle voice system, these middle markers are obligatory with some verbs, i.e., they do not contrast with an active voice and are hence "non-oppositional", while they are non-obligatory with other verbs. In this latter group, they are "oppositional", as they show a functional contrast with the active voice. We follow this definition in our analysis.

Across the Dravidian languages of South India, we see evidence for such a system to varying degrees. According to the etymological dictionary by Burrow & Emeneau (1984, entry 2151), the verb *koḷ-* (also: *koḷḷu*) 'to seize, receive, buy, acquire, etc.' is an auxiliary with reflexive meaning in Tamil, Malayalam, Kannada, Telugu, and other languages. However, upon closer examination, the picture is somewhat more complex in that *koḷ-* brings about readings which are not restricted to the reflexive, e.g. in Kannada, *koḷḷu* can serve as a reflexive, reciprocal, self-benefactive and anticausative marker and, e.g., obligatorily occurs in *is(a)koḷḷu* 'to take' (Bucher 1899: 55).

In our talk, we will illustrate in which ways the middle marker in Kannada competes with other constructions. With "competition" we mean that a marker other than the middle marker is also available to encode the same function. This is in line with Inglese (2022), who compares different interaction scenarios of middle markers with other valency-reducing constructions. For example, in Kannada, *koḷḷu* is used as reciprocal marker with the verb *haṃcu* 'to share' resulting in *haṃc-i-koḷḷu* [share-LNK-MM] 'to share with each other'. Reciprocity can, however, also be expressed without *koḷḷu* by expressions such as *obba-r-ig' obba-ru* [one.HUM-HON-DAT one.HUM-HON] 'to each other' (Sridhar 1990: 124-125). Among others, the question remains whether one strategy is unmarked or whether both express slightly different semantics. It is also interesting to note that the primary function of the middle voice in Dravidian appears to be the self-benefactive, unlike e.g. the data discussed in Inglese (2021: 506), but in line with its origin in Dravidian from a self-benefactive construction.

All modern Kannada examples are drawn from two sources: An annotated corpus, which is the result of continuous joint work, and questionnaire-based fieldwork. Data from Middle Kannada are also included to trace this development in Kannada since the 12th century. The annotation is being done in Toolbox and FLEx (SIL) at the morphosyntactic level and a translation of every sentence is provided.

At present, the approach is chiefly descriptive but closely linked to recent typological studies. Future work will also include other aspects of middle marking, e.g., non-valency related functions, and will aim to contribute to a better understanding of the middle voice cross-linguistically.

[495 words]

References

Barber, Elizabeth Jane Wayland. 1975. Voice–beyond the passive. In Cathy Cogen, Henry Thompson, Graham Thurgood, Kenneth Whistler & James Wright (eds.), Proceedings of the First Annual Meeting of the Berkeley Linguistics Society, 16–24. Berkeley: Berkeley Linguistics Society.

Bucher, Johannes. 1899. *A Kannaḍa-English School-Dictionary. Chiefly Based on the Labours of Rev. Dr. F. Kittel*. Mangalore: Basel Mission Book & Tract Depository.

Burrow, Thomas & Murray Barnson Emeneau. 1984. *A Dravidian Etymological Dictionary*. 2nd ed. Oxford: Clarendon Press.

Inglese, Guglielmo. 2021. Towards a typology of middle voice systems. *Linguistic Typology* 26(3), 489-531. https://doi.org/10.1515/lingty-2020-013.

Inglese, Guglielmo. 2022. How do middle voice markers and valency reducing constructions interact? Typological tendencies and diachronic considerations. *Folia Linguistica* 56(2), 239-271. https://doi.org/10.1515/flin-2022-2019.
Kemmer, Suzanne. 1993. The Middle Voice (Typological Studies in Language, 23). Amsterdam & Philadelphia: John Benjamins.
Klaiman, Miriam Holly. 1991. Grammatical voice (Cambridge Studies in Linguistics, 59). Cambridge: Cambridge University Press.
Sridhar, Shikaripur Narayanarao. 1990. Kannada (Descriptive Grammars). London & New York: Routledge.

Noncausal events and anticausative marking in Modern Kannada – a first overview

John Peterson & Lennart Chevallier
Kiel University

In our talk we offer a preliminary overview of noncausal events and their marking in Modern Standard Kannada. Kannada has various strategies to express noncausal events, none of which, however, is productive.

The first strategy is through P-preserving labile verbs such as *tere* 'open (ITR/TR)', i.e., Letuchiy's (2009) "anticausative lability" class. Despite several excellent Kannada-English dictionaries (e.g., Bucher, 1899; KSP, 2019; Kittel, 1894 [2020]; Učida et al., 2018), these tend to incorporate entries from all periods of Kannada, so that reliable data for Modern Kannada must come from corpus data and interviews. Table 1, based on Nichols et al. (2004), presents a (non-exhaustive) overview of 18 sample verbs from a questionnaire for a cross-linguistic study of basic transitivity from field work.

As Table 1 shows, labile verbs are rather seldom in Modern Kannada, e.g. in only three of the 18 verb pairs from this sample (11-13). However, in Middle Kannada (e.g., 12th century), this strategy was somewhat more common: E.g., *to*ː*ru* 'appear; show', *ba*ː*gu* 'bend (ITR/TR)', etc., while according to the speakers we consulted, in the modern language the induced form of these verbs is expressed through the causative suffix *-isu*.

As Table 1 also shows, 10 of the 18 pairs discussed there derive the respective induced form from the plain form through this same causative suffix *-isu* (grey background in Table 1), making overt transitivization the basic pattern. Interestingly, however, *-isu* can also serve as an anticausative marker with some verbs, e.g., *ka*ː*ɳ-isu* [see-CAUS] 'appear; become visible' (*'show') and *ke*ː*ɭ-isu* [hear-CAUS] 'cause to hear' but also 'become heard/audible', a strategy also found in some IA languages, e.g. Sadri (Peterson & Baraik 2022: 216-217).

A further, highly restricted anticausative marking strategy is the middle marker *koḷḷu*. This is found with the verbs *muccu* 'close (ITR/TR)' and *tere* 'open (ITR/TR)' but here apparently only with the subject *ba*ː*gilu* 'door'. Although these are both ambitransitive verbs, for some speakers this appears to be the preferred means to express a non-causal event of opening or closing, while the active form can denote an unexpected, uncontrolled event; cf. example (1).

(1)

Active:	*ba*ː*gilu*	*mucc-itu*	Middle:	*ba*ː*gilu*	*mucc-i-koṃɖ-itu*
	door	close-PST.3SG.N		door	close-LNK-MID.PST-PST.3SG.N
	'The door closed (unexpectedly).'			'The door closed.'	

Finally, the status of the so-called "passive" constructions will be discussed, in which the main predicate is marked as an infinitive and followed by *paḍu* 'experience, undergo' or *a*ː*gu* 'become'. While these can usually be interpreted as passive, examples such as (2) show that they can (marginally) also be used for situations where no external agent is assumed, a development which has also been attested for Vedic (Kulikov, 2011) and elsewhere (e.g., Peterson, submitted).

(2) (from a children's comic book)

*idannu okkuː*ʈ*aveṃdu kareyala*ː*yitu.*

id-annu	*okkuː*ʈ*a-v*	*eṃdu*	*kare-y-alu*	*a*ː*-y-itu.*
3SG.PROX-ACC	alliance-EUPH	QUOT	call-EUPH-INF	become-EUPH-PST.3SG.N

It was called / known as "the Alliance".

We discuss these constructions further and suggest a possible analysis of them as *nonactive* (Kalluli, 2007; 2009), a category found in some languages which corresponds to both the passive and the anticausative of English.

[499 words]

Table 1: Plain and induced forms in Modern Standard Kannada (after Nichols et al., 2004)

Pair	**Plain**	**Induced**
1. laugh / make laugh, amuse, strike as funny	*nagu*	*nag-isu*
2. die / kill	*saːyu*	*saːy-isu; kollu* 'kill'
3. sit / seat, have sit, make sit	*kuːru*	*kuːr-isu*
4. eat / feed, give food	*tinnu*	*tinn-isu*
5. learn, know / teach	*kali*	*kali-su*
6. see / show	*noːḍu*	*toːr-isu*
7. be, become angry / anger, make angry	*koːpa baru* [anger come]	*koːpa tar-isu* [anger bring-CAUS]
8. fear, be afraid / frighten, scare	*hedaru*	*hedar-isu*
9. hide, go into hiding / hide, conceal, put into hiding	*aḍag-i-koḷḷu* [hide.ITR-LNK-MID]	*mucc-i iḍu* [hide.TR-CVB put]
10. (come to) boil / (bring to) boil	*kudi*	*kud-isu*
11. burn, catch fire / burn, set fire	*suḍu*	*suḍu*
12. break / break	*muri*	*muri*
13. open / open	*tere*	*tere*
14. dry / make dry	*oŋagu*	*oŋag-isu*
15. be, become straight / straighten, make straight	*saraḷavaːgu (= saraḷa aːgu* [straight become])	*saraḷa maːḍu* [straight make]
16. hang / hang (up)	*joːtu biːḷu* [hang.CVB fall]	*joːtu haːku* [hang.CVB put]
17. turn over / turn sth. over	*tiragu (tirugu, tirigu)*	*tirag-isu* (etc.)
18. fall / drop, let fall	*biḷu*	*biḷ-isu*

References

Bucher, Johannes. 1899. *A Kannaḍa-English school-dictionary. Chiefly based on the labours of the Rev. Dr. F. Kittel.* Mangalore: Basel Mission Book & Tract Depository.

Kannada Sahitya Parishattu [KSP]. 2019. *Abridged Kannada-English Dictionary.* Bangalore: Kannada Sahitya Parishattu.

Kittel, Ferdinand. 1894. *Kannada English Dictionary.* Mangalore. [New Delhi: Gyan Publishing House, 2020].

Kulikov, Leonid. 2011. Passive to anticausative through impersonalization. The case of Vedic and Indo-European. In: Andrej Malchukov & Anna Siewierska (eds.), *Impersonal constructions. A cross-linguistic perspective.* Amsterdam / Philadelphia: Benjamins. 229-254.

Kallulli, Dalina. 2007. Rethinking the passive/anticausative distinction. *Linguistic Inquiry* 38(4): 770-780.

Kallulli, Dalina. 2009. (Non-)Blocking in the voice system. *Language typology and universals / STUF - Sprachtypologie und Universalienforschung*, 62(4): 269-284.

Letuchiy, Alexander. 2009. Towards a typology of labile verbs: Lability vs. derivation. Patience Epps & Alexandre Arkhipov (eds.), *New challenges in typology. Transcending the borders and refining the distinctions.* [Trends in linguistics, Studies and monographs, 217]. Berlin / New York: Mouton de Gruyter. 247-268.

Nichols, Johanna, David A. Peterson & Johnathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8(2): 149-211.

Peterson, John. submitted. Noncausal events and anticausative marking in Modern Irish – synchronic and diachronic aspects.

Peterson, John & Sunil Baraik. 2022. *A grammar of Chotanagpuri Sadri. An Indo-Aryan lingua franca of eastern central India.* Mysore: Central Institute of Indian Languages.

Učida, Norihiko, B.B. Rajapurohit & Jun Takashima. 2018. *Kannada-English Etymological Dictionary*. Tokyo: Institute for Languages and Cultures of Asia and Africa.

Gerunds as Mixed Categories in Hindi and Bangla

Soumini Banerjee, Debolina Bhattacharjee
University of Delhi

Mixed categories are constructions combining the distributional and morphological properties of two distinct syntactic categories headed by a single phrasal constituent.

We analyse gerunds in two Indo-Aryan languages, Bangla and Hindi, to establish cross-linguistic criteria for mixing in nominalisations.

Existing accounts claim full nominalisations are n-based and have a full φ-feature set. D-based nominalisations are mixed or defective. Their lack of a phrasal n-head blocks a full φ-feature set realisation leading to default φ-agreement. Additionally, they disallow flexible determiners which correlates to their lack of adjectival modification.

However, gerunds in Hindi and Bangla can take flexible determiners, while not allowing adjectival modification, making the previous account seem inadequate in a cross-linguistic analysis.

Example:-
Everyone saw [her **this/that** making the *rotis* quickly/***quick**].

(1) Hindi

səb-ne us-ka: ***ye*** **/ *vo*** *jʰəʈ-pəʈ*$_{*i/j}$**/*gərəm*$_{i/*j}$** *[roʈi*$_i$ *bəna:-na]*$_j$
all-ERG 3SG-GEN.M **DEM.PROX /DEM.DIST** quickly / **warm** [bread make-NMLZ]

ɖekʰ-a *hɛ*
see-PRF.3MSG be.PRS.3SG

(2) Bangla

ʃobai o-ɾ ***ei*** **/ *oi*** **/ *ʃei*** *jʰɔʈ-pɔʈ*$_{*i/j}$**/*gɔrom*$_{i/*j}$** *[ruʈi*$_i$
everyone.NOM 3SG-GEN **DEM.PROX /DEM.DIST /DEM.DIST** quickly /**warm** [bread

bana-no]$_j$ *-ʈa* *dekʰ-ecʰ-e.*
make-NMLZ]-CLF see-PRF-PRS.3SG

We find that gerunds in Hindi and Bangla allow flexibility in determiners in terms of proximal information (distal vs proximal). These determiners do not hold any nominal φ-feature information of the head noun. While Iordăchioaia's (2020) account of mixed nominalizations having default features remains consistent in our observation, her account of flexibility in determiners is contrasted with mixed nominalizations in the languages studied

here. We show languages show a different axis of flexibility in determiners, not involving φ-features.

Moreover, to delineate a system capturing mixed status containing both nominal and verbal properties, we divide "internal syntax" into two distinct categories:-

i. **Projection:** licensing of phrasal heads like adjectives, determiners, and adverbs.

ii. **Valuation:** operations that include feature valuation of a probe to a goal like Agree of φ-features and Case valuation.

In a mixed structure, we claim that both valuation and projection must contain nominal and verbal properties. Thus, when the n-head defectively nominalises, we see evidence of mixing within the internal syntax in both projection and valuation.

For our data, the projection includes the ability to flexibly take determiners (nominal) and adverbs (verbal). The valuation includes a genitive subject (nominal), accusative object (verbal), and default φ-features (nominal). The mixed properties in both categories establish the gerunds as mixed nominals.

This simultaneous mixing is our syntactically grounded definition of a mixed category. The paper will further test the proposed criteria in the internal syntax of other nominalisations in these South Asian languages to strengthen our analysis.

References:

Alexiadou, A. (2020). D vs. n nominalizations within and across languages. In Oxford University Press eBooks (pp. 87–110). https://doi.org/10.1093/oso/9780198865544.003.0005

Borsley, R. D., & Kornfilt, J. (1999). Mixed extended projections. In BRILL eBooks (pp. 101–131). https://doi.org/10.1163/9781849500098_006

Chomsky, N. (2000) Minimalist Inquiries: The Framework. In: Martin, R., Michaels, D. and Uriagereka, J., Eds., Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik, MIT Press, Cambridge, 89-155.

Iordăchioaia, G. (2020). D and N are different nominalizers. Glossa a Journal of General Linguistics, 5(1). https://doi.org/10.5334/gjgl.1111

Lowe, J. (2016). Participles, gerunds and syntactic categories. Proceedings of the International Conference on Head-Driven Phrase Structure Grammar. https://doi.org/10.21248/hpsg.2016.21

Lowe, J. J. (2019). Mixed projections and syntactic categories. Journal of Linguistics, 56(2), 315–357. https://doi.org/10.1017/s0022226719000100
Panagiotidis, P. (2014). Categorial features. https://doi.org/10.1017/cbo9781139811927

The semantic contribution of complement clauses in Hindi verbs of perception and cognition

Lucrezia Carnesale
(University of Pavia)

This paper is a corpus-based study of complement clauses of Hindi verbs of cognition, perception and volition. These two classes show a tendency to occur with Stimuli instantiated by complement clauses rather than NPs. Building upon previous studies that propose that the choice of complement types is semantically motivated (Bolinger 1968, Dik and Hengeveld 1991, Dixon 1984, Goddard 2007, Luraghi 2020, Wierzbicka 1988), I argue that Hindi complement clauses show a distributional preference for verbs with meanings closely related to that of the construction.

For the data retrieval, I rely on the hiTenTen corpus, available on SketchEngine[1]. I extract a random sample of 100 occurrences for 20 verbs belonging to the two classes and I manually scrutinize each occurrence. Relying on a constructional approach (Goldberg 1995, 2006; Barðdal 2008), I assess the number of verbs that occur with a given construction to determine the type-frequency of the construction. I then analyze the semantic properties of the verbs occurring with a given construction to establish its semantic coherence.

The results show that verbs belonging to these two classes exhibit some similarities in their complement type distribution. For example, they seem to occur with a comparatively similar frequency with finite complement clauses (1). However, the analysis also shows significant differences. For example, Hindi verbs of (visual) perception exhibit a high frequency with predicative participles (2), while verbs of cognition rarely allow for such constructions. On the other hand, both volition and cognition verbs often allow non-finite constructions (3), which are never found with perception verbs.

Additionally, I argue that verbs with a different meaning can still occur with a given construction if there is a way of integrating the semantics of the verb into that of the construction (Perek 2015). In such cases, the construction contributes its meaning and changes the verb's interpretation. For example, the predicative participle construction typically expresses direct perception of states of affairs, due to its ability to represent simultaneity between two events (Noonan 2007). Some Hindi cognition verbs, however, may sometimes occur with a predicative participle clause: in such cases, they result in a different interpretation, which stands somewhere between cognition and perception (4).

1. *maiṁ* *kabhī-kabhī* *soc-t-ī* *hūṁ,* *ki* *tum*
   1SG.NOM sometimes think-IPRF-F.SG be.1SG.PRS that 2PL.NOM
   yahāṁ na *ā-te,* *to* *acch-ā ho-t-ā.*
   here not come-IPRF-M.PL then good-M.SG be-IPRF-MSG
   'Sometimes I think it would have been better if you hadn't come here.'

2. *mujhe* *apne* *ghar=meṁ=se* *ek* *aurat* *bāhar*
   1SG.DAT REFL house=in=INS one woman(F.SG.NOM) out
   nikal-t-ī *hu-ī* *dikhāī* *d-ī*
   come_out-IPRF-F.SG be.PRF-F.SG seeing(F) give-PRF.F.SG
   'I saw a woman coming out of my house.'

3. *tum* *bahut* *jaldī* *bambaī* *choṛ=kar* *dillī* *ā-n-e=kī*
   2PL.NOM very soon Bombay leave=CP Delhi come-INF-OBL=GEN-F
   soc *rah-e*

[1] https://www.sketchengine.eu/hitenten-hindi-corpus/

think PRGR-PRF.M.PL
'You are thinking of leaving Bombay very soon and coming to Delhi.'

4. *jālpā=ko* *in* *śabd-oṁ=meṁ* *sneh* *aur*
jalpa=DAT these.OBL word(F)-PL.OBL=in affection(M.SG.OBL) and
sahānubhūti=kā *ek* *sāgar* *umaṛ-t-ā*
sympathy(F.SG.OBL)=GEN one ocean(M.SG.NOM) rise-IPRF-M.SG
hu-ā *jān* *paṛ-ā.*
be.PRF-M.SG knowledge(M.SG.NOM) fall-PRF.M.SG
'Jalpa felt an ocean of affection and sympathy rising in these words.'

Acknowledgments

Research for this paper has been supported by European Union funding – NextGenerationEU – Missione 4 Istruzione e ricerca - componente 2, investimento 1.1. Fondo per il Programma Nazionale della Ricerca (PNR) e Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN)" progetto 20223XH5XM "Verbs' constructional patterns across languages: a multi-dimensional investigation" CUP F53D23004570006.

References

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Bolinger, Dwight L. 1968. Entailment and the meaning of structures. *Glossa* 2.2: 119-127.

Dik, Simon Cornelis and Kees Hengeveld. 1991. The Hierarchical Structure of the Clause and the Typology of Perception-verb Complements. *Linguistics* 29: 231–259.

Dixon, Robert M.W. 1984. The semantic basis of syntactic properties. *Berkeley Linguistics Society, Proceedings* 10: 583-595.

Goddard, Cliff. 2007. A 'lexicographic portrait' of forgetting. In Mengistu Amberber (ed.) *The language of memory in crosslinguistic perspective*, pp. 119-138. Amsterdam: John Benjamins.

Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Luraghi, Silvia. 2020. *Experiential Verbs in Homeric Greek*. Leide: Brill.

Perek, Florent. 2015. *Argument Structure in Usage-Based Construction Grammar. Experimental and corpus-based perspectives*. Amsterdam: John Benjamins.

Noonan, Michael. 2007. Complementation. In Shopen T. (ed.) *Language Typology and Syntactic Description,* 52–150. Cambridge: CUP.

Wierzbicka, Anna. 1988. *The Semantics of Grammar*. Amsterdam: John Benjamins.

Complex predicate and associated motion in Bengali

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

Introduction

Complex predicates or compound verbs are among the most extensively studied phenomena in South Asian languages, analyzed from both synchronic and diachronic perspectives (Hook 1974; Butt 1995, 2010; Slade 2013, 2021, among others). This paper describes and analyzes the Bengali COME/GO Construction. In this construction, the first verb (V1) expresses a prior non-motion event, and the second verb (V2) expresses a subsequent deictic directional motion, as shown in (1) and (2). This construction formally resembles the typical vector verb constructions of Bengali.

(1) dada [khe-e]$_{V1}$ [eś-ech-en]$_{V2}$
brother eat-CVB come-PRF-HON
'Brother, have you eaten?' (mzamin.com)
Literal meaning: 'Brother, did you eat and come?'

(2) æk hajar ṭaka [di-e]$_{V1}$ [gech-e]$_{V2}$
one thousand money give-CVB go.PRF-3
'He left 1,000 Tk.' (prothom-alo.com)
Literal meaning: 'He gave 1,000 Tk and went.'

The Bengali COME/GO Construction is used to express associated motion (AM), a grammatical category that has only recently been recognized (Guillaume & Koch 2021). AM is defined here as a grammatical category that serves to associate translational motion to a non-motion verb event (Guillaume & Koch 2021: 3).

Despite its formal resemblance to vector verb constructions, the Bengali COME/GO Construction has not received sufficient attention in the literature on complex predicates. It has been regarded as a biclausal converbal clause chaining construction (Klaiman 1981: chapter 4), rather than a monoclausal complex predicate. Furthermore, given the lower degree of semantic bleaching of V2, V2 in the same type of construction found in other New Indo-Aryan languages is positioned between lexical verbs and vector verbs in the grammaticalization cline (also called "factor verbs" in this context) (Hook 2013, Pardeshi 2021: 235).

This paper argues that the Bengali COME/GO Construction is a subsequent AM complex predicate, in which V2 remains semantically less bleached, yet shows signs of grammaticalization. By analyzing the construction from the perspective of AM, this paper aims to position it within the broader landscape of complex predicates of Indo-Aryan languages.

Method

The data for this study comes from the bnTenTen17 corpus on SketchEngine and from elicitation with consultants. My consultants are from Dhaka (born in 1993) and Kolkata (born in 1984).

Analysis and findings

Functionally, the Bengali COME/GO Construction expresses a subsequent AM. AM typology distinguishes the type of AM based on the temporal relation between translational motion and a non-motion event. In this construction, V1, a converbal form, expresses a prior non-motion event, to which V2 adds subsequent translational motion, hence expressing a subsequent AM as a whole.

Morphosyntactically, this construction is monoclausal. The monoclausality is evidenced by a shared temporal or locational reference between the verbs, a single referential subject, and a negative polarity item licensing. Being monoclausal, this construction constitutes a complex predicate.

Lastly, deictic directional V2 is shown to be grammaticalized even though it is semantically less bleached than typical vector verbs. Following the definition of grammaticalization by Boye (2023), V2's grammaticalization is confirmed by its loss of discourse prominence, in that the V2 usually does not fall within the scope of negation. In other words, it is outside of the actual focus domain (cf. Van Valin & LaPolla 1997: 212). Additionally, V2 displays properties frequently observed in grammaticalized expressions, specifically concerning bondedness and syntagmatic variability (cf. Bisang et al. 2020).

Conclusions

The Bengali COME/GO Construction is a monoclausal subsequent AM construction in which the deictic directional V2 is grammaticalized. By incorporating a typological perspective on associated motion, this study sheds light on an understudied type of complex predicate in Bengali. While this construction has been largely marginal in the literature on complex predicates in Indo-Aryan, it plays an important role in expressing associated motion. This study enhances our understanding of the grammaticalization of V2, as it demonstrates that V2 can be grammaticalized even when it does not undergo semantic bleaching.

References

Bisang, Walter, Andrej Malchukov, Iris Rieder & Linlin Sun. 2020. Measuring grammaticalization: A questionnaire. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization Scenarios: Volume 1 Grammaticalization Scenarios from Europe and Asia*, 89–104. Berlin: De Gruyter Mouton.

Boye, Kasper. 2023. Grammaticalization as conventionalization of discursively secondary status: Deconstructing the lexical–grammatical continuum. *Transactions of the Philological Society*

121 (2) 270–292.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI Publications.

Butt, Miriam. 2010. The light verb jungle: still hacking away. In Brett Baker, Mark Harvey & Mengistu Amberber (eds.), *Complex Predicates: Cross-Linguistic Perspectives on Event Structure*, 48–78. Cambridge: Cambridge University Press.

Guillaume, Antoine & Harold Koch. 2021. Introduction: Associated motion as a grammatical category in linguistic typology. In Antoine Guillaume & Harold Koch (eds.), *Associated Motion*, 3–30. Berlin: De Gruyter Mouton.

Hook, Peter Edwin. 1974. *The Compound Verb in Hindi*. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.

Hook, Peter Edwin. 2013. Competition between vectored verbs and factored verbs. *NINJAL Project Review* 3 (3) 178–189.

Klaiman, M. H. 1981. *Volitionality and Subject in Bengali: A Study of Semantic Parameters in Grammatical Processes*. Bloomington: Indiana University Linguistics Club.

Pardeshi, Prashant. 2021. Classification of complex verbs and the evolution of the compound verb in Marathi. In *Verb-Verb Complexes in Asian Languages*, 223–248. New York: Oxford University Press.

Slade, Benjamin. 2013. The diachrony of light and auxiliary verbs in Indo-Aryan. *Diachronica* 30(4). 531–578.

Slade, Benjamin. 2021. Development of verb-verb complexes in Indo-Aryan. In *Verb-Verb Complexes in Asian Languages*, 249–274. New York: Oxford University Press.

Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.

Clefting in the Northern Kasaragod Variety of Malayalam: an exoskeletal approach

This paper examines cleft constructions in the Northern Kasargod Variety of Malayalam(NKV-M), a contact language spoken in the multilingual regions of the Kerala-Karnataka border, through the theoretical lens of the Exoskeletal Frame Model (XSM). It is a generative competence model based on a neo-constructivist (Borer, 2003), non-lexicalist approach (Riksem et al., 2019). XSM holds that syntactic structures are generated independently of both functional and lexical content items. The model distinguishes between the underlying abstract syntactic structure of a language and its phonological manifestation. The paper analyses the specific features of clefting in NKV-M alongside other spoken varieties of the region, such as Kannada and Tulu, focusing on the structural adaptations at the syntax-phonology interface, to elucidate the influences of language contact.

In Standard Malayalam, the copula /aaɳə/ serves as a crucial grammatical marker in cleft constructions. The finite verb is replaced by its nominalised form (finite verb + affix -aṯ̪ə), and adding /aaɳə/ yields the cleft structure. The focused element and the copula must be adjacent; violating this adjacency leads to a different interpretation. According to Swenson's (2019) Functional Syntactic Analysis, Malayalam clefts are built on a little pP-under-TP spine, with a CP projection above for the focused constituent. No new projection is introduced for /aaɳə/, and it is inserted as a vocabulary item into T to rescue T's stranded tense feature when the little p-head lacks the agreeing [TENSE] feature.

For the sentence, '*It was Ramiza who went to Dubai yesterday*', the equivalent cleft constructions are cited below.

1.	***iṉṉale***	***ḏubai-kkə***	***poy-aṯ̪ə***	***rəmiiza***	***aaɳə***	**(Malayalam)**
	yesterday	Dubai-DAT	go.PST-NMLZ	Ramiza	COP	
2.	***ṉenne***	***ḏubai-ge***	***hooḏavɭu***	***rəmiizaa***		**(Kannada)**
	yesterday	Dubai-DAT	go.PST-REL-3.SG.F	Ramiza		
3.	***koɖe***	***ḏubaigə***	***poinavu***	***ramiiza***		**(NKV-Tulu)**
	yesterday	Dubai-DAT	go.PST-REL-3.SG.F	Ramiza		
4.	***iṉṉaala***	***ḏubai-kkə***	***poy-aṯ̪ə***	***rəmiiza***		**(NKV-Malayalam)**
	yesterday	Dubai-DAT	go.PST-NMLZ	Ramiza		

Cleft constructions in NKV-Malayalam follow the same functional architecture (little pP-under-TP-under-CP) as Std. Malayalam. However, the copula is entirely omitted, and the position of the focused element becomes structurally fixed. The finite verb is substituted with its nominalised form (finite verb + the affix '-aṯə'), and the constituent to be focused on is consistently positioned sentence-finally. Unlike Std. Malayalam, focus in NKV-M is marked purely through syntax, and the stranded T is not phonologically realised via the copula. This modification in NKV-M is likely due to prolonged diachronic contact with Kannada and Tulu. In both these languages, there is no overt copula in cleft constructions, and focus is grammaticalised through word order, with the focused XP appearing in sentence-final position.

Discussion:

The XSM is utilised to understand the influence of the underlying syntactic frames of Malayalam, Kannada and Tulu on the clausal architecture of NKV-M. An XSM analysis indicates that NKV-M most likely derives its underlying syntactic skeleton from Malayalam, and the observed shift in focus marking from a morphological strategy (copula insertion) to a syntactic strategy (sentence-final positioning) is understood as a restructuring of the phonological form (PF) rules governing the expression of focus.

Keywords: diachronic language contact, structural adaptations, PF level restructuring

Total word count: 512 (including title and keywords)

Key References

Borer, H. (2003). Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In J. Moore & M. Polinsky (Eds.), The nature of explanation in linguistic theory (pp. 31–67). Stanford: CSLI Publications.

Riksem, B. R., Grimstad, M. B., Lohndal, T., & Åfarli, T. A. (2019). Language mixing within verbs and nouns in American Norwegian. *The Journal of Comparative Germanic Linguistics*, *22*(2), 189–209. https://doi.org/10.1007/s10828-019-09109-6

Swenson, A. (2019). 6 . The Syntax and Semantics of Copulas in Malayalam. In *Malayalam Verbs: Functional Structure and Morphosemantics* (pp. 213-246). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9781501510144-006

Verbal grammatical reduplication in Sino-Tibetan languages of the western Himalayas: between genealogy and typology

Lars Borin University of Gothenburg
Anju Saxena Uppsala University

Reduplication is probably universal in language, at least in the form of iconical use of full reduplication of words or phrases for expressing intensity or emphasis.

It is also a very common grammatical device, found in languages all over the world where it is employed to express specific morphosyntactic features (Rubino 2005). However, even in such cases it is frequently iconical, prototypically encoding meanings such as plurality, intensity, repeated or distributed action, reciprocicality, etc. We find this kind of reduplication in all the major language families of South Asia (Abbi 1992). The form and function of verbal grammatical reduplication in the Sino-Tibetan (ST; also: Trans-Himalayan) languages of the western Himalayas are the focus of this presentation. In these languages, reduplication appears in the verbal morphology, forming categories such as converbs, nominalizations, and participles, which often probably are historically related, but also perfectives in some languages.

As far as our investigation shows, this kind of reduplication is not an areal phenomenon of the region: notably, these constructions are largely lacking from the Indo-Aryan (IA) and other languages spoken in this region. At the same time, even in cases where the same verbal morphosyntactic feature involves reduplication in different ST languages, they may employ formally different reduplicative constructions (1–3), which may simply be due to a typological tendency towards reduplication in converbs (Haspelmath 1995: 9). This naturally raises questions about the genealogical status of verbal grammatical reduplication in the ST languages of the western Himalayas, and in particular, how useful these constructions – sometimes quite specific (4) – are in teasing out the genealogical relationships among the ST languages of the region.

(1) Rongpo (rnp) (ST; Sharma 2001b: 37)
jəka jəka dhe di
eating eating he go
'While eating he went.'

(2) Byangsi (bee) (ST; Sharma 2001a: 108)
uo dza:laŋ dza:laŋ laiya
he eating eating slept
'While eating he slept.'

(3) Raji (rji) (ST; Rastogi 2012: 100, ex. (130))
əi ləgya jaja [ja~ja] rukka
3SG bread eat go
'While eating bread he went.'

(4) Kinnauri (kfk) (ST; Saxena 2022: 104, ex. (220))
gə dilli bjo-mu suntse~tse to-k
1SG.NOM p.name go-INF think~PFV AUX-1SG
'I have thought of going to Delhi.'

As part of a wider investigation of the areal, genealogical, and typological relationships among the languages of the western Himalayas, we have been collecting data on grammatical reduplication in verbs (for 26 ST languages and 42 IA languages, as well as Kurux [Dravidian] and Kusunda [language isolate]). The following table skeleton shows the formal properties of reduplication found in the ST languages of the area (column labels), as well as the verbal morphological features expressed by it (row labels).

form ⇒ ⇓ **FUNCTION** |	*full identical*	*full modified prefixed*	*full modified suffixed*	*partial prefixed*	*partial suffixed*
NOMINALIZATION					
PARTICIPLE					
CONVERB					
PERFECTIVE					

In this presentation, we will focus our attention on the data, describing and exemplifying the details of verbal grammatical reduplication hiding behind the column labels of the table above, and fill out the cells of the table with languages. Finally, we will also endeavor to discern and describe possible broader tendencies of the distribution of reduplicative constructions within and across languages, and discuss the genealogical implications of our findings.

References

Abbi, Anvita. 1992. *Reduplication in South-Asian languages: An areal, historical and typological study*. New Delhi: Allied.

Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective*, 1–55. Berlin: Mouton de Gruyter.

Rastogi, Kavita. 2012. *A descriptive grammar of Raji*. Delhi: Aviram Prakashan.

Rubino, Carl. 2005. Reduplication. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Available online at `http://wals.info/chapter/27`, Accessed on 2022-08-07. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Saxena, Anju. 2022. *The linguistic landscape of the Indian Himalayas: Languages in Kinnaur*. Leiden: Brill.

Sharma, Suhnu Ram. 2001a. A sketch of Byangsi grammar / Byangsi-English, English-Byangsi glossary. In Yasuhiko Nagano & Randy J. LaPolla (eds.), *Bon studies 3: New research on Zhangzhung and related Himalayan languages* (Senri Ethnological Reports 19), 271–341. Osaka: National Museum of Ethnology.

Sharma, Suhnu Ram. 2001b. A sketch of Rongpo grammar / Rongpo-English, English-Rongpo glossary. In Yasuhiko Nagano & Randy J. LaPolla (eds.), *Bon studies 3: New research on Zhangzhung and related Himalayan languages* (Senri Ethnological Reports 19), 195–270. Osaka: National Museum of Ethnology.

Loanword morphology as a window on linguistic connections: Sino-Tibetan and Indo-Aryan in the western Himalayas

Anju Saxena Uppsala University
Lars Borin University of Gothenburg

The western Himalayas are a multilingual region, where a large number of languages belonging to at least two language families are spoken, mutually unrelated Sino-Tibetan (ST; also known as Trans-Himalayan) and Indo-Aryan (IA). It is not unusual that languages belonging to both language families are spoken in the same village. Historical documents, too, record that speakers of ST languages have long been in contact with IA languages – through religion and trade. Since IA languages often have been and are sociolinguistically dominant in the region, IA loans are regularly found in the local ST languages.

Some unique linguistic adaptations of IA loans (nouns and verbs), whose origin is unclear and which are present in only a subset of ST languages (Kanashi, Kinnauri, Chhitkuli, Shumcho, Chaudangsi, Byangsi and Raji), give us some illuminating insights into the prehistory of this region, including indications of plausible migratory paths. This is the focus of this presentation.

The present study is based on data from primary fieldwork and secondary sources on 40 Sino-Tibetan language varieties. In addition Indo-Aryan (20 varieties), Dravidian (2 varieties), and Austro-Asiatic languages (Munda: 2 varieties) as well as the language isolates Burushaski and Kusunda spoken in the region are also included in this study, in order to see how widespread these phenomena are, and also what could plausibly be their source.

We have examined the following features in these languages: (i) the adaptive marker (also: nativizing morpheme) *-aŋ* on IA loans (e.g. Kanashi *kasaŋ* ‘bronze’; cf. Hindi *kās*, *kā̃sā*; Nepali *kā̃so* ‘bronze, pewter, white metal’); and (ii) the valency increasing marker *-ja* and the valency decreasing marker *-ed/-i* on IA loans (Kanashi: *kaʈ-jaː-* [cut-TR-], *kaʈ-e-* [cut-INTR-], Kinnauri: *kaʈ-jaː-* [cut-TR-], *kaʈ-ed-* [cut-INTR-], Rongpo *bhəbəṛ-* ‘to roast (IA.INTR)’, *bhəbəṛya-* ‘to roast (IA.TR)’).

Based on these empirical data, we can distinguish a set of ST languages (Kanashi, Kinnauri, Raji and, to some extent, Shumcho, Chaudangsi and Byangsi) that share these features, which suggests that they are either closely related or have been in close contact at some point in the past. Note in this connection that Kanashi, Kinnauri and Raji are today not spoken in close vicinity of each other, and further that they are not classified within the same sub-group of Sino-Tibetan.

In addition to the mentioned linguistic features, observations reported in the literature on local architecture, characteristic socio-cultural practices, and population genetics will be drawn upon to support the suggestion that these languages share a common past which singles them out from other ST languages of the region.

Telugu SL Gesture Recognition System

Atreyee Sharma (EFLU, Hyderabad)

In the paper I propose a continuous gesture recognition system of/for Telugu Signers of Hyderabad. Continuous sign language recognition poses significant challenges due to overlapping and transitional hand movements. To address this, I employed a gradient-based approach for extracting key frames, which enables effective segmentation of continuous signing into discrete units while filtering out redundant frames. After splitting of gestures each sign has been treated as an isolated gesture. The primary aim of the project is to develop a communication aid for Telugu signers which can be implemented in a mobile phone.

The Indian Sign Language dataset contains samples captured under diverse lighting environments and with varied hand orientations, ensuring robustness of the training data. Several machine learning models were explored for sign detection, including SVM, Logistic Regression, KNN, and deep learning methods such as CNN. To address the issue of limited training samples, the DeepSign framework was leveraged for improved learning efficiency

Next up was extracting features from voice. In speech processing, feature extraction involves reducing the raw audio signal into a compact representation that preserves distinguishing characteristics of each spoken unit. Then the result of this process is sent to Feature matching Hand gesture recognition which is of great importance for human-computer interaction (HCI). The dataset to train the software was gained from an online repository and also self-designed. During this process, Image Augmentation was also used to improve the efficiency. During segmentation, irrelevant background elements and noise are filtered out so that only the Region of Interest (ROI)—the hand gesture—is retained for further processing. Among various segmentation techniques, k-means clustering has proven to be an effective method for isolating meaningful regions in image-based classification tasks.

This Recognition system is capable of numerical sign with high accuracy using SVM (Support Vector Machine) Classifier. The SVM algorithm achieved good performance compared to other algorithms. The four parameters are used to calculate the accuracy of the trained dataset, i) True Positives (TP): where actual class and predicted class are yes. ii) True Negatives (TN): where the actual value and predicted value are both no. iii) False Positives (FP): where the actual class are no and predicted class are yes. iv) False Negatives (FN): where actual classes are yes and predicted classes are no. The formula for accuracy is:

$$\textbf{Accuracy} = \frac{TP+T}{FP+FN+TN}$$

Combining HOG feature descriptors with the SVM classifier produced the highest accuracy, reaching 97.1% in experimental evaluation. Beyond mapping gestures to words, the system also integrates video and text processing through Python's OpenCV library.

Future improvements include extending the system into an Android application and developing a multilingual sign language translator that incorporates sign dictionaries from various countries, thereby enabling broader communication among signers worldwide.

Key words: SVM, K-means Clustering, HOG features, CNN, KNN, DeepSign.

Some References

Akanksha Singh, Saloni Arora, Pushkar Shukla and Ankush Mittal, "Indian Sign Language Gesture Classification as Single or Double Handed Gestures", International Conference on Image Information Processing, 2015.

Ananya Choudhury, Anjan Kumar and Kandarpa Kumar, "A Conditional Random Field based ISL Recognition System under Complex Background". International Conference on Communication Systems and Network Technologies, 2014.

Divya Deora and Nikesh Bajaj, "Indian Sign Language Recognition", International Conference on Emerging Technology Trends in Electronics, Communication and Networking, 2012.

Kanchan Dabre and Surekha Dholay, "Machine Learning Model for Sign Language Interpretation using Webcam Images", International Conference on Circuits, System, communication and Information Technology Applications, 2014.

Sunny Patel, Ujjayan Dhar, Suraj Gangwani, Rohit Lad and Pallavi Ahire, "Hand-Gesture Recognition for Automated Speech Generation", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May, 2016.

S. Yarisha Heera, Madhuri K Murthy, Sravanti and Sanket Salvi, "Talking Hands- An ISL to Speech Translating Gloves", International Conference on Innovative Mechanisms for Industry Applications, 2017.

Variability in (non-)canonical word order and focus marking in Hindi/Urdu

The canonical word order in Hindi/Urdu is SOV but the constituents can scramble to encode information structure (Gambhir, 1981; Kidwai, 2000; Manetta, 2012). Existing studies on word order and focus marking in Hindi/Urdu are rife with contradictions. Based on examples from news articles and personal conversations, Gambhir (1981) claimed that new information is placed sentence finally. However, using semi-spontaneous speech Jabeen (2022) showed Hindi/Urdu speakers' preference to keep new information focus (NIF) in-situ. Jabeen reported that this preference was affected by the grammatical role (subject, (in)direct object) of the focused word. Importantly, prior research has not accounted for speaker-based differences in NIF marking. We argue that the apparent contradiction in Gambhir and Jabeen's claims can be attributed to individual differences in the preference for canonical (in-situ) vs. non-canonical (scrambled) position for NIF. This study explores how grammatical role (direct vs. indirect objects) and word order (in-situ, preverbal, postverbal) interact while accounting for speaker-based variation.

We conducted a production study using four dialogues (direct objects = 2, indirect objects = 2) extracted from a Bollywood movie script. Twenty-four Hindi/Urdu speakers were presented with dialogue contexts followed by the target sentence presented in different word orders. The alternatives differed in the position of the focused objects that were in-situ (preverbal for indirect objects, immediately preverbal for direct objects) or placed at a non-canonical position (preverbal for direct objects, immediately preverbal for indirect objects, postverbal for both). Focus was determined using the QUD framework (Riester, Brunetti and Kuthy, 2018). Participants were asked to read the context, select a word order they deemed appropriate in the context, and read out their chosen sentence. For the analysis, the position of the focused word in the target sentences was manually annotated. We ran Logistic Mixed Effects Regression to analyse if grammatical role can predict Hindi/Urdu speakers' preferred word order for focused (in)direct objects. Participants and dialogues were used as crossed random effects to allow for speaker- and item-based differences.

The analysis showed no effect of grammatical role on the position of the focused word ($p = 0.2$). 87% of participants preferred to keep the focused (in)direct objects (n=83) in-situ, thus confirming Jabeen's findings. Within these, speakers either preferred the word order with every constituent placed in-situ (64%) or the word orders where the non-focused object was scrambled (23%). However, Gambhir's claim can partially be upheld as 13% of focused words (n=13) were placed at postverbally. The investigation of speaker-based differences showed that 17% of the participants (n=4) used canonical as well as both the non-canonical positions at least once. Moreover, 71% of the participants (n=17) postposed the focused words. We also observed item-based variation as they were scrambled preverbally or postverbally with differing frequency.

Our findings indicate Hindi/Urdu speakers' differential use of in-situ vs. postverbal position to mark NIF. These results also illustrate that individuals vary in their preference for the non-canonical word order, thus adding nuance to the discussion of word order and focus marking in the so-called free word order languages.

References

Gambhir, Vijay. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi*. Ph. D.thesis, University of Pennsylvania.

Jabeen, Farhat. 2022. Word order, intonation, and prosodic phrasing: Individual differences in the production and identification of narrow and wide Focus in Urdu. *Languages* 7(2), 103.

Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. Oxford: Oxford University Press.

Manetta, Emily. 2012. Reconsidering rightward scrambling: Postverbal constituents in Hindi-Urdu. *Linguistic Inquiry* 43(1), 43–74.

Riester, Arndt, Brunetti, Lisa and Kuthy, Kordula De. 2018. Annotation guidelines for questions under discussion and information structure. *Linguistic Annotation Journal* 12(3), 45–67.

Hybrid language and multilingual pedagogy: The case of Eastern India

Long-term language contact created several hybrid languages in different patches of eastern India, which has been the homeland of multiple indigenous speech communities and attested melting pot zones. The hybrid languages are not found on a continuum, as the contact between different communities is different in nature, and the languages are found in different geographical areas . Rather, they are found in a vast area where pockets of resistance are found where indigenous languages are spoken and maintained, at times coexisting with the hybrids. These hybrid or mixed languages are found inside forest hamlets, in secluded hilly zones, and in a crowded economic hub, where different structures are evident and different varieties co-exist at times.

The paper aims to:

a) Discuss the case of Northern Bengal, India, which has borders with two neighbouring states and three countries where multiple hybrids are present— both named and unnamed.
b) Argues that the agreement patterns of the hybrids are affected by the choice of language in the official/public domain (Bangla, Nepali, Santali, or Hindi) and the solidarities of the speech communities.
c) Finally, the paper examines if the education policies affect the everyday use of the hybrid in the target area in formal schools, where they are either maintained or marginalised

Data:

a.	**hamin**	**sɔbeǰʰan**	**bʰɑt̪**	**kʰɑ-i-ŋ**
	1.pl	all	rice	eat-IMPV-1P
	‘We all eat rice’			

b.	**horʃiṭa**	**ar moiri**	**kali**	**bʰaṭ**	**kʰa-l-i/kʰa-l-ek**
	Harshita and Mauri		yesterday	rice	eat-PST-3P
	‘Harshita and Mauri ate rice yesterday’				

In the examples above, we find two different types of hybrid structures. The paper argues that these varieties are found not only in different locations; it depends on the population dynamics and the choice of language for the speakers of different hamlets. The names of the hybrids vary. In some cases, similar hybrids are put under a single umbrella. Otherwise,the same hybrids are considered a part of the same language.

The paper uses the concept of 'language making' (Kramer et al., 2022) as a major theoretical tool for analysis. I will analyse the agreement patterns of some hybrid language data, which we collected through extensive fieldwork. Finally, the use of the hybrids in formal schools of the target areas will be discussed, keeping in mind the new language-in-education policy recommendations in India that push multilingual pedagogy.

Key References:

Irvine, J. T. and Gal, S. (2000). Language ideology and linguistic differentiation. (P. V. Kroskrity, Ed.) Regimes of language : Ideologies, politics, and identities, 35-84.

Kramer, P., Vogl, U., & Kolehmainen, L. (2022, March 11th). What is language making'? International Journal of the Sociology of Language, 2022(274), 1-27. doi:https://doi.org/10.1515/ijsl-2021-0016

Nichols, J. (1992). Linguistic Diversity in Space and Time. Chicago: University of Chicago Press. doi: https://doi.org/10.1075/dia.11.1.11nie

A Crosslinguistic Study of Modal Expressions in Hindi, Marathi, Kannada, and Tamil

Riddhi Nirkhe

Keywords: Modality, Semantics, Polysemy, Hindi, Marathi, Kannada, Tamil, Typology

This research focuses on the patterns found in the verbal and grammatical expressions of selected modalities in Hindi, Marathi, Kannada, and Tamil. The objective of this paper is to contribute to the semantic and morphological understanding of these modal expressions in the selected South Asian languages and provide a base for further research. The analysis covers a total of 9 modalities including imperative, hortative, optative, obligative, ability, permissive, conditional, and prospective. Additionally, the research also discusses the polysemy patterns of modality found across these languages.

Marathi and Kannada were chosen for this study due to their contact history, as highlighted in the works of Southworth (1974, 2005), with a hypothesis of finding a connection between two languages in terms of modal expressions. Hindi and Tamil were chosen as the languages spoken in the neighbouring regions of Maharashtra and Karnataka which belong to the same language family respectively.

Linguistic expression of modality varies across different languages. The methods used to express modal categories are diverse and depend on the linguistic features of each language. Narrog (2016, 90-97) describes these expressions as explicit (overt) and implicit (covert). Languages that have lexical and grammatical means to express them are said to have explicit methods while languages that do not, are said to have implicit methods. Verbs/verbal auxiliaries, affixes, nouns, adverbs, and particles come under explicit expressions, and mood, voice, possession, aspect, etc. for modality are the implicit ways to express. As mentioned by Bhat (1999), the language will have well-defined paradigms for the prominent category while mood/modality will be expressed through loosely grammaticalized forms or through a combination of tense and aspect or tense and mood.

The analysis is done with the help of descriptive grammars of all four languages; Kachru (2006) for Hindi, Dhongde and Wali (2009) and Pandharipande (1999) for Marathi, Sridhar (1990) for Kannada, and Asher (1985) and Lehman (1989) for Tamil as well as the language family descriptions by Masica (1991) for Indo-Aryan and Krishnamurti (2003) for Dravidian. In addition to this, native speakers' judgments were also taken to cross-check the data and description.

References

Annamalai, E. and Steever, Sanford B. 2020. 'Modern Tamil', in S. B. Steever (ed.), *The Dravidian Languages,* 2nd edition. London and New York: Routledge, 104-130.

Asher, R.E. 1985. *Tamil.* London: Routledge.

Bhat, D. N. S. 1999. *The Prominence of Tense, Aspect and Mood.* Amsterdam: John Benjamins.

Dhongde, Ramesh Vaman and Wali, Kashi. 2009. *Marathi*. Amsterdam and Philadelphia: John Benjamins.

Kachru, Yamuna. 2006. *Hindi.* Amsterdam and Philadelphia: John Benjamins.

Krishnamurti, Bhadriraju. 2003. *The Dravidian Languages.* Cambridge: Cambridge University Press.

Lehmann, Thomas. 1989. *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Language and Culture.

Masica, Colin. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Narrog, Heiko. 2012. *Modality, Subjectivity, and Semantic Change: A Cross-Linguistic Perspective.* Oxford: Oxford University Press.

Narrog, Heiko. 2016. 'The Expression of Non-Epistemic Modal Categories', in J. Nuyts and J. van der Auwera (eds.), *Oxford Handbook of Mood and Modality.* Oxford: Oxford University Press, 89-116.

Nuyts, Jan and van der Auwera, Johan (eds.). 2016. *Oxford Handbook of Mood and Modality.* Oxford: Oxford University Press.

Pandharipande, Rajeshwari V. 1997. *Marathi*. London: Routledge.

Southworth, Franklin C. and Apte, M. L. (eds.). 1974. *Contact and Convergence in South Asian Languages,* International Journal of Dravidian Linguistics, Pp. 245.

Southworth, Franklin C. 2005. *Linguistic Archaeology of South Asia.* London and New York: Routledge.

Sridhar, S. N. 1990. *Kannada.* London and New York: Routledge.

Full Reduplication in Sanskrit:
A corpus study of verbal *āmreḍitas in Vedic and Classical Sanskrit*

Full or lexical reduplication involves the complete repetition of an entire prosodic word, including its inflectional morphology (Rubino 2005). This process is well attested in South Asian languages (Abbi 1992), where it serves multiple aspectual and discourse functions. In ancient Indo-European languages, it is reported to be a productive process only in Sanskrit (de Vaan 2015, Klein 2003). In his *Aṣṭādhyāyī* (ca. 400 BC), Pāṇini defines full reduplication as *sarvasya dve* 'two occur in place of one whole form' (P. 8.1.1). The final member of the iterated sequence is called *āmreḍita* (P. 8.1.2). According to the grammatical tradition, verbal *āmreḍitas* convey *nitya* 'over and over again' (i.e. *ābhīkṣṇya* 'continue repetition'), while case-marked forms express *vīpsa* 'pervasion' (i.e. distributive) (P. 8.1.4):

(1) a. *pacati* *pacati*
cook.PRS.3.SG cook.PRS.3.SG
'Cooks and cooks.'

b. *puruṣaḥ* *puruṣo* *nidhanam* *upaiti*
person.NOM.SG person.NOM.SG death.ACC.SG reach.PRS.3.SG
'Each and every person reaches death.'

Previous research has focused on the *Ṛgveda*, where verbal *āmreḍitas* are extremely rare, except for the imperative *píba-piba* 'drink, drink!'. Klein explained this absence as resulting from a functional overlap with the intensive, which encodes both iterative and intensive meanings through morphological reduplication (Whitney 1889: 363). Grieco (2023) extended this observation to a sample of *brāhmaṇa* prose, where only a small number of reduplicated verbs occur, mainly absolutives.

The present paper expands the scope in two directions. First, it enlarges the dataset to a digital corpus of about 5.5 million words spanning more than 1,600 years, from Vedic mantras and prose to a wide range of Classical Sanskrit genres. Second, it integrates corpus evidence with the Indian grammatical tradition in order to reassess the meaning and function of verbal *āmreḍitas*. The study extracted 605 verbal *āmreḍitas* using regular expressions and Python tools: 82 absolutives, 35 finite forms, 488 imperatives. Absolutives in *-tvā* and *ya* most clearly illustrate the *nitya/ābhīkṣṇya* meaning described in the tradition. Finite forms, by contrast, are concentrated in Vedic prose, especially in the *brāhmaṇas*, and frequently occur in clause final position. Their function is often emphatic or stylistic, sometimes even marking textual closure, which complicates the simple equation of reduplication with habituality. Imperatives dominate in Classical texts, especially in ritual and dramatic dialogue. In these contexts reduplication underscores urgency, intensity, or rhetorical force, rather than purely habitual repetition.

These findings suggest that full verbal reduplication in Sanskrit developed along two principal lines: (i) habitual or iterative marking, especially in absolutives; and (ii) emotive, rhetorical, or stylistic uses, especially in imperatives and certain finite forms. From the perspective of the grammatical tradition, it sits at the intersection of rule-governed *āmreḍita* and non-technical repetition, raising the question of whether these are distinct phenomena or different manifestations of a single underlying process. By combining corpus-based methods with traditional analysis, this study shows that full verbal reduplication in Sanskrit is not marginal, but a productive resource that adapts across genres and periods to serve different semantic and pragmatic functions.

References

Abbi, Anvita. 1992. *Reduplication in South Asian Languages. An Areal, Typological and Historical Study*. New Delhi: Allied Publishers Limited.

Cardona, G. 1996. 'Āmreḍita compounds?', *Vedavyākaraṇa-vyākhyāna: Festschrift Paul Thieme zum 90*. Geburtstag am 18. März 1995, in SCHMIDT, H. and WEZLER, A. (1996, Hrsg.), *Studien zur Indologie und Iranistik*, 20, Verlag für Orientalische Fachpublikationen, Reinbek, pp. 67-72.

De Vann, Michiel. 2015. 'Reduplicated Demonstratives in Ancient Indo-European Languages'. *Transactions of the Philosophical Society* 113(1), pp. 38–52.

Ditrich, Tamara. 2011. 'The typology of *āmreḍita* compounds in the *Ṛgveda*', *Acta linguistica asiatica* 1(1). 71–83.

Grieco, Beatrice. 2023. 'On *āmreḍitas* in Sanskrit: Evidence from Vedic prose', *Studi e Saggi Linguistici*.

Klein, Jared S. 2003. '*Āmreḍitas* and Related Constellations in the *Rigveda*'. *Journal of the American Oriental Society* 123, pp. 773–802.

Rubino, C. 2005. 'Reduplication: Form, function, and distribution'. In B. Hurch (ed.), Studies on Reduplication, Berlin/New York: Gruyter, pp. 11–30.

Sharma, R. N. 2003. *The Aṣṭādhyāyī of Pāṇini*. Vol. 6: English Translation of *Adhyāyas* Seven and Eight with Sanskrit Text, Transliteration, Word-Boundary, Anuvṛtti, Vṛtti, Explanatory notes, Derivational History of Examples, and Indices, Munshiram Manoharlal Publishers, New Delhi.

Speijer, J. S. (1886), Sanskrit Syntax, Motilal Banarsidass Publishers Private Limited, Delhi.

Whitney, W. D. 1889. *A Sanskrit Grammar*. Breitkopf and Härtel, Leipzig

Yates, A. D. 2014. 'The Anatolian āmreḍitas: Distribution, Function and Prehistory', paper presented at the 224th Annual Meeting of the American Oriental Society (Phoenix, March 15, 2014).

Revisiting Nominalization Patterns: Gender and Classifier Marking in Halbi and Bhatri

Husendra Sethiya
Adam Mickiewicz University in Poznań

Keywords: Classifier, Gender markers, Nominalizers, relativization, Relative clause.

This study investigates whether the classifier and gender marking patterns in Halbi and Bhatri conform to Shibatani's proposed nominalization hierarchy, and if not, how they diverge from the hierarchy. Our primary research question examines how these related but distinct languages challenge or support the universality of Shibatani's hierarchy (2023)

NP use: NUM > DEM > GEN > ADJ > V-based NMLZ

Mod use: NUM > DEM > GEN > ADJ > V-based NMLZ

Shibatani (2023) proposes a framework treating gender/classifier marking as a classifying nominalization phenomenon that constrains synchronic distribution patterns and predicts diachronic developments across languages. This hierarchy suggests that numerals are most susceptible to classifier/gender marking, followed by other structures in descending order.

Our data was collected through extensive fieldwork in the Bastar region of Chhattisgarh, and bhubhaneswar, Jharsuguda and navrangpur district of Odisha, India following Shibatani's (2023) methodological protocols for eliciting nominalization constructions.

Halbi and Bhatri are Indo-Aryan languages belonging to the Halbic subgroup of the eastern Indo-Aryan family. sharing approximately 60-65% lexical similarity and having over 76% bilingualism in the Bastar district (Beine 2017),

In Halbi, classifiers are restricted to numerals only (e.g., *dui-ʈʰən beʈa* 'two-CLF son'), while gender markers appear in the remaining categories except numerals. For adjectives and V-based nominalizations, gender marking is obligatory, In NP use where the adjective cannot appear without gender marking. Bhatri, conversely, employs classifiers across all adnominal categories, with gender marking possible everywhere except with numerals. Crucially, both languages exhibit mandatory marking for adjectives and V-based nominalizations, as in the Bhatri example (1) adjectives (2) V-based nominalization.

where the classifier/gender marker is required.

(1) niko **ʈa/biʈ̪-i** aj
good CLF/NMLZ-NM be.PRST
(this one) is good.

(2) renu-ər [anla **ʈa/biʈ̪-ɪ**]$_{NMLZ}$ ke]$_{NP}$ rədʒəʈ̪ kʰaj-la
Renu-GEN bring.PFV.3SG CLF/NMLZ-NM ACC Rajat eat-PFV.3SG
Whatever Renu brought, Rajat ate.

These patterns can be schematized as follows:

Halbi / Bhatri

- NP Use: NUM>^{OPT}DEM>^{OPT}GEN>**ADJ**>**V-BASED**
- Modification use: NUM>^{OPT}DEM>^{OPT}GEN>^{OPT}ADJ>^{OPT}V-BASED

(No mark= CLF/NMZL marking obligatory; ^{OPT}=optional marking; *=marking impossible)

In both Halbi and Bhatri languages, nominalization markers adhere to Shibatani's hierarchy (2023) up to the genitives. However, the mandatory use of nominalization markers in the NP-use with adjectives and v-based nominalization violates this hierarchy, which posits that nominalization should be optional or not possible beyond the genitives. We extend our comparative investigation to include Odia, Bengali and Assamese (Bez, 2022) Data offering a synchronic perspective on these variations.

References

Beine, D. 2017. *A Sociolinguistic Survey of the Bhatri-Speaking Communities of Central India*. SIL Electronic Survey Report.

Shibatani, M. (2023). Towards Dynamic Functional Typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan*, *12*, 1–52. https://doi.org/10.57565/hlj.12.0_1_3

Atypical compound verbs in IA – diachronic and typological insights

Krzysztof Stroński, Riddhi Nirkhe, and Husendra Sethiya
Adam Mickiewicz University in Poznań

Abstract

The present study focuses on the 'atypical' compound verb (CV) construction ('V1+V2'), where the V2 element is an intransitive light verb with a detransitivizing force (1).

1) Punjabi (own and Bhatia 1993: 252)

a.	*billī*	*ne*	*dud*	*piyā*
	cat.F	ERG	milk.NOM.M.SG	drink.PAST.M.SG
	'A cat drank milk.'			
b.	*billī*	*sarā*	*dud*	*pī* *gaī*
	cat.F	all.NOM.N.SG	milk.NOM.M.SG	drink go.PAST.F.SG
	'The cat drank up all the milk.'			

Given the fact that constructions of this type have a very low frequency, it is not surprising that they have not received much attention in the linguistic literature until recently. Drocco and Tivari (2020) is the first attempt to analyze 'atypical' compound verbs in Hindi from a typologically oriented perspective.

The aim of the present paper is, therefore, to demonstrate that:
a) there is notable variation in the syntactic behaviour of atypical CVs, which chiefly depends on the transitivity status of either V1 (2), V2 (3), or both (4);
b) atypical CVs can be traced back to early New Indo-Aryan (NIA) (5).

Moreover, atypical CVs can be analyzed as non-prototypical or antipassive lookalikes. The antipassive-like voice is marked analytically, which is a rare phenomenon (Zuñiga and Kittilä 2019: 104, 113). However, the flagging alternation observed in (1a) and (1b), as well as the loss of indexation, can only be attested in those Indo-Aryan (IA) languages that show a split ergative pattern. We suggest that the assumed antipassive-like status of the construction requires further verification based on a larger sample of modern New Indo-Aryan languages.

2) Nepali (Kantipur Daily 12/03/2017 accessed on 02.01.2024)

varṣauṁ-samma	*du:kha gar-ī*	*chorā-harū*	*hurkā-ẽ,*
year-till	sorrow do-CVB	boys-PL	raise.up-1SG.PAST
buḍhesakāl-mā	*euṭā-le*	*chāḍ-i-gay-o*	
old age-LOC	one-ERG	leave-CVB-go.PAST	

'I raised my sons with sorrow for years, one left (me) in old age.'

3) Marathi

tyā-ne *dzhop-ūn* *ghet-l-e*

he-ERG sleep-CVB take-PAST-3SG.N

'He slept.'

4) Hindi

dhobī *vahɑ̃* *se* *cal* *diyā*

washerman.NOM.M.SG there from go give.PAST.M.SG

'A washerman has gone from there.'

5) Old Rajasthani, A.D. 1450 (Smith 1976: 11, 62)

rŏva-t-ī *melh-i* *g-aŭ* *dhaṇa* *kau* *[re]*

cry-IPVF-F.SG leave-CVB go-PST.M.SG lady.F.SG GEN.M.SG [oh]

nāha-∅

husband-M.SG.**NOM**

'[Oh] the lady's husband left the weeping one (='left (her) weeping').'

Our preliminary analysis shows that a set of motion verbs, attested from the early periods of NIA, contributed to the emergence of atypical compounds. This process, however, has not yet been fully realized in several IA languages, such as Marathi or Kumaoni.

The research is based on corpora of early NIA varieties—such as Braj, Dakkhini, Awadhi, Rajasthani, Gujarati, Kumaoni, and Nepali—as well as modern NIA corpora, such as the Nepali National Corpus and Hindi Web 2012 (HindiWaC v. 4).

References

Bhatia, Tej. 1993. *Punjabi*. London and New York: Routledge.

Drocco, Andrea and Neha Tivari. 2020. Atypical Compound Verb Constructions in Hindi/Urdu The Case of Transitive Polar Verbs with Intransitive Vector Verbs. *Annali di Ca' Foscari. Serie orientale* Vol. 56. 307-334.

Zúñiga, Fernando, and Seppo Kittilä. 2019. *Grammatical Voice*. Cambridge: Cambridge University Press.

Smith, John D. 1976. *The Vīsaḷadevarāsa. A restoration of the text*. Cambridge: Cambridge University Press.