

Bhāṣā

Journal of South Asian Linguistics,
Philology and Grammatical Traditions

Vol. 4 – Num. 1

April 2025

e-ISSN 2785-5953



Edizioni
Ca' Foscari



e-ISSN 2785-5953

Bhasha

Journal of South Asian Linguistics,
Philology and Grammatical Traditions

Editor-in-chief
Andrea Drocco

Edizioni Ca' Foscari - Venice University Press
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL <https://edizionicafoscari.unive.it/it/edizioni/riviste/bhasha/>

Bhasha

Journal of South Asian Linguistics, Philology and Grammatical Traditions

Semestral journal

Editor-in-chief Andrea Drocco (Università Ca' Foscari Venezia, Italia)

Advisory board Gregory Anderson (Institute for Endangered Languages, Salem, USA) E. Annamalai (University of Chicago, USA; Central Institute of Indian Languages, India) Vit Bubenik (Memorial University of Newfoundland and Labrador, Canada) Paolo Di Giovine (Sapienza Università di Roma, Italia) Hans Henrich Hock (University of Illinois Urbana-Champaign, USA) Leonid Kulikov (Universiteit Gent, België) Malhar Kulkarni (Indian Institute of Technology, Mumbai, India) Silvia Luraghi (Università degli Studi di Pavia, Italia) Tiziana Pontillo (Università degli Studi di Cagliari, Italia) Adriano Valerio Rossi (Università degli Studi di Napoli «L'Orientale», Italia) Krzysztof Stroński (Adam Mickiewicz University, Poland) Boris Zakharyin (Moscow State University, Russia)

Editorial board Valentina Barnabei (Università Ca' Foscari Venezia, Italia) Erica Biagetti (Università di Pavia, Italia) Lucrezia Carnesale (Università di Pavia; Università degli Studi Bergamo, Italia) Maria Casadei (Jagiellonian University, Kraków, Poland) Bryan De Notariis (Università Ca' Foscari Venezia, Italia)

Managing editor Andrea Drocco (Università Ca' Foscari Venezia, Italia)

Head office Ca' Foscari University of Venice | Department of Asian and North African Studies | Dorsoduro 3462, 30123 Venice, Italy | bhasha_journal@unive.it

Publisher Edizioni Ca' Foscari | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venice, Italy | ecf@unive.it

The ISMEO co-financing was provided by Progetto MUR Storia, lingue e culture dei paesi asiatici e africani: ricerca scientifica, promozione e divulgazione CUP B85F21002660001.



ISMEO

© 2025 Ca' Foscari University of Venice

© 2025 Edizioni Ca' Foscari for the present edition



Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale
This work is licensed under a Creative Commons Attribution 4.0 International License



Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.

Scientific certification of the works published by Edizioni Ca' Foscari: all essays published in this issue have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Advisory Board of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Table of Contents

ARTICLES

**Revisiting Passivity and Politeness
in Selected Indic Languages**

Satyam Kumar, Chinmay Dharurkar 7

**Kirāti Toponyms: Semantic, Cultural
and Ecological Interpretations**

Bhim Lal Gautam, Nirajan Rai, Tara Mani Rai 31

**The Emergence of Rhymed Meters
in the Indo-Aryan Prosody**

Ryosuke Masaoka 51

**Code-Switching in South Asia:
Comparing the Equivalence Constraint
and Matrix Language Frame Models with Hinglish**

Anmol Sahni 77

Sonorant Gemination in Old Tamil

Ramprashanth Venkatakrishnan, Somdev Kar 95

**Semantics of Sanskrit *Eva*:
A Cognitive Linguistics Approach**

Yuliang Zhou 123



REVIEWS

Alyssa Ayres

***Speaking Like a State. Language
and Nationalism in Pakistan***

Marta Varini

157

Articles

Revisiting Passivity and Politeness in Selected Indic Languages

Satyam Kumar

IIT Kanpur, India

Chinmay Dharurkar

IIT Kanpur, India

Abstract This paper examines politeness and its realisation through passive constructions in Hindi, drawing on established theoretical frameworks. Using fresh survey data from near-native Hindi speakers whose first languages are Eastern Indic languages (EILs), it explores how passive constructions may get differently ranked in the politeness hierarchies. This differential ranking of the passives has been argued to be correlated to the first languages (the EILs) – a reason for the regional variation in politeness judgements, suggesting that speakers of EILs – familiar with explicit honorific markings – often perceive modal forms as more polite than passives. It also discusses the grammaticalisation of the motion verb *jānā* ('to go') in passive modal constructions, proposing that indirectness via passivisation can signpost politeness, though this perception varies across language backgrounds.

Keywords Passive structures. Realisation of politeness strategy. Politeness hierarchy. Grammaticalisation. Modal construction.

Summary 1 Introduction. – 2 Passives in Indic Languages. – 3 Passives on the Politeness Scale. – 4 Reconsidering the Judgements on the Passive. – 5 Further Grammaticalisation of *jānā* in EILs. – 6 Conclusion.



Peer review

Submitted 2025-02-06
Accepted 2025-05-12
Published 2025-07-01

Open access

© 2025 Kumar, Dharurkar | © 4.0



Citation Kumar, S.; Dharurkar, C. (2025). "Revisiting Passivity and Politeness in Selected Indic Languages". *Bhasha*, 4(1), 7-30.

DOI 10.30687/bhasha/2785-5953/2025/01/001

1 Introduction

The passive voice refers to a grammatical construction in which the subject of a sentence or clause refers to the action's recipient rather than the performer.¹ Passivisation involves addition of passive markings and agreement rules. At the same time, it also boosts the underlying object and demotes the underlying subject. From a pragmatic perspective, as discussed by Brown and Levinson (1978), passivisation functions primarily to minimise negative face by depersonalising the action and diffusing potential negative attributions towards the actor. In simpler terms, when the focus is on the action, the passive voice is used. Who or what is executing the action is unimportant or unknown or is deliberately avoided. Besides the passive constructions in Hindi, this paper presents active and passive modal constructions from Thethi, Magahi, and Angika. The data is collected from Jamui, Bihar (for Thethi), Jehanabad, Bihar (for Magahi), and Godda, Jharkhand (for Angika). These languages belong to the Eastern Indic language (henceforth EIL) family.

In Indic languages, the connection between language and politeness unfolds through various linguistic mechanisms, and passivisation is one among many realisations of politeness strategies. This linguistic phenomenon, while rooted in grammatical structures, extends far beyond syntax, serving as a subtle yet potent means of expressing politeness. Passive constructions enable speakers to shift the focus of a sentence, subtly emphasising the action or event rather than the agent responsible. This strategic shift in perspective aligns with cultural norms of humility and deference, essential elements of politeness across a range of languages, the focus here is on the selected Indic languages.

2 Passives in Indic Languages

Across diverse Indic languages, the use of passive structures as a realisation of politeness strategy is particularly pronounced in formal or deferential contexts. In languages like Hindi, Tamil, Bangla, Bhojpuri, Maithili, or Telugu, this politeness-driven passivisation is evident in various instances. For example, in Hindi, according to Kachru (2006), the passive construction is indicated by the inclusion

¹ Drawing on diverse perspectives, we examined various definitions of passives found in the works of Postal 1985; Givón 2006; Bresnan 1982. While passivisation demonstrably influences argument prominence, as highlighted by Blevins (2006), this effect can be viewed as a secondary consequence of the grammatical shift inherent in passive constructions. Consequently, defining passives solely based on their ability to invert prominence relations may be unnecessary.

of the passive auxiliary *jā* which is accompanied by the past participle of the principal verb. The motion verb *jā* has been grammaticalised to perform several functions in Indic languages, one significant function among those, arguably,² is expressing the passive sense. The tensed or modal form of *jā* carries the grammatical features of person, number, and gender, besides arguably partly the sense of passive itself. These are exemplified in the following sentences:

- (1) *kitabē* *paṛhī* *gāī.*
book.F.PL read.PERF.F PASV.F.PL³
'The books were read.'
(Kachru 2006, 93)

Similarly, a sentence like

- (2) *mair̥-ne* *kām* *ki-yā* *h-ai*
1.SG-ERG work do-PFV.M.SG be-PRS.SG
'I have done the work/task.'

It can be rendered in a more polite and indirect form through passivisation, as in (3):

- (3) *kām* *ki-y-ā* *ga-y-ā* *h-ai*
work do-PFV-M.SG GO_{PASS.} 4PFV-M.SG be-PRS.SG
'The task has been done.'

Including the works of Srivastava and Pandit (1988) and the works referred to therein, mainly Pandharipande (1979), the conventional understanding posits that the employment of the passive voice is commonly associated with creating a sense of distance between

2 This is being called arguably because, though, synchronically *jānā* seems to be the motion verb undergoing grammaticalisation, historically, it seems that it is related to Sanskrit-Prakrit passive base *ya* which has later changed to *jja* which then eventually started sounding like *jānā*. For instance, *chid + ya + ti > chijjati* (to pierce or to break) is a correspondence between Sanskrit and Pali. See Bubenik 1996, 118.

3 The sentence, glossing, and the meaning has been taken from Kachru 2006. Going by the glossing followed in this paper, this would look like:

kitāb-eṁ *paṛh-ī* *ga-īm̐*
book-F.PL read-PFV.F.SG GOPASS-PFV.F.PL

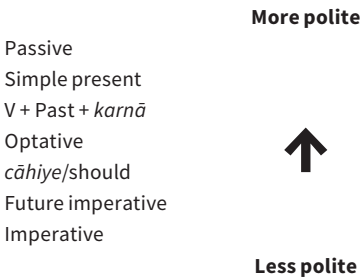
4 Here, it is understood to be the perfective passive form of the verb itself. Historically, it has been understood to be a passive participle form of the verb by which the verb gets gender marking which is typologically unusual. It should also be noted that passivity can be understood to be expressed at once by *kiyā* and *gayā* both.

the speaker and the listener. Consequently, it is inferred that its application is likely to be perceived as more courteous, especially in scenarios involving interaction between individuals of disparate hierarchical positions, where the speaker holds a lower status compared to the listener.

3 Passives on the Politeness Scale

To measure the degree of politeness, scholars such as Brown and Gilman (1960) and Brown and Levinson (1978, 79-89) have extensively explored the maintenance of distance between the addressor and the addressee. This inquiry unfolds along two primary dimensions: the aspect of ‘power’, encompassing respect derived from factors like status, authority, and age, and the examination of discourse strategies. Within the realm of discourse strategies, the employment of passive voice emerges as a notably effective technique, as elucidated by Lakoff (1977) among others. The distancing quality of passives aligns with the concept of politeness, where passives are deemed suitable for situations in which the speaker and listener lack intimacy, and the speaker seeks to refrain from intruding upon the hearer’s privacy. Notably, Pandharipande (1979) has proposed a politeness hierarchy for Hindi, positioning passives at the apex as the most polite forms, a categorisation illustrated in Table 1.

Table 1 Hierarchy of degree of politeness in Hindi, as per Pandharipande 1979



Pandharipande seems to argue that in Hindi, there is a special, polite way to suggest to an elderly person that they might be spoiling their children. This special way involves using passive construction. Other ways of saying it might be acceptable, but not as polite. Any ranking of preferences to be polite can be made more convincing by offering some rationale, other than the one emerging from the empirical survey, may be through the form-function interaction in (a) language. While the connection between passive and indirectness is found in many languages, it does not seem to be a necessary one (as

would be evident in the hierarchy presented by us which matches with that of Srivastava and Pandit). The argument presented in this paper is suggestive of the relation between the honorific agreement richness in languages and their interpretation of imperatives vis-à-vis passives. To truly understand how polite different sentence structures are, we need to consider the intuitions people follow when they talk to each other. These rules are not written down, but they are important for showing respect and making sure the message gets across in the right way. As Srivastava (1977) points out, the social meaning behind a sentence is just as important as the words themselves.

Revisiting Pandharipande's study and the dataset of Hindi speakers from the EIL family is crucial for two key reasons. First, it allows us to survey (more or less similar sentences) for a different group of participants, providing valuable comparative insights. This reinforces the generalisability of existing findings and expands our understanding of construction-based language variation within the EIL family. Second, while previous research, like Pandharipande's, has focused primarily on sociolinguistic variation – how social factors like age, gender, and social class influence language use – our study specifically addresses the understudied area of pragmatic variation. This refers to how the context of communication and the speaker's intentions shape their language choices.

So, Pandharipande's observation about polite ways to talk about grandparents/children in Hindi presents a fascinating starting point. Her work would benefit from a more detailed explanation of how she arrived at her conclusions, as well as a consideration of the social context that influences these conversations. This leaves some important questions unanswered and invites further research on how politeness and social meaning work together in a language or its varieties. This was a motivation for reconsidering some sentences and identifying the responses of speakers of EILs to the Hindi sentences.

The following is a brief survey, which may more accurately be described as a consultation with a reasonably varied group of individuals representing three Eastern Indic languages. Their responses to selected Hindi sentences are used to gather their judgements. The aim was to offer insights that may contribute to a more nuanced comparative analysis of regional variation in a language as expansive as Hindi.

4 Reconsidering the Judgements on the Passive

Taking into consideration Pandharipande's (1979) study, we conducted a small survey with similar sentence constructions of Hindi on twenty native speakers of Thethi, Magahi, and Angika (all of them are near-native speakers of Hindi). The survey was conducted primarily in-person and the participants were given a set of Hindi sentences, the details of which follow. For all participants – and indeed, for the region where these languages are spoken – Hindi is a second language. They were asked to rank the sentences using square boxes placed beside each. In addition to this ranking task, the surveyor engaged in discussions with the participants to understand their rationale for ranking the sentences and the contexts in which they would use them. For participants who were unable to read the sentences, responses were gathered through verbal discussions and an interrogation-based approach, ensuring that their perspectives were also considered. Additionally, to provide context, participants were presented with an imagined scenario for the utterance of sentences in the survey. For instance: 'Imagine that you need to admonish an elder; which sentence would you choose to sound polite?' The set of sentences used for the survey are as follows:

Passive

- (4) *vidyārthī-yom⁵ ko aise sajā⁶ nahīn diyā⁷ jā-t-ā*
student-PL.OBL DAT this way punishment NEG give.PFV.M.SG go-IPFV-M.SG
'Students are not to be punished like this.'

Habitual/⁸Simple present

- (5) *vidyārthī-yom ko aise sajā nahīn de-t-e*
student-PL.OBL DAT this way punishment NEG give-IPFV-M.PL
'(Teachers/people) do not punish students like this.'

⁵ Here the word *vidyārthī* has been considered as default singular form and hence the plurality is glossed against the oblique marker *yom*.

⁶ In standard Hindi-Urdu, the word is typically pronounced as *sazā*. However, while collecting this data, the participants consistently rendered the same word as *sajā* instead. It aligns with broader features of the regional varieties found in Bihar, Jharkhand, and Uttar Pradesh. Such shifts might reflect historical linguistic influences or phonological differences in the registers, diglossic usages, or some other axes of variation.

⁷ Complex predication including examples like *becnā*, *diyā*, *denā*, *ānā*, etc. do not go for any gender agreement and by default, these always remain as masculine singular in this variety. Hence, these are expressed in M SG.

⁸ We get a more habitual sense from this and the following examples, but this has been termed as simple present in Pandharipande's paper.

V+karnā

- (6) *vidyārthī-yorñ ko aise sajā nahīm dī-yā kar-t-e*
student-PL.OBL DAT this way punishment NEG give-PFV do-IPFV-M.PL
'(Teachers/people) do not (usually) punish students like this.'

cāhiye/should

- (7) *vidyārthī-yorñ ko aise sajā nahīm de-nā⁹ cāhiye*
student-PL.OBL DAT this way punishment NEG give-INF wish.SBJV
'(Teachers/people) should not punish students like this.'

Optative

- (8) *vidyārthī-yorñ ko aise sajā nā d-eñ*
student-PL.OBL DAT this way punishment NEG give-PL.OPT¹⁰
'(Please) do not punish the students like this.'

Imperative

- (9) *vidyārthī-yorñ ko aise sajā nā dī-jiye*
student-PL.OBL DAT this way punishment NEG give-IMP.H¹¹
'(Please) do not punish the students like this.'

Pandharipande's (1979) examples:¹²

Passive

- (10) *baccon ko istarah bigaaraa nahiin jaataa.*
Children- acc this way spoil neg passive marker
'Children are not to be spoilt like this.'

Simple present

- (11) *baccon ko istarah nahiin bigaarate.*
Children- acc this way neg spoil-imp
'(People) do not spoil children like this.'

9 Refer to fn. 7.

10 OPT stands for optative.

11 H stands for honorific.

12 The data and glossing in these examples are as per Pandharipande (1979) as quoted in Srivastava, Pandit 1988.

V + karanaa

- (12) baccon ko is tarah nahiin bigaaṛaa karate.
Children- acc this way neg spoil do
'(People) do not (usually) spoil children like this.'

Caahiye/ 'Should'

- (13) baccon ko is tarah nahiin bigaaṛannaa caahiye.
Children- acc this way neg spoil (non-finite) modal
'(People) should not spoil children like this.'

Optative

- (14) baccon ko is tarah na bigaaren.
Children- acc this way neg spoil (plural-imp)
'Please do not spoil the children like this.'

Imperative

- (15) baccon ko is tarah na bigaariye.
Children- acc this way neg spoil-imp (honorific)
'Please do not spoil the children like this.'

In the sentence set provided by Pandharipande, the propositional content remains unchanged ('someone will spoil the children in this manner'). In contrast, our study includes structurally similar sentences but with a different propositional meaning ('someone will punish the students in this manner'). The participants' engagement in ranking the given sentences according to their individual judgements and perceptions of politeness reveal noteworthy and systematic patterns. The responses exhibit how politeness is conceptualised, with respect to these constructions, across speakers of the EI languages. The following sub-section presents a detailed analysis of the survey findings, including the derived politeness hierarchy, through both tabular and figurative representations for clarity and comparative interpretation.

4.1 Tabulation of the Data

We found that among the sentences in our survey, participants ranked ‘passives’ at the bottom on the politeness scale, while ‘imperatives’ were placed at the top. The tables given below depict the rankings allotted by the participants to the various sentence structures. We have mentioned the age and language proficiency of the speakers and ranking allotted by them for the respective structures. Table 2 represents the participants and their ranking for different sentence structures.

Table 2 Participants and their ranking for different sentence structures

Sentence structures	Ranking by the participants (with gender, age and 1 st -2 nd language proficiency)																			
	A ^m	B ⁿ	C ^f	D ⁿ	E ^f	F ^f	G ⁿ	H ^m	I ^m	J ^f	K ^f	L ^m	M ^f	N ^f	O ⁿ	P ^m	Q ^f	R ^f	S ^m	T ^f
	26	29	32	27	63	58	41	22	47	59	33	41	19	25	27	36	39	44	76	68
	TH	TH	TH	TH	MH	TH	TH	TH	TH	TH	AH	TH	AH	TH	TH	AH	MH	MH	TH	TH
Passive	6	5	6	5	6	6	6	5	5	6	6	5	6	6	6	5	6	6	5	6
Habitual/Simple present	5	6	5	6	5	5	5	6	6	5	5	6	5	5	5	6	5	5	6	5
V+Past <i>karnā</i>	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Optative	2	1	2	2	1	3	3	2	1	3	2	3	2	2	2	3	2	1	3	1
<i>cāhiye/should</i>	3	3	3	3	2	2	3	3	2	3	2	3	2	3	3	2	3	3	2	3
Imperative	1	2	1	1	2	1	1	1	2	1	1	1	1	1	1	1	2	2	1	2

In Table 2, the abbreviations T, M, A, and H represent the participants’ proficiency in Thethi, Magahi, Angika, and Hindi, respectively. The first letter in the sequence denotes the participant’s first language, while the second letter indicates their second language, which is Hindi in all cases. For example, read TH as, the participant’s first language is Thethi and second language is Hindi. Additionally, superscripted ‘m’ or ‘f’ signifies the gender of the participants. However, since this is not a unified or homogeneous group of people but a diverse and variegated one, the idiosyncratic details on the individuals have not been specified. It has been ensured that diversity of the respondents has been retained on each parameter, i.e., age, sex, and language background. The numbers placed above each language proficiency label represent the age of the participants, providing a clearer demographic overview of the surveyed group.

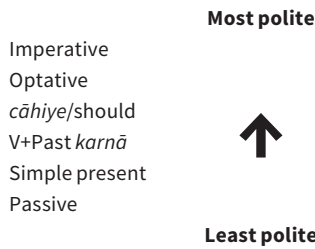
Table 3 Preferences of the participants for different sentence structures

Sentence structures	1st preference*		2nd preference (if any)		3rd preference (if any)		% of first preference
	Number of participants	Rank	Number of participants	Rank	Number of participants	Rank	
Passive	13	6	7	5	-	-	Least polite (65%)
Simple present	13	5	7	6	-	-	
V+Past <i>karnā</i>	20	4	-	-	-	-	
Optative	9	2	6	3	5	1	
<i>cāhiye/should</i>	14	3	6	2	-	-	Most polite (70%)
Imperative	14	1	6	2	-	-	

* Read the columns as: 13 out of 20 participants ranked passives on number 6.

The rankings assigned by participants in Table 2 are summarised in Table 3. Notably, the passive construction is ranked as the least polite by 65% of participants, while the imperative is considered the most polite by 70% of participants. Based on these survey results, the corresponding politeness scale is represented in Table 4.

Table 4 Politeness scale based on the result of this survey



Pandharipande (1979) presented the hierarchy based on the sentences of her survey and the responses of the consultants. The hierarchy she proposes is indeed in consonance with the principle of indirectness and the principle of optionality.¹³ We have conducted the survey by offering similar sentences to our consultants who are speakers of Thethi and some Bihari languages, and are near-native speakers of Hindi. The same has been dealt with in detail in the following section. Indeed, their perception seems to differ from the perception of Pandharipande's consultants.

13 These two principles have been elaborated upon in Section 4.2.

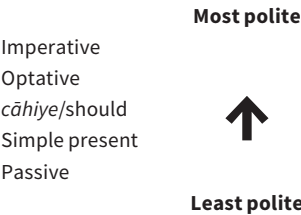
Our research examined how people politely suggest to elders that children should be taken care of in a better way. Interestingly, we found that direct requests phrased with respect ('Please don't punish them like this') were perceived as more polite than indirect expressions using the passive voice ('Students should not be punished like this'). This preference for the direct but respectful requests seems to have stemmed from a desire to avoid implying criticism or judgement. The passive voice, while traditionally considered polite, can carry an undercurrent of disapproval, suggesting that the elder is violating some unspoken precept. A direct request, when softened with a 'please' and appropriate honorifics, becomes more about seeking understanding and cooperation.

Furthermore, social hierarchy appears to play a role. Even individuals of lower social standing felt comfortable suggesting restraint through respectful requests, finding them less controlling than the indirectness of the passive voice.¹⁴ This finding contradicts traditional politeness guidelines, which often prioritise indirectness. However, this research suggests that when dealing with sensitive topics like precept, prioritising respect and avoiding any hint of censure can outweigh established grammatical conventions.

Interestingly, this trend is supported by another study, Srivastava and Pandit (1988), investigating similar dynamics in situations with unequal social status. Their finding aligns with ours, suggesting that the preference for respectful directness might be more widespread than initially assumed. Our politeness scale prepared based on the survey, almost completely matches the politeness scale of Srivastava and Pandit (1988). A slight difference is that we have also considered the 'V + Past *karnā*' constructions which are absent in their scale. Also, we have taken into consideration both status equals as well as status unequals. Table 5 shows the politeness scale proposed by Srivastava and Pandit.

¹⁴ See Ogiermann 2009, 191; Decock, Depraetere 2018, 36; Ruytenbeek 2020 for discussions on the relationship between indirectness and politeness.

Table 5 Politeness scale proposed by Srivastava, Pandit 1988



Subbarao et al. (1991), in line with Srivastava and Pandit (1988), examine syntactic strategies for politeness in Indo-Aryan and Dravidian languages, arguing that passive constructions are generally not employed as a primary means of achieving politeness.¹⁵ However, they note that in certain mundane contexts, particularly negative statements (it is customary in Telugu to say, *uppu ivavā* literally, ‘won’t you give the salt’ i.e., ‘please pass the salt on’), passives may contribute to mitigating face-threat or indicating politeness. In contrast, the present study demonstrates that in Thethi, Magahi, and Angika, passive constructions can function as realisations of politeness strategy in specific contexts, such as calls to joint action. This suggests that while passives may not be a conventional politeness device across Indo-Aryan languages, their pragmatic role can vary depending on sociolinguistic and contextual factors. Thus, the findings of this study both align with and extend the observations of Subbarao et al. (1991) by highlighting how passive structures can serve politeness functions in different discourse settings.

In conclusion, when offering suggestions about precept/child-rearing, particularly to elders, respecting the addressee and avoiding even unintentional objection may be more important than adhering to strict grammatical rules of politeness. This finding underscores the importance of considering social context and implicit assumptions when studying the nuances of language and communication.

4.2 **Reconsidering Indirectness and Optionality**

In the above sections, we saw that passive sentences have been ranked at the top of the politeness scale in the data surveyed by Pandharipande (1979), through principles such as indirectness and

15 Their classification of certain constructions as ‘syntactic strategies’ or ‘politeness strategies’ – without the qualifier ‘realisations of’ – appears to diverge from Brown and Levinson’s (1978) proposed universals.

optionality, whereas what Srivastava and Pandit (1988) found, was the other way round. Our participants too seem to align with the ranking presented by Srivastava and Pandit.

Brown and Levinson (1978) discuss several strategies, the one among them relevant for this paper is the impersonalisation mechanism discussed therein using the passive constructions, see Brown, Levinson 1978, 278-81. The principled understanding of the indirectness and optionality may be understood in the following way:

Indirectness is a strategy where direct expressions (like imperative forms used while giving commands) are avoided and some indirect ways of suggesting carrying out an action are preferred. An example in Thethi: *ab sute ke cāhī* 'Let us sleep now'.

Optionality is a strategy where the listener is offered a choice and thus softens the imperative force of a command or an appeal to carry out some action. This reduces the perceived imposition compared to direct commands and generates a sense of mutual respect. An example in Thethi: *ab khāyl jāy ki?* 'Shall we eat now/Let us eat now?'

By adhering to these principles, speaker balances between effectively conveying their message and maintaining positive social interactions. According to Leech (1983), by using indirect speech acts and offering more options, the speaker minimises the potential for conflict or negative interpretation by the hearer. This creates a more cooperative and considerate atmosphere in communication.

Passive constructions are central to this paper as it engages with the ranking of the passives on the politeness scale. The way passives have been defined also holds significance in relation to which aspect of passive construction does a definition hold central. One way is to focus on the influence of passivisation on argument prominence, as done in Blevins 2006, or the other is to focus on the grammatical shift inherent in passive constructions, and view argument prominence as a mere consequence of it (the same has been pointed out in Blevins 2006, 236). Whichever way does one define passive, these grammatico-pragmatic shifts do hold significance in relation to the sense of indirectness that passives bring to the table.¹⁶

16 The grammatico-pragmatic shifts mentioned here can be seen as the alternate forms carrying more or less the same denotative meanings. Although forms adopting new functions (such as becoming more indirect and thus more polite) may correlate with argument prominence, we do not explore that connection to make any definitive claims about it. For the very reason that when it comes to placing the constructions in the overall hierarchy, many other formal (of or related to the form of language) parameters come to converge, and each language or variety comes to assume (i.e., grammaticalise or pragmaticise) the nuanced senses of politeness in different ways. So, while the grammatico-pragmatic shifts may be correlated to the forms and their dynamics, what stands central to this paper is the difference among the varieties in relation to how they interpret the hierarchy in their own way(s).

Passives used in EILs to reduce face threats or enhance politeness often appear as impersonal constructions. The impersonal constructions exemplified in Section 5 make an interesting case regarding the degree of indirectness and therefore, politeness. We have referred to Blevins (2006) which offers a general encyclopaedic survey of core ideas related to passive and impersonal constructions. Impersonal constructions differ from passives in a number of significant respects, which have been summarised in Blevins 2003. We are inclined to identify example (16) and similar ones as impersonal constructions, not only because of the absence of the overt subject, but also because they seem to be historically related to the impersonal or *bhāve* constructions in Sanskrit.¹⁷

Our surveyed data shows differences in the arguments suggested by these two principles. For example, sentence number (9), which presents an imperative structure, is a direct speech act in itself. Despite this, due to the verbal inflection of respect and request used here, this sentence is interpreted by the speakers as a polite sentence and a polite speech act. Because of this, despite being a direct speech act, it emerges as a more polite sentence than other sentences. On the other hand, we can also see that through passive sentence, as in (4), a speaker, without imposing his/her command or intention on the listener, also provides him/her an option not to do that thing. Due to the absence of an explicit marker conveying respect or request, this sentence is interpreted as less polite by the consultants in this survey, and therefore, it got the lowest position on the politeness scale in our survey.

One generalisable insight that emerges through this is – if honorificity is morphologically encoded in the pronominal system of a language and the honorifics have corresponding imperative forms, then, despite being direct expressions, the directness is subdued by the explicitly marked honorific imperative forms of the verbs. Why

Given the multiple grammatico-pragmatic shifts offered by a language's grammar or morphology, alternate forms are likely to develop new or different functions. This can lead to a specific mapping between form and pragmatic interpretation, meaning these forms may undergo grammaticalisation in various ways across language varieties. The autonomy of these varieties lies in their distinct choices of grammaticalisation pathways; some may adopt one pathway, while others adopt another.

17 *bhāve prayoga* refers to a construction where the verb is in the impersonal or passive form, and the action is emphasised rather than the agent. This is commonly referred to as impersonal construction or passive construction (when the agent is omitted or deemphasised) or a less precise middle voice construction, in some contexts. The *bhāve prayoga* typically uses the third person singular verb form (often in the *ātmanepada* middle voice or passive) and does not specify an agent. For example, the active voice of an intransitive verb like 'laugh' in present tense construction *saḥ hasati* –he laughs. The impersonal form of this is – (*tena*) *hasyate* – can be roughly rendered in English as –(by him) it is laughed or laughing is done. For details, see Pāṇini sutra 3.4.69, for English translation and explanation, see Sharma 1995, 637-9.

would some languages/varieties choose indirect forms to be more polite while others would go for directness and clarity, seems to be one of the questions of typology and historical pragmatics at once.

It is interesting to note that the EILs encode honorificity even on the tensed forms of the verbs. And therefore, perhaps for the speakers of these languages, when they see concrete markings for each level of honorificity in the modal forms, they are more inclined to interpret the modal forms, despite (or rather, because of) their directness to be more polite than the indirect expressions like passives, and in addition to it, causatives, interrogatives-exclamatives, etc. (see Kumar 2024). From a syntactic perspective, the availability of multiple domains – such as DP, vP, and ForceP or CP – for encoding politeness or checking honorific feature (see Hong 2013 on similar line) allows EILs to express politeness mainly through imperatives and/or other direct constructions, rather than relying on indirect forms like passives, interrogative-exclamatives, or causatives. Additionally, distinctive features such as MAP and addressee/allocutive agreement¹⁸ further support this tendency in these languages. In other words, when politeness is explicitly marked in the structure of a language, it takes precedence over the indirect forms of expressing politeness. As we come westward (i.e., in the area of Central and Western Indic languages), such distinctions in the modal forms seem to decrease, and due to overall lack of morphological richness of the corresponding modal forms, the speakers (as in Pandharipande's set of consultants) would be more inclined to interpret passives as more polite.

Our findings suggest that, for a language like Hindi, which has a huge geographical expanse, there cannot be an absolute ranking of sentence structures that are rigidly associated with indirectness and optionality. Which structures are to be ranked more polite or less polite, would vary from variety to variety or language to language. Explicit modal inflections seem to be interpreted as more polite vis-à-vis the passive constructions. In some contexts, directness and clarity might be preferred. As far as the speakers of these languages of Bihar and Jharkhand are concerned, as has been discussed in the preceding paragraph, it seems that since verbal inflections and honorific markers are available in abundance in these languages to show respect towards an addressee, they tend to keep it specific and clear. Perhaps this is why these speakers neither hesitate nor see the need for using passive structures or indirect methods over active structures or direct methods.

18 See Bhattacharya 2016; Alok 2021; Kumar 2024; Raina 1994; Antonov 2015 for details on these phenomena across world languages

Pandharipande's ranking of passives based on these politeness principles may not fully reflect how speakers of EILs and nearby regions perceive and use them. While these principles offer a theoretical framework, our survey suggests that the speakers of these languages prioritise different factors when expressing and judging politeness through passives. For them, the entire passive construction seems to act as a single politeness marker rather than individual grammatical elements like verbal inflections and vocatives. This can be understood as an instance of periphrasis.¹⁹ This suggests that passivisation inherently expresses politeness (albeit to a slightly lesser extent than direct honorific imperatives, particularly in contexts where direct address is deliberately avoided), without needing to adhere to specific rankings based on individual features. This flexibility allows speakers to adapt the form and level of politeness depending on the context.

Our findings indicate that the passives, previously suggested as the most polite option based on the discussed principles and Pandharipande's regional data, may not universally translate to greater perceived politeness. This is further supported by similar observations in Srivastava and Pandit 1988, 199. These variations suggest that politeness interpretations and applications can differ across varieties or languages or contexts. Since we are looking at the hierarchy as a spectrum, it may be observed that for the Bihari speakers, passives could still, despite being on the other end of the spectrum, be a general politeness tool. Further research is needed to understand the specific factors influencing their perception and use of passives for expressing politeness.

In the above sections so far, we have seen how the survey conducted by Pandharipande is based on two principles and we have also seen that when we surveyed the similar data set on the speakers of EILs, the results are different. Since Hindi is a second language of these speakers, we wanted to study how they use passives in their mother tongue, i.e., in Thethi, Magahi, and Angika. The following section is not presented as a further proof for how the informants treat passive constructions differently from standard Hindi speakers. The idea here is to see, rather in what ways, the sense of politeness is visible in certain passive/impersonal constructions at once in

19 It is interesting to note that the indirectness conveyed through passives may be leading to the sense of politeness. It is further interesting to note that apart from the indirectness in the sense of the passive constructions, the constructions themselves exhibit an instance of longer phrasing, and in some cases the passivity is not just marked on one word but is understood through a combination of words in a phrase. For instance, in *kiyā jānā* – *kiyā* is a perfective form followed by the motion verb *jānā* which has been grammaticalised to convey passivity. We call such passive construction analytic for the passivity is conveyed here through a combination of words in the phrase.

EILs and the Hindi spoken in and around this region. No empirical survey is conducted for this, mainly because this is presented as a similarity that may be further investigated. One of the authors of this paper is a native speaker of Thethi and also speaks Hindi spoken in this region, and so the suggestive direction is based on the native speaker's intuition which is indeed open for investigation. The merit of this correlation lies in the formal similarity of the modal passive construction which is generally related to grammaticalisation of *jānā*. Let us take a look at the passive constructions in these languages.

5 Further Grammaticalisation of *jānā* in EILs

Passives fall low on the politeness scale as is evident in our survey. Still, being low on the scale means a certain degree of politeness, though the lowest, is associated with them. The relationship between passives and politeness cannot be denied altogether. To understand this better let us have a look at the EILs which use passive constructions to express politeness. For this, we took three languages: Thethi, Magahi, and Angika. In all these languages, grammaticalisation of *jānā* 'to go' partly contributes to passive constructions. The grammaticalised motion verb also shows up in modal constructions which have a passive-like syntax and appear to be impersonal, more like the Sanskrit *bhāve* mentioned earlier. The same have been exemplified in sentences (16) onwards. While we are concerned mainly with the passive forms, we are deliberately positing the corresponding active constructions to make the contrast evident, especially with respect to the presence of 'GO_{PASSIVE}' and the degree of politeness.

5.1 Thethi

Thethi shows synthetic and analytic realisations of passives. Verbs are first changed into their participle form by adding a marker *-al-* or some of its variants like *-yl-* (the palatalised form), that is how the passives are partly synthetic, and then followed by an auxiliary motion verb *jāy* (the analytic realisation) 'to go' at the end. This process makes the expressions passive, which start functioning like a polite expression without overtly marking subject in the constructions. Following are some examples where the passive constructions have been listed first, followed by the probable corresponding active constructions.

- (16) *khel-al* *jā-y*²⁰
play-PST PTCP GO-OPT²¹
'Let us play.'²² or '(Please) come to play.' [literally, 'let it be played now.']
- (17) *ab* *kha-yl* *jā-y*
now eat-PST PTCP GO-OPT
'Let us eat now.' or '(Please) have a meal.'

Corresponding active constructions of (16) and (17):

- (18) *cal* (*kirket*) *khel-iy-ay*
come.IMP.1.NH cricket play-1-D₁.²³SBJV
'Let us play cricket.'
- (19) *he-re,* *cal* *kha-iy-ay*
VOC.2-M.SG.NH come.IMP.1.NH eat-1-D₁.SBJV
'Hey (friend/younger brother), let us eat.'

Let us look at the passive of tensed construction in Thethi. These are the typical canonical passives where, like Hindi, the motion verb is grammaticalised for the sense of passivisation, as shown in (20):

- (20) *cor-ā* *pakr-āy* *ge-l-ay*
thief-M.SG catch-PFV GO_{PASS}-PFV-3.NH
'The thief got caught.'

Constructions (16) and (17) show evidence of synthetic as well as analytic passives where both the realisations are carried out by tweaking the main verb into its past participle form and adding

20 It is common in Sanskrit to use imperative passive (*ājñārtha* or LOT passives. For instance, *karotu vis-à-vis kriyatām* (i.e., *Do* vs. *May it be done*). Similarly, *bhāve* constructions are also unique to Sanskrit. Indeed, in Sanskrit too, these are understood to be more polite than their active counterparts.

21 We are calling it a passive optative because the structure *khelal jāy* also appears in passive constructions. For instance, in example *ī khel hamrā taraph khelal jāy hai* (This game is played in my region).

22 In this sentence (and also in the following ones), even though the verbs appear in their singular form, the way they are used together implies that there are multiple participants (usually 1P and 2P) involved. This makes a direct word-for-word or literal translation tricky, so we have focused on conveying the overall meaning instead.

23 *-ay* in Thethi (and Angika, and *-ai* in Magahi), appearing here and at other places in this paper, due to their consistent occurrence with first person and third person non-honorific subjects, have been glossed as default (D1) (see Kumar 2024 for more details).

an auxiliary *jāy* after it. Examples (18) and (19) present the active counterparts of the previous two passive constructions. (20) is an example to show how the morph *-ay* as the past participle marker to the verb root is used to transform the verb as well as the construction into a passive one.

5.2 Magahi

Magahi shows a similar way to construct passives by adding the auxiliary motion verb *jāy*, as in (21), after the past participle form of the verb in the constructions. Like Thethi, *-al-* is the most usual morphological suffix to be added to verb, converting it to past participle form, as can be seen in the constructions below. This makes the process synthetic as well. For example:

- (21) *baiṭh-al* *jā-y*²⁴
sit-PST PTCP go-OPT
'(Please) be seated.' (Singh et al. 2014, 115)

Example (22) presents the active counterpart of the sentence in (21):

- (22) *cal* *baiṭh-i-ai*
come.IMP.1.NH sit-1- D₁.SBJV
'Let us sit.'

In the following examples, (24) presents the active counterpart of the passive construction with the past participial form of the verb 'sleep', as shown in (23):

- (23) *ab* *kirkeṭ* *khel-al* *jā-y*
now cricket play-PST PTCP GO-OPT
'Let us play cricket now.' or '(Please) play cricket now.'

24 Singh et al. (2014) offer a different gloss for the sentence:

- baiṭhal* *jāy*
sit-pass go-PASS
'Please be seated.'

They seem to consider *-al* and *jā-y* together working towards the sense of passive. However, our gloss here differs from theirs because we associate *baiṭh-al* or *-al* ending forms of nouns with the participial forms and *jāy* as an optative form. We consider passivity as a consequence of these morphological details. *baiṭh-al* appears in other Magahi or Thethi sentences, for example *ayse nāy baiṭhal jāy hai* 'This is not the right way to sit' and so on in other tensed constructions. It would be inconsistent to gloss it as participle in these constructions and as passives in (21).

- (24) *cal* *kirkeṭ* *khel-i-ai*
come.IMP.1.NH cricket play-1- D₁.SBJV
'Let us play cricket.'

5.3 Angika

Synthetic and analytic realisations of passives are present in Angika as well. Like Thethi and Magahi, it adds the form *jā-y* (auxiliary verb of motion), as in (25) and (26), after the past participle form of the active verb to make the expression passive. Marker for the past tense in Angika is usually *-l-* and then a following morph *-ô* forming the participle, as evident in the following constructions:

- (25) *cal-l-ô* *jā-y*
come-PST-PTCP GO-OPT
'Let us go.' or '(please) Come.'
- (26) *sut-l-ô* *jā-y*
sleep-PST-PTCP GO-OPT
'Let us sleep.' or '(Please) sleep.'

Following is the corresponding active construction of (25):

- (27) *cal*
come.IMP.1.NH
'Let us go.' or 'Come.'

In all the three languages we discussed in this section, we observed that speakers generally use passive constructions (which have an impersonal thrust in modal constructions) either to express respect or to avoid the risk of disrespecting the addressee. The active counterparts of these constructions also exist very commonly, but they do not seem to be as polite as the passive ones. This feeds into saving the 'face' of an addressee. Such constructions are also generally used in situations when the speaker is not familiar with the person he/she is addressing, and he/she has to express respect towards them.

Table 3 presents the observation regarding passives in Thethi, Magahi, and Angika with clear distinctions between participial suffixes, the auxiliary verb of motion added, and the type of passives:

Table 6 Passives in Thethi, Magahi, and Angika.

	Participial Suffix	Auxiliary Verb	Form of Passives
Thethi	-al/-yl	<i>jāy</i>	Synthetic and Analytic
Magahi	-al	<i>jāy</i>	Synthetic and Analytic
Angika	-l-ô	<i>jāy</i>	Synthetic and Analytic

Table 6 shows that all three varieties exhibit the synthetic and analytic realisations of passives. The auxiliary verb of motion, i.e., *jāy* appears across the three languages. Participial suffixes with morphological variations precede the auxiliary verb of motion which makes this pattern quite similar in these varieties.

6 Conclusion

Let us have a look at the general overriding concerns discussed in this paper.

In the hierarchy presented by Pandharipande (1979), the passive has been placed at the top on the basis of the indirectness and optionality principles. The result presented by Srivastava and Pandit (1988) more or less inverts the hierarchy. When we surveyed similar data on speakers of EILs, the passive was found to be at the bottom of the hierarchy, which matches with the result of Srivastava and Pandit's. This paper, however, does not intend to falsify the research of Pandharipande nor verify the study of Srivastava and Pandit, rather, the idea is to see how the judgements on the politeness hierarchies vary based on the region and the first languages of the near-native speakers of Hindi.

We also saw that in languages like Thethi, Magahi, and Angika, some modal constructions, going by their morphological make-up, appear to be passives. In addition to the fact that the morphological or the grammaticalised form conveys passivity, we have also discussed how some constructions could be understood as their active counterparts. These active counterparts, in turn, make the indirectness evident in the passive constructions, and the ensuing politeness of the modal constructions accrued historically through grammaticalisation.

It is interesting to note that the passive construction was ranked lowest by speakers of EILs in Hindi. However, in their first languages, the passives of modal constructions appear to be more polite than their active counterparts. So, passivity indeed, has a considerable connection to indirectness, and therefore politeness; only that in comparison to other non-passive constructions – its position on the spectrum of politeness varies.

One typological and perhaps areal implicational feature or a correlation that may emerge from this paper, which of course needs to be fortified by further empirical research, is the presence of personal pronominal honorificity and the corresponding verbal modal imperative forms lead to diminishing of the sense of politeness expressed by the passive constructions.

In future studies, it can be seen what kind of behaviour this passive has in other languages of the world. It can also be studied whether there is a typological implicational relation between indirectness and politeness. The way in which passive constructions have been studied in relation to politeness, more such constructions, like causatives, interrogatives-exclamatives, past tense, etc., other than passives too, could be studied in relation to politeness or other strategies, and more such differences across the regions could be mapped to study several aspects of pragmatic variation. It is the intuitions of the speakers that seem to cause the variation, in our case, for example, the ranking of passive constructions to be lowest, or as the set of consultants in the study of Pandharipande, the ranking of passive constructions is highest. What are the underlying factors that dictate these intuitions – whether it is the overall inclination of the community to interpret a construction with a certain degree of politeness, whether there are any historico-morphological reasons that guide the speakers’ intuitions, whether there are any typological implicational universals behind this or whether there is some areal linguistics at work here are the areas that may elucidate our understanding of such variation. It is in this sense that this paper does not intend to falsify or verify objectively the conclusions of the previous studies but holds that variation is inevitable and that the causes for the same should be extensively studied.

The discussion in this paper underscores that politeness hierarchies are not static but may vary according to region and first languages of the speakers. The modal passive constructions (and several other constructions, as well) in Thethi, Magahi, and Angika serve as invitations to joint action and are thus perceived to be polite. However, for the same speakers, the passive construction in Hindi which evokes the sense of reproach are ranked lowest on the politeness scale. This contrast highlights the complex interaction between morphosyntax and pragmatics. This variation suggests that politeness is not solely a matter of indirectness but is mediated by how specific constructions are conventionally understood within a linguistic community. Future research comparing semantically equivalent passive constructions across these languages could provide a clearer picture of how linguistic background influences politeness perception, further refining our understanding of pragmatic variation in Indic and/or non-Indic languages.

Bibliography

- Blevins, J.P. (2003). "Passives and Impersonals". *Journal of linguistics*, 39(3), 473-520.
- Blevins, J.P. (2006). s.v. "Passives and Impersonals". Brown, K., *Encyclopedia of language and linguistics*, vol. 1. Amsterdam: Elsevier.
- Bresnan, J. (1982). "The Passive in Lexical Theory". Bresnan, J. (ed.), *The Mental Representation of Grammatical Relations*. Cambridge (MA): MIT Press, 3-86.
- Brown, P.; Levinson, S. (1978). "Universals in Language Usage: politeness phenomena". Goody, E.N. (ed.), *Questions and Politeness: Strategies in Social Interaction*. Cambridge: Cambridge University Press, 56-289.
- Brown, P.; Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Brown, R.; Gilman, A. (1960). "Pronouns of Power and Solidarity". Sebeok, T.A. (ed.), *Style in Language*. Cambridge (MA): MIT Press, 253-76.
- Bubenik, V. (1996). *The Structure and Development of Middle Indo-Aryan Dialects*. New Delhi: Motilal Banarasidass.
- Decock, S.; Depraetere, I. (2018). "(In)directness and Complaints: A Reassessment". *Journal of Pragmatics*, 132, 33-46.
- Givón, T. (2006). "Grammatical Relations in Passive Clauses". *Passivization and Typology*, 337-50.
- Hong, S. (2013). "Feature-Based Honorifics and Politeness in English and Korean at a Syntax-Pragmatics Interface". *Studies in generative Grammar*, 23(2), 199-219.
- Kachru, Y. (2006). *Hindi*. Amsterdam: John Benjamins.
- Kumar, S. (2024). *Politeness and Honorificity in Thethi, Magahi, and Angika: Towards Pragmaticisation* [Doctoral thesis]. Kanpur: IIT Kanpur.
- Lakoff, R. (1977). "What You Can do with Words: Politeness, Pragmatics and Performatives". Rogers, A.; Wall, B.; Murphy, J.P. (eds), *Proceedings of the Texas conference on performance, presupposition and implicatures*. Charlottesville (VA): Centre for Applied Linguistics.
- Leech, G.N. (1983). *Principles of Pragmatics*. London: Longman.
- Ogiermann, E. (2009). "Politeness and In-Directness Across Cultures: A Comparison of English, German, Polish and Russian Requests". *Journal of Politeness Research*, 5, 189-216.
- Pandharipande, R. (1979). "Passive as an Optional Rule in Hindi Marathi and Nepali". Kachru, B.B. (ed.), *South Asian Languages Analysis*. Urbana (IL): University of Illinois, 88-106.
- Postal, P.M. (1985). *Studies of Passive Clauses*. Albany (NY): State University of New York Press.
- Ruytenbeek, N. (2020). "Do Indirect Requests Communicate Politeness? An Experimental Study of Conventionalized Indirect Requests in French email Communication". *Journal of Politeness Research*, 16(1), 111-42.
- Sharma, R.N. (1995). *The Aṣṭādhyāyī of Pāṇini*, vol. 3. New Delhi: Manushiram Manoharlal.
- Singh, S.; Kumar, R.; Atreya, L. (2014). "Politeness in Language of Bihar: A Case Study of Bhojpuri, Magahi, and Maithili". *International Journal of Linguistics and Communication*, 2(1), 97-117.
- Srivastava, R.N.; Pandit, I. (1988). "The Pragmatic Basis of Syntactic Structures and the Politeness Hierarchy in Hindi". *Journal of Pragmatics*, 12(2), 185-205.
- Subbarao, K.V.; Agnihotri, R.K.; Mukherjee, A. (1991). "Syntactic Strategies and Politeness Phenomena". *International Journal of the Sociology of Language*, 35-53.

Kirāti Toponyms: Semantic, Cultural and Ecological Interpretations

Bhim Lal Gautam

Tribhuvan University, Nepal

Nirajan Rai

Mahendra Multiple Campus-Dharan, Tribhuvan University, Nepal

Tara Mani Rai

Central Department of Linguistics, Tribhuvan University, Nepal

Abstract This paper explores various Kirāti toponyms in various ways i.e. within and beyond linguistic interpretations by following the semantic, ecological and ideological interpretations based on the socio-ethnographic method of data collection where we have used observation, narratives and storytelling tools among the Bāntāwā and Dungmāli people. The data were collected from the extensive field visit in the four cultural regions of the central part of eastern hilly region, namely Amchok, Dilpā, Hatuwā and Dungmā, the homelands of Bāntāwā and Dungmāli languages speaking people. The analysis of the data shows that Kirāti toponyms have various types like khā, lā, tang, wāng and other toponyms that represent various semantic interpretations. From the narratives of Bāntāwā and Dungmāli people, we have discovered different morphemes and their place names that were connected to culture and identity. It is concluded that Kirāti toponyms have been used, interpreted and changed randomly that needs a detailed study Kirāti based on cultural and ecological perspectives.

Keywords Kirāti. Dungmāli. Bāntāwā. Khuwālung. Identity.

Summary 1 Introduction. – 2 The Context. – 3 The Methods. – 4 Analysis and Discussions. – 5 Shifting Trends in Toponyms and Language Ideology. – 6 Conclusion and implications.



Peer review

Submitted 2025-02-24
Accepted 2025-05-23
Published 2025-07-01

Open access

© 2025 Gautam, Rai, Rai | 4.0



Citation Gautam, B.L.; Rai, N.; Rai, T.M. (2025). "Kirāti Toponyms: Semantic, Cultural and Ecological Interpretations". *Bhasha*, 4(1), 31-50.

1 Introduction

Place names or toponymies are human creations to show the identity of people and culture (Kapur 2019). They reflect the migrations of people, their religious and cultural traditions, local languages, conquests, fortifications, topography and the developments (Everett-Heath 2000; Gautam, Giri 2024). The eastern hilly region of Nepal has been inhabited by Kirāti people since the prehistoric era. The toponyms show multiple evidence to show the Kirāti people as the first settlers of the region. Limbu, Yakkha, Rai and Koits (Sunuwar) are identified as Kirati people with distinct linguistic identities. However, Rai is the collective name of twenty-five distinct mother tongue speaking communities including Bantawa, Chamling, Koyee, Thulung, Yamphu and Lohorung. The Mundhum ritual language is the cultural foundation of Kirāti people. The land inhabited by Kirāti people spreads between Tamakoshi river in the west to Nepal's eastern border in the east. The toponyms found in this region exhibit consistent linguistic patterns and significant cultural nuances of the Kirāti people. The rivers, hillocks, high hills, mountains, cliffs, forests and terrains are identified not only with their spatial dimensions but also their connection with the livelihood related ethnic peoples' past activities. Moreover, the physical dimension of the places, perception towards land and vegetation of the ethnic Kirāti people and a legacy of their cultural and historical activities are interwind in the toponyms. A significant discourse on toponyms has emerged in the country recently, especially during the naming of local administrative divisions, however Kirati toponyms have not got systematic inquiry except few, such as Rai and Chamling (2017), Rapacha (2024) and Rai (2024). However, these inquiries also exhibit methodological limitations. As a result, they are largely neglected, misinterpreted and influenced by the Nepali language and its dominance in the past (Gautam, Poudel 2022) and being replaced with fancy names coined from Nepali and English languages at present. Moreover, like other ethnic communities in Nepal, Kirāti people are shifting towards various new identities.

However, Kirāti toponyms have not got systematic inquiry yet. As a result, they are largely neglected, misinterpreted and colonized in the past and being replaced with fancy names coined from Nepali and English languages at present.

Kirāti toponyms share significant similar structures and semantic properties among different Kirāti groups of languages across the region. They can be grouped and classified in terms of their meaning-based affixation. Rai and Chamling (2017) have presented a structural study of few toponyms of each 20 different Kirāti languages ranging from Limbu in the east to Sunuwar in the west which discusses structural properties of the toponyms followed by

their general classification based on their structure-led meaning. Rapacha (2024) discusses the deep-rooted identity of Kirāti-Koits people in the toponymic lexicons of the Kirāti-Koits language as archeological evidence of first settlers in the Ollo Kirat. On the other hand, Rai (2024) has questioned the practice of etymological analysis of Kirāti toponyms, where he argues that etymological analysis of Kirāti toponyms is incomplete, misleading and erroneous because it lacks ecological analysis and interpretations. Rai (2024) has analyzed a few Kirāti toponyms from the ecological and cultural perspectives for the first time.¹ After the promulgation of the Constitution of Nepal (2015), the country was restructured in 753 municipalities. The issue of naming the administrative division in the new federal country Nepal has recently been a central debate in the political sphere. The issue led to a political and ideological conflict between the state and the various ethnic people of the country. The discourse of identity movement in the country led to the naming of some municipalities based on the linguistic and cultural identity of ethnic people.

In the present study, we sought to interpret Kirāti toponyms based on cultural and ecological perspectives. Specifically, the paper focuses the following questions:

- a. What are the different Kirāti toponyms based on different semantic properties?
- b. Why do Kirāti toponyms are explained from cultural and ecological perspectives?
- c. How are the Kirāti toponyms shifting in the modern time?

2 The Context

The study of toponyms and its context has not been well established in Nepal. A place name is more than just a word or words that describe a type of feature (Randall 2001, 4). Within Kiranti culture, the case of ‘Khuwālun’,² a remarkable historical place located in the confluence of Arun, Tamor and Sunkoshi River relates to the historical upward movement (migration) of all Kirāti people that shows a trend of Kirāti toponymic interpretation. Khuwālun, a holy place, which frequently comes in the recitation of *Mundhum*³

1 Recent discourses such as Maharjan 2024; Maden 2024; Rai 2020; Maharjan 2019 have articulated different dimensions of toponyms of indigenous people with connection of pre-historic identity, renaming or replacing names and its relation with the linguistic vitality of the related languages.

2 The holy stone is located at the confluence of Arun, Tamor and Dudhkoshi River that carries various historical and cultural information about Kirāti people especially about their migration to the northern hilly regions from the south plain region.

3 The ritual oral text of Kirāti people.

has now become a cultural identity of Kirāti people. However, its etymological interpretations given by scholars do not agree with other. As Rai (2022) suggested, Khuwālung is affixed with *khu-*, *-wā-* and *-lung* and refer ‘river’, ‘water’ and ‘rock’ respectively, likewise Rai (2021) also insists that interpretation should be sought with these three affixes. On the other hand, Mukarung (2024) mentions that *khuwā* and *lung* refer to ‘water’ and ‘rock’ respectively. Chamling (2021) has presented five various interpretations of Khuwālung characterized the same ‘rock’ from different angles. The range of *-khu* extends from the eastern part of Nepal to the central Kathmandu Valley, transitioning to *-ti* and *-di* in western Nepal, particularly in the Magarat territories (Rai 2024, 19-26). Toponymic study has not been introduced in any university course in Nepal. It is not even introduced as a topic of any curriculum and syllabus. Thus, issues of toponyms have remained largely as discourses rather than scholarly works. In this context, the studies carried out till date are limited to etymology-based general explanations. Interpretation of few remarkable toponyms in the central part of the eastern hilly region such as *ten + ke* (village + hill) for *Temke* (Rai 2021), *ha + tuwā* blood + ‘mixing up for *hatuwā* (Rai 2017), and *am + chak* (your + border) for *Amchok* (Amchok Municipality 2017) and *khikā + mākhchā* (*khikā* + son-in-law) for *khikāmakchā* (Rai 2023) are not only erroneous but also harmful in many ways because they potentially ruin the history and worldviews of the local people. On the other hand, stories and narratives of different place names have been eroded, lost and extinct; majority are deviated due to lack of scholarly interpretations, such as ‘sāmsilā’ (the grassland) has been interpreted as *shyām* (proper male name) + *shilā* (proper female name) and *Waisipākhā* (raspberry land) as the *bhaisipeptā* (killing buffalo) and *haktukwā* (windy place) as *hatuwā* (mixing up the blood of two parties in the war) etc. These interpretations do not carry convincing and logical essence of Kirāti toponyms. In this context, this study is concerned with Bāntāwā and Dungmāli people, their habitats, languages and cultures and other various activities within their communities

3 The Methods

This small case study is based on field observations and the collections of narratives, stories and the interpretations of Bāntāwā and Dungmāli people from socio-ethnographic methods. The field visits and observations were made during July 2022-March 2024 in different times by two of the authors belonging to same ethnic group. Observations of ecological situatedness of the places were done multiple times to assess its relation with semantic meaning.

Furthermore, people's activities or movements to the particular places were also observed to examine cultural connection of toponyms. Additionally, multiple unstructured interviews were carried out with the local people, who were chosen from snowball and judgmental sampling to explore cultural connection of the toponyms.

3.1 Site

The selected sites for data collection are Amchok, Hatuwā, Dilpā and Dungmā which are the popular cultural areas of eastern hilly regions of Nepal. Amchok, Hatuwā, Dilpā are original homelands of Bāntāwā language speaking community which is the largest language speaking community within Rai-Kirāti ethnic group of Tibeto-Burman language family in Nepal. Additionally, the researchers have visited Dungmā, a cultural and original homeland of Dungmāli mother tongue speaking people who share identical cultural and ritual beliefs with Bāntāwā people and live geographically contiguous. The sites lie in the Bhojpur and Khotang districts of eastern Nepal. However, the cultural geography is extended even further to some parts of Khotang, Udaypur and Dhankuta districts. Altogether 462 toponyms were collected during the field visit and a significant amount of ethnographic data and narratives were also recorded and collected. The data was collected manually, in an audio recording and photographs. The reason for choosing this site was to be the significant wide range of the Bantawa speaking region in comparison to other Kirati cultural regions. The key sites of the data collection were Bālānhā of the Amchok region, Bhuruchok of the Dilpā region, Ghoretār of the Hatuwā region, and Thulo Dungmā of the Dungmā region. All kinds of data such as, collection of toponyms and ethnographic information, observation and interviews were taken but not limited to the centre of these points.

3.2 Procedure

Myths and narratives of all cultural practices have remained a distinct identity of Kirāti people. For this study, those myths and narratives about toponyms were recorded. Forty adult and elderly people were interviewed by using open-ended questions. The nature of the interview was unstructured, and the research participants were sampled by using snowball and judgmental sampling methods. The majority of interviews took place in the field, such as forest, nearby paddy fields, roadsides and participants' working stations. Interviews were recorded manually and in an audio recorder device.

Later, only significant narratives were transcribed into English for thematic analysis.

3.3 Analysis

The majority of remarkable lands did not correlate with their multiple narratives and general interpretation. In that case, the socio-cultural approach by Steward (1955) was adopted to analyze the toponyms, which provides holistic lens for toponymic analysis. Further, ecological perspective by Hunn (1994) was used as a lens for analyzing toponyms from the ecological perspectives, which informs us that indigenous and local knowledge systems shape the toponymic pattern. Some of the remarkable places had already been interpreted in different ways, such as name of the rural municipalities had got explained in their administrative profile. They are assessed and aligned with the situatedness, ecology and environmental orientation of the places and made conclusions. Hence, the analysis was done from taxonomic semantic classification to theory-governed discursive interpretation.

In short, the method of this study is presented below [tab. 1].

Site	Four cultural areas; Amchok, Hatuwā, Dilpā (Bāntāwā) and Dungmā of eastern hilly region
	↓
Procedure	Collection of toponyms (40 Adults) Recordings of Myths and narratives Snowball and judgmental sampling
	↓
Analysis	Transcription of narratives. Based on thematic analysis.

Table 1 Methods of data collection and analysis

4 Analysis and Discussions

This section outlines the various aspects of toponymy and its thematic analysis based on the data we have collected. We have tried to be less erroneous and more realistic. We noticed that the Kirāti languages are derived from verbs, such as *yungkhā* ‘a place to sit’ is derived from *yungmā* ‘to sit’, *imkhā* ‘bed’ is derived from *immā* ‘to sleep’ and *chākhā* ‘a plate’ is derived from *chāmā* ‘to eat’ in Bāntāwā. These verbs remain equally meaningful with the alternation of every vowel in the language, such as *emmā*, *ommā*, *āmmā*, *ummā* refer distinct meanings, such as ‘freeze’, ‘whiten’, ‘shoot’ and ‘hold’ respectively; having respective nominal derivation to each. Thus, extracting

etymological meanings of toponyms in Kirāti languages requires distinct approach because names generally embody action within it. Similarly, Kirāti toponyms carry various dimensional properties of the land, such as its size, appearance, direction, height, peoples' cultural activities and vegetation regarding different ecological belt. Due to this reason, remarkable place names in the region embody adjectival property. Some thematic descriptions are presented below.

4.1 Semantic Interpretation of Kirāti Toponyms

Kirāti toponyms are interpreted in different ways by different people and places. From our observation of around two years, we have noticed the following toponyms.

4.1.1 The *khā* Toponyms

Kirāti toponyms appear with obvious word structures and meanings. In a small territory of southern part of the Dhintāng hill of the Amchok region, some place names are found identical in suffixation such as, thidinkhā, chemkhā, chongkhā, bālankha, bechukkhā, dāngkha, hopsikhā, lungkongkhā etc. The first four names are even situated between the 2.5 kilometers of distance in the horizontal rural pathway.

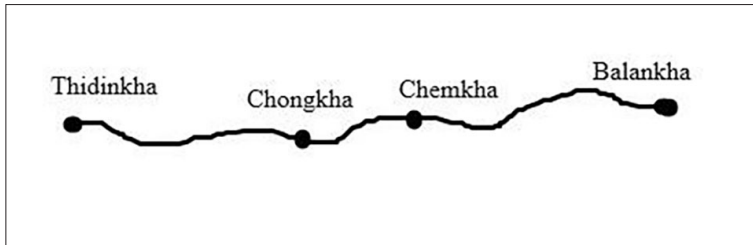


Figure 1 Name of places: villages in Amchok region within five kilometers

The names with identical suffixation continue in the other Hatuwā and Dilpā region as well, such as bākchākhā, langkhā, pichākhā, wāplukhā, sumlikhā, khambukhā, bāhāpkhā, chhukkhā etc. This kind of place name constitutes a quarter of Kirāti toponymic taxonomy. The suffix <-khā> denotes a place occupying a relatively small territory that has been cultivated or used by the people. The suffix can also express very small places like *imkhā* 'bed', *yungkhā* 'seat' and *ekhā* 'toilet' etc. Human activity is mandatorily connected with

these place names. For example, *chongkhā*, *dāngkhā*, *bechukkhā* refer to the farmlands of wheat, maize and ginger respectively. On the other hand, *bākhākhā* denotes the boar hunting place, *bālānkhā* denotes boar chasing place and *pichākhā* denotes the cow grazing land. Although this dominant suffix refers to the small territory, the toponyms may later relate to the whole village. This kind of toponym is largely formed with noun + suffix and then verb + suffix and lesser with noun + verb + suffix.

4.1.2 The *lā* Toponyms

lā is also a suffix which denotes 'grove' in Bāntāwā and some other Kirāti languages including Dungmāli, Puma, Nachhiring, Limbu and Kulung. Due to this reason, place names with the suffixation of <-lā> are found in many areas of eastern hilly reasons, such as *bāksilā* (Khotang), *thoksilā* (Udaypur), *syāmsilā* (Bhojpur), *yāngsilā* (Morang) etc. It refers to a small group of trees which grow close together. Within the Northern hill range of Amchok region 7 villages are found to be named of their kind [fig. 2].

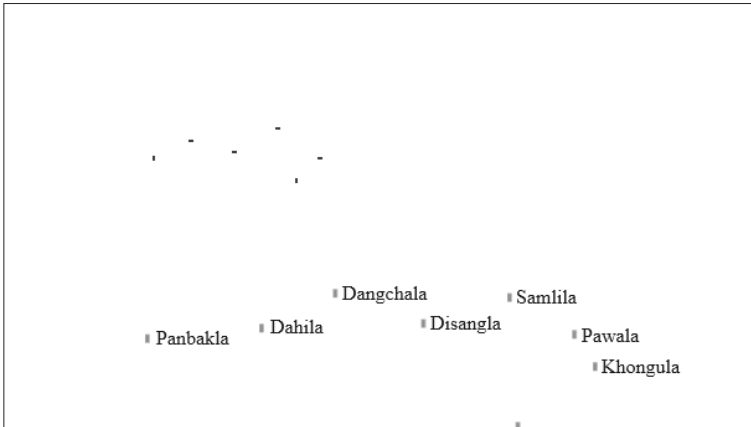


Figure 2 Name of the villages in Amchok highlands

Kirāti people hold rich experiences of recognizing the functionality of the trees for their livelihood. Due to this reason, they name the trees based on their use, shape, size and functionality. Eventually, they call the place with the name of the tree and its grove, such as *makuplā* 'a grove of Himalayan bamboo', *wāsānglā* 'a grove of *Alnus Nepalensis*', *yāngsanglā* 'a grove of *Schima Wallichii*', *dhisanglā* 'a grove of big trees' etc. The other frequent toponyms of these categories that

are found many areas are *peklā*, *dangchhalā*, *khakkhānlā*, *bilimlā*, *chimpulā*, *sukumlā*, *pānbāklā*, *khailā*, *suplā*, etc. This kind of toponyms are formed with ‘noun + suffix’ structure.

4.1.3 The *tāng* Toponyms

tāng is an independent word in Bāntāwā and Dungmāli. It is alternatively used as *rāng*. *tāng/rāng* refers to a single but a larger tree in the area. After all, the area is identified by the tree, and ultimately it becomes the name of the village. Kirāti people use various trees or vegetation in their rituals; thus, they recognize most of the trees in nearby forests. The *Tāng* toponyms also found in dominant numbers such as *chumārāng* ‘Sal tree’, *wāsirāng* ‘Fig tree’, *homtang* ‘Siris tree’ (scientific name, *Albizia Lebbeck*), *phinturāng* ‘Hog plum tree’, *khairāng* ‘walnut tree’ etc. Some other toponyms are *phukwentāng*, *khongtāng*, *dhenutāng*, *nikuwātāng*, *mukhitāng*, *wālāptāng*, *gurātāng*, *khairāng*, *wākhāngrāng*, *haikharāng* etc. Due to the erosion of culture and language, multiple *tang* toponyms have lost their meaning in the language because of shifting identity (Gautam, Sapkota 2024). The formation of toponyms of this kind is ‘noun + suffix’.

4.1.4 The *wāng* toponyms

wāng is an independent word in Bāntāwā and denotes the large farmland of the village usually cultivated by the villagers. However, it does not necessarily refer to the communal land, instead denotes a part of that land. Literal meaning of *wāng* is ‘paddy’ or ‘maize field’, however, it relates the situatedness of the field, such as *dhiwang* ‘the big farm field’, *chungwāng* ‘the shady farm field’, *sichchewāng* ‘the fertile farm field’, *hyāuwāng* ‘the next farm field’, etc. The frequency of this kind of toponyms is less but connects the intense human activity, even it relates the collective activity of the people. The suffix is attached to the adjective for the formation of the toponyms. The main interpretation of this classification is the attachment of agriculture and the production of crops in the traditional life of Bāntāwā, and Dungmāli people.

4.1.5 *ten* and *hon* Toponyms (Ethnic Villages/Directions)

In Kirāti toponymy study, few toponyms are found having suffixed <-ten> ‘village’ and <-hon> ‘direction’. The *ten* inherits the physical and structural features of the village and *hon* solely denotes the direction from the point of space, such as *chongten* ‘the upper village’,

hyauten ‘the next village’, *hyuten* ‘the southern village’, *dhāuten* ‘the northern village’, *tumiten* ‘the village where Tumi clan people reside’, *pāniwangten* ‘the village where people of Chhetri community practicing Hindu religion-based Indo-Aryan culture, reside’ and so on. On the other hand, *dhāuhon*, *hyuhon*, *busuhon*, and *densuhon* refer to the north, south, east and the west. They may name the whole sphere of land based on the direction as presented briefly [tab. 2].

Table 2 *ten* and *hon* toponyms

Suffix	Interpretation
- <i>khā</i>	A place where human activity occurred, such as <i>kāyākhā</i> ‘paddy field’
- <i>lā</i>	Grove of trees, grew naturally, but not in the typical forest, such as <i>angmālā</i> ‘a grove of pine trees’
- <i>tāng</i> (alternatively <i>rāng</i>)	A remarkable and usually huge tree in the village (not in the forest), such as <i>dākbugrāng</i> ‘a rhododendron tree’
- <i>wāng</i>	A collective farmland of the village, such as <i>dhiwāng</i> ‘bigger farmland’
- <i>ten</i>	Village or part of village based on societies or communities, such as <i>tumiten</i> ‘a village where Tumi clan Kirāti people live’
- <i>hon</i>	The whole area/direction one can see in front of him or her from a point of standing, such as <i>busuhon</i> ‘the east’

From the table we can easily guess that these toponyms are based on the ethnicity of Kirāti people and the direction of landscape in the area we studied. The construction of identity is a very powerful tool to be visible in national and international forums.

4.2 Cultural and Ecological Interpretation of Kirāti Toponyms

Toponyms have remained as linguistic artifacts and archaeological evidence of Kirāti people in their homelands. Various myths and narratives have been passed down to the generations regarding toponyms. However, many such myths and narratives embody the deep emotional attachment of people with the land. A narrative (75-year-old male) says:

The ancestor of the Māngphāng people used to live in the Chhābung cave nearby the Bongwā stream. He had twelve sons. Later, they cultivated the whole area bordering themselves by planting bamboos. Thus, the name of the whole cultural area named later Amchok; owns border.

The next narrative (68-year-old male) says about the deep-rooted socio-cultural practice and emotional attachment with the relatives or people:

Golme Raja and Golme Rani were the powerful rulers in the southern part of the Pikhuwā River. They had a son-in-law named Khikā who was the resident of northern part of the Pikhuwā River. Khikā was loving son-in-law of Golme Raja and Golme Rani. Due to this reason Golme Raja and Golme Rani used to see the Khikā's village every day. Later, the region called later Khikāmakchhā (the land of Khikā son-in-law), the earliest name of the Bhojpur.

From the above narratives we can easily guess that Kirāti people are the first settlers of the region, and they have developed different place names during the course of time. Their daily rituals, practice, culture and ecological perception are deeply rooted in the toponyms.

4.2.1 Concept of Horizontal Ecological Space

Kirāti toponyms show a space distinction in terms of their horizontal situatedness. The smallest space is suffixed with <-khā>; generally termed as space. It is applicable to all human activity related places, such as *wāchākkhā* 'the place for bathing', *nāmtāngkhā* 'the place for sunbathing' and *ogikhā* 'the field of sweet yam', *sākikhā* 'potato field', *thenyungkhā* 'the resting place' etc. Later, these names may be coined to refer large spaces like hamlets and villages.

Although the *tāng/rāng* refers to a single tree, it denotes a larger space than <-khā>, because it is associated only with a bigger and significant tree. For example, Banyan and Walnut trees cover the big areas of land. Toponyms related with this category are very common in the region. For example, *fukwentāng*, *rāksirāng*, *sāmbārāng*, *haikharāng*, *khongtāng*, *rāgātāng*, *bāmrāng*, *hontāng* are the name of the villages in the region. Furthermore, some of these toponyms used to be the name of village development committees, an administrative division of the nine villages until the nation adopted the federal system. A brief linguistic interpretation of some Kirāti toponyms is presented in the figure below [fig. 3].

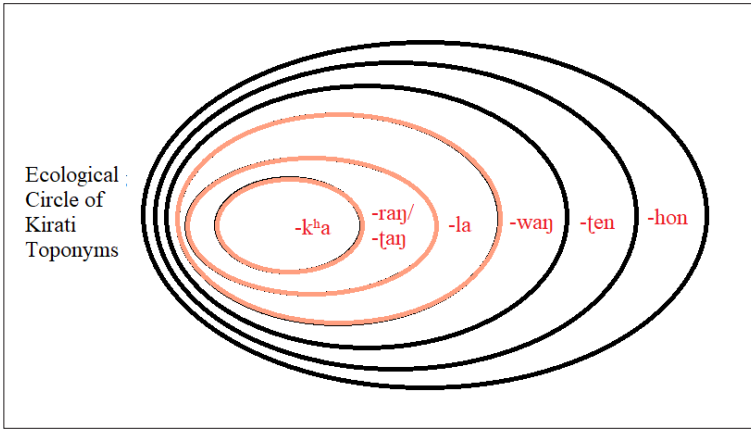


Figure 3 Ecological circle of Kirant Toponyms

In the figure, the *lā* toponym even relates the bigger land space. Although it refers to the group of trees, it does not refer the forest of same kind of trees. A forest is perceived as the place where different kind of trees are grown. Thus, the forest of pine trees is perceived as *angmālā* ‘a grove of pine trees’ in Bāntāwā. *khongulā* ‘a grove of Maleto trees’, *bhāmālā* ‘a grove of oak trees’ *dangchhalā* ‘a grove of Himalayan bamboo’ ultimately refer the name of the villages now, but are connected with the different vegetations.

wāng, *ten* and *hon* refer even largest and widest places. *wāng* usually relates a part of a village where villagers own their farmland in a place. They share the land and cultivate it by owning a part of land in a particular side of the village. They do so because this land is most fertile and they build a collective farmland. Due to this, they help each other in farming. On the other hand, *ten* refers to an area of village residing a group of societies, such as in a village, there may exist multiple *ten* based on the clans, caste and geography. The *hon* is the largest and widest area, one can see in front of him/her.

4.2.2 The Concept of Vertical Ecological Space

The concept of land orientation in Kirāti community is unique. The vertical concept of land which is also ecologically based in Kirāti toponymy system can be schematized in table 2 below:

Table 2 Vertical toponyms

Suffix	Interpretation
-lung	The high points of land in the earth, such as <i>chomolung</i> ‘Mt. Everest’, <i>bhāktānglung</i> ‘Mt. Kumbhakarna’, <i>sewālung</i> ‘Mt. Makalu, etc.’
-chong	The highest point of earth seen from the territory, such as <i>rimāchon</i> , the highest visible point of Dilpā and Hatuwā region, <i>bomākkhāchon</i> , the highest point visible from Amchok region, <i>dhiwāchon</i> , the highest point visible from villages around <i>bongwā</i> stream etc.
-bhara	The hillocks around the regions, such as <i>dhintāngbhara</i> , <i>thongbhara</i> etc.
-chok	The pointed part of the land, but not a typical hillock, such as <i>bhuruchok</i> , <i>bhimāchok</i> , <i>chhināmchok</i> etc.
-chhan	The upper part of the land called by the people who live in the bottom part, such as <i>hongkuchhān</i> , <i>lambichhān</i> etc.
-khuk	The bottom part of the land called by the residents of upper land such as <i>khetikhuk</i> , <i>bhirākhuk</i> , <i>kholikhuk</i> , <i>lāmkhuk</i> etc.

Kirāti people never describe their places as the down part of the higher land, instead they call it the higher part of the down part. This unique concept of vertical ecological concept represents different semantic interpretations and place names.

4.2.3 Developmental Concept of Toponymy

Many Kirāti toponyms have been developed and conceptualized during the development and modernization processes. Bhuruchok is the center of the Dilpā region, where the offices of the government bodies have been established for a long time. A government funded high school has been established, where many students from far and different villages studied and still study. The place connects all the villagers because most of the grocery shops and other local markets are also available for local business in this place. At present, this is the headquarter of the Temke Maiyung Rural Municipality. Moreover, People across the region meet up there every day for various business-like politics, social service and other activities. Western side of the Bhuruchok is a remarkable high hill, famously known as *temke* which is 3010 meters high from the sea level and the landscape is not much vertical from the southern side. It has only 15-degree slope angle from the Bhuruchok which is also measured around 20 degrees from Hatuwā and the Bhojpur Bazar, the center of Bhojpur district. Due to this physical orientation of the land, people from Dilpā region called it *temyāngko bhara* that means ‘the flat hill’. Later, it is shortened to *temke bhara* to *temke*.

Likewise, ‘Hatuwā’ a cultural region of Hatuwāli dialect of Bāntāwā language has been interpreted as *hā + tuwa* ‘the place where blood of Gorkhali and Kirāti armies were mixed up during

war of pre-unification era’. However, it does not provide convincing logic because verb is never suffixed with the toponyms. Instead, a narrative mentioned by Rai (2024) complements the authentic meaning of this remarkable toponym:

Long ago, a Kirāti girl from highland of Amchok region married to a boy of lowland of Hatuwā region. Due to the altitude, the girl felt heavy hot in Hatuwā. One day, she came back to her birth home Amchok and complained her brothers: “You allow me to marry a boy from lowland, so I need to bear hot every day, I am suffering from the hot air and the environment’. When she was going back to her home, her brothers provided her a special kind of box telling her: ‘Don’t open it on the way, just open it when you reach back your home’. However, she could not resist and opened the box near Ghoretār. When she opened the box, a big pouchful air flew away making her chill at the moment. This popular narrative says that from that time a heavy air blows every day in the afternoon in the Ghoretār region. (74-year-old female)

This interesting narrative of Amchok, Hatuwā and Ghoretār seems to be unique and mysterious. By observing the ecology of the Ghoretār which is the center market of Hatuwā region these days have been modified and shifted from Hatuwā. Hatuwā is connected with the movement of the air in the daytime in the region and it is located at the top of the hill on Arun River basin. Later people started to call it as Haktukwā, ‘a place where air blows in every season’.

Likewise, the toponym ‘Amchok’ relates to the orientation of the land rather than the narratives of twelve sons of Chhābung. ‘Amchok’ is in the high altitude so people from around called it ‘Khamchok’ ‘the highland’. On the other hand, *khikāmākchhā* does not relate the son-in-law of Golme Raja and Golme Rani, instead it relates the court or place to debate and fight from *khikhāmukhā* of Bāntāwā people. The earlier *khikhāmukhā* used to be the center of all administrative works in the region and people had to go to Bhojpur time and again, but the meetings were not usually pleasant. So that Bāntāwā people might have called *khikhāmukhā* to current Bhojpur as a place for conflict and dissatisfaction.

5 Shifting Trends in Toponyms and Language Ideology

Over time, many toponyms underwent changes, replaced and disregarded resulting in the distortion of multiple historical legacies of the region including oral histories, narratives and deep-rooted indigenous emotions. It happened in two phases in the recent past; before and after the administrative restructuring of the country.

Before the recent political reconstruction of the country (Nepal became a federal republic after the 2006 political change), the changes of place names were less intense and less deliberate. However, major reasons for changes, modification or replacement of toponyms always remained the dominant Nepali cultural and linguistic ideology. Earlier language and education policies of the country (NNEPC 1956; NESP 1971) highlighted Nepali and English language and culture by ignoring minority languages, cultures and ethnicities (Gautam, Poudel 2022; Gautam 2021; 2025). As a result, minority ethnic and linguistic communities ought to shift and fit in dominant Nepali language and culture (Gautam, Giri 2024). Kirāti toponyms were colonized even after the 1990s political change in the country. Colonization of ‘Bomākkhāchon’ (Perilla Frutescens field) to ‘Bhanjyāng Kharkha’ (pasture land), ‘Dhintāng’ (tree of Himalayan Holly) to ‘Thām Dāndā’ (high hillock), ‘Danchhalā’ (grove of Himalayan bamboo) to ‘Pasal Bhanjyāng’ (shop located pass), Dimālūng (grandmother rock) to Purne Bazar (Fortnightly market), Bululūmā (a place to sing) to Bhimeswori (name of school that relates the Hindu goddesses) and Chhongkhā (wheat farmland) to Chandi Dānda (Chandi hillock), Wāsiyong (place of fig trees) to Wāsing Tharpu (Wasing Court), Bhimāchok (upper land of Bhima (Rai)) to Raketār (‘land of Rāke fair’) show the tremendous shifting trends in toponyms based on ideology in the first phase.

After promulgation of the Constitution of Nepal (2015), a huge administrative change took place in the country. The Monarchy was abolished by the peoples’ revolution and the country became federal republic state or ‘New Nepal’. Five regional development regions, 75 districts, 130 municipalities and 3,833 Village Development Committees (VDCs) were restructured into 7 provinces, 77 districts and 753 local governments’ bodies. During this transitional period (2007-15), naming of provinces and local governments had aroused ideological conflict between government and the ethnic communities across the country. The government was forced to implement geography-based naming, which could weaken the aspiration of identity-based naming articulated by different ethnic and linguistic minorities in the country. Provincial governments endorsed geography-based names which could again strengthen the dominant Nepali cultural and linguistic ideology by adopting English as an alternative language in education. As a result, historically suppressed ethnic and linguistic communities protest against governments’ move for naming the provinces, however, they remained fail but even persisting in the case of Koshi Province. The name ‘Koshi’, coined from a name of the river is passed by the provincial parliament. Nevertheless, Kirāti people of the province are still carrying out protest rally demanding dismiss ‘Koshi’ and rename it with historical identity of the Kirāti people. Meanwhile, during this transitional

period, issue of naming of local municipalities also became a national debate. However, majority of municipalities got named from coinage and compounding of physical geographical remarks, such as hills and rivers overlooking historical and cultural geography. An old man (86) narrates the story before:

Bululumā, a pleasant place of Amchok region was a place of occasional gathering of young villagers. They used to sing and dance long ago there before radio and music system introduced in the village. It was their part of culture, so they could stay happy and joyous. Bululumā refers singing in Bāntāwā language. In course of time, a temporary primary school was established with the name Bhimeshwori, a name of Hindu Goddess. Sometime later, the school was relocated to Bāsikhorā village, three hours walking distance away from Bululumā. However, the place has been famously called Bhimeshwori then after. The community lost the legend of Bululumā along with its name. No one except few elderly people even recall the history of this place. No one will stop by singing again there. (86-year-old male)

From this narrative, we can easily say that many toponyms have been shifted, changed and modified because of the power, policy and the assimilation ideology of past and present. The trend indicates the unprecedented erosion and loss of the Kirāti toponyms. Earlier colonization of Kirāti toponyms and suppression of dominant ideology ultimately distorted the Kirāti peoples' deep-rooted human emotions with their homeland and broke the ancestral legacy to the new generation.

6 Conclusion and Implications

Place names or toponyms have important functions having possible relations between language, culture and geographical space. They are more than just labels on a map—they are stories, memories, and reflections of the people who have lived in a region. Names can be powerful tools for political control that indicate various things. While a place is renamed it often clarifies a statement of ownership, or some kind of hegemony related to language politics which is used either by community or government to mark the history.

The findings of the study indicate three major themes of Kirāti toponyms. Firstly, Kirāti toponyms are largely inspired by human activity and nature. In this case, Kirāti people perceive similar places differently based on human activity. Majority of the toponyms are perceived by vegetation, such as trees and groves, inspired by their functionality and uses. Secondly, sole linguistic interpretations of

Kirāti toponyms are always incomplete and may be harmful for multiple reasons including distortion of indigenous history and legacy. Thus, Kirāti toponyms require holistic interpretation complemented by cultural and ecological perspectives. Thirdly, the tendency of shifting toponyms has been grounded in the historical domination of the Nepali language and dominant Hindu culture. However, during and after the transition of the country from a Monarchy to a federal republic there are many tendencies of shifting trends in Kirāti toponyms. This is a good indication that Kirāti people are struggling and negotiating their historical identity while naming the administrative divisions.

This study has two major implications. Firstly, it provides a different perspective to the researchers while carrying out study in the same kind of field. Secondly, it gives a new insight to observe the toponyms. Overall, it informs academic and non-academic communities to see the toponyms in a holistic way, which leads objective interpretation of the place names. This study also broadens the ways of doing further research in the similar field. Kirāti toponyms across the eastern hilly region requires documentations and interpretations. Moreover, toponymic studies of indigenous ethnic settlements should be carried out to broaden the horizon of the knowledge in the field. Studying Kirāti toponymic requires comprehensive field-based approach with specific theoretical perspectives. Linguistic and etymological study carried out so far do not cover societal, cultural and ecological aspects of Kirāti toponyms.

Acknowledgements

We would like to express our sincere acknowledgments to the Bāntāwā and Dungmāli people who supported us during our field visit and observation. Various writers and experts cited in this article are also duly acknowledged.

Bibliography

- Amchok Rural Municipality (2018). *Amchok gaupalikako bastugat bibaran* (Profile of Amchok Rural Municipality). Bhojpur: Amchok Rural Municipality.
- Chamling, B. (2021). “Khuwālūng: uhile yatharth, ahile mithak” (Khuwālūng: Reality in the ancient time but now the myth). Chamling, B. (ed.), *Ek Dhunga Ek Badapatra*. 1st ed. Kathmandu: Bela Publications, 29-35.
- Everett-Heath, J. (2000). *Place Names of the World: Historical context, Meanings & Changes*. New York: Macmillan.
- Gautam, B.L. (2021). “Language Politics in Nepal: A Socio-Historical Overview”. *Journal of World Languages*, 7(2), 355-74.
Available at <https://doi.org/10.1515/jwl-2021-0010>
- Gautam, B.L. (2025). “English Language Politics in Nepal: A Socio-Historical Perspective”. Giri, R.A.; Gnawali, L. (eds). *Handbook of English Language Education in Nepal*. London: Routledge, 34-45.
Available at <https://doi.org/10.4324/9781003508540>
- Gautam, B.L.; Giri, M. (2024). “Trajectories of Language and Culture of Gopali Community in Chitlang Valley”. *Glottodidactica*, 51(1), 7-26.
Available at <https://doi.org/10.14746/gl.2024.51.1.1>
- Gautam, B.L.; Paudel, P. (2022). “Diversity, Multilingualism and Democratic Practices in Nepal”. *Bandung Journal of Global South* 9, 80-102.
- Gautam, B.L.; Sapkota, S. (2024). “Shifting Identity in Kurmali: A Preliminary Observation”. *Nepalese Linguistics*, 38(1), 30-9.
Available at <https://doi.org/10.3126/nl.v38i1.71551>
- Government of Nepal (2015). *The Constitution of Nepal 2015*. Kathmandu: Nepal Government, Ministry of law and Justice.
- Hatuwāli, C.K. (2017). “Hatuwāgadhi darbarko aitihasikta ra bartaman awastha” (History of Hatuwāgadhi Darbar and its present condition). *Ancient Nepal*, 195, 47-52.
- Hunn, E.S. (1994). “Place-names, population density, and the magic number 500”. *Current Anthropology*, 35(1), 81-5.
Available at <https://doi.org/10.1086/204240>
- Jorden, P. (2012). “Place names as ingredients of space-related identity”. *Oslo Studies in Language* 4(2), 117-31.
- Kapur, A. (2019). *Mapping Place Names of India*. London: Routledge.
- Maden, K. (2018). “Isthannamma manabiya pahichan” [Human identity in Toponyms]. *Kantipur Daily*, 10 March.
Available at <https://ekantipur.com/opinion/2018/03/10/20180310073014.html>
- Maharjan, B. (2018). “Itihaska sakshi sthannam” (Toponyms as witnesses of History). *Kantipur Daily*, 31 March.
Available at <https://ekantipur.com/bibidha/2018/03/31/152246346610799246.html>
- Maharjan, B. (2019). “Sankatma aitihasik isthannam” (Historical toponyms in danger). *Himal Khabarpatirka*, 17 March. www.himalkhabar.com

- Ministry of Education (1971). *Report of National Education System Plan*. Kathmandu: Ministry of Education.
- Mukarung, R. (2024). “Sakela/Sakenwama Khuwālūṅ” (Khuwālūṅ in Sakenwa/Sakela). *Khuwālūṅ*, 1(1), 3-4.
- National Education Planning Commission (1956). *Education in Nepal*. Kathmandu: Nepal Education Planning Commission.
- Rai, Bh. (2022). “Kolambus Paila: Raithane Gyanko Ananta Dukha” (Columbus Footstep: Tragedy of Indigenous Knowledge). *Kantipur Daily*, 16 July.
Available at <https://ekantipur.com/koseli/2022/07/15/165786918694336308.html>
- Rai, Bh. (2023). *Kirāti sabhyatako kendra Khuwālūṅ* (Centre of Kirat civilization: Khuwālūṅ). Kathmandu: Nepal Academy.
- Rai, K. (2021). “Temkemaṭiyung gāupālikāko chināri” (An introduction to temkemaṭiyung Rural Municipality). Lumumbu, S.; Basnet, G. (eds), *Bhojpur Darpan*. 1st ed. Kathmandu: Bhojpur Matribhumi Jorpati Boudha Ekata Samaj, 925-60.
- Rai, N. (2024). “An Analysis of Kiranti Toponyms from Etymological and Perspectives”. *International Conference of Linguistic Society of Nepal, Kathmandu*, 26-7.
- Rai, N. (2024). “Kirāwāhon khonki sākenwā” (Kirāt region and Sākenwā). *Khuwālūṅ* 1(1), 23-32.
- Rai, T.M. (2020). “Isthānnāmko bhāshya” (Narratives of toponyms). *Nayapatrika Daily*, 7 November.
Available at <https://jhannaya.nayapatrikadaily.com/news-details/1170/2020-11-07>
- Rai, T.M. (2021). “Khuwālūṅko rakshārtha pratirodh” (Protest against the Khuwālūṅ). Chamling, B. (ed.), *Ek Dhungā Ek Badāpatra*. 1st ed. Kathmandu: Bela Publications, 29-35.
- Rai, T.M. (2024). “Settlement patterns in the Kirāt and Magar (Magarat) territories: Insights from the hydronyms”. *Magar Studies*, 1, 19-25.
- Rai, T.M.; Chamling, B. (2017). *Kirat isthannam: prageitihāsik ra bhasik adhhyān* (Kirat toponyms: prehistoric and linguistic study). Lalitpur: National Foundation for Development for Indigenous Nationalities.
- Randall, R. (2001). *Place Names: How They Define the World and More*. Lanham: Scarecrow Press.
- Rapacha, L.S. (2024). “Decolonizing Nepal’s Kiranti-Köits Grammar and Lexicon”. *Journal of Nepalese Studies*, 16(1), 74-101.
Available at <https://doi.org/10.3126/jns.v16i1.71783>
- Steward, J.H. (1955). *Theory of Culture Change: The Methodology of Multilinear Evolution*. Champaign: University of Illinois Press.

Code-Switching in South Asia: Comparing the Equivalence Constraint and Matrix Language Frame Models with Hinglish

Anmol Sahni

Emory University, USA

Abstract This article examines code-switching in South Asia by applying Shana Poplack's Equivalence Constraint Model and Carol Myers-Scotton's Matrix Language Frame Model to Hinglish, a hybrid of Hindi and English. Through syntactic and sociolinguistic analysis, it argues that the Matrix Language Frame Model better accounts for South Asian patterns of intra- and inter-sentential switching. The article also advocates translanguaging in English classrooms, linking structural theory to pedagogical practice in India's multilingual context.

Keywords Code-switching. Hinglish. South Asia. Equivalence constraint model (ECM). Matrix language frame model (MLFM). Transformational generative grammar. Bilingualism.

Summary 1 Introduction. – 2 Aims and Objectives. – 3 Literature Review of CS with an Illustrative Hinglish Example. – 4 Sociolinguistic and Structural Approaches to Code-Switching. – 5 Structural Dimensions and Constraint Approaches to Code-Switching. – 6 The Equivalence Constraint Model. – 7 Matrix Language Frame Model. – 8 Conclusion.



Peer review

Submitted 2025-03-25
Accepted 2025-05-07
Published 2025-07-01

Open access

© 2025 Sahni |  4.0



Citation Sahni, A. (2025). "Code-Switching in South Asia: Comparing the Equivalence Constraint and Matrix Language Frame Models with Hinglish". *Bhasha*, 4(1), 51-76.

DOI 10.30687/bhasha/2785-5953/2025/01/003

1 Introduction

Over the past three and a half decades (1990-2024), scholarly interest in code-switching (CS) has grown significantly, challenging earlier negative perceptions of language mixing. This phenomenon is particularly salient in the Indian context, where Hindi, the dominant language of northern India, coexists with English – a language shaped by its colonial past and its global significance as the language of trade and commerce. Hindi is primarily spoken across the northern Indian states collectively referred to as the ‘Hindi Belt’, which includes Uttar Pradesh, Bihar, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, and Rajasthan. Within this region, Hindi encompasses various dialects that reflect its linguistic diversity. For example, the Awadhi dialect spoken in Lucknow, the capital of Uttar Pradesh, is distinct from the Bhojpuri dialect prevalent in Bihar.

The 2001 Census of India reported that Hindi was spoken by approximately 422 million individuals, including speakers of various dialects classified under Hindi (Koul 2008, 1). By the 2011 Census, this figure had risen to approximately 528 million, reflecting a significant increase in the demographic representation of Hindi speakers. Beyond India, the online journal *Ethnologue*¹ identifies countries with substantial Hindi-speaking populations due to immigration, including Nepal, Australia, Bhutan, Canada, Mauritius, Germany, and the United States, among others.

Hindi is traditionally classified into two main groups: Eastern Hindi and Western Hindi. Eastern Hindi includes dialects such as Awadhi, Bagheli, and Chhattisgarhi, while Western Hindi encompasses dialects like Haryanvi, Brajbhasha, Bundeli, Kanauji, and Khariboli. Additional dialects such as Maithili, Bhojpuri, and Magahi in Bihar, and Marwari, Jaipuri, and Malvi in Rajasthan, contribute to Hindi’s linguistic richness. Some dialects from northwestern Uttar Pradesh and Himachal Pradesh, previously excluded from earlier classifications, are now recognized under the broader category of ‘Hindi’. As one of India’s official languages, Hindi serves as a vital medium for communication among diverse communities. Its historical roots in the Indo-Aryan language family are intertwined with the socio-cultural and political developments of the region. Over centuries, Hindi has been shaped by languages such as Sanskrit, Persian, Arabic, and, more recently, English, reflecting its adaptability to the dynamic landscape of Indian society.

Hindi’s literary tradition is equally diverse, spanning religious poetry, prose, dramas, and novels written in dialects like Awadhi, Braj Bhasha, and Khariboli. Foundational figures like Kabir (fifteenth

1 *Ethnologue*: <https://www.ethnologue.com/>.

century) and Tulsidas (sixteenth century), deeply influenced by the Bhakti movement, laid the groundwork for a rich literary heritage. Kabir's *dohas* (couplets) and Tulsidas's *Ramcharitmanas* in Awadhi not only exemplified devotional literature but also elevated vernacular Hindi as a medium for artistic and spiritual expression. This tradition evolved further in modern Hindi literature, marked by luminaries such as Munshi Premchand, known for his social-realist novels (*Godaan*), Jaishankar Prasad, celebrated for his prose and dramas (*Chandragupta*), and Ramdhari Singh Dinkar, revered for his patriotic and philosophical poetry. During the colonial era, Hindi literature emerged as a powerful medium for resisting colonial hegemony, with nationalist themes prominently featured in the works of Bharatendu Harishchandra and Maithili Sharan Gupt.

The history of English in India parallels this complex and rich evolution of Hindi, originating during British colonization when it became a language of administration and education. Over time, English developed into a distinct variety known as Indian English, marked by unique phonetic, grammatical, and lexical features shaped by local languages, particularly Hindi and other regional tongues. This blend of native elements with English reflects India's diverse cultural and social fabric. In contemporary India, Hindi functions as a lingua franca for many speakers, while English, often used in formal contexts, education, and media, complements this linguistic landscape.

The interplay between Hindi and English has given rise to Hinglish – a hybrid language that fuses elements of both languages. The linguistic tapestry of Hinglish not only facilitates communication among diverse groups but embodies the speakers' cultural identities and social affiliations. Despite Hinglish's rich character and the insights it provides into code-switching and sociolinguistic dynamics, research on Hinglish remains limited. This article aims to deepen the understanding of Hinglish by exploring the structural principles underlying code-switching through the frameworks of Shana Poplack's Equivalence Constraint Model (ECM) and Carol Myers-Scotton's Matrix Language Frame Model (MLFM). By examining Hinglish, this study seeks to highlight its significance within the broader context of bilingualism and the sociolinguistic complexities of contemporary India.

2 Aims and Objectives

Research on code-switching has evolved into two primary approaches: the sociolinguistic approach, which examines code-switching at the discourse level, and the structural or syntactic approach, which focuses on code-switching below the sentence level. This article provides a comprehensive literature review of generativist approaches to code-switching, with a particular focus on two influential theories: Shana Poplack's Equivalence Constraint Model (ECM), developed from a Spanish-English corpus (Poplack 1980), and Carol Myers-Scotton's Matrix Language Frame Model (MLFM), based on Swahili-English CS data (Myers-Scotton 1993).

The article in part three begins by exploring key debates and challenges that animate CS research, including the difficulty of establishing a singular, definitive definition of CS. Part four of the article outlines the sociolinguistic context of CS research, tracing how the increasing body of work in this field has led to the development of structural approaches to CS. Part five then contextualizes these structural approaches – specifically, constraint-based theories of CS – within Noam Chomsky's influential theory of transformational generative grammar. Parts six and seven apply the ECM and MLFM to Hinglish, providing insights into the structural dynamics of code-switching and a comparative evaluation of the models' effectiveness in analyzing Hinglish. Finally, the article concludes by pointing to future directions for the development of CS research.

The primary objectives of my analysis are twofold: (a) to analyze Hinglish through the lenses of both ECM and MLFM; and (b) to demonstrate how ECM and MLFM contribute to the scholarship on bilingualism and code-switching, particularly in South Asia. Overall, the article seeks to address the underexplored application of the ECM and the MLFM to Hinglish within CS research, thus bridging a significant gap in the field.

3 Literature Review of CS with an Illustrative Hinglish Example

3.1 Challenges in Defining Code-Switching

Despite the significant contributions of scholars such as Ad Backus (1992), Susan Berk-Seligson (1986), Carol Myers-Scotton (1993a), and Shana Poplack (1980), there remains a lack of consensus regarding the definition and governing principles of CS. Poplack (1980) characterizes CS as the alternation between two languages within a discourse, sentence, or constituent, resulting in a hybrid utterance. Poplack describes CS as the

alternation of two languages within a discourse, sentence or constituent [...] [CS] is categorized according to the degree of integration of items from one language (L1) into the phonological, morphological, and syntactic patterns of the other language (L2). (Poplack 1980, 583)

The challenge of providing a singular, definitive explanation of CS is compounded by the ongoing debate among linguists regarding its distinction from lexical borrowing. Two primary viewpoints emerge: one group, including Poplack and her colleagues at the University of Ottawa Sociolinguistics Laboratory, advocates for differentiating between isolated foreign-language items and more extensive instances of code-switching. Conversely, Myers-Scotton and her supporters argue against treating CS and lexical borrowing as separate phenomena within bilingualism.

In her work *Borrowing* (2018), Poplack meticulously delineates code-switching, loanwords, and nonce borrowings as distinct phenomena of language mixing, based on their synchronic or diachronic manifestation in a speech community. Nonce borrowing refers to the temporary use of a word from another language on an ad-hoc basis to express a particular idea, without it becoming a permanent part of the borrowing language. For example: “I start my day with *chai* (tea) and *samachar* (news)”. Here, *chai* and *samachar* are Hindi words temporarily borrowed to be used in an otherwise English sentence to convey specific meanings in the context of the sentence. In contrast, loanwords are more permanent lexical borrowings that result from sustained cultural contact. A good example is the anglicized term *cummerbund*, which can be traced back to its origins in Hindustani and Persian. Over time, *cummerbund* has been fully adopted into English and is commonly used without reference to its foreign roots. Furthermore, Poplack categorizes established loanwords – such as the English word ‘country’, which etymologically originates from French, according to

the Oxford English Dictionary – as a diachronic phenomenon. Poplack differentiates loanwords from CS by emphasizing that loanwords undergo a process of diachronic integration, wherein they lose their original linguistic attributes and assimilate into the morphology and, often, the phonology of the recipient language.

3.2 Distinction Between Code-Switching and Lexical Borrowing

Poplack argues that the fundamental distinction between CS and lexical borrowing lies in their structural integration. CS primarily involves intra-sentential mixing at the constituent level, where the grammar of the donor language is operational. Conversely, lexical borrowing involves the integration of words from a donor language into a recipient language, where the grammar of the recipient language is adhered to, maintaining grammaticality (Poplack 1980).

Poplack provides a framework for categorizing these phenomena, as illustrated in Figure 1 of her 1980 publication titled “Sometimes I’ll start a sentence in Spanish Y TERMINO EN ESPAÑOL” [fig. 1].

Type	Levels of Integration Into Base Language			Code-Switching?
	Phonological	Morphological	Syntactic	
1	✓	✓	✓	No
2	×	×	✓	Yes
3	✓	×	×	Yes
4	×	×	×	Yes

Figure 1 Poplack’s (1980) analysis of code-switching using threefold criteria of lexicon integration into the base language

Poplack evaluates the integration of a foreign word into the base language according to three criteria: phonological, morphological, and syntactic integration. In Figure 1, she identifies ‘Type 1’ cases, which do not qualify as CS but as loanwords. These instances involve foreign lexicon from (L2) that have been fully integrated into the phonology, morphology, and syntax of the recipient language (L1), shedding their donor language properties [fig. 1]. In contrast, cases classified as ‘Type 2’ to ‘Type 4’ are considered instances of CS due to their partial integration into the base language. In other words, CS occurs when a speaker alternates between two or more languages within the same conversation or sentence, but the switch remains consistent with the grammatical rules of each language.

For example, speaker code-switching between Spanish and English is demonstrated as follows:

I was talking to María, y ella me dijo que v. a la tienda.
'I was talking to María, and she told me she's going to the store.'

In this sentence, the speaker switches from English to Spanish without blending the grammar of the two languages. The two languages – namely, Spanish and English – are clearly distinguished. In contrast, Poplack would consider an utterance a case of borrowing when a word from one language is adopted into another language and is often integrated into the borrowing language's grammar and pronunciation. The borrowed word might be slightly modified to fit the phonological or grammatical rules of the new language, as evidenced in the following example:

Voy a parquear el coche.
'I am going to park the car.'

In this second example, the word 'parquear' is a borrowing from the English verb 'to park', but it has been adapted to Spanish grammar by taking on the Spanish verb ending '-ar'. Over time, such borrowings become fully integrated into the borrowing language.

Figure 2, as adopted from Poplack's coauthored work (1989), further distinguishes between established loanwords and nonce borrowings by examining parameters such as the degree of phonological integration and the role of the lexicon as a content word [fig. 2]. This nuanced approach highlights the complexity of language-mixing phenomena and underscores the importance of context in understanding their manifestation and classification.

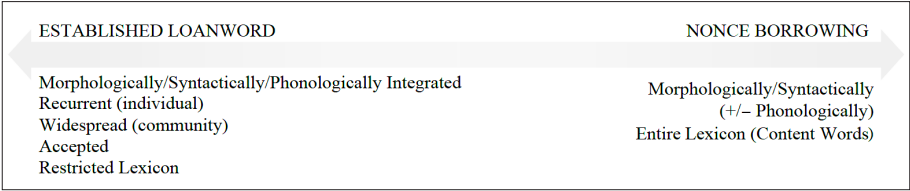


Figure 2 The continuum for levels of borrowing in code-switching utterances (Poplack, Wheeler, Westwood 1989)

3.3 Divergent Perspectives on Code-Switching

Poplack's conceptualization of code-switching, characterized by the partial integration of phonological, morphological, and syntactic elements, contrasts sharply with Myers-Scotton's perspective, which challenges the differentiation between CS and lexical borrowing as distinct phenomena within bilingualism. Myers-Scotton has described

code-switching as “the selection by bilinguals or multilinguals of forms from an embedded language (or languages) in utterances of a matrix language during the same conversation” (1993b, 4).

Myers-Scotton’s model posits that the languages or codes involved in CS hold equal status but fulfill distinct roles within the process, specifically within the framework of the Matrix Language (ML) and the Embedded Language (EL). According to this model, the ML is responsible for establishing the morpho-syntactic structure of the clause and supplying both the system morphemes and the majority of the content morphemes. In contrast, the EL contributes by inserting constituents and individual content morphemes into the ML’s structural framework. Myers-Scotton thus asserts that “there is no need to make the borrowing vs. code-switching distinction” (2002, 153). This position of Myers-Scotton is reaffirmed in a later work, where she argues that both CS and borrowing “undergo the same morphosyntactic procedures during language production” (Myers-Scotton 2003, 8). This foundational divergence in the definition of CS is critical, as it shapes the subsequent discussion of the ECM and the MLFM, developed by Poplack and Myers-Scotton, respectively.

3.4 Code-Switching in Hinglish

An example of CS, as defined by both Poplack and Myers-Scotton, is demonstrated in 1(a), where the speaker alternates fluidly between two typologically distinct languages: Hindi, which follows an SOV word order, and English, characterized by an SVO word order. This type of language mixing is commonly referred to in the literature as Hinglish, denoting the blending of Hindi and English during conversation. Similar to other contact-blend languages such as Franglais (a mixture of French and English, prevalent in Canada), Taglish (a blend of Tagalog and English, common in the Philippines), and Portunhol (a mixture of Portuguese and Spanish), Hinglish incorporates both intra-sentential CS and borrowed loanwords from Hindi and English.

1 (a) *Bhai yeh tera hard - earned money hai isliye isko udana mat.*

*1 (b) *Bhai this your hard - earned money isliye ise udana mat.*
‘Brother this is your hard-earned money therefore don’t squander it.’

While (1a) illustrates a typical example of Hinglish, (1b) is deemed ungrammatical due to its violation of certain implicit rules or constraints governing CS, which will be discussed in greater detail later in the article. This example highlights that, despite appearing spontaneous or arbitrary, code-switched sentences are not random

assemblages of words from different languages. Instead, CS follows systematic patterns, occurring at intra-sentential sites – specific points within a sentence, such as major constituency breaks – and at inter-sentential sites, where switching occurs at the discourse level. Although bilingual speakers often engage in inter-sentential code-switching, this type of switching has received relatively less attention from syntacticians compared to intra-sentential mixing, as syntactic analysis typically focuses on grammatical relationships within the sentence.

4 Sociolinguistic and Structural Approaches to Code-Switching

This section provides a historical overview of code-switching research by examining the development of two parallel approaches: sociolinguistic and structural. The structural approach focuses on the morphosyntactic constraints governing CS, addressing issues related to the grammaticality of code-switched utterances. In contrast, the sociolinguistic approach explores the functional aspects of CS at the discourse level, elucidating how meaning is constructed when two languages are juxtaposed within a conversation. Historically, both approaches have evolved significantly; whereas early studies of bilingualism often viewed CS as indicative of a speaker's linguistic inadequacy, contemporary perspectives recognize it as a demonstration of proficiency in the grammars of the involved languages.

4.1 Historical Perspectives on Code-Switching

In the early stages of bilingualism research, beginning around the 1950s with foundational works such as Weinreich's *Languages in Contact* (1953), CS was largely viewed as detrimental to the grammar of mixed languages. This perspective was endorsed by prominent linguists like Leonard Bloomfield and Uriel Weinreich. During this early period, Weinreich distinguished between the “ideal bilingual” and the “imperfect bilingual”, reflecting the prevailing erroneous attitudes towards grammaticality concerning research on CS. According to Weinreich, an ideal bilingual was a speaker who seamlessly switched languages based on changes in the speech context (e.g., interlocutors, topics) without disrupting the flow of conversation (1953). In contrast, the imperfect bilingual was perceived as someone with uneven proficiency across the languages being switched.

Weinreich's dichotomy between ideal and imperfect bilingualism, arising from the lack of a theoretical framework for CS grammar, contributed to the development of pejorative attitudes towards CS. Notable among these were the concepts of semilingualism, the deficit hypothesis, and prescriptivism. Semilingualism posited that CS indicated a bilingual's lack of proficiency in the languages being used, suggesting that mid-sentence switches reflected linguistic inadequacy. The deficit hypothesis, closely related to prescriptivism, asserted that speakers of non-standard dialects, such as African American English (AAE), were linguistically deficient and incapable of conveying complex ideas. In tandem, prescriptivism endorsed the supremacy of standard dialects, like Standard American English (SAE), and promoted linguistic uniformity within the education system, marginalizing non-standard dialects and reinforcing hierarchical views of language competence.

4.2 Code-Switching in Postcolonial and Multilingual India: The Case of Hinglish

From the 1950s to the 1990s, the absence of a robust theoretical framework for grammar in CS led many linguists to perceive CS as a threat to the linguistic integrity and grammatical structures of blended languages. This view, grounded in the erroneous belief that CS negatively impacts the grammars of the involved languages, prompted some prescriptive grammarians to caution against CS, fearing it would lead to language convergence and the bastardization of grammatical norms. This negative perception of CS was compounded by widespread linguistic insecurity concerning the lingua franca of newly formed nation-states during decolonization. For instance, post-1947, after India achieved independence from British colonial rule, there were a series of public and intellectual debates that attempted to renegotiate the hegemonic status of English in India. Notable among these debates was the controversy surrounding the 'three-language formula' introduced by the Indian central government under the 1968 National Policy of Education. This policy mandated that English be taught alongside Hindi and a modern Indian language to bridge the gap between the educated elite and the broader population (Krishnaswamy 2006).

Krishnaswamy observes that

recognizing the role of English in globalization, India has shed its colonial complexes towards English and has come to terms with the language; Indians have separated the English language from the English. (2006, 153)

Consequently, Indian English has been adopted as one of India's official languages alongside Hindi. Recent research (Kothari 2011) indicates that code-switching between Hindi and English, exemplified by Hinglish, is a natural linguistic phenomenon facilitated by the bilingual nature of the Indian speech community. There is no empirical evidence to suggest that the use of Hinglish has negatively impacted the grammars of English and Hindi. Instead, Hinglish has enriched the linguistic repertoire of its speakers while adhering to the constraints of grammaticality associated with CS, as this paper demonstrates.

4.3 The Ebonics Debate's Influence on Sociolinguistic Views of Code-Switching

In the 1990s, while linguistic prescriptivism was increasingly being discredited as a valid theoretical framework, it continued to maintain marginal influence, particularly among linguists who prioritized the notion of linguistic 'purity'. From a contemporary perspective, this insistence on purity can be understood as a form of linguistic racism. As prescriptivism's credibility waned, descriptivism gained widespread acceptance within the linguistic community. This transition signified a shift from prescribing rigid linguistic norms to objectively identifying and documenting the intra-sentential sites of CS within speech communities.

The paradigmatic shift towards descriptivism can be traced back to the 1960s, with the pioneering work of linguists such as William Labov. Labov and others challenged the prevailing attitudes toward AAE, which had previously been treated as a deviation from SAE. Instead, AAE came to be recognized as a legitimate dialect with its own grammatical rules. The descriptivist approach in U.S. linguistics fostered research on code-switching between AAE and SAE, further establishing the legitimacy of AAE as a dialect.

A key moment in this shift was the Ebonics debate, sparked by the 1996 resolution passed by the Oakland, California Board of Education (CBE). This resolution called for schools across the U.S. to recognize "Ebonics" as the "primary language" of African American students, with the goal of enhancing their proficiency in SAE (Young 2014, 40). In response to this resolution, a wave of opposition emerged, with critics dismissing AAE as an improper or 'slang' version of English. This ignited a national debate in which both linguists and legislators grappled with the question of whether AAE constituted a separate dialect or merely a corrupted form of standard English. Central to the argument for recognizing AAE as a distinct dialect was Labov's seminal 1972 work *Language in the Inner City*. Labov testified before a Senate subcommittee in January 1997, at the height of the Ebonics

controversy, where he presented evidence that AAE is indeed a distinct dialect of English. Labov's research demonstrated that AAE possesses unique phonological and grammatical features, such as negative concord and contractions and deletions in the copula (e.g., "Sam here" instead of "Sam is here") (Labov 1972, 65). These findings provided a robust defense of AAE's systematic and rule-governed nature, solidifying its status as a legitimate variety of English. From the perspective of research into the social and structural role of CS, Labov was one of the first to suggest that CS between AAE and SAE should not be viewed as a linguistic deficit but rather as a manifestation of bilingualism. In his testimony, Labov advocated for educators to support and cultivate their students' proficiency in AAE, rather than stigmatizing the natural alternation between AAE and SAE. Labov's stance marked an important shift in recognizing the legitimacy of language mixing as part of a broader sociolinguistic competence.

In subsequent years, U.S. linguists further developed Labov's ideas, with sociolinguist Suresh Canagarajah playing a prominent role in popularizing CS and other forms of language mixing, such as translanguaging and code-meshing, as effective tools for rhetorical education (Canagarajah 2013). In his work, Canagarajah has conceptualized translanguaging as the ability of a speaker to seamlessly transition between languages. According to Canagarajah (2011), translanguaging can serve as a pedagogical approach that encourages and supports this linguistic fluidity in educational settings. Through translanguaging, students are able to engage in cognitive processes across multiple languages concurrently, utilizing their home language as a tool for acquiring and mastering academic English. Canagarajah's work builds upon the legacy of linguists like Labov, emphasizing the pedagogical value of harnessing multilingual students' full linguistic repertoire.

4.4 Hinglish as a Catalyst for Translanguaging in the Multilingual Context of India

As code-switching between Hindi and English, Hinglish naturally promotes translanguaging by enabling speakers to draw from their entire linguistic repertoire for communication and learning. By combining vocabulary, syntax, and idiomatic expressions from both languages, Hinglish allows speakers to convey nuanced meanings and cultural references that may be inaccessible in either language alone. If fostered in pedagogical settings and classrooms, Hinglish can promote inclusivity by creating a shared communicative space in the multilingual context of India, bridging gaps between speakers with varying proficiencies in Hindi, English, and other marginalized

regional languages. Hinglish can further facilitate learning in classrooms by integrating familiar linguistic elements with complex or challenging content, fostering active engagement and deeper understanding. Since Hinglish challenges linguistic hierarchies that privilege English over regional languages like Hindi, the use of code-switching in the form of Hinglish can ultimately promote a more equitable linguistic environment central to translanguaging pedagogy. By embracing Hinglish, educators and learners can create flexible, inclusive, and culturally relevant frameworks for communication and education.

4.5 From Sociolinguistic Foundations to Generative Approaches in Code-Switching

In addition to Labov's contributions, the pioneering research of Myers-Scotton significantly advanced the public and scholarly understanding of CS. Through works such as *Social Motivations for Code-Switching* (1993b), which focused on multilingualism in African contexts, Myers-Scotton demonstrated that CS is a skilled and purposeful linguistic practice in which bilinguals strategically draw upon their linguistic resources from different speech communities. Myers-Scotton's research highlighted the social and communicative motivations behind CS, positioning it as a sophisticated form of language use rather than a sign of linguistic deficiency. The sustained growth of sociolinguistic scholarship on CS led to a parallel expansion of research into its grammatical structures. Early studies by sociolinguists, particularly between the 1950s and 1970s, sought to challenge the prevailing negative attitudes toward CS, which often regarded language mixing as indicative of a lack of proficiency in either language. From the 1990s onward, researchers such as Poplack (1980) and Backus (1992) contributed to the development of structural approaches to CS. Their work helped to neutralize linguistic prejudices against code-switching by showing that it follows systematic grammatical rules and reflects a high degree of linguistic competence.

4.6 Generative Grammar's Influence on Structural Studies of Code-Switching

In light of the significant influence of generative syntax, recent scholarship has proposed conflicting constraint models of CS as part of the expanding body of structural approaches to CS. Before applying the ECM and the MLFM to analyze Hinglish data, it is essential to address three key inquiries that have driven much of

the research in this area. These questions, outlined by Bullock (2009), are central to the generative approach to CS: (a) How should CS be defined, and should nonce borrowing be treated as distinct from CS? (b) What rules, if any, constrain CS? (c) How can a constraint model for CS be formulated in alignment with generativist grammar theory? The prevailing trend in CS scholarship is to develop a constraint model capable of accounting for the spontaneous and dynamic nature of bilingual language use. Among these models, Poplack's ECM initially gained prominence. However, it was later challenged by Myers-Scotton, whose research on Swahili-English CS identified instances that could not be explained by ECM. In response, Myers-Scotton introduced the MLFM as an alternative framework to account for these linguistic phenomena. The following section defines both the ECM and MLFM models, beginning with a detailed explanation of ECM. Subsequently, the section applies both models to the analysis of a representative Hinglish sentence, contrasting their explanatory power in this context.

5 Structural Dimensions and Constraint Approaches to Code-Switching

5.1 Historical Foundations of the ECM

The publication of the ECM by Poplack in 1978 marked a pivotal development in the study of CS. The ECM's major contribution was its explanation of CS within the framework of universal grammar, applying the same principles that govern monolingual syntax to bilingual speech. This represented a significant departure from earlier explanations of CS, which focused primarily on surface-level grammatical structures where the syntax of the two languages coincided, thereby permitting a switch. Such approaches lacked a theoretical basis grounded in the universal principles of generative grammar, as advanced by Chomsky (1965).

Prior to the ECM, most CS research analyzed language alternation at the discourse level, providing a macro-level understanding of how blending languages could create meaning but offering little insight into the micro-level grammatical constraints that governed permissible switch sites within a sentence. For instance, in *Discourse Strategies*, Gumperz (1982) proposed that CS was impossible in specific syntactic environments, such as between verbs and pronominal subjects or between conjunctions and conjuncts. While Gumperz's work was instrumental in offering a discourse-based analysis of CS, his treatment of the grammatical aspects of CS remained limited and lacked the formal rigor found in generative approaches. Poplack's ECM filled this gap by aligning CS analysis with generativist syntactic

theory, allowing for a more nuanced and theoretically grounded understanding of the constraints governing bilingual speech. This shift enabled researchers to move beyond descriptive accounts of language alternation toward a more systematic explanation of CS rooted in universal grammar.

5.2 **Generativist Foundations of the ECM**

Building on Chomsky’s transformational-generative theory, first introduced in *Aspects of the Theory of Syntax* (Chomsky 1965), Poplack’s ECM pioneered a constraint-driven approach to studying CS. Two key insights from Chomsky’s work have become foundational in contemporary CS scholarship (Newmeyer 1986): (a) all human languages are governed by constraints, and (b) syntax analysis consists of two components: base structure and transformational rules. Figure 3, reproduced from *Aspects* (Chomsky 1965), illustrates Chomsky’s framework for syntax analysis [fig. 3].

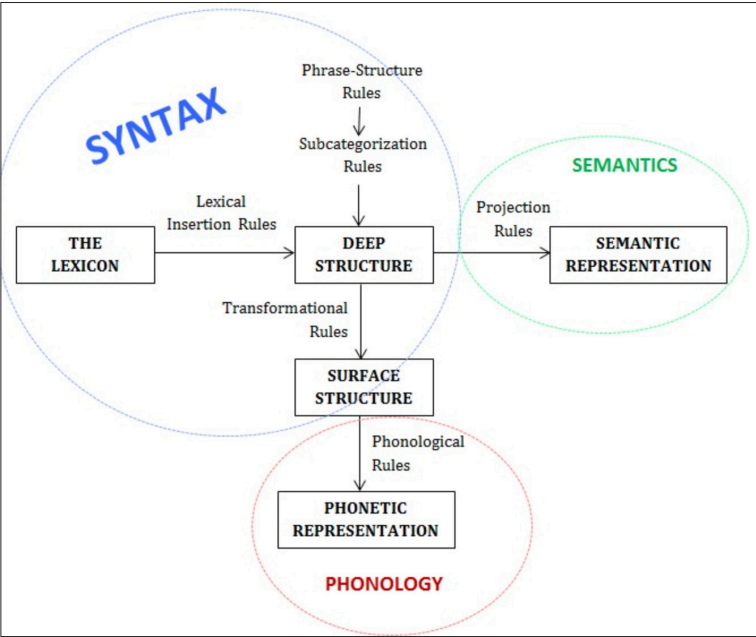


Figure 3 The generativist-transformational model presented in Chomsky 1965

Explaining this framework, Chomsky wrote:

The syntactic component consists of a base and a transformational component. The base, in turn, consists of a categorial subcomponent and a lexicon. The base generates deep structures. A deep structure enters the semantic component and receives a semantic interpretation; it is mapped by transformational rules into a surface structure, which is then given a phonetic interpretation by the rules of the phonological component. (1964, 151)

The impact of Chomsky's transformational model on CS research is notable. Chomsky's generative-transformational model introduced the possibility of analyzing bilingual speech through finite syntactic rules and constraints, applicable across all languages, regardless of their typological differences or surface structures. The emergence of all constraint models of CS, including Poplack's ECM and Myers-Scotton's MLFM, can be traced back to the Chomskyan revolution in linguistics. Despite their differences in methodology, both models are predicated on the linear and hierarchical analysis of sentence structures. The following section first discusses ECM with an illustrative example and then elucidates the workings of MLFM.

6 The Equivalence Constraint Model

Poplack defines the ECM and its related Free Morpheme Constraint as follows:

- The equivalence constraint states that code-switched sentences are made up of concatenated fragments of alternating languages, each of which is grammatical in the language of its provenance. The boundary between adjacent fragments occurs between two constituents that are ordered in the same way in both languages, ensuring the linear coherence of sentence structure without omitting or duplicating lexical content (Poplack 2001, 2062).
- The Free Morpheme Constraint: Codes may be switched after any constituent in discourse provided that the constituent is not a bound morpheme (Poplack 1980, 585).

For example, code-switching between 'eat', an English verb stem, and '-iendo', the Spanish present progressive, is impossible unless the former is phonologically integrated into Spanish (Sankoff Poplack 1981). In summary, ECM suggests that CS only occurs at those sites where the surface order of constituents surrounding the switch point aligns with the order in both participating languages. Poplack

provides an example of a sentence that is considered ungrammatical due to its violation of ECM.

Example	Gloss	Translation
*told <i>le</i>	told to-him	“(I) told him”.
<i>le</i> told	to-him I-told	
him <i>dije</i>	him I-told	
<i>dije</i> him	I-told him	(Poplack 1981, 176)

6.1 Application of ECM to Hinglish

The article now applies ECM to analyze the Hinglish sentence 2(a), which comprises two clauses conjoined by the coordinating conjunction ‘and’. The sentence 2(a) reads as follows:

- 2 a Ram highly qualified *hai* and he can find a job *poori duniya mai*.
‘Ram is highly qualified, and he can find a job anywhere in the world’.

This sentence contains two instances of CS, each occurring within one of the clauses. To analyze these CS instances, we can represent the sentence using two distinct phrasal trees: Tree 1 and Tree 2. Each tree separately depicts the grammatical structure of the clauses surrounding the CS points, allowing us to evaluate whether the switches conform to ECM’s stipulations. This approach enables a detailed examination of how the CS instances align with the syntactic rules of the participating languages and whether they maintain the linear coherence required by ECM.

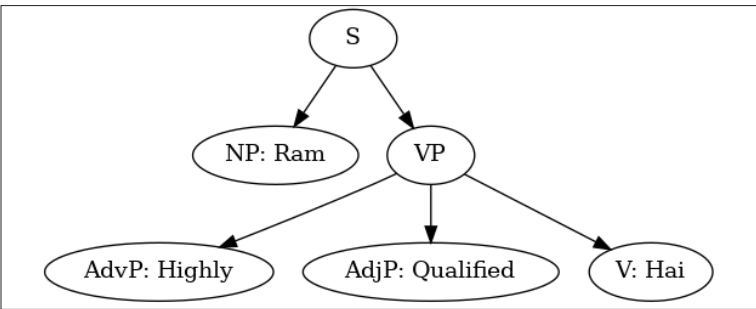


Figure 4 Tree 1

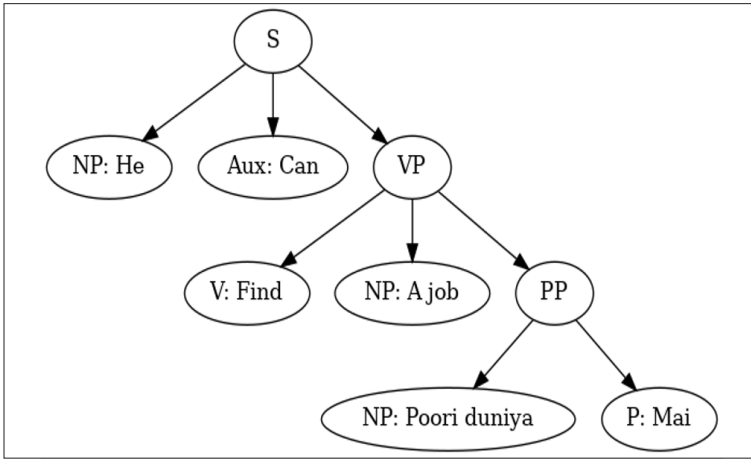


Figure 5 Tree 2

6.1.1 Analysis of the First Clause (Tree 1)

The sentence “Ram highly qualified hai” provides a compelling example of CS within the framework of ECM. In this instance, the speaker incorporates the English phrase “highly qualified” into an otherwise Hindi sentence. This switch occurs at the level of the Adjective Phrase (ADJP), where the integration of English elements is seamless due to the structural compatibility between Hindi and English. Both languages share a similar pattern for adjectival modification, where an adverb modifies an adjective (ADVP → ADV ADJ). This structural equivalence aligns with the ECM’s principle that code-switching is permissible at nodes where the grammars of the participating languages are compatible. The ADJP node satisfies this criterion, facilitating the smooth incorporation of English into the Hindi sentence.

As the sentence progresses, the speaker transitions back to Hindi with the verb ‘hai’, illustrating the morpho-syntactic rules specific to Hindi. The verb phrase (VP) node adheres to Hindi’s syntactic constraints, requiring the use of Hindi lexical items to maintain grammaticality. This transition exemplifies another key tenet of the ECM: while code-switching is allowed at structurally compatible nodes, constraints arise at points dictated by the morpho-syntactic rules of each language. The VP node, for instance, enforces these constraints, ensuring that the Hindi verb ‘hai’ occupies its expected position at the end of the clause.

This example highlights the broader implications of CS in Hinglish and other bilingual contexts. Adjective phrases often serve as frequent

sites of CS, as their structural alignment across languages enables smooth transitions. By demonstrating how the ECM applies to this example, the analysis underscores the model's utility in explaining the grammatical boundaries and possibilities of code-switching in bilingual speech.

6.1.2 Analysis of the Second Clause (Tree 2)

In the second clause, the speaker employs code-switching, transitioning from Hindi to English within the structures represented by the (S), (VP), and (NP) nodes. This phenomenon is facilitated by the shared phrase-structure rules governing these constituents across both languages, allowing for seamless integration of English elements without disrupting the hierarchical organization of the sentence. For example, the NP node contains the English pronoun 'He', while the VP node includes the auxiliary 'can' and the verb 'find', demonstrating a coherent English predicate.

However, a notable divergence arises in the treatment of postpositional phrases (PPs). Hindi typically utilizes postpositions - such as 'mai' (in) - which occur after the noun phrase, as seen in the Hindi phrase 'poori duniya mai'. In contrast, English generally employs prepositions, leading to structural incompatibilities when integrating PPs from both languages. While English can use postpositions in certain contexts (e.g., 'miles away'), these cases do not apply to the current example. Consequently, the speaker strategically reverts to Hindi for the phrasal structure below the PP node, maintaining adherence to Hindi's morpho-syntactic conventions.

This bilingual construction allows for a nuanced expression of meaning while preserving the grammatical integrity of the sentence. The linear arrangement of constituents reflects the hierarchical structure, where the English elements align with their corresponding Hindi counterparts. The combined analysis of the first and second clauses illustrates how the ECM can be effectively applied to Hinglish sentences. Figure 6 visualizes this ECM analysis by mapping the grammatical constraints and code-switching sites, highlighting the shared and language-specific structures identified within the clauses [fig. 6].

Through this analysis, we observe that code-switching in Hinglish is not merely a random blending of languages; rather, it follows a systematic approach governed by the syntactic frameworks of both Hindi and English. By carefully navigating these frameworks, speakers can create complex sentences that reflect their bilingual competence, ultimately enriching the communicative potential of Hinglish.

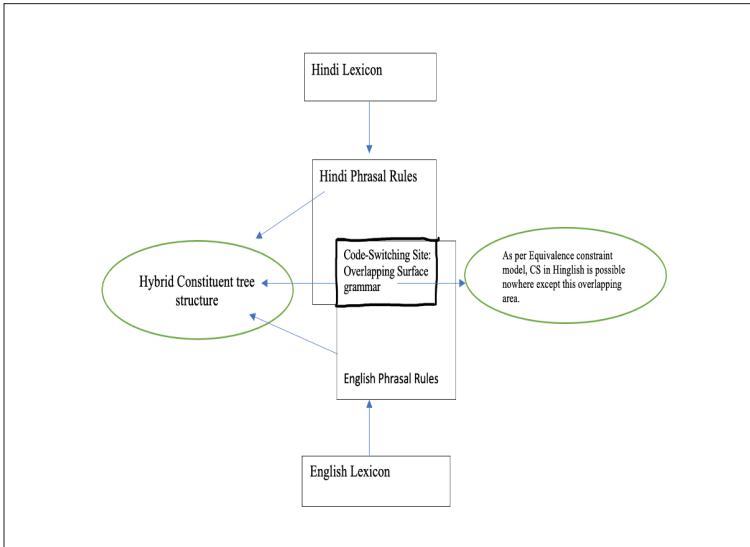


Figure 6 A graphical representation of Poplack's Equivalence Constraint Model of CS used to analyze the Hinglish example (2a)

7 Matrix Language Frame Model

7.1 Critique of the Equivalent Constraint Model

Myers-Scotton challenged Poplack's ECM on both empirical and theoretical grounds. Empirically, Myers-Scotton argued that while ECM showed promise for studying CS between typologically distinct languages, Poplack's ECM failed to account for CS data from the Swahili-English corpus, thus lacking empirical validation. Theoretically, Myers-Scotton contended that ECM did not address the concept of asymmetry, which is prevalent in most language contact scenarios. Joshi (1985) was among the first to highlight the disproportionate distribution of lexicon between the languages in a code-switched sentence, referring to this as the asymmetry of language involvement in bilingual speech. Joshi observed that "speakers and hearers generally agree on which language the mixed sentence is coming from. We can call this language the *matrix language* and the other language the *embedded language*" (Joshi 1985, 198; emphasis in original). Myers-Scotton's MLFM emerged as an alternative to ECM, incorporating Joshi's observation by distinguishing between the matrix and embedded languages. Unlike ECM, which treats loanwords and code-switching similarly, MLFM offers a framework that recognizes the asymmetry in language usage.

7.2 Influence of Psycholinguistic Theories

The development of the MLFM was also influenced by psycholinguistic theories. MLFM presupposes that speakers acquire their linguistic repertoire through a process of lemmatization, involving both conceptual and grammatical functions. This approach parallels cognitive theories of concept acquisition, such as those proposed by philosophers of mind like Jerry Fodor (1998, 40-9). The lemma-based approach in linguistics posits that language acquisition involves cognitive processes such as prototypicality and categorization, providing a foundation for understanding how speakers navigate and utilize multiple languages in code-switching contexts.

7.3 Definitions of Code-Switching

To illustrate Myers-Scotton's critique of ECM's inability to account for asymmetry in CS, it is important to contrast the definitions of CS proposed by Poplack and Myers-Scotton. Unlike Poplack, who treats CS and lexical borrowing as similar phenomena, Myers-Scotton argues that ECM fails to address the critical aspect of asymmetry in lexicon distribution within CS. Myers-Scotton defines CS as follows: "CS is the selection by bilinguals or multilinguals of forms from an embedded language (or languages) in utterances of a matrix language during the same conversation" (1993b, 11). This definition underscores asymmetry as a central feature of CS.

7.4 Key Principles of MLFM

To address the shortcomings of ECM, Myers-Scotton introduced the MLFM. The MLFM posits that in intra-sentential CS, speakers utilize the two languages asymmetrically. The first language – termed the matrix language (ML) – serves as the primary language of discourse, characterized by a higher frequency of morphemes and lexical items from that language. Conversely, the other language, known as the embedded language (EL), contributes fewer morphemes and lexical items relative to the ML. In essence, the MLFM captures the asymmetrical nature of language use in CS by distinguishing between the ML and EL.

According to the premise that the ML has a higher frequency of morphemes and lexicon distribution in code-switched sentences, Myers-Scotton proposed three key principles for conducting an analysis using the MLFM:

1. the Morpheme Order Principle: in ML+EL constituents consisting of singly occurring EL lexemes and any number of

- ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML.
2. the System Morpheme Principle: in ML+EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e., which participate in the sentence's thematic role grid) will come from the ML.
 3. the Blocking Hypothesis: in ML+EL constituents, a blocking filter blocks any EL content morpheme which is not congruent with the ML with respect to three levels of abstraction regarding subcategorization (Myers-Scotton 1993a, 83-120).

Simply put, according to the first principle of the MLFM, the ML dictates the ordering of elements within ML+EL constituents. MLFM differentiates between content morphemes (e.g., nouns and verbs) and system morphemes (e.g., articles and inflections), with the second principle asserting that function morphemes must originate from the ML. The third principle restricts the EL to contributing only specific content morphemes within the mixed constituents.

7.5 MLFM Analysis of Hinglish Sentence

To illustrate these principles, the article reanalyzes the Hinglish sentence 2(a), this time employing the MLFM. This analysis provides a comparative framework to evaluate the strengths and limitations of both the ECM and MLFM approaches. To reiterate, Hinglish sentence 2(a) is as follows:

- 2 a Ram highly qualified *hai* and he can find a job *poori duniya mai*.
'Ram is highly qualified, and he can find a job anywhere in the world'.

As the following analysis will demonstrate, the ECM insufficiently accounts for the presence of asymmetry in code-switched Hinglish sentences, whereas the MLFM demonstrates relatively effective performance in this context. Additionally, while the ECM aids in understanding the grammaticality of the code-switched sentence, it fails to address whether the speaker is inserting Hindi words into an English sentence or vice versa. The MLFM, on the other hand, can address this ambiguity concerning the predominance of one language over the other within a code-switched utterance. Myers-Scotton (1993a) proposes a discourse-oriented approach to identify the ML, suggesting that it is the language containing more morphemes in each utterance. However, identifying the ML can be complicated as it can shift within a single conversation. Alternative methods for determining the ML from a structural perspective have also been proposed by Joshi (1985) and Treffers-Daller (1994). Structural

analysis of the Hinglish sentence 2(a) reveals that Hindi provides the grammatical framework (such as the syntactical scaffolding of verbs and prepositional phrases), whereas English contributes key lexical elements (such as adjectives, pronouns, and verb phrases). Because the core grammatical framework of sentence 2(a) is provided by Hindi, one can infer Hindi to be the ML. This analysis can be explained as follows: First, ‘Ram’, a proper noun, is used consistently in both Hindi and English. Second, the verb ‘hai’ in Hindi, which roughly translates as ‘is’ in English, shows the ML structure. Lastly, the Hindi phrase “poori duniya mai”, translating as “anywhere in the world”, indicates location. In short, with grammatical elements like verbs and prepositional phrases being furnished from Hindi in Hinglish sentence 2(a), Hindi is inferred to be the ML in this case.

On the other hand, English serves as the EL inserted into the Hindi structure. The English phrase ‘highly qualified’ in the sentence describes Ram’s qualifications. The coordinating conjunction ‘and’ in English connects the two clauses. The subject pronoun ‘he’ in English refers to Ram. Another English phrase ‘can find a job’ is used by the speaker to describe Ram’s ability. Lastly, an English noun phrase, ‘a job’, is used to refer to the high chances of Ram’s employability. With the contribution of the ML (Hindi) and EL (English) identified, the speaker’s use of code-switching in the sentence can be further parsed as follows:

Matrix Language Frame (Hindi Structure)

Component	Description
Subject	“Ram” (remains constant as it’s a proper noun used universally)
Predicate	“highly qualified hai” (mix of English adjective and Hindi verb)
Conjunction	“and” (English conjunction used to link clauses)
Additional Clause	“he can find a job poori duniya mai” (mixed structure with English embedded phrase)

Embedded Language Usage (English)

Component	Description
Clause	“highly qualified” and “can find a job” are English phrases embedded within the Hindi framework.
Pronoun	“he” is an English subject pronoun embedded within the sentence.
Prepositional Phrase	“poori duniya mai” (anywhere in the world) uses Hindi for the locative context.

7.6 Summary of Analysis

Matrix Clause 1: Ram highly qualified hai

Component	Description
Matrix Language	“Ram” (noun) + “hai” (verb)
Embedded Language	“highly qualified” (adjective phrase)

Matrix Clause 2: and he can find a job poori duniya mai:

Component	Description
Matrix Language	“and” (conjunction) + “poori duniya mai” (prepositional phrase)
Embedded Language	“he can find a job” (subject-verb-object phrase)

As demonstrated above, in this Hinglish sentence, Hindi supplies the grammatical framework through its use of verbs and prepositional phrases. In contrast, English introduces essential lexical components, including adjectives, pronouns, and verb phrases. This blending reflects a common pattern in Hinglish where English vocabulary is inserted into the grammatical framework of Hindi.

8 Conclusion

The preceding analysis compared two prominent constraint models of CS: the Equivalent Constraint Model (ECM) and the Matrix Language Frame Model (MLFM). As noted by scholars such as Nishimura (1985; 1986) and Chan (2009), ECM has been notably influential in the study of CS between typologically similar language pairs, such as English and Spanish. ECM’s symmetrical approach to CS, which distinguishes between loanwords and code-switching, has garnered popularity among linguists focusing on such language pairs. In contrast, MLFM is favored for analyzing typologically distinct language pairs, such as English-Hindi (Pandit 1986) and Japanese-English (Nishimura 1985, 1986). The popularity of MLFM in these contexts can be attributed to its emphasis on asymmetry and its application of Occam’s Razor to distinguish between borrowing and code-switching.

Recent scholarship has expanded the understanding of code-switching by incorporating diverse perspectives and methodologies, significantly contributing to the study of Hinglish. For instance, research by scholars like Kachru (2006) and Kothari (2011) emphasizes the socio-cultural dimensions of code-switching, highlighting how Hinglish serves as a marker of identity and cultural hybridity among speakers. Furthermore, the integration of

psycholinguistic theories into CS studies, as suggested by Wei (2009), enhances our understanding of how bilingual speakers navigate between languages in real-time, revealing cognitive processes that underpin code-switching behavior.

Emerging approaches that explore the neuro-linguistic dimensions of bilingual cognition (Kutas 2009) also provide valuable insights into how language processing occurs in bilingual individuals, further enriching the discourse on Hinglish. These developments signal a vibrant and evolving field, offering new insights into the complexities of bilingual language usage, such as Hinglish, and their broader implications for linguistic theory. By drawing from interdisciplinary research, scholars can deepen the analysis of code-switching phenomena, shedding light on the intricate relationships between language, identity, and cognition in bilingual contexts.

Bibliography

- Backus, A. (1992). *Patterns of Language Mixing: A Study in Turkish-Dutch Bilingualism*. Wiesbaden: Harrassowitz.
- Berk-Seligson, S. (1986). "Linguistic Constraints on Intrasentential Code-Switching: A Study of Spanish/Hebrew Bilingualism". *Language in Society*, 15(3), 313-48.
Available at <https://doi.org/10.1017/S0047404500011799>
- Bhatia, T.K.; Ritchie, W.C. (eds) (2013). *The Handbook of Bilingualism and Multilingualism*. 2nd ed. Hoboken: Wiley-Blackwell.
- Bullock, B.E.; Toribio, A.J. (eds) (2009). *The Cambridge Handbook of Linguistic Code Switching*. Cambridge: Cambridge University Press.
- Canagarajah, S. (2011). "Translanguaging in the Classroom: Emerging Issues for Research and Pedagogy". *Applied Linguistics Review*, 2, 1-28.
Available at <https://doi.org/10.1515/9783110239331.1>
- Canagarajah, S. (2013). *Translingual Practice: Global Englishes and Cosmopolitan Relations*. London: Routledge.
- Chan, B. (2009). "Code-Switching Between Typologically Distinct Languages". Bullock, Toribio 2009, 182-98.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge (MA): MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge (MA): MIT Press.
- Fodor, J.A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.
- Gumperz, J.J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.
- Joshi, A. (1985). "Processing of Sentences with Intrasentential Code-Switching". Dowty, D.R.; Kattunen, L.; Zwicky, A. (eds), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Cambridge: Cambridge University Press, 90-205.
- Kachru, Y. (2006). *Hindi*. Amsterdam: John Benjamins Publishing Company.
- Kothari, R. (2011). *Chutnefying English: The Phenomenon of Hinglish*. Gurgaon: Penguin Books India.
- Koul, O. (2008). *Modern Hindi Grammar*. Hyatsville: Dunwoody Press.
- Krishnaswamy, N.; Krishnaswamy, L. (2006). *The Story of English in India*. New Delhi: Foundation Books.

- Kutas, M.; Moreno, E.; Wicha, N. (2009). "Code-Switching and the Brain". Bullock, Toribio 2009, 289-306.
- Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- Myers-Scotton, C. (1993a). *Duelling Languages: Grammatical Structure in Code-Switching*. Oxford: Oxford University Press.
- Myers-Scotton, C. (1993b). *Social Motivations for Code-Switching: Evidence from Africa*. Oxford: Oxford University Press.
- Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
- Myers-Scotton, C. (2003). "What Lies Beneath: Split (Mixed) Languages as Contact Phenomena". Matras, Y.; Bakker, P. (eds), *The Mixed Language Debate: Theoretical and Empirical Advance*. Berlin; New York: De Gruyter Mouton, 73-106.
Available at <https://doi.org/10.1515/9783110197242.73>
- Newmeyer, F.J. (1986). *Linguistic Theory in America*. London; Oxford; Boston; New York; San Diego: Academic Press.
- Nishimura, M. (1985). *Intrasentential Code-Switching in Japanese and English* [PhD Dissertation]. Philadelphia: University of Pennsylvania.
- Nishimura, M. (1986). "Intra-Sentential Code-Switching: The Case of Language Assignment". Vaid, J. (ed.), *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*. Mahwah: Lawrence Erlbaum, 123-43.
- Pandit, I. (1986). *Hindi-English Code-Switching: Mixed Hindi-English*. New Delhi: Datta Book Centre.
- Poplack, S. (1978). *Quantitative Analysis of Constraints on Code-Switching*. New York: Centro de Estudios Puertorriqueños.
- Poplack, S. (1980). "Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a Typology of Code-Switching". *Linguistics*, 18(7-8), 581-618.
- Poplack, S. (2001). "Code-switching (Linguistics)". Smelser, N.; Batles, P. (eds), *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Elsevier Science Ltd, 2062-5.
- Poplack, S. (2018). *Borrowing: Loanwords in the Speech Community and in the Grammar*. Oxford: Oxford University Press.
- Poplack, S.; Wheeler, S.; Westwood, A. (1989). "Distinguishing Language Contact Phenomena: Evidence from Finnish-English Bilingualism". *World Englishes*, 8(4), 389-406.
Available at <https://doi.org/10.1111/j.1467-971X.1989.tb00677.x>
- Sankoff, D.; Poplack, S. (1981). "A Formal Grammar for Code-Switching". *Papers in Linguistics*, 14(1), 3-45.
- Treffers-Daller, J. (1994). *Topics in Sociolinguistics*. Vol. 9, *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin: Walter de Gruyter.
- Wei, L. (2009). "Code-Switching and the Bilingual Mental Lexicon". Bullock, Toribio 2009, 270-88.
- Weinreich, U. (1953). *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.
- Young, V.A.; Lovejoy, K.B.; Barrett, R.; Young-Rivera, Y. (2014). *Other People's English: Code-Meshing, Code-Switching, and African American Literacy*. New York: Teachers College Press.

The Emergence of Rhymed Meters in the Indo-Aryan Prosody

Ryosuke Masaoka

University of Tokyo, Japan

Abstract Indo-Aryan poetry began to incorporate the end rhyme as an obligatory metrical rule from Apabhraṃśa literature in the eighth century, although the precise reasons for this development remain uncertain. The emergence of the end rhyme is a multifactorial event, influenced by not only phonological but also sociolinguistic factors, such as the impact of more prestigious literary traditions. From a morphophonological perspective, Arjunwadkar suggests that the simpler variations in word endings in later languages facilitated the end rhyme. This paper evaluates his hypothesis by calculating the actual difficulty of rhyming with textual data of Sanskrit, Middle Indo-Aryan, and New Indo-Aryan and argues that less difficulty of rhyming in later languages contributed to the adoption of the end rhyme.

Keywords Middle Indo-Aryan. Apabhraṃśa. End Rhyme. Phonology. Prosody.

Summary 1 Introduction. – 2 Background. – 3 Test Design. – 4 Results. – 5 Conclusion.



Peer review

Submitted 2025-03-25
Accepted 2025-05-07
Published 2025-07-01

Open access

© 2025 Masaoka | © 4.0



Citation Masaoka, R. (2025). "The Emergence of Rhymed Meters in the Indo-Aryan Prosody". *Bhasha*, 4(1), 77-94.

1 Introduction

Peust (2014, 343) defines the concept of ‘rhyme’ as “the phonological identity of substrings of lines”. This feature is prevalent in poetic texts across various literary traditions. Among different types of rhyme, one notable type is the ‘end rhyme’. In this study, the term ‘end rhyme’ is used to refer to the phonological identity of the final (typically more than one) substrings of metrical units.

New Indo-Aryan poetical traditions generally employ the end rhyme as a mandatory metrical rule. Typically, the final two syllables or four morae (*mātrās*) of *pādas* are matched as the end rhyme. (1) is an example of Braj Bhāṣā poetry.

(1) *caraṇa kamala, baṃdaṃ hari rāi*
jākī kṛpā paṃgu giri laṃghai, aṃdhe kaṃ saba kachu
darasāi
bahirau sunau gūṃga puni bolai, raṃka calai sira chitra
dharāi
sūradāsa svāmī karunāmaya, bāra bāra baṃdau tihim pāi
(Sūrdas, *Sūr-sāgar*, 1, ed. Vājpeyī 1949, 1)

I salute the lotus feet of Hari, the king,
through whose mercy the lame will cross a mountain, the blind
[may all things see.
the deaf hear, the dumb speak again, the pauper go with a
[canopy held above his head;
Sūrdās’ lord is composed of compassion, again and again I
[salute his feet.
(Snell 1991, 85)

However, in the early verse literature of Indo-Aryan languages, that is, in most of the Sanskrit verse literature following the *R̥gveda*, the end rhyme is not a metrical rule.¹ Furthermore, the end rhyme is not used in early Middle Indo-Aryan (Prakrit) texts such as the Pāli Buddhist canons.

It is from the second half of the first millennium AD that Indo-Aryan prosody began to use the end rhyme as a metrical rule. According to Ollett (2017, 102), meters using the end rhyme called

¹ Indeed, the end rhyme is used as an optional stylistic device already in the early Indo-Aryan literature, too. Klein (2002, 221) indicates that a rhyme-matching rhetorical figure called homoioteleuton serves phonetic and grammatical functions in *R̥gveda*. The end rhyme is also used as one of the *alamkāras* in classical Sanskrit poetry. However, such rhetorical uses of the end rhyme should be distinguished from the end rhyme as an obligatory metrical rule. As a metrical rule indivisibly connected with the prosody, the end rhyme is not found in early Indo-Aryan literature.

khañjaka appeared partially in the Sanskrit-Prakrit mixed drama of the seventh century, such as Harṣa's *Ratnāvalī*. Furthermore, in the verse literature of Apabhraṃśa from the eighth century onwards, rhymed metrical patterns dominated the entire body of literature. (2) is an example of the *khañjaka* verse from Harṣa's *Ratnāvalī*. (3) is an example of the Apabhraṃśa verse.

(2) kusumāuhapiadūao maulāiabahucūao |
siḍhiliamāṇaggahaṇao vāai dāhiṇapavaṇao
(Harṣa, *Ratnāvalī* 13, ed. Kale 1921, 14)

The southern breeze is here, bringing buds to
the mango, the dear messenger of the God of Love,
slackening anger and quarrels,
(Ollett 2017, 101, ed. Bhayani 1953, 1)

(3) paṇaveppiṇu āibhaḍārāhō | saṃsārasamudduttārāhō
(*Paūmacariu* I, 1, 1)

Bowing before Rishabha, the first lord,
who escaped from the ocean of samsara;
(De Clercq 2018, 9)

As described above, the end rhyme was not systematically used in the early Indo-Aryan languages. Still, from the seventh century, it emerged as a rule of metrical composition, then prevailed in the entire Late Middle Indo-Aryan literature in the eighth to tenth centuries, and has continued to be used up to the present day.

The reason for the emergence of the end rhyme is an important yet unsolved problem in many languages that have rhymed poetry. In particular, the development of the end rhyme in Indo-Aryan, which we cannot explain solely by language contact, has not been thoroughly studied.

The emergence of rhyme is a multifactorial phenomenon. Aside from language-internal factors such as sound change, there are factors like a cultural, musical, or stylistic factor to consider. This study examines one of the factors behind the emergence from a morphophonological point of view.

In this paper, I argue that the diversity of word-final syllables diminished in the language changes from Sanskrit to later languages, which reduced the difficulty of matching the end rhyme in Sanskrit and created the environment more suitable for the emergence of the end rhyme in later languages. I would like to argue that the simplification of the word-final patterns contributed to the emergence of the end rhyme in Indo-Aryan prosody.

2 Background

In this section, we organize the types of languages and the periods of verse texts in the Indo-Aryan languages. Thereafter, we will review previous research and examine the hypotheses on the emergence of the end rhyme.

The earliest verse literature in Indo-Aryan languages is the Vedic literature, including the *Ṛgveda*. The language of the Vedic literature is Old Indo-Aryan (OIA), and the later Classical Sanskrit is the standardized Old Indo-Aryan. As for verse literature in Sanskrit, the two epics, *Mahābhārata* and *Rāmāyaṇa*, were composed around the beginning of the Common Era, and *Kāvya* literature flourished from the fourth century AD onwards.

From around the third century BC, verse literature in the Middle Indo-Aryan (MIA, Prakrit) began to emerge. Middle Indo-Aryan languages are divided into three stages. The oldest stage includes Pāli with Buddhist canon. The middle stage (Middle Middle Indo-Aryan) consists of Prakrits used in Jain texts, such as *Ardhamāgadhī*, and Prakrits used in drama and lyrical poetry, such as *Māhārāṣṭrī* and *Śaurasenī*, with works influenced by the Sanskrit *Kāvya* literary tradition. The newest stage is *Apabhraṃśa* (Late Middle Indo-Aryan), with verse para-canonical works by Jain authors.

From the thirteenth century onwards, verse literature began to emerge in the New Indo-Aryan languages (NIA), including Bengali, Marathi, Braj Bhāṣā, Awadhi, and others. In addition, Urdu verse literature, exemplified by works such as *Kadam Rāo Padam Rāo*, which was heavily influenced by Islamic Persian poetry, also developed around the fourteenth to fifteenth centuries.

The history of Indo-Aryan poetries, as described so far, is summarized in Table 1.

Table 1 Indo-Aryan verse literature and the use of the end rhyme

Period	OIA	Early MIA	Middle MIA	Late MIA	NIA
BC 1600-500	Vedic Literature	-	-	-	-
BC 500-AD 700	Epic Literature Kāvya Literature	Pāli Canonical Texts	Prakrit Drama(hybrid with Sanskrit) Lyric Poetry Jain Canonical Texts	-	-
AD 700-1200	-	-	-	Apabhraṃśa Tantric Literature Apabhraṃśa Jain Paracanonical Literature	-
AD 1200-	<i>Gīta Govinda</i>	-	-	-	NIA literature

As shown in section 1, literature with the end rhyme throughout the entire composition as a metrical rule in Indo-Aryan languages spread from the eighth century AD onwards, namely in Late Middle Indo-Aryan verse literature and subsequent modern Indo-Aryan verse literature.² Meanwhile, even works created after the eighth century AD in older languages such as Sanskrit and Pali did not use rhyme, probably because of the literary tradition. The exception is Sanskrit lyrical poetry in the twelfth century, *Gīta Govinda*, which features rhymed verses, presumably influenced by the Middle Indo-Aryan and New Indo-Aryan poetic traditions that used the end rhyme.³

Then the question naturally arises: why did the end rhyme emerge in Indo-Aryan languages? This phenomenon can be attributed to multiple factors. The influence of other literary traditions, particularly the Perso-Arabic prosody, which uses the end rhyme, is one possible factor. However, the use of the end rhyme in Apabhraṃśa was from the eighth century, a period when works composed in the Persian language and with the Persian metrical system were not prevalent.⁴ In addition, contact with languages whose poetry use the end rhyme does not necessarily lead to the adoption of the end rhyme. For example, despite the extensive contact with Classical Chinese poetry, which uses elaborate end rhyme, Classical Japanese poetry never introduced the end rhyme. Therefore, the influence of Perso-Arabic poetry alone does not provide a convincing explanation, unless some poets deliberately imported Perso-Arabic styles.

Mahesh (1964, 182) suggests that “the common folk, who continuously use rhymes in everyday life, are the originators of this [rhyme]”. A stylistic reason such as Mahesh’s theory could be one factor behind the adoption of the end rhyme. It is possible that Indian poets gradually came to appreciate the aesthetic appeal of end rhyme from folk songs and began to use it, but we cannot prove or refuse this theory.

From linguists’ point of view, as stated by Fabb and Sykärä (2022, 6), “There is a widespread view that a language offers ‘affordances’ which make rhyme, or a particular kind of rhyme, possible in its verbal arts, and that languages differ in their affordances”. This concept is referred to as the ‘development hypothesis’ in Fabb 2010. From this perspective,

2 However, the earliest meters of Apabhraṃśa poetry do not use the end rhyme as a mandatory rule. One example is *gāhā* meter, whose origin is the earlier Prakrit prosody.

3 Cf. Kurkarni 1965, xii-xiii. Therefore, I do not consider rhyming verses in the *Gīta Govinda* as a counterevidence of the primary assertion that language change promoted the emergence of the end rhyme.

4 Nagasaki (2012) also discusses the origin of the end rhyme in *Savaiyā* meters in Hindi prosody and states: “Perso-Arabic metre as applied to New Indo-Aryan was not yet in widespread use except among the court poets of the Mughal Dynasty during the Bhakti period in which *Savaiyā* gained popularity” (Nagasaki 2012, 123).

it is plausible that the affordances of rhyme evolve alongside language change. Consequently, it is reasonable to consider language change as a factor in the emergence of the end rhyme.

In relation to the development hypothesis, some scholars argue that stress accent facilitates the development of rhyming poetry with rhyme typically including a stressed syllable. For example, Sedgwick (1924, 335) states in his paper on the origins of Medieval Greek and Latin rhyme: “As long as stress was ignored in verse, modern rhyme, which depends on it, was obviously impossible; it is recurrence of stress, not of verse-ictus which causes rhyme”.

Certainly, there is a possibility that the Late Middle Indo-Aryan language had a stress system. A long vowel before or after an accented syllable in Old Indo-Aryan is reduced to a short vowel in Apabhramśa (De Clercq 2005, 2099),⁵ suggesting that Apabhramśa may have inherited a stress system from the Old Indo-Aryan pitch accent system. However, this reduction is already found in the Middle Middle Indo-Aryan languages⁶ and is inconclusive evidence for a synchronic stress accent system in Late Middle Indo-Aryan. Furthermore, even if such a stress system did exist, Late Middle Indo-Aryan end rhyme sometimes lacks a syllable that would be considered stressed if it had a stress system.⁷

Fabb and Sykäre (2022, 7) also note that “the vocabulary of a language can make certain kinds of rhyme more easy, or less easy”. Some might attribute the origin of the end rhyme to this, but they oppose this view in the same work.

It is worth noting that the development hypothesis also makes a possibly incorrect assumption about aesthetics, which is that a verbal art should be constrained so that it is easy to produce poetry relative to the resources of the language. But we should also remember that verbal arts often gain their value by their difficulty, and by the skill of the composer in overcoming those difficulties. (Fabb, Sykäre 2022, 7)

I fully accept that the difficulty of rhyming can be overcome by poets’ creativity; but in traditions like Indo-Aryan poetry where

⁵ For example, OIA *ūlūka* ‘owl’ > Late MIA *uluhaṃ* and OIA *pravāhā-* ‘stream’ > Late MIA *pavaha-*

⁶ Using the examples above, Māhārāṣṭrī has such forms with short vowels as *uluga-*, and *pavaha-*.

⁷ Using the example of OIA *ūlūka* ‘owl’, the initial syllable of Late MIA *uluhaṃ* would be stressed. However, *Paūmacariu* has a verse whose rhyme (*-uhaṃ*) lacks the stressed syllable.

(4)*nāya-naūlayaṃ kāyaloluhaṃ | hatthi-ajayaraṃ dava-mahīruhaṃ ||*
(PC LXXXI, X, 3, ed. Bhayani 1960, 206)

the end rhyme became pervasive, the development hypothesis is convincing, and I would like to pursue that possibility by analyzing actual language data to determine whether the easiness of the end rhyme correlates with its occurrence.

As for Indic languages, Arjunwadkar (1985, 197-8) argues that the simplification of the morphology of later languages is a factor behind the development of the end rhyme. According to him, Sanskrit has a more complex inflectional system than Marathi, resulting in a vast number of combinations of phonemes in the word-final environment, making it unsuitable for rhyming. Unlike the theories of Mahesh (1964) I mentioned above, it is possible to verify this theory with linguistic data, which Arjunwadkar has not done, however.

It is highly likely that the difficulty of rhyme, which is inseparably linked to the end rhyme, plays a role in its emergence. I would like to pursue the line of Arjunwadkar's hypothesis that the phonological variation at the end of words decreased over time in Indo-Aryan languages, resulting in ready rhymability and the emergence of the end rhyme.

3 Test Design

This section presents the methods and results of the data survey.⁸ The survey utilizes seven texts as shown in Table 2 below.

Table 2 Materials used

Abbreviation	Title	Author/Editor	Language	Attested period
RV	<i>R̥gveda Saṃhitā</i>	-	OIA	twelfth-tenth century BC
SS	<i>Gāhā Sattasāi</i>	Hāla	Māhārāṣṭrī (Middle MIA)	first-second century
SB	<i>Setubandha</i>	Pravarsena		fifth century
GV	<i>Gaūḍavaho</i>	Vākpatirāja		eighth century
PC	<i>Paūmacariu</i>	Svayambhūdeva	Late MIA	seventh-tenth century?
PD	<i>Pāhuḍadohā</i>	Rāmasiṃha	(Apabhraṃśa)	tenth century
DV	<i>Dohāvalī</i>	Tulsīdāsa	NIA (Braj Bhāṣā)	sixteenth century

These texts are significant and representative of the poetical traditions in each language. *R̥gveda Saṃhitā* marks the beginning of the Old Indo-Aryan poetry. *Gāhā Sattasāi* is the starting point of the Middle Middle Indo-Aryan literary tradition (Ollett 2017, 7).

⁸ Codes and result data are published in GitHub. <https://github.com/ryosuke-masaoka/The-Emergence-of-Rhymed-Meters-in-the-Indo-Aryan-Prosody>.

Setubandha and *Gauḍavaho* are prominent works among Māhārāṣṭrī court epics. *Paūmacariu* is one of the earliest examples of Late Middle Indo-Aryan literature. Although *Pāhuḍadohā* is not as important as *Paūmacariu*, it remains one of the main Apabhraṃśa works identified by Tagare (1987, 16-20). *Dohāvalī* is an influential work in the NIA poetry.

Furthermore, all these texts are available in electronic form, except for *Pāhuḍadohā*. As for the *Ṛgveda Saṃhitā*, the Lubotsky's (1997) text⁹ is available for download from VedaWeb.¹⁰ Electronic texts of *Gāhā Sattasāi*, *Paūmacariu*,¹¹ and *Dohāvalī* are accessible on GRETIL.¹² Electronic texts of *Setubandha* and *Gauḍavaho*¹³ can be obtained from Prakrit Digital Text Project.¹⁴ As for *Pāhuḍadohā*, I created text data based on Śāstrī's edition (1998).

As shown in section 2, Old Indo-Aryan texts do not match the end rhyme. Some Middle Middle Indo-Aryan texts use the end rhyme as Ollett (2017, 102) reports, but *Gāhā Sattasāi* doesn't seem to employ

9 Lubotsky's (1997) text presents each word with sandhi effects eliminated, resembling the *padapāṭha*, but in the underlying representation. Final /-r/ is retained in its original form to distinguish it from final /-s/. However, for the purpose of this study, I changed the final /-r/ into -ḥ because the distinction is unnecessary in *pausa*, which corresponds to the rhymed position in later texts. Metrically lengthened forms, such as *vedā+*, were replaced by short forms, such as *veda*, in this analysis. From a philosophical perspective, the *Ṛgveda Saṃhitā* text we have now is not entirely identical to the original composition in the earliest stage of Old Indo-Aryan because of the editing process which replaced older forms with younger, more familiar forms similar to those in Classical Sanskrit. For example, the oldest text contained at least two genitive plural endings (Arnold 1905, 104), *-ām* and **-am* (< PIIr. **-aHam* < PIE **-oHom*), but the later editing process replaced **-am* forms into *-ām*, which became standard in Classical Sanskrit. To access the text before the editing process, one can use restored text by van Nooten and Holland (1994). However, in this analysis, I used Lubotsky's (1997) text as it approximates the Old Indo-Aryan text closely enough and the language of the text after the editing process is more closely related to the fundamental language of later Old Indo-Aryan and Sanskrit poetic traditions.

10 VedaWeb. <https://vedaweb.uni-koeln.de>.

11 *Ya-śruti* was eliminated from the text of *Paūmacariu* in this survey. In *Paūmacariu*, *ē* and *ō* in open syllables are orthographical variants of *i* and *u*, respectively, and are thus replaced by *i* and *u*.

12 GRETIL – Göttingen Register of Electronic Texts in Indian Languages. <https://gretil.sub.uni-goettingen.de/gretil.html>.

13 Orthographical irregularities sometimes pose challenges when working with Middle Middle Indo-Aryan texts. Vowels *e* and *o* in closed syllables are sometimes written as *i* and *u*. I was unable to fully correct these irregularities, so I conducted a survey using the text with the original orthography, and these results are presented as the main findings. Also, I conducted a survey using the text where all instances of *e* and *o* in close syllables were replaced with *i* and *u*, and these results are provided in the footnotes. The true values likely lie between the values obtained from these two surveys. External sandhi in the Middle Middle Indo-Aryan texts was unapplied by the author.

14 Ollett, A. *Prakrit Digital Text Project*. https://github.com/aso2101/prakrit_texts.

rhyme because this text is from the earliest stage of Middle Indo-Aryan. *Setubandha* and *Gaṇḍavaho* have a few rhyming verses, but they do not match the end rhyme for the most part. *Paūmacariu*, *Pāhuḍadohā*, and *Dohāvalī* show the use of the end rhyme in the entire text.

The purpose of this survey is to compare the diversity of word-ending patterns in each text. First, let us clarify the definition of the ‘word-ending pattern’ (hereafter referred to as the ‘rhyme pattern’). This definition can only be established when the definition of the end rhyme in Late Middle-Aryan texts is clearly defined.

According to Yamahata (2009, 854), there were several kinds of end rhymes in Late Middle Indo-Aryan poetries. These included end rhymes with only one syllable, called *puruṣatuk*, and end rhymes with two or three syllables, called *komalatuk*. Given the variety of rhyming schemes in Late Middle Indo-Aryan poetry,¹⁵ I conducted a preliminary survey to identify the predominant rhyming structures. Among rhyming couplets of *Paūmacariu*, 94.56% matched pāda-final -VC₀VC₀#, 56.24% matched pāda-final -C₀VC₀VC₀#, 50.13% matched pāda-final -VC₀VC₀VC₀# and 12.61% matched pāda-final -CVC₀VC₀VC₀#. These results indicate -VC₀VC₀# and -VC₀VC₀VC₀# as the main rhyming schemes. Therefore, I conducted separate surveys targeting -VC₀VC₀# as the rhyme pattern and targeting -VC₀VC₀VC₀# as the rhyme pattern. As for imperfect rhymes, several patterns are observed (for example, /nt/ rhymes with /nd/); however, only the perfect matches are considered for this study. Monosyllabic words were excluded from this investigation.

The second question concerns how we should compare diversity. There are several diversity indices available, one of which is Gini-Simpson’s Diversity Index. This is used in biology to calculate ecological diversity and is considered the most independent diversity index from the sample size (Kunakh et al. 2023, 136-8). According to Kunakh et al. (2023, 134), the definition of Gini-Simpson’s Diversity

15 In introducing the rhyme pattern, I describe how rhyme in Late Middle Indo-Aryan works. According to Tagare (1987, 39), Late Middle Indo-Aryan had 9 vowels (*a, ā, i, ī, u, ū, ṛ, e, o*) and 31 consonants (*k, kh, g, gh, c, ch, j, jh, ñ, ṭ, ṭh, d, ḍh, ṇ, t, th, d, dh, n, p, ph, b, bh, m, r, l, v, ś, s, h, ḷ*). The vowel *ṛ* appeared only in loanwords from Sanskrit, and *ś* was found only in eastern Apabhraṃśa, a different variety from the language of *Paūmacariu* and *Pāhuḍadohā*. All vowels except for *ṛ* could be nasalized. Unlike Sanskrit, Apabhraṃśa did not have diphthongs such as *ai* and *au*. According to De Clercq (2018, xx), the metrical structure of *Paūmacariu* is as follows: “The *Paūmacariu* is divided into five books (*kāṇḍas*), totaling ninety chapters (*sandhis*), each consisting of on average sections (*kaḍavakas*). A section is made up of a body of rhyming couplets (*yamakas*) followed by a stanza (*ghattā*) of two, four, or six quarter-verses with complex rhyming scheme. [...]. Each chapter opens with a stanza usually in the same meter as the closing stanzas of the sections of that chapter. Sometimes the sections commence with a longer stanza of two or four quarter-verses before the body of rhyming couplets”. For further details on the complex rhyming scheme of *ghattās*, see Bhayani 1953, 78-92.

Index is as follows:

$$\text{Gini – Simpson's Diversity Index, } D_{S1} = 1 - \sum_{i=1}^S P_i^2$$

(P_i = proportion for the i th species, S = total number of species in the community)

The complement of Gini-Simpson's Diversity Index (that is, $\sum_{i=1}^S P_i^2$) can be interpreted as the probability that two randomly selected tokens happen to belong to the same type. If we apply this idea, we can get the probability of a coincidental rhyme match in each text.

The frequency of a given rhyme pattern (r_i) can be considered as the probability $\{P(r_i)\}$ of (r_i) occurring at any given word-final location in the text. This $\{P(r_i)\}$ is acquired by dividing the frequency of (r_i) by the total number of rhymes in the text. Then, the probability of (r_i) occurring at another location is $\{P(r_i)\}$, too. Hence, the probability of the same (r_i) happening to occur at any two-given word-final locations is $\{P(r_i)\}^2$. By calculating this probability for all rhyme patterns and summing them, we obtain the probability that any given rhyme pattern will match at the two locations. Let us call this Rhyme Match Probability (RhMP).

$$(RhMP) = \sum_{i=1}^S \{P(r_i)\}^2$$

(S = the total number of rhyme pattern types)

The Rhyme Match Probability calculated in this way does not serve only as an indicator of diversity in word-endings, but also as an indicator of how easy it is to match the end rhyme in a language. If it is high, it is easier to match the end rhyme.

To avoid overrepresentation of rhyme patterns in rhyming texts such as *Paūmacariu*, *Pāhuḍadohā*, and *Dohāvalī*, multiple identical rhyme patterns that arise due to the metrical rhyming rule are counted only once.¹⁶ For example, two occurrences of -VC₀VC₀# rhyme patterns (iṇu, āhō) are extracted from (3), not three (iṇu, āhō, āhō), because the second āhō in 'saṃsārasamudduttārāhō' is present due to its rhyming position with āhō in 'āibhaḍārāhō'.

¹⁶ If rhyme patterns do not match at the rhyming positions, I counted all patterns because they do not appear because of the rhyming rule.

(3) paṇaveppiṇu āibhaḍārāhō | saṃsārasamudduttārāhō
(*Paūmacariu* I, I, 1 reappeared)

4 Results

As for the results of the survey targeting -VC₀VC₀#, Table 3 shows the token count, the number of rhyme pattern types, and the Rhyme Match Probability.

Table 3 Rhyme Match Probability of -VC₀VC₀#

Language	Vedic Skt.	Māhārāṣṭrī Prakrit ¹⁷	Late Middle Indo-Aryan		NIA(Braj)		
Text	RV	SS	SB	GV	PC	PD	DV
Token count	134,443	9,399	10,024	8,496	72,730	2,456	7,754
Types	5,506	837	659	644	1,184	359	852
Average Frequency	24.42	11.23	15.21	13.19	61.43	6.84	9.10
RhMP	0.321 %	1.212 %	1.183 %	1.464 %	2.079 %	2.572 %	1.163 %

According to the data, it is considered that as time progressed, the variation in rhyme pattern types decreased, and the diversity of rhyme patterns increased, making it progressively easier to follow a rhyming rule.

In the *Ṛgveda*, 80% of the cumulative total of all rhyme patterns are covered by 707 patterns, 160 in the *Sattasāi*, 140 in the *Setubandha*, 125 in the *Gauḍavaho*, 95 in the *Paūmacariu*, 77 in the *Pāhuḍadohā*, and 173 in the *Dohāvalī*. Fewer rhyme patterns account for the 80% in newer texts except for the *Dohāvalī*.

Interestingly, *Dohāvalī*, the newest text of the materials used, exhibited its Rhyme Match Probability close to those of the Middle MIA texts, yet lower than those of the Apabhramśa texts. This observation suggests that the emergence of the end rhyme is a one-way, irreversible phenomenon. Despite the increasing difficulty in the end rhyme, they continued to compose rhymed poetry due to the rhyming tradition.

¹⁷ The results of the survey replacing *e* and *o* in closed syllables with *i* and *u* are as follows:

Texts	SS	SB	GV
Token count	9,399	10,024	8,496
Types	809	634	631
Average Frequency	11.51	15.81	13.46
RhMP	1.213 %	1.186 %	1.465 %

As for the results of the survey targeting $-VC_0VC_0VC_0\#$, Table 4 shows the token count, the number of rhyme pattern types, and the Rhyme Match Probability.

Table 4 Rhyme Match Probability of $-VC_0VC_0VC_0\#$

Language	Vedic Skt.	Māhārāṣṭrī Prakrit ¹⁸				Late Middle Indo-Aryan	NIA (Braj)
Text	RV	SS	SB	GV	PC	PD	DV
Token count	70,058	6,500	8,245	6,724	54,947	1488	3,638
Types	15,201	3,116	2,984	3,058	8,316	811	1,492
Average Frequency	4.61	2.09	2.76	2.20	6.61	1.83	2.44
RhMP	0.0483%	0.101%	0.109%	0.109%	0.121%	0.312%	0.673%

These results are similar to those of the survey targeting $-VC_0VC_0\#$ in that later rhyming texts show higher values of the Rhyme Match Probability. It is noteworthy that the Rhyme Match Probability of *Paūmacariu* is close to those of *Sattasāi*, *Setubandha*, and *Gāṇḍavaho*. One possible explanation is that the end rhyme of the final two syllables ($-VC_0VC_0\#$) was established first, and more complex $-VC_0VC_0VC_0\#$ end rhyme developed later.

In order to compare the data above with those of other languages, I conducted the same survey in Italic languages. The materials are the first volume of *Aeneis* in Classical Latin (non-rhyming text) and *Yvain ou le Chevalier au Lion* in Old French (rhyming text). In this survey, rhyme pattern was defined as the part after the stressed or accented vowel in each word, which is the French rhyming scheme (Peust 2014, 371). Table 5 shows the results:

18 The results of the survey replacing *e* and *o* in closed syllables with *i* and *u* are as follows:

Text	SS	SB	GV
Token	6,500	8,245	6,724
Types	2,882	2,835	2,830
Average Frequency	2.56	2.91	2.38
RhMP	1.118%	0.115%	0.110%

Table 5 Rhyme Match Probability in Italic Languages

Languages	Classical Latin	Old French
Texts	the 1st volume of <i>Aeneis</i>	<i>Yvain ou le Chevalier au Lion</i>
Token counts	2,285	28,401
Types	1,334	820
Average Frequency	1.71	34.64
RhMP	0.141%	2.620%

This data on Italic languages supports the main claim of this paper that the development of the end rhyme is influenced by the increase of the Rhyme Match Probability. Other factors are said to contribute to the emergence of the end rhyme in Italic languages, and I am not claiming that the increase of the Rhyme Match Probability is the sole factor. However, from a comparative perspective, it is probable that an environment conducive to rhymed poetry is a necessary condition for the acceptance of the end rhyme in prosody.

Based on the data of Indo-Aryan languages (especially those with rhyme patterns of $-VC_0VC_0\#$) and Italic languages, I propose 1.5 ~ 2.0 % as a hypothetical threshold of rhyme emergence. In future research, this threshold should be refined with the data of more languages. Theoretically speaking, the more language data we examine, the closer we will inductively get to the ‘true’ threshold. In fact, non-linguistic factors play important roles in the rhyme emergence of every language, so it may be impossible to reach a unique threshold.

The survey conducted thus far has assumed only the language model, disregarding the meters. However, the end rhyme is a part of the metrical schemes and several major rhyming meters match the bisyllabic end rhyme in the Guru-Laghu (heavy-light, henceforth GL) rhythm. Therefore, I also extracted only the GL rhyme patterns and calculated the Rhyme Match Probability in each text. The results are as follows:

Table 6 Rhyme Match Probability of GL rhyme patterns¹⁹

Language	Vedic Skt.	Māhārāṣṭrī Prakrit			Late Middle Indo-Aryan		NIA(Braj)
Text	RV	SS	SB	GV	PC	PD	DV
Token count	56,016	3,847	3,407	3,487	22,309	742	2,669
Types	2,117	345	235	224	563	179	329
Average Frequency	26.46	11.15	14.50	15.57	39.63	4.07	8.11
RhMP	0.962%	3.28%	2.94%	4.32%	1.68%	1.73%	1.95%

Interestingly, the values of the Rhyme Match Probability in Middle Middle Indo-Aryan texts are significantly higher than those of the later rhyming texts. This observation suggests two possibilities. One possibility is that the end rhyme has a different origin from meters, which is unlikely given their inseparability. The second possibility is that the end rhyme may not be adopted if rhyming is too easy. Thus, there may not be only a minimal threshold for the rhyme emergence but also a maximal one.

As the factors of the increase of the Rhyme Match Probability, we can name mainly two factors: morphological simplification and syllable structure simplification. The latter is more important. Because of phonological changes, the restrictions on consonants that can appear in the coda of the syllables became stricter in MIA. For example, there are 161 different consonant clusters that appear in the middle of words 10 times or more in *Ṛgveda* according to Kobayashi (2004, 185-91), while there are only 44 clusters that appear 10 times or more in *Paūmacariu*. Less contrasted phonemes in the coda led to a reduction in the variety and deviation of rhyme patterns.

5 Conclusion

This paper examined the factors contributing to the emergence of the end rhyme in the Indo-Aryan prosody by a morphophonological approach. The results in section 4 revealed that the number of rhyme

¹⁹ The results of the survey replacing *e* and *o* in closed syllables with *i* and *u* are as follows:

Text	SS	SB	GV
Token	3,847	3,407	3,487
Types	325	222	219
Average Frequency	11.84	15.35	15.92
RhMP	3.28%	2.96%	4.32%

pattern types decreased over time and the diversity in their frequency increased, making it easier for poets to use the end rhyme. With the Rhyme Match Probability, we can observe distinct differences between older languages and Late Middle Indo-Aryan; it became easy enough to rhyme in Late Middle Indo-Aryan, while it was too difficult in Old Indo-Aryan. The Rhyme Match Probability increased in later languages and exceeded a certain threshold (here hypothetically 1.5 ~ 2.0%) in Late Middle Indo-Aryan. It is highly likely that the increased rhymability of literary languages, a factor closely related to the end rhyme, contributed to the development of the end rhyme.²⁰

As a reason for the absence of the systematic end rhyme (especially in classical Greek and Latin), it is sometimes claimed that it is too easy to match word-endings because many words share the same inflectional endings. While the main claim of this paper might appear to contradict this line of argumentation, in fact it aligns with it, although we need a detailed survey into Greek and Latin. In section 4, I suggested the possibility that both excessive ease (too high Rhyme Match Probability) and excessive difficulty (too low Rhyme Match Probability) can prevent the adoption of the end rhyme. As for Indo-Aryan languages, however, the minimal threshold is the primary concern.

As I stated in sections 1 and 2, the emergence of the end rhyme is a phenomenon that might have been triggered by multiple factors. Various elements, such as the cultural influence from Perso-Arabic poetry or purely stylistic reasons, also played roles in the end rhyme development. It is not easy to clarify the entire process by which the end rhyme became an obligatory element of Indo-Aryan poetry.

A culture-sensitive phenomenon, such as the emergence of the end rhyme, cannot be completely analyzed by a simple linguistic model. However, there is a correlation between the morphophonological difficulty and the presence of the end rhyme in each language, which may be identified as a contributing factor in its emergence. An increase in the Rhyme Match Probability means relaxing the verbal restriction to the end rhyme and therefore may be one of the necessary conditions for the emergence of the end rhyme. The development of the end rhyme is influenced not only by traditional and cultural factors but also by changes in the language, which form the basis of poetic composition. This study, which correlates the ease of the end rhyme with its emergence, provides an argument in favor of the development hypothesis.

20 An alternative hypothesis is also possible: the reduction in word endings made 'interesting' end rhyme more difficult due to the oversimplicity of word endings, thereby making the end rhyme a more fascinating ornament. However, this paper does not adopt this hypothesis because rhyming words with the same suffixes are frequently used and probably not considered 'less interesting' in Late Middle Indo-Aryan rhymed poetry.

Appendix: Rhyme Patterns (-VC₀VC₀ #) with High Frequencies

Text	RV	SS	SB	GV	PC	PD	DV
Most frequent RP	asya	ai	ai	ai	ai	ai	ana
Frequency rate (%)	2.62	5.01	5.48	5.47	6.28	9.52	4.66
2nd frequent RP	āya	eṇa	iaṃ	anti	au	au	ata
Frequency rate (%)	1.25	4.13	3.02	4.11	6.21	8.35	3.74
3rd frequent RP	ate	aha	aaṃ	ēṇa	ahu(ahō)	ahaṃ	āma
Frequency rate (%)	1.22	3.23	2.89	3.73	5.02	3.83	3.43
4th frequent RP	ati	assa	ēṇa	āṇa	iu	aṇu	asī
Frequency rate (%)	0.97	3.10	2.80	3.55	5.01	3.26	3.29
5th frequent RP	ayaḥ	āi	assa	āō	aṇu	ahim	ala
Frequency rate (%)	0.91	3.03	2.31	3.11	3.30	3.01	2.91
6th frequent RP	ataḥ	ammi	ammi	ammi	aī	aim	ara
Frequency rate (%)	0.87	2.02	2.11	2.79	2.90	2.57	2.91
7th frequent RP	avaḥ	ia	aha	ia	ia	iu	ati
Frequency rate (%)	0.87	2.01	2.03	2.45	2.53	2.48	1.97
8th frequent RP	āsaḥ	iaṃ	anti	assa	eṇa	iya	ita
Frequency rate (%)	0.85	1.98	1.88	2.42	2.51	2.38	1.70
9th frequent RP	asaḥ	iō	iō	iha	ihī(ēhī)	ahi	ahim
Frequency rate (%)	0.83	1.55	1.71	1.90	2.24	2.36	1.69
10th frequent RP	ānaḥ	āṇa	aṇaṃ	ēsu	ivi(ēvi)	avi	ama
Frequency rate (%)	0.82	1.54	1.70	1.81	2.08	1.87	1.52

Abbreviations

Abbreviations	Corresponding Names
DV	<i>Dohāvalī</i>
GV	<i>Gāḍḍavaho</i>
MIA	Middle Indo-Aryan
NIA	New Indo-Aryan
OIA	Old Indo-Aryan
PC	<i>Paūmacariu</i>
PD	<i>Pāhuḍadohā</i>
PIE	Proto-Indo-European
PIIr	Proto-Indo-Iranian
RhMP	Rhyme Match Probability
RV	<i>Ṛgveda Saṃhitā</i>
SB	<i>Setubandha</i> <i>Gāhāsattasāī</i>

Texts and Textual Sources

- Niedersächsische Staats- und Universitätsbibliothek (2010). GRETEL - Göttingen Register of Electronic Texts in Indian Languages.
Available at <https://gretel.sub.uni-goettingen.de/gretel.html>
- Bhayani, H.C. (1953-1960). *Paūmacariu of Kaviraja Svayambhudeva: A Pre-Tenth Century Puranic Epic in Apabhraṃśa*. Bombay: Bharatiya Vidya Bhavan.
- Kale, M.R. (1921). *The Ratnāvalī of Śrī Harṣa-Deva*. Delhi: Motilal Banarsidass.
- Kölligan, D.; Neuefeind, C.; Reinöhl, U.; Sahle, P.; Casaretto, A.; Fischer, A.; Borge K.; Korobzow, N.; Rolshoven J.; Halfmann, J.; Mondaca, F. (2017-2025). *Vedaweb: Online Research Platform for Old Indic Texts*.
Available at <https://vedaweb.uni-koeln.de>
- Lubotsky, A. (1997). *A Ṛgvedic Word Concordance*. New Haven (CT): American Oriental Society. American Oriental Series 82-3.
- Muni R.; Śāstrī, D. (eds) (1998). *Pāhuḍadohā (Apabhraṃśa kā rahasyavādī kāvyā)*. Nayī Dillī: Bhāratīya Jñānapīṭha.
- Ollett, A. *Prakrit Digital Text Project*.
Available at https://github.com/aso2101/prakrit_texts
- Vājpeyī, N.D. (ed.) (1949). *Sūr-sāgar, khand 1*. Vārāṇasī: Nāgarī Pracārīṇī Sabhā.

Bibliography

- Arjunwadkar, K.S. (1985). "Rhyme in Sanskrit and in Marathi". *Annals of the Bhandarkar Oriental Research Institute*, 66(1-4), 93-201.
- Arnold, E.V. (1905). *Vedic Metre in Its Historical Development*. Cambridge: Cambridge University Press.
- De Clercq, E. (2005). *Kritische studie van Svayambhûdeva's Paûmacariu* [PhD Dissertation]. Gent: Universiteit Gent.
- De Clercq, E. (2018). *The Life of Padma / Svayambhudeva*, vol. 1. Cambridge: Harvard University Press.
- Fabb, N. (2010). "Is Literary Language a Development of Ordinary Language?". *Lingua*, 120(5), 1219-32.
- Fabb, N.; Sykärî, V. (2022). "Rhyme in the Languages and Cultures of the World". Fabb, N.; Sykärî, V. (eds), *Rhyme and Rhyming in Verbal Art, Language, and Song*. Helsinki: Finnish Literature Society, 11-44.
- Klein, J. (2002). "Homoioteleuton in Rigveda". Southern, M.R.V. (ed.), *Indo-European Perspectives*, monogr. no., *Journal of Indo-European Studies Monograph*, 43. Washington, D.C.: Institute for the Study of Man. Inc., 199-222.
- Kobayashi, M. (2004). *Historical Phonology of Old Indo-Aryan Consonants*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. Study of Languages and Cultures of Asia and Africa Monograph Series 42.
- Kulkarni, V.M. (1965). *Jayadeva's Gîta Govinda with King Mânârka's Commentary*. Ahmedabad: Lalbhai Dalpatbhai Bharatiya Sanskriti Vidyamandira. Lalbhai Dalpatbhai Series 7.
- Kunakh, O.M.; Volkova, A.M.; Tutova, G.F.; Zhukov, O.V. (2023). "Diversity of Diversity Indices: Which Diversity Measure is Better?". *Biosystems Diversity*, 31(2), 131-46.
- Mahesh, M.S. (1964). *The Historical Development of Mediaeval Hindi Prosody: Rāmānanda-Keśava, 1400-1600 AD*. Bhagalpur: Bhagalpur University Publication.
- Nagasaki, H. (2012). "Hindi Metre: Origins and Development". Nagasaki, H.; Ronald I.K. (eds), *Indian and Persian Prosody and Recitation*. Delhi: Saujanya Publications, 107-29.
- Ollett, A. (2017). *Language of the Snakes: Prakrit, Sanskrit, and the Language Order of Premodern India*. Oakland: University of California Press.
- Peust, C. (2014). "Towards a Typology of Poetic Rhyme". Grossman, E.; Haspelmath, M.; Richter, T.S. (eds), *Egyptian-Coptic Linguistics in Typological Perspective*. Berlin: De Gruyter Brill, 341-85.
- Sedgwick, W.B. (1924). "The Origin of Rhyme". *Revue Bénédictine*, 36(4), 330-46.
- Snell, R. (1991). *The Hindi Classical Tradition: A Braj Bhāṣā Reader*. London: School of Oriental and African Studies, University of London. SOAS South Asian Texts 2.
- Tagare, G.V. (1987). *Historical Grammar of Apabhraṃśa*. Delhi: Motilal Banarsidass.
- van Nooten, B.A.; Holland, G.B. (eds) (1994). *Rig Veda: A Metrically Restored Text with an Introduction and Notes*. Cambridge (MA): Harvard University Press.
- Yamahata, T. (2009). "Indo Inritsushi ni okeru apaburanshago no yakuwari". *The Japanese Association of Indian Buddhist Studies*, 57(2), 848-54.

Sonorant Gemination in Old Tamil

Ramprashanth Venkatakrishnan

Madurai Kamaraj University, India

Somdev Kar

Indian Institute of Technology Ropar, India

Abstract Old Tamil exhibits two kinds of sonorant gemination, which were previously explained by two separate, linear phonological rules. We postulate that the rules explaining both these ‘separate’ phenomena are part of the same conspiracy. This study is devoted to exploring the optimality theoretic analysis of the phenomenon of morphologically derived geminates in Old Tamil, as depicted from Sangam poetic texts and the traditional Tamil Grammars (*Tolkāppiyam* and *Naṇṇūl*). We use optimality theoretic markedness or well-formedness constraints like Prosodic Word constraint (ProsWd) and faithfulness constraints like DEP-IO, MAX-IO, NoCoda, etc., as part of the explanatory apparatus. The analysis focuses on a unified ranking of constraints explaining the gemination phenomenon in Old Tamil.

Keywords Sonorant Gemination. Old Tamil phonology. Optimality theory. Prosodic word constraint. Conspiracy. Morphology.

Summary 1 Introduction. – 1.1 Old Tamil. – 1.2 Notes on Transliteration and Examples. – 1.3 Segment inventory of Old Tamil. – 1.3.1 Vowels. – 1.3.2 Consonants. – 2 The Two Rules. – 2.1 Sonorant Consonant Gemination: Type A (SCG-A). – 2.2 Sonorant Consonant Gemination – Type B (SCG-B). – 2.3 Phonotactics of Sonorant Consonants in Old Tamil. – 2.4 Prohibition of Gemination. – 2.5 Sonorant Consonants that Do not Geminate. – 3 Optimality Theoretic Analyses. – 3.1 A Preliminary Analysis. – 3.2 The Prosodic Word Constraint. – 3.3 Why This Particular Notion of Mora? – 3.4 Evidence of Moraic Sensitivity in Tamil. – 3.4.1 Traditional Poetics. – 3.4.2 Sanskrit Loan Words. – 3.4.3 Compensatory Vowel Lengthening. – 3.5 Analysis. – 3.5.1 Verb Roots of the Shape CV. – 3.5.2 Shortest Old Tamil Words. – 3.5.3 Comparative Dravidian Data. – 3.5.4 Sonorant Gemination in Modern Tamil. – 3.5.5 The Role of Onset Constraint. – 4 Relic Forms Without Gemination. – 5 Do the Obstruents Geminate in Similar Morphophonological Contexts? – 6 Conclusion.



Peer review

Submitted 2025-01-19
Accepted 2025-06-03
Published 2025-07-01

Open access

© 2025 Venkatakrishnan, Kar | © 4.0



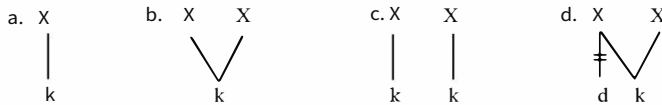
Citation Venkatakrishnan, R.; Kar, S. (2025). “Sonorant Gemination in Old Tamil”. *Bhasha*, 4(1), 95-122.

1 Introduction

Sonorant consonants are a subcategory of consonants that include the nasals, liquids (laterals and trills) and glides. These are distinguished from obstruents (stops, affricates, and fricatives) because of their relatively high sonority value. Geminates take a longer span of time to produce consonant sounds than singleton consonants. In Old Tamil, we can see two types of sonorant geminate formation. We have considered these two phonological phenomena to be part of the same conspiracy, i.e. one that prefers the optimal output of a sonorant geminate. This alternative explanation is parallel, i.e., in the framework of Optimality Theory (Kager 1999; McCarthy, Prince 1993; 1994; 1995), as opposed to linear phonological rules. The Sonorant consonants in Old Tamil are the Nasals (bilabial, dental, alveolar, palatal, retroflex and velar), Liquids (alveolar lateral, retroflex lateral, alveolar tap, alveolar trill and retroflex approximant) and Glides (labiodental and palatal). The Sonorant consonants that do not undergo this gemination are the velar nasal, alveolar tap/flap, alveolar trill and retroflex approximant.

Old Tamil (as well as Modern Tamil) is one of the unique languages that have morphologically derived geminates. A language like Tashlhiyt Berber has four kinds of geminates (which don't cover the type in Old Tamil, which we will discuss in this paper). It has (1a) singleton stops, (1b) lexical geminates, (1c) geminates due to morpheme concatenation, and (1d) geminates due to assimilation (Ridouane 2010).

(1)



1.1 Old Tamil

Tamil is a South Dravidian language (Krishnamurti 2003, 22) spoken in the state of Tamil Nadu in Southern India. It is spoken in southern India and north-eastern Sri Lanka since prehistoric times (Annamalai, Steever 1998, 100). It is spoken by 53 million speakers in India, according to data from the 2011 census (Ministry of Finance 2012). It has the status of an Official and National language in Sri Lanka. It has the status of an Official language only in India (Choudhry et al. 2016) and Singapore (Tan 2005), while it is recognised as a minority language in Mauritius (Meetarbhan 2017), South Africa (Kamusella, Ndhlovu 2018) and Malaysia (Schiffman 1999, 1). Tamil's closest relatives are Malayalam (spoken in the neighbouring state of

Kerala) and Irula (a tribal language spoken in the Nilgiris district of Tamil Nadu) (Annamalai, Steever 1998, 101). Tamil is classified as one of the South Dravidian languages (Krishnamurti 2003). Tamil is diglossic, the formal or literary variety still essentially conforming to standards set in the thirteenth century by the Tamil grammarian Pavananti, who composed a treatise called *Naṇṇūl* (“the good book”) on Tamil grammar. This formal variety (High Tamil) is used in almost all written media and for certain high-register functions. In all other situations, colloquial Tamil (Low Tamil) is used and is characterised by considerable regional and social variation (Schiffman 2017). Old Tamil is said to have been spoken from 300 BCE to 700 CE, followed by Middle Tamil (700 CE to 1600 CE) and Modern Tamil (1600 CE to the present) (Lehmann 1998, 75). Such a periodisation, though convenient, is ‘necessarily distortive’ as many classical texts cover several strata of time and are not easily pinned down to one point in time (Wilden 2018, 4). High Tamil is considerably based on Old Tamil grammatical features and vocabulary.

All of the data on Old Tamil for this paper are obtained from the Sangam poetic texts and the traditional Tamil Grammars (*Tolkāppiyam* and *Naṇṇūl*). The Sangam corpus is the oldest extant literature on Old Tamil. It consists of text corpora like the *Eṭṭutokai* (‘Eight Anthologies’) and *Pattuppāṭṭu* (‘Ten Idylls’), which are collections of Old Tamil poems of *Akam* (love poems) and *Puram* (heroic poems) genres.

(2) Sangam corpora: *Eṭṭutokai* (‘Eight Anthologies’) and *Pattuppāṭṭu* (‘Ten Idylls’)

a. Eight texts of *Eṭṭutokai*

- *Kuṟuntokai*(*kuṟu*);
- *Aiṅkuṟunūru*(*aiṅ*);
- *Narriṇai*(*nar*);
- *Kalittokai*(*kali*);
- *Patirruppattu*(*pati*);
- *Paripāṭal*(*pari*);
- *Akanāṇūru*(*aka*);
- *Puraṇāṇūru*(*pura*).

b. Ten texts of *Pattuppāṭṭu*

- *Tirumurukāṟruppaṭai*(*tiru*);
- *Poruṇarāṟruppaṭai*(*poru*);
- *Cirupāṇāṟruppaṭai*(*ciru*);
- *Perumpāṇāṟruppaṭai*(*perum*);
- *Mullaippāṭṭu*(*mull*);
- *Maturaikkāñci*(*matu*);
- *Neṭunalvāṭai*(*neṭu*);
- *Kuṟiñcippāṭṭu*(*kuṟi*);
- *Paṭṭiṇappālai*(*paṭṭ*);
- *Malaipaṭukaṭām*(*malai*).

The word Sangam refers to the Sangam period or age, which is the historical period of ancient Tamil Nadu, Kerala and parts of Sri Lanka (then known as Tamiḻakam) spanning from 300 BCE to 700 CE (Parameswara Iyer 1953). It was named after the famous ‘Sangam’ (meaning congregation) academies of poets and scholars. Tolkāppiyam is the earliest extant grammatical treatise on Old Tamil that discusses grammar and poetics. A particular chapter called *Eḷuttatikāram* (*Eḷuttu* means ‘letter’ and can be interpreted as meaning a phoneme) in Tolkāppiyam deals with the phonology and phonotactics of Old Tamil (Rangan 2012, 5-6). We are informed about much of Old Tamil phonology through these traditional grammars. For example, one of the phonological rules that we deal with in § 2 which is part of a conspiracy, is taken from a *Naṇṇūl* aphorism.

There have been several treatments of the Morphophonology of Tamil in the past. These include descriptive studies by Balasubramaniam (1989) and Subramoniam (2003), a study on generative phonology by Vasanthakumari (2000), a work on lexical phonology by Christdas (2013), and constraint-based studies by Beckman (2009) and Ramasamy (2011). There has been a brief sketch of Old Tamil Phonology in Lehmann (1998). But there has not been any explanatory work on Old Tamil phonology. All the literature on Old Tamil phonology is a descriptive treatise of traditional grammar. This study may be considered a pioneering attempt to address this gap in the literature.

1.2 Notes on Transliteration and Examples

All of the names and terms from Old Tamil texts are transliterated by using the convention called International Alphabet of Sanskrit Transliteration (IAST), which is used to transliterate the names and texts written in Indic scripts. However, the data from Old Tamil is given in IPA. The phonetic realisations of Old Tamil words and allophony are gauged by the aphorisms on articulation and phonology in the traditional Tamil grammars. The glossing convention (including the abbreviations) used in this paper is as per The Leipzig Glossing Rules.

Most of the examples are cited, i.e. their source texts are mentioned. However, there are a few examples that are hypothetical. [kell-il] is given as an example in (8). While the word /ka/ ‘stone’ and the locative suffix /-il/ find mention in the Sangam texts, their combination is not. We have utilised such hypothetical but very possible Old Tamil examples in our paper.

1.3 Segment inventory of Old Tamil

Old Tamil has ten vowels, five short and five long: /a/, /ā/, /i/, /ī/, /u/, /ū/, /e/, /ē/, /o/, /ō/. The seventeen consonants include six stops: /k/, /c/, /t/, /ṭ/, /p/, /b/; five nasals: /ñ/, /ṇ/, /ṡ/, /n/, /m/; two laterals: /l/, /ḷ/; two glides /y/, /v/; one tap /r/; and one approximant /ɻ/ (Lehmann 1998, 76).

1.3.1 Vowels

Surface realization of Old Tamil vowels with their approximate positioning as per the cardinal vowel chart.

Chart 1 Cardinal vowel chart (reconstructed by the authors)

ɪ/i:		ʊ/u: ʉ
ɛ/e:		o/o:
	ə/a:	

Here, the short vowels phonetically have a slightly lower pronunciation when compared to their cardinal, longer counterparts. For example, /i/ would be realized as [ɪ] while its longer version would be [i:]. Of the short vowels mentioned above, only /u/ has an allophone – an unrounded counterpart [ʉ].

1.3.2 Consonants

Old Tamil has a six-way contrast in articulation. This is seen in the surface realization of six stops and nasals. Besides these, Old Tamil possesses laterals, rhotics and glides. As for stop segments, there is no contrast of voice feature.

Table 1 Consonant inventory

	Labial	Dental	Alveolar	Retroflex	Palatal	Velar
Stops	p	t̪	t	ʈ	c	k
Nasals	m	n̪	n	ɳ	ɲ	(ŋ)
Taps			r			
Laterals			l	ḷ		
Glides	ʋ				j	
Approximant				ɻ		

Note: (Lehmann 1998, 77)
() indicates allophone

The velar nasal is an allophone of other nasals. It occurs only before the velar stop. There is a rare fricative [h] called *āytam*, which can be considered as an allophone of the glide /j/. All the consonants, except the stop class pattern as sonorants in Old Tamil.

2 The Two Rules

2.1 Sonorant Consonant Gemination: Type A (SCG-A)

Let us consider this phonological rule in Old Tamil: Given a monosyllabic root (noun or verb) of the shape (C)VC where the Consonant in round brackets is optional and the vowel is a short one, when the coda sonorant consonant forms a juncture with another concatenating morpheme that begins in a vowel, the consonant geminates. The important condition is that this geminating consonant has to be a sonorant (i.e. a nasal, a liquid or a glide). Even though in Old Tamil, the flap/tap and retroflex approximant count as sonorants, they do not geminate at all at the surface level. We propose to term this specific pattern of gemination ‘Sonorant Consonant Gemination: Type A’(henceforth, SCG-A). The following is the formal representation of the above-stated SCG-A:

- (3) [+con, +son] → [+con, +son][+con, +son]/#C₀[-con][+con, +son]_____#[-con]

The above phonological rule is expressed in the form of an aphorism in the traditional grammar *Tolkāppiyam* (aphorism no. 161, line no. 2) (*Tolkāppiyam* 2021, 86):

- (4) குறியதன் முன்னர்த் தன் உரு இரட்டலும்
kuṛiyataṇ munṇart taṇ uru iraṭṭalum

This can be translated as: “the letter (coda consonant) geminates before a short syllable (of the form (C)VC”. It is also mentioned in the other traditional grammatical treatise *Naṇṇūl* (aphorism no. 205) (Vijayavenugopal 1968, 210):

- (5) தனிக் குறில் முன்னற்று உயிர்வரின் இரட்டும்
taṇikkuril munṇorru uyirvarin iraṭṭum

This can be translated as: “The letter (the coda consonant) in a short syllable (of type (C)VC) doubles before an oncoming vowel (of the vowel-initial concatenating morpheme)”.

This gemination occurs across a wide range of morphophonological contexts, namely noun derivation, verbal inflection, case suffixation, compound noun formation, etc. Let us see some examples of this phenomenon:

(6)	Noun derivation	/min/ root noun	+ /al/ noun-forming suffix	> [minnel] 'lightening'
(7)	Verb inflection	/cej/ 'do'	+ /um/ 3.SG.N.FUT	> [cejjum] '(it) will do'
(8)	Case suffixation	/kal/ 'stone'	+ /il/ LOC	> [kellɪl] 'in the stone'
(9)	Compound Noun	/muɻ/ 'thorn'	+ /ilaj/ 'leaf'	> [muɻ[ɪləj] 'thorny leaf'
(10)	Noun + Finite Verb	/pal/ 'tooth'	+ /uɻaj-ŋɻ-at/ 'break'-PST-3.SG.N	> [pelluɻeɻŋɻeɻu] '(its) tooth broke'

The following are examples of sonorant consonants geminating on concatenation with a vowel-initial morpheme. The first morpheme in the examples is mostly nouns and pronouns (except in the case of gemination of labiodental glide, where we have given only examples of a deictic root plus a vowel-initial noun), which are being suffixed by case markers. The following examples are grouped by the kind of sonorant consonant that is geminating (Podhuvan 2025):

(11) Bilabial nasal

a.	/em/ 1PL.EXCL.O	+ /ot/ SOC	> [ɛmmoɻu] 'with us'	(aka 11.7)
b.	/ɻam/ 3PL.R.O	+ /aj/ ACC	> [ɻemmej] 'them'	(kali 85.30)
c.	/ŋam/ 1PL.INCL.O	+ /in/ GEN	[ŋemmin] 'our'	(narr 245.9)
d.	/ŋum/ 2PL.O	+ /ot/ SOC	> [ŋuɻmmoɻu] 'with you all'	(aka 168.1)

(12) Alveolar nasal

a.	/ɻan/ 3SG.R.O	+ /il/ LOC	> [ɻennɪl] 'in himself/herself'	(tiru 33.1)
b.	/ŋin/ 2SG.O	+ /aj/ ACC	> [ŋinnej] 'you(acc.)'	(aiñ 420.3)
c.	/en/ 1SG.O	/aj/ ACC	> [ennej] 'me'	(aka 212.12)

- | | | | | |
|----|--------|--------|-----------------|------------|
| d. | /min/ | + /ot/ | > [minnoɖw] | (aka 43.2) |
| | ‘star’ | SOC | ‘with the star’ | |

(13) Retroflex nasal

- | | | | | |
|----|---------|--------|----------------|---------------|
| a. | /maŋ/ | + /aj/ | > [mɐŋŋɐj] | (aka 24.12) |
| | ‘earth’ | ACC | ‘earth(acc.)’ | |
| b. | /eŋ/ | + /il/ | > [ɛŋŋɪl] | (puṛa 213.15) |
| | ‘fate’ | LOC | ‘in fate’ | |
| c. | /kaŋ/ | + /ot/ | > [keŋŋoɖw] | (aka 164.9) |
| | ‘eye’ | SOC | ‘with the eye’ | |
| d. | /paŋ/ | + /in/ | > [peŋŋɪn] | (aka 352.15) |
| | ‘song’ | GEN | ‘of the song’ | |

(14) Alveolar lateral

- | | | | | |
|----|---------|--------|----------------|--------------|
| a. | /kal/ | + /in/ | > [keɭɪn] | (paṭṭ 11.2) |
| | ‘stone’ | GEN | ‘of the stone’ | |
| b. | /pal/ | + /in/ | > [peɭɪn] | (kali 21.10) |
| | ‘tooth’ | GEN | ‘of the tooth’ | |
| c. | /pul/ | + /in/ | > [peɭɪn] | (kali 94.19) |
| | ‘grass’ | GEN | ‘of the grass’ | |
| d. | /ŋel/ | + /in/ | > [ŋeɭɪn] | (aka 46.14) |
| | ‘paddy’ | GEN | ‘of the paddy’ | |

(15) Retroflex lateral

- | | | | | |
|----|----------|----------|-----------------|---------------|
| a. | /kaɭ/ | + /in/ | > [keɭɪn] | (aka 256.4) |
| | ‘toddy’ | GEN | ‘of the toddy’ | |
| b. | /mul/ | + /uɕaj/ | > [muɭɪɕɐj] | (naṛṛ 203.2) |
| | ‘thorn’ | ‘having’ | ‘thorn-having’ | |
| c. | /uɭ/ | + /il/ | > [uɭɪɭ] | (puṛa 160.21) |
| | ‘inside’ | LOC | ‘in the inside’ | |

(16) Palatal glide

- | | | | | |
|----|--------|--------|----------------|--------------|
| a. | /kaj/ | + /il/ | > [kejɟɪl] | (kali 33.18) |
| | ‘hand’ | LOC | ‘in the hand’ | |
| b. | /ŋej/ | + /ot/ | > [ŋeɟjoɖw] | (aiṇ 211.1) |
| | ‘oil’ | SOC | ‘with the oil’ | |
| c. | /maj/ | + /il/ | > [mejɟɪl] | (aiṇ 235.1) |
| | ‘ink’ | LOC | ‘in the ink’ | |

- | | | | | |
|----|--------|--------|--------------|-------------|
| d. | /mej/ | + /aj/ | > [mɛjjej] | (aka 310.3) |
| | 'body' | ACC | 'body(acc.)' | |

(17) Labiodental glide

- | | | | |
|----|--------------------------------|----------|--------------|
| a. | /au/ | + /iṭam/ | > [ɐuɪdɐm] |
| | distal demonstrative prefix | 'place' | 'that place' |
| b. | /iu/ | + /iṭam/ | > [ɪuɪdɐm] |
| | proximate demonstrative prefix | 'place' | 'this place' |

The sonorants that do not form this kind of geminate are also ones that do not occur as a coda in the (C)VC roots. Those are the dental, palatal and velar nasals. It may be noted that in the first set of forms with sonorant consonant geminates is that it is the sonorant coda in the first morpheme that geminates.

2.2 Sonorant Consonant Gemination – Type B (SCG-B)

Apart from SCG-A, another pattern of gemination involving sonorant consonants is observed in Old Tamil. This pattern, where the sonorant that geminates is not part of a (C)VC morpheme, but the juncture is one of a monosyllabic short vowel and a sonorant. We propose to term this “Sonorant Consonant Gemination: Type B” (hereafter, SCG-B). Some of the examples in (18-21) are from (Vijayakrishnan 1982, 54) and represent the pattern of SCG-B:

- | | | | | |
|------|-------------------|--------|-----------|-------------|
| (18) | Labiodental glide | /i/ | + /ʋalaj/ | > [ɪʋʋelej] |
| | | 'this' | 'net' | 'this net' |

- | | | | | |
|------|--------------|--------|------------|--------------|
| (19) | Dental nasal | /i/ | + /ṇe:ram/ | > [ɪṇṇe:rem] |
| | | 'this' | 'time' | 'this time' |

- | | | | | |
|------|----------------|---------|----------|---------------|
| (20) | Bilabial nasal | /mu/ | + /maṭi/ | > [mummedɪ] |
| | | 'three' | 'folds' | 'three folds' |

- | | | | | |
|------|---------------|--------|------------|--------------|
| (21) | Palatal nasal | /a/ | + /ṇa:lam/ | > [ɐṇṇa:lem] |
| | | 'that' | 'earth' | 'that earth' |

The crucial thing that needs to be noted here is that it is the initial sonorant consonant of the second morpheme that geminates. The velar nasal, alveolar nasal and retroflex nasal that do not form the word-initial onset of the second morpheme do not geminate this way.

(22) Illustrates the pattern of gemination discussed above.

(22) [+con, +son] → [+con, +son][+con, +son]/#([+con])([-con])_____#[+con, +son]

2.3 Phonotactics of Sonorant Consonants in Old Tamil

Crucial to our point about sonorant consonant gemination is a discussion on what sonorant consonants can and cannot occur morpheme-initially or as a coda of a monosyllabic morpheme, hence having implications as to whether they geminate as elaborated above. If a sonorant consonant can occur morpheme-initially, it can geminate as per Rule 2 (§ 2.2). If a sonorant consonant can occur as a coda consonant of a monosyllabic morpheme with a short vowel, it can geminate as per Rule 1 (§ 2.1). Below is a summary of this phonotactic information:

Table 2 Conditions governing the occurrence of sonorant consonants

Sonorant Consonants	Can occur as a coda	Can occur Morpheme-initially
Bilabial nasal	YES	YES
Alveolar nasal	YES	NO
Retroflex nasal	YES	NO
Alveolar lateral	YES	NO
Retroflex lateral	YES	NO
Palatal glide	YES	YES
Labiodental glide	YES	YES
Palatal nasal	NO	YES
Dental nasal	NO	YES
Velar nasal	NO	NO
Alveolar tap/flap	YES	NO
Retroflex approximant	YES	NO
Alveolar trill	NO	NO

The following table contains information about what sonorant consonants can occur as coda and can geminate as per Rule 1 (column 1) and what sonorant consonants can occur as coda but cannot geminate as per Rule 1 (column 2).

Table 3 Conditions governing gemination

Sonorant Consonants that can geminate as per Rule 1	Sonorant Consonants that cannot geminate as per Rule 1
Bilabial nasal	Alveolar tap/flap
Alveolar nasal	Retroflex approximant
Retroflex nasal	
Alveolar lateral	
Retroflex lateral	
Palatal glide	
Labiodental glide	

The following table contains information about what sonorant consonants can occur morpheme-initially and can geminate as per Rule 2 (column 1) and what sonorant consonants can occur morpheme-initially but cannot geminate as per Rule 2 (column 2):

Table 4 Conditions governing gemination (morpheme-initially)

Sonorant Consonants that can geminate as per Rule 2	Sonorant Consonants that cannot geminate as per Rule 2
Bilabial nasal	–nil–
Palatal glide	
Labiodental glide	
Palatal nasal	
Dental nasal	

2.4 Prohibition of Gemination

There is a blocking of gemination of the sonorant consonant when the vowel in the first morpheme is a long one instead of a short one. So there is a blocking of gemination in the following examples (Vijayakrishnan 1982, 55):

(26) /t̪i:/ + /ṇa:ttam/ > [t̪i:ṇa:ttam] * [t̪i:ṇṇa:ttam]
bad ‘smell’ ‘bad smell’

(27) /ma:/ + /ṇilam/ > [ma:ṇilem] * [ma:ṇṇilem]
big ‘land’ ‘earth’

2.5 Sonorant Consonants that Do not Geminate

There are a few sonorant consonants that do not geminate in the above elaborated morphophonological context and otherwise. These are the velar nasal ṁ /ŋ/, alveolar tap ṛ /ɾ/, alveolar trill ṛṛ /r/ and retroflex approximant ṣṣ /ɻ/.

The velar nasal neither occurs as the coda of a monosyllabic root (with a short vowel) nor does it occur morpheme-initially. Hence, it doesn't occur at all in the above-mentioned morphophonological context to geminate. You do not find lexical geminates involving a velar nasal. "The geminated velar nasal appears only in words *iñṇaṇam*, *añṇaṇam* and must represent a dialectal form" (Vacek 2019, 99). There are surface constraints in Old Tamil against the occurrence of geminates of certain sonorant consonants. This includes the alveolar tap and retroflex approximant. The geminate of an alveolar tap or flap is articulatorily not feasible. Hence, you do not find lexical geminates of alveolar tap or retroflex approximant. As for the alveolar trill, it is an allophone of the voiceless alveolar plosive. By complementary distribution, the alveolar trill occurs only intervocalically, while the voiced alveolar plosive occurs post-nasally, and the voiceless alveolar plosive occurs as a geminate alone.

(28)	Root + Verbal noun suffix	/ʋar/	+ /al/	> [ʋerel]	*[ʋerrel]
		'come'		'the coming'	

(29)	Root + Noun formative	/ʋa:ɭ/	+ /a:/	> [ʋa:ɭa:]	*[ʋa:ɭa:]
		'live'		'not live'	

3 Optimality Theoretic Analyses

3.1 A Preliminary Analysis

Let us go into OT analyses of some of the examples cited previously. First, we consider a case that illustrates the inadequacy of the existing constraints like Onset, NoCoda, and *Gem in choosing the optimal candidate with sonorant consonant gemination. Previous literature shows that the constraint Onset ranks high above a constraint like NoCoda (Ramasamy 2020). Hence, we have chosen the constraint hierarchy, as seen in Tableau 1. The failure of our preliminary analysis spurs us to postulate a new constraint (§ 3.2).

Let us define the constraints to be used in the preliminary analysis.

(30) Onset

Every syllable must have an Onset.

This constraint penalises any syllable with no onset consonant (Kager 1999). A candidate like [kel.ɭ] (see (33)) would incur one violation of the Onset constraint as its second syllable has no onset consonant.

(31) NoCoda

Each syllable should not have any Coda.

This constraint penalises any candidate that has a syllable with a coda consonant (Kager 1999). A candidate like [kel.lɪl] (see (33)) would incur two violations of the NoCoda constraint as both its syllables have coda consonants.

(32) *Gem

No Geminate consonants are allowed in a word.

This constraint penalises any candidate with geminates (Pajak 2009). A candidate like [kel.lɪl] (see (33)) would incur one violation of *Gem constraint as it has a geminate alveolar lateral.

Let us consider a preliminary OT analysis of example (8).

(33) Tableau 1: Onset >> NoCoda >> *Gem

	/kal/ + /il/	Onset	NoCoda	*Gem
a.	☉ [ke.lɪl]		*	
b.	[kel.lɪl]	*!	*	
c.	[kel.lɪl]		**!	*

In Tableau 1 (33), candidate (b) violates the highest-ranking constraint, Onset, in addition to NoCoda. This fatal violation rules out candidate (b). On the other hand, both candidates (a) and (c) violate the high-ranked constraint NoCoda. While candidate (a) incurs one violation of this constraint, candidate (c) violates it twice and, hence, is ruled out. As a result, candidate (a) emerges as the optimal candidate. However, empirically, we know that the optimal candidate should be [kel.lɪl], as in candidate (c). So, the OT analysis fails to identify the correct candidate as the winning one. Between [ke.lɪl] and [kel.lɪl], if there is a higher constraint prohibiting the former, the latter (known empirically as the optimal one) will win. In the following section (3.2), we propose one such constraint that will help in identifying the optimal candidate.

3.2 The Prosodic Word Constraint

In all the instances of sonorant consonant gemination illustrated above, we see that there is a preference in Old Tamil for words to begin with a heavy (bimoraic) syllable. Hence, we propose a Prosodic Word constraint, ProsWd. It is defined as follows:

(34) ProsWd (Prosodic Word)

Words must begin with a heavy (bimoraic) syllable.

The constraint in (34) requires that a prosodic word in Old Tamil begin with a heavy syllable. The heaviness or lightness of a syllable is measured in terms of the number of morae in it. An open syllable of the shape CV with a short vowel would be one mora, whereas a syllable of the shape CVC or CV: (with a long vowel) would consist of two morae. A syllable of one mora is called a light syllable, and that of two syllables is called a heavy syllable. A super-heavy syllable is of the shape CVCC or CV:C. It becomes clear that vowel length and coda consonantal length are what count as significant in measuring morae. This ProsWd constraint explains the resultant forms in derivations such as the following:

- (35) /col/ + /in/ > [col.lɪn] *[co.lɪn]
'word' GEN 'of the word'

- (36) /e/ + /il/ > [e.lɪl] *[e.ɪlɪ]
'sesame seeds' LOC 'in the sesame seeds'

The gemination in the above examples results in a word-initial heavy syllable of two morae. They contain a short vowel followed by a coda consonant, each contributing one mora. The ungrammatical forms *[co.lɪn] and *[e.ɪlɪ], where there is no gemination, initially have only a light syllable word.

It must also be mentioned that this prosodic word constraint requires a word-initial heavy or bimoraic syllable. It doesn't lay any constraint on the second or any other syllable apart from the first. The second syllable (or any other) can either be monomoraic (as in [col.li] 'having said'), bimoraic (as in [col.lɪn] 'of the word') or trimoraic (as in [col.la:lɪ] 'with the word'). A similar constraint has been proposed to account for consonant gemination in Japanese loan words, albeit in the form of a prohibition of trimoraic syllables if not a preference for bimoraic syllables word-initially (Ito et al. 2017).

3.3 Why This Particular Notion of Mora?

Here in this section, we argue for why this particular notion of mora is important in analysing Tamil phonology. What we need to prove here is why the length of the vowel nucleus and the number of segments in the coda contribute to the syllabic weight, in Old Tamil. In all the

examples considered above, we can easily see that the non-obligatory onset of the monosyllable has nothing to do with sonorant gemination. Consider the examples [pel.lil] ‘in the tooth’ (< /pal/ + /il/) and [pa:.lil] ‘in the milk’ (< /pa:l/ + /il/). These should be the optimal candidates in any correct OT analysis since they are the empirically attested ones, right down to their syllabifications. There is gemination in the first case, while it is blocked in the latter. Both are identical forms except for their first syllable. The first syllable of the first form has a short vowel and a coda lateral. The first syllable of the latter has a long vowel. In what ways can we structurally unify, call as identical a VC and a V:? That is by way of introducing the notion of syllabic weight, where a short vowel and a single coda consonant contribute one mora each, while a long vowel would contribute two morae. So the first syllables of the first and second example would be each two morae. The case is solved. Now, both the examples considered are identical mora-wise. We can go further and claim that there is a requirement in Old Tamil for prosodic words to begin with a bimoraic syllable. That is the trigger for gemination in the first example.

3.4 Evidence of Moraic Sensitivity in Tamil

In this section, we argue in favour of Tamil (Old and Modern) as a moraic-sensitive language. Having proven that Tamil is moraic sensitive, one can accept the validity of the ProsWd constraint, i.e. the notion of mora is important in Tamil, and such a constraint on the initial moraic structure of prosodic words is acceptable. We also look at a few arguments that directly substantiate the postulation of the ProsWd constraint in §§ 3.5.1-3.5.3.

3.4.1 Traditional Poetics

In the traditional poetic grammar, there is a notion of மாத்றிரை / *māttiray*/ [ma:tt̪irɐj], i.e. syllable weight. According to this concept of measuring syllable weight, both the open syllables [ɐ] அ and [kɐ] க would be of one *māttiray* or one mora. Also, the open syllables [a:] ஆ and [ka:] கா would be of two *māttiray* or two morae. Hence, we can understand that the notion of *māttiray* is the same as mora. We can also see that any onset consonant does not contribute to the syllabic weight. A coda consonant, however, contributes one mora or one *māttiray*.

The basic units of Tamil metre are the எழுத்து *eḷuttu*’s, which are the letters or characters. In the abugida system that is Tamil script, each consonant grapheme, /e/, along with its vowel diacritic, denotes an onset consonant plus the nucleus vowel. For example, கி

[kɪ] represents the consonant /k/ with the diacritic for the vowel /i/. Such consonant-plus-vowel letters are classified into two: குறில் *kuṛil* ‘letters with a short vowel diacritic’ and நெடில் *neṭil* ‘letters with a long vowel diacritic’. Besides these consonant-plus-vowel letters, there are pure vowel letters like அ [e] and ஆ [a:], and pure consonant letters like க [k] and ங [j]. They are called as உயிர் *uyir* and ஒற்று *orru* letters respectively. The *uyir* or pure vowel letters always form the nuclei of syllables, while the ‘*orru*’ letters always form the coda consonant of syllables. Only these four, namely ‘*kuṛil*’, ‘*neṭil*’, ‘*uyir*’ and ‘*orru*’ letters, are the basic building blocks of Tamil metre. We can clearly see that only the vowel length matters when it comes to measuring *māttiray*, ‘*kuṛil*’ of one *māttiray* and ‘*neṭil*’ of two *māttiray*. While the ‘*uyir*’ and ‘*orru*’ contribute one *māttiray* each.

3.4.2 Sanskrit Loan Words

There is the case of Sanskrit loan words in Old Tamil. The Sanskrit words that end with a long vowel like [a:] are nativised as [ej]. This preserves the syllable weight or the number of morae of the loaned word. The long vowel [a:] is of two morae, while its nativised form [ej] is also of two morae where the short vowel contributes one mora and the coda semivowel contributes another mora. The examples in (35) may be inferred from the above discussion:

- (37) a. Sanskrit [leṅka:] a place name > Old Tamil [ɭeŋgej]
b. Sanskrit [ʊpema:] ‘metaphor’ > Old Tamil [ʊvemej]

This phenomenon of mora preservation as a nativisation strategy in Sanskrit loan words has been formulated as an aphorism in *Nannūl* (Ganeshasundaram, Vaidyanathan 1958). Some more examples that can be cited are *gaṅgā* > *kaṅkai*, *cōdanā* > *cōtanai*, *icchā* > *iccai*, *sīmā* > *cīmai*. In the above examples, we have transcribed the Old Tamil examples with the diphthong *ai*, as that is how it is written in the Old Tamil Sangam corpus. However, phonologically, there is no diphthong in Old Tamil, and the underlying representation of *ai* is /aj/.

3.4.3 Compensatory Vowel Lengthening

There is a historical sound change in Tamil, which is the preservation of syllabic weights or morae, even in native words that undergo a deletion of onset and nucleus (a combination of consonantal and following vowel). When an intervocalic consonant (onset of the second syllable) is deleted along with the short vowel following it, the result is a lengthening of the first vowel. This could be termed

compensatory vowel lengthening. Hence, there is a preservation of the number of morae in the word that underwent a deletion of the onset and nucleus.

(38)	(a)	[e. geṭ. ṭu]	>	(b)	[a:ṭ. ṭu]
Number of morae		1 + 2 + 1 = 4			3 + 1 = 4
		‘of the house’			‘of the house’

In (38a), [e. geṭ. ṭu] is a trisyllabic wordform having a total of four morae. The process of compensatory vowel lengthening follows the deletion of the onset and nucleus. The resulting form (b) [a:ṭ. ṭu] retains the same number of morae and is in existence only in the Brahmin sociolect of Tamil. However, this new form has a super-heavy first syllable with a long vowel as the nucleus, followed by a coda. Let us consider a few more examples of the historical process of compensatory vowel lengthening:

(39)	(a)	[pɛ.jeɾ]	>	(b)	[pe:r]
Number of morae		1 + 2 = 3			3
		‘name’			‘name’


(40)	(a)	[po.ṭu.ḍu]	>	(b)	[po:.ḍu]
Number of morae		1 + 1 + 1 = 3			2 + 1 = 3
		‘time’			‘while’

(41)	(a)	[ɛ.ṭuṇ.ḍi.ru]	>	(b)	[e:ṇ.ḍi.ru]
Number of morae		1 + 2 + 1 + 1 = 5			3 + 1 + 1 = 5
		‘arise, wake up’			‘arise, wake up’

3.5 Analysis

Based on the discussion in § 3.4 on the moraic sensitivity of Old Tamil syllables, a specific ranking of relevant constraints is being adopted, leading to the emergence of the geminated form [kel.lɪl], from the example illustrated in (8).

(42) Tableau 2: /kal/+ /il/ ‘in the stone’ (Onset >> ProsWd >> NoCoda >> *Gem)

/kal/+ /il/	Onset	ProsWd	NoCoda	*Gem
a. [ke.lɪl]		*!	*	
b. [kel.ɪl]	*!		*	
c.  [kel.lɪl]			*	*

The OT analysis in (42) offers a tentative ranking of Onset >> ProsWd >> NoCoda >> *Gem. The ProsWd constraint is ranked next to Onset because we envision candidates like (b) [kel.lɪ] that fulfil ProsWd by robbing the second syllable of its onset consonant. Hence, a highly ranked Onset would penalise such a candidate. The high-ranked constraint ProsWd prefers only a heavy syllable word initially, while the lowest-ranked constraint *Gem prevents any gemination. In this tableau, the first candidate (a) [kɐ.lɪ] has a light (L) syllable as its first syllable and hence violates ProsWd. This candidate is ruled out for the fatal violation of the high-ranking constraint ProsWd. Candidate (b), even though it satisfies the ProsWd constraint by having an initial bimoraic syllable, violates the highest-ranked constraint, Onset, and gets out of the competition. The constraint ProsWd triggers gemination in candidate (c), [kel.lɪ], resulting in an initial heavy syllable that satisfies the constraint. Although candidate (c) violates NoCoda and *Gem, the violations are not fatal as higher-ranked ProsWd and Onset dominate these two lower-ranked constraints. Candidate (c) [kel.lɪ] emerges as the optimal output. Hence, it is not an optimal candidate. Therefore, the third candidate wins even though it violates the lowly ranked *Gem constraint.

The OT analysis is an instance of sonorant consonant gemination as per the condition discussed in § 2.1. Let us consider another example from (20), /mu/ + /maʈi/ ‘three folds’, which satisfies the other type of Sonorant Consonant Gemination, which is (SCG-B), as discussed in § 2.2.

(43) Tableau 3: /mu/ + /maʈi/ ‘three folds’ (Onset >> ProsWd >> NoCoda >> *Gem)

	/mu/ + /maʈi/	Onset	ProsWd	NoCoda	*Gem
a.	[mʊ.me.ɖɪ]		*!		
b.	[mʊm.e.ɖɪ]	*!		*	
c.	ಮುಮ [mʊm.me.ɖɪ]			*	*

The first candidate [mʊ.me.ɖɪ] (candidate (a)) is ruled out because of the lack of a bimoraic initial syllable (as required by constraint ProsWd to be satisfied). On the other hand, candidate (b), [mʊm.e.ɖɪ], even though it doesn’t violate ProsWd, as it contains an initial bimoraic syllable, violates the highest-ranked constraint, Onset. Since candidate (c) is the only constraint left with violation of two low-ranked constraints, NoCoda and *Gem, after the elimination of candidates (a) and (b), it emerges as the optimal candidate in Tableau 3. It may be noted that the optimal candidate (c) is the one with the sonorant consonant gemination [mʊm.me.ɖɪ].

In (43), the first morpheme has a short vowel. We attempt to analyse another example, from (27), where the first morpheme consists of a long vowel, as illustrated in Tableau 4 (44).

(44) Tableau 4: /ma:/ + /ṇilam/ ‘earth’ (Onset >> ProsWd >> NoCoda >> *Gem)

	/ma:/ + /ṇilam/	Onset	ProsWd	NoCoda	*Gem
a.	மா [ma:.ṇi.lɐm]			*	
b.	[ma:ṇ.ṇi.lɐm]		*!	**	*
c.	[ma:.ṇi.lɐm]			**!	*
d.	[ma:ṇ.i.lɐm]	*!	*	**	

In tableau 4, the first candidate (candidate (a)), [ma:.ṇi.lɐm], starts with a heavy syllable (having a long vowel) and hence satisfies the high-ranking constraint, ProsWd. It is evident that it doesn’t violate *Gem, as there is no gemination visible in it. Candidate (b), [ma:ṇ.ṇi.lɐm], has a word-initial super-heavy syllable (a long vowel followed by a coda consonant). Hence, it violates ProsWd, in addition to NoCoda (twice) and *Gem. The third candidate (c) has a morpheme-internal gemination, which leads to the violation of *Gem, in addition to the double violation of NoCoda, which is a fatal violation. The fourth candidate (d), [ma:ṇ.i.lɐm], fatally violates the highest-ranked constraint Onset as the second syllable lacks an onset, and violates the constraint ProsWd as it has a super-heavy initial syllable and also violates NoCoda twice. Hence, it gets eliminated. The winning candidate is the first one (candidate (a), [ma:.ṇi.lɐm], with minimal violations of high-ranked constraints.

3.5.1 Verb Roots of the Shape CV

Does this preference for a word-initial heavy syllable (bimoraic) mean that all Old Tamil words start only with a heavy syllable? No. The preference for a word-initial heavy syllable is exercised only when the morphophonemics allow it to. We need to add a caveat to the definition of ProsWd constraint, i.e. it plays a role only in morphological contexts involving a monosyllable with a short vowel.

Consider the case of verb roots of the shape CV in the following examples:

- (45) /va-ṇṭ-a:n/ (kali 70)
come-PST-3.Sg.M
‘(he) came’

- (46) /ṭa-ṇṭ-a:l/ (kali 84)
give-PST-3.Sg.F
‘(she) gave’

The verbal roots in the above examples are /va/ and /ṭa/ respectively. The verbal bases in Dravidian also function as imperative singular, as clear by the following quote from *The Dravidian Languages* by Bh. Krishnamurti:

Synchronically, a verbal base (root with or without formatives) is said to be identified by its form in the imperative singular, e.g. wā ‘come’, koy ‘cut’ in most languages. (2003, 278)

So the imperative singular forms for the above-mentioned examples ought to be /va/ and /ṭa/ respectively. But what we find is that the imperatives are [va:] ‘come’ and [ṭa:] ‘give’. This is to fulfil the ProsWd constraint, which acts on these monosyllabic roots with a short vowel, as mere /va/ and /ṭa/ are of one mora only. Here we find that the constraint ProsWd is fulfilled by vowel lengthening rather than consonant gemination. Also, the strategy of vowel lengthening to fulfil the ProsWd constraint is employed here as such a lengthening doesn’t change the meaning conveyed by the root.

3.5.2 Shortest Old Tamil Words

“If we compare the shortest Tamil words, we can see that no V or CV represents a free word” (Vacek 1969, 91). Hence, the shortest words in Old Tamil are of the shape VC or V:. On observation, one can see that these are monosyllabic words of two morae precisely. They are the shortest words possible that obey the ProsWd constraint. Words shorter than these, such as shape (C)V (V or CV), are not feasible in Old Tamil. Hence, we can see that the individual short vowels do not make up a free word, while the long vowels like /a:/ ‘cattle’, /i:/ ‘housefly’ do make up the shortest words in Old Tamil. Examples of the shortest words in shape VC would be /aj/ ‘lord, master’, etc. Similarly, a consonant followed by a vowel, namely CV, doesn’t make up the shortest word in Old Tamil.

3.5.3 Comparative Dravidian Data

There is evidence that the kind of sonorant gemination seen in Old Tamil also happened in Old Kannada (which has inscriptional and literary records). This means that this sonorant gemination must have started at the branching node of Tamil-Kannada in the South Dravidian family tree or higher up than that. Let us consider a few examples from Old Kannada (Shastri 2015):

(47) /pal/ + /ili/ > [peɭɭɪɪɪ]
'tooth' 'one without' 'tooth-less person'

(48) /kaŋ/ + /iŋɖa/ > [keŋŋiŋɖe]
'eye' ABL 'from the eye'

Illustrative examples for the operation of the ProsWd constraint are the monosyllabic deictic roots /i/ and /a/, which are proximate and distal, respectively. These short monosyllabic roots surface as [i:] 'this' and [a:] 'that' in Kannada (also Malayalam). This is to fulfil the requirement of two morae imposed by the constraint, ProsWd.

3.5.4 Sonorant Gemination in Modern Tamil

Certain word forms like /kal/ 'stone' have a surface realisation as [kel.lu] in Modern Tamil, whereas it was pronounced as just [kel] in Old Tamil (as seen in the Sangam corpus). This is because of the constraint in Modern Tamil against consonant-ending words. Also, if one observes, there is a gemination of the coda lateral in the word /kal/. This is triggered by the presence of the ProsWd constraint in Modern Tamil too. However, like in Old Tamil, there is blocking of gemination when monosyllabic words with a long vowel take the epenthetic [u] to satisfy the no-consonant-ending constraint in Modern Tamil. Hence, the Old Tamil word /pa:l/ 'milk' manifests as [pa:.lu] and not *[pa:l.lu] in Modern Tamil. There is a gemination of the lateral coda here.

3.5.5 The Role of Onset Constraint

The role of the Onset constraint in Sonorant Gemination is crucial. Since it is highly ranked, it prevents the C2 in a C1V1C2V2 from being parsed as a coda of the first syllable rather than the onset of the second syllable. This, with the ProsWd constraint, triggers the Sonorant Gemination. Hence, you get /kal/ > [kel.lu], *[kel.u] or *[ke.lu].

4 Relic Forms Without Gemination

The *Eḷuttatikāram* chapter of *Tolkāppiyam* has an aphorism (no. 162, lines 1-2) which says that there is an ‘exception’ to the kind of Sonorant consonant gemination (*Tolkāppiyam* 2021, 87) that has been dealt with in this paper (§ 2). According to this aphorism, there is no gemination of a sonorant consonant when the ‘sixth’ or genitive case is affixed to a (C)VC noun root.

- (49) ஆறான் உருப்பினும் நான்கன் உருப்பினும்
கூறிய கூற்றொற்று இரட்டல் இல்லை
ārāṇ uruppiṇum nāṅkaṇ uruppiṇum
kūriya kūṛroṛru irattal illai

The above aphorism can be translated as “the sixth and the fourth case, when added to a short syllable (noun root of form (C)VC) the ‘single letter’ (the coda consonant) doesn’t geminate”.

Some oblique pronouns suffixed with the dative or ‘fourth’ case are /ṇama-kk/ [ṇemekku] ‘to him/her’ and /ṇama-kk/ [ṇemekku] ‘to us’. The traditional grammars wrongly consider the epenthetic [e] between the pronominal root and case suffix as part of the case suffix itself. For the traditional grammarians, the dative case suffix is hence /-akk-/ (hence it is treated as an exception to gemination triggered by a vowel-initial suffix). However, modern analysis says that the Old Tamil dative case marker is just /-kk-/ and it doesn’t fall under the morphophonological context described by Rule 1 (§ 2.1).

The sixth or genitive case takes the form /aṭ/ [eḍu]. It is to be noted here that what is considered as the ‘sixth’ case marker is only /aṭ/ [eḍu] and not other genitive case markers like /in/ [in] or /uḍaj/ [uḍej]. Let us consider the suffixation of this particular case to a few pronominal roots:

- | | | | | |
|------|-------|--------|------------|------------|
| (50) | /ṇam/ | + /aṭ/ | > [ṇemedu] | *[ṇemmedu] |
| | 1.PL | GEN | ‘our’ | |
| (51) | /en/ | + /aṭ/ | > [enedu] | *[ennedu] |
| | 1.SG | GEN | ‘mine’ | |
| (52) | /un/ | + /aṭ/ | > [unedu] | *[unnedu] |
| | 2.SG | GEN | ‘your’s’ | |

Even though the above examples satisfy the morphophonological criteria for gemination, as mentioned in § 2, there is no gemination of the sonorant consonant coda. Hence, this forms a seeming ‘exception’ to the phenomenon under study. Let us consider an OT analysis of one of the examples mentioned above:

(53) Tableau 5: /ṇam/ + /aṭ/ ‘our’ (Onset >> ProsWd >> NoCoda >> *Gem)

	/ṇam/ + /aṭ/	Onset	ProsWd	NoCoda	*Gem
a.	[ṇe.me.ḍu]		*!		
b.	[ṇem.e.ḍu]	*!		*	
c.	ṇem.me.ḍu			*	*

In Tableau 5 (53), candidate (a) violates a high-ranking constraint, ProsWd, and hence, is ruled out. Candidate (b) fatally violated the highest-ranking constraint, Onset, which leads to the elimination of this constraint. The winning candidate (c), [ṇem.me.ḍu], violates only the lower-ranked constraints, NoCoda and *Gem, it is the optimal choice in this tableau, as the other two were already ruled out. However, candidate (c) is not the empirically attested one. The first candidate (a), [ṇe.me.ḍu], is the observed form in Old Tamil, even though it is not the optimal candidate, according to our constraint ranking. Hence, it forms an exception.

The reason for these seeming ‘exceptions’ could be attributed to them being relic forms, i.e. these words are frozen from a time when such a morphophonological gemination of sonorant consonants was not active in the language.

5 Do the Obstruents Geminate in Similar Morphophonological Contexts?

One is confronted with the question: Is it only the Sonorant consonants that geminate in the morphophonological contexts discussed above (§ 2)? What about the Obstruents? Consider the following examples:

(54) /iṭ/ + /il/ > [iṭṭil] * [iṭṭil]
‘this’ LOC ‘in this’

(55) /vit/ + /a:/ > [viṭa:] * [viṭṭa:]
‘leave’ NEG ‘not leave’

(56) /ket/ + /in/ > [keṭin] * [keṭṭin]
‘spoil’ COND ‘if (it) spoils’

Even though the morphophonological context seems to be similar to the cases where sonorant consonants geminate, there is no gemination in the example above because the coda consonant of the first morpheme is a dental plosive – an obstruent and not a sonorant consonant. Obstruents in the same morphophonological contexts do not geminate because obstruent gemination has a grammatical

meaning, i.e. they create the oblique root forms. Oblique root forms are forms to which case suffixes, apart from the nominative (null), are appended. The mere oblique root can function as the genitive.

(57) [a:ɖu] /a:t/ ‘goat’ > a:tɖu /a:tɖ/ ‘of the goat’

However, sonorant gemination has no such grammatical function and hence can be employed to fulfil prosodic preferences.

We introduce another highly ranked markedness constraint *Gem(Obs), which prohibits obstruent geminates in a word. Consider the following analysis of example (55):

(58) Tableau 6: /vɪt/ + /a:/ ‘not leave’ (Onset >> *Gem(Obs), ProsWd >> NoCoda >> *Gem)

	/vɪt/ + /a:/	Onset	*Gem(Obs)	ProsWd	NoCoda	*Gem
a.	❌ [vɪ.ɖa:]			*		
b.	[vɪt.ɖa:]		*		*!	*
c.	[vɪɖ.ɖa:]	*!				

However, obstruents seem to geminate in the context described in § 2.2 (Type B) as seen in example (59). 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 (60).

(59) /mu/ + /kulam/ > [mukkulem]
‘three’ ‘clan’ ‘three clans’

(60) /ɕi:/ + /paɭakkam/ > [ɕi:ppɛɭekkem]
‘bad’ ‘habit’ ‘bad habit’

This obstruent gemination also happens in contexts not involving a monosyllable. Hence, its motivation is to preserve voicelessness and not to fulfil any prosodic constraint. Consider the following examples:

(61) /ŋila:/ + /ki:tt/ > [ŋila:kki:ttu]
‘moon’ ‘rays’ ‘moon rays’

(62) /pala:/ + /kant/ > [pela:kkendu]
‘jackfruit’ ‘sapling’ ‘sapling of jackfruit’

6 Conclusion

This study has described and offered an exposition for a phonological problem in Old Tamil using the explanatory apparatus of Optimality Theory. The problem of sonorant consonant gemination was erstwhile regarded as two separate linear phonological rules that were unified as a single conspiracy using the formalism of OT. Only an Optimality theoretic analysis could have done justice in explaining this phenomenon, as it has the formalism to unify separate rules as a single conspiracy. A prosodic word constraint was postulated for Old Tamil (i.e. a preference for word-initial heavy syllables). Using this and other constraints, a ranking was constructed which explained the occurrence and blocking of sonorant consonant gemination. The OT ranking (Onset >> *Gem(Obs), ProsWd >> NoCoda >> *Gem) emerged as a robust ranking of constraints that works for ‘almost’ every scenario, with a few exceptions. More data needs to be analysed to refine the constraint ranking that can explain the exceptions as well.

Bibliography

- Annamalai, E.; Steever, S.B. (1998). "Modern Tamil". Steever, S.B. (ed.), *The Dravidian Languages*. 2nd ed. London: Routledge, 100-28.
- Balasubramaniam, P. (1989). *Descriptive Study of Standard Spoken Tamil* [PhD Dissertation]. Kuala Lumpur: University of Malaya.
- Beckman, J.N. (2009). "On the Status of Codacond in Phonology". *International Journal of English Studies*, 4(2), 105-34.
- Choudhry, S.; Khosla, M.; Mehta, P.B. (2016). *The Oxford Handbook of the Indian Constitution*. Oxford: Oxford University Press.
- Christdas, P. (2013). *The Phonology and Morphology of Tamil*. New York: Routledge.
- Ganeshasundaram, P.C.; Vaidyanathan, S. (1958). "An Evaluation of Sanskrit Loan-Words in Tamil from the Point of View of Nannūl". *Indian Linguistics*, Turner Jubilee, vol. 1, 63-70.
- Ito, J.; Kubozono, H.; Mester, A. (2017). "A Prosodic Account of Consonant Gemination in Japanese loanwords". Kubozono, H. (ed.), *The Phonetics and Phonology of Geminate Consonants*, vol. 1. New York: Oxford University Press.
Available at <https://doi.org/10.1093/oso/9780198754930.003.0013>
- Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.
- Kamusella, T.; Ndhlovu, F. (eds) (2018). *The Social and Political History of Southern Africa's Languages*. London: Palgrave Macmillan.
- Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge: Cambridge University Press.
- Lehmann, T. (1998). "Old Tamil". Steever, S.B. (ed.), *The Dravidian Languages*. 2nd ed. New York; Oxford: Routledge, 75-99.
- McCarthy, J.; Prince, A. (1993). *Prosodic Morphology: Constraint Interaction and Satisfaction*. Amherst (MA): University of Massachusetts; Amherst and Rutgers Centre for Cognitive Science (RuCCS).
- McCarthy, J.; Prince, A. (1994). "The Emergence of Unmarked: Optimality in Prosodic Morphology". Gonzalez, M. (ed.), *Proceedings of the North East Linguistics Society* 24. Amherst (MA): GLSA, 333-79.
- McCarthy, J.; Prince, A. (1995). "Faithfulness and Reduplicative Identity". Beckman, J.; Urbanczyk, S.; Walsh Dickey, L. (eds), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*. Amherst (MA): GLSA, 249-384.
- Meetarbhan, M.J.N. (2017). *Constitutional Law of Mauritius: Constitution of Mauritius with Commentaries*. Bell Village: Milan JN Meetarbhan.
- Ministry of Finance (Govt of India) (2012). *Economic Survey 2011-12*. New Delhi: Oxford University Press.
- Pajak, B. (2009). "Contextual Constraints on Gemimates: The Case of Polish". *Annual Meeting of the Berkeley Linguistics Society*, 269-80.
- Parameswara Iyer, U.S. (1953). *Kerala Sahithya Charithram*. Part 1, *The History of Kerala's Literature*. Thiruvananthapuram: University of Kerala.
- Pillai, Na.K. (1998). *Tamil Moli Marathi: Tamil – Tamil Dictionary*. Madras: Asian Educational Services.
- Podhuvan, S. (2025). தமிழ்த்துளி *Tamil Drops*.
Available at <https://vaiyan.blogspot.com>
- Ramasamy, M.D. (2011). *Topics in the Morphophonology of Standard Spoken Tamil (sst): An Optimality Theoretic Study*. Newcastle: Newcastle University.
- Ramasamy, M.D. (2020). "Irregular Behaviours of Coda /M/ in Tamil Reduplications". *GEMA Online Journal of Language Studies*, 20(2), 57-82.
Available at <https://doi.org/10.17576/gema-2020-2002-04>

- Rangan, K. (2012). *Toward Formulating Formal Phonological Rules of Tolkāppiyam-Eḷuttatikāram*. Chennai: Central Institute of Classical Tamil.
- Ridouane, R. (2010). “Geminates at the Junction of Phonetics and Phonology”. Fougeron, C.; Kühnert, B.; D’Imperio, M.; Vallée, N. (eds), *Laboratory Phonology 10*. Berlin: De Gruyter Mouton, 61-90.
Available at <https://doi.org/10.1515/9783110224917.1.61>
- Schiffman, H.F. (1999). *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- Schiffman, H.F. (2017). “Diglossia as a Sociolinguistic Situation”. Coulmas, F. (ed.), *The Handbook of Sociolinguistics*. 1st ed. Oxford: Blackwell, 205-16.
Available at <https://doi.org/10.1002/9781405166256.ch12>
- Shastri, T.V.V. (2015). *Shabdamanidarpanam*. Kannada Sahitya Parishattu.
Available at <https://archive.org/details/bmshri.shabdamanidarpan0000tvve/page/n208/mode/1up?q=ಪಲ್ಲವಿ>
- Subramoniam, V.I. (2003). *A Descriptive Analysis of a Dialect of Tamil*. Thiruvananthapuram: Dravidian Linguistics Association.
- Tan, K. (2005). *An Introduction to Singapore’s Constitution*. Singapore: Talisman.
- Tolkāppiyam: Text (with Non-Metrical Segmentation), Transliteration and Translations in English Verse and Prose* (2021). 1st ed. Chennai: Central Institute of Classical Tamil.
- Vacek, J. (1969). “Some Problems of Tamil Phonological Structure”. *Orientalia Pragensia*, Charles University, 6, 87-103.
- Vasanthakumari, T. (2000). *Progress of Phonology*. Thiruvananthapuram: Dravidian Linguistics Association.
- Vijayakrishnan, K. (1982). *The Tamil Syllable* [PhD Dissertation]. Hyderabad: The English and Foreign Languages University.
<http://hdl.handle.net/10603/130417>
- Vijayavenugopal, G. (1968). *A Modern Evaluation of Nannul, Eḷuttatikāram*. Chidambaram: Annamalai University.
- Wilden, E. (2018). *Grammar of Old Tamil for Students*. 1st ed. Pondichery: École Française D’Extrême-Orient, Institut Français de Pondichéry.

Semantics of Sanskrit *Eva*: A Cognitive Linguistics Approach

Yuliang Zhou

Shanghai Conservatory of Music

Abstract This paper delves into the semantic analysis of the Sanskrit emphatic particle *eva* using the framework of Cognitive Grammar. Traditionally, grammarians have provided a definition of *eva* that is concise and lacks clarity, while logicians offered a ‘binary’ explanation based on two-valued logic. However, a more nuanced linguistic analysis is essential for a deeper understanding. Through an examination of various examples from Sanskrit literature, this paper posits that all instances of *eva* can be distilled into a single schematic meaning: the speaker’s cognitive process of selecting one candidate from a pool of options, while remaining neutral about other alternatives. Additionally, this paper explores the conjunction of *eva* with the discourse particle *tu*, asserting that *eva tu* and *tu eva* convey distinct nuances. In the former, the discourse connection is on the component marked by *tu*, whereas in the latter, it is on the group of candidates indicated by *eva*.

Keywords *Eva*. Emphatic expression. Focus particle. Cognitive linguistics. Sanskrit grammar.

Summary 1 Introduction. – 2 Emphatic Expression. – 3 Semantics of *Eva*. – 4 Conclusion.



Peer review

Submitted 2024-06-09
Accepted 2025-05-07
Published 2025-07-01

Open access

© 2025 Zhou | © 4.0



Citation Zhou, Y. (2025). “Semantics of Sanskrit *Eva*: A Cognitive Linguistics Approach”. *Bhasha*, 4(1), 123-154.

1 Introduction

Eva is one of the most frequently used particles in Sanskrit. Its usage can be roughly described as laying emphasis on the preceding word and thus is categorized as what I call an ‘emphatic particle’ later. Ancient scholars in India have been annotating the usage of *eva* for a long time. Opinions on it can be generally divided into two categories: those of grammarians and those of logicians.

The first definition of *eva* comes from Pāṇini. *Aṣṭādhyāyī* 8.1.62 goes, *cāhalopa eve ’ty avadhāraṇam*, which, according to S.C. Vasu (1898, 1523), means that when particles *ca* and *aha* are omitted, the first finite verb retains its accent if there is an *eva* with the meaning of restriction (*eve ’ty avadhāraṇam*) taking their place.¹ A similar definition can also be found in *Mahābhāṣya* on sūtra 5,3,58. Here, Patañjali comments, “What is the meaning of the word *eva*? It means limitation (*niyama*)”.²

This definition can be regarded as the standard among traditional grammarians and is inherited in later mainstream literature in Sanskrit grammar. Nevertheless, we are not sure about what traditional grammarians mean precisely by ‘restriction’ (*avadhāraṇa*) or ‘limitation’ (*niyama*). A piece of evidence suggests that they may not be taken too literally, i.e., they should not be understood as equivalent to English words like *only* or *alone*. In verse 3,4,15 of *Nāmaliṅgānuśāsana*, Amarasiṃha defines five particles as ‘word of restriction’ (*avadhāraṇavācaka*), including *evam*, *tu*, *punar*, *vā*, and *eva*.³ Obviously, all the words other than *eva* will never be taken as meaning ‘only’. Also proving this point is Maheśvara’s commentary on this verse, by which these five words are defined as indicating ‘ascertainment’ (*niścaya*).⁴

In contrast to the complexity and variety of its practical uses, traditional grammarians have explained *eva* in a very concise manner, which appears to be oversimplified to a contemporary eye. It is until the rise of the great Buddhist philosopher Dharmakīrti that we finally have our first detailed description of *eva*. In *Pramāṇavārttika*,

Unless otherwise stated, all translations of the original Sanskrit texts included in this paper are by the author.

¹ According to *Vaiyākaraṇasiddhāntakaumudī* on this sūtra, *eva* has another usage expressing ‘impossibility’ (*anavakīpti*). The example sentence given here is *deva kve ’va bhokṣyase*, which is interpreted as “Oh Devadatta, you are not going to eat anywhere” (*na kvacid ity arthaḥ*). Nevertheless, to my knowledge, this meaning of *eva* is not attested anywhere in practical uses and thus might be highly contextually limited.

² *evakāraḥ kimarthaḥ. niyamārthaḥ* | (*Mahābhāṣya* on *Aṣṭādhyāyī* 5,3,58).

³ *syur evaṃ tu punar vai ’ve ’ty avadhāraṇavācakāḥ* | (*Nāmaliṅgānuśāsana* 3,4,15).

⁴ *evaṃ tu punar vai ’va pañca niścayārthakāḥ syuḥ* | (*Māheśvarī Ṭikā* on *Nāmaliṅgānuśāsana* 3,4,15).

Dharmakīrti states the following:

It is the non-connection, the connection with others, or the absolute non-connection with a property that the limitative particle [i.e., *eva*] excludes when uttered together with a non-verbal predicate, a modified nominal, or a verb. Even when not said [explicitly], every one of these meanings can be recognized [merely] by [speaker's] intention to utter [*eva*] because the utterance [itself] has the effect of exclusion. For examples, *Caitraḥ dhanurdharaḥ [eva]*; *Pārthaḥ [eva] dhanurdharaḥ*; *nīlaṃ [bhavati eva] sarojam*.⁵

Here, Dharmakīrti enumerates three different usages of *eva* together with corresponding example sentences. Manorathanandin further comments on them in detail, according to which they can be interpreted as:

a. The exclusion of connection with others (*aparayoga-vyavaccheda*):

Example: *pārthaḥ eva dhanurdharaḥ*

‘It is Pārtha who is an archer’.

This sentence is uttered with regard to the doubt of whether the property of being an archer also belongs to someone else (*tādṛśamanyasyāpi kimasti*). The possibility of a positive answer (thus, the **connection with others** of the property of the predicate) is excluded.⁶

b. The exclusion of non-connection (*ayoga-vyavaccheda*):

Example: *caitraḥ dhanurdharaḥ eva*

‘Caitra is indeed an archer’.

This sentence is uttered with regard to the doubt of whether Caitra has the property of being an archer (*caitre dhanurdharatvam*). And the possibility that Caitra is not an archer (thus, the **non-connection** of the property) is excluded.⁷

⁵ *ayogaṃ yogam aparair atyantāyogam eva ca | vyavacchinatti dharmasya nipāto vyatirecakah | viśeṣaṇaviśeṣyābhyāṃ kriyayā ca sahoditah | vivakṣāto prayoge 'pi sarvo 'rtho 'yaṃ pratiyate | vyavacchedaphalaṃ vākyaṃ yataś caitro dhanurdharaḥ | pārtho dhanurdharo nīlaṃ sarojam iti vā yathā |* (Pramāṇavārttika 4,190-2). Ganeri (2011, 238-40) has also addressed the same passage and given his own translation. His quotation has in verse 191 *viśeṣaṇaviśeṣyābhyāṃ* and *tasyārtho 'yaṃ* instead of *viśeṣaṇaviśeṣyābhyāṃ* and *sarvo 'rtho 'yaṃ*.

⁶ *kintu tādṛśam anyasyā 'pi kim astī 'ti sandehe 'nyayogavyavacchedaphalaṃ viśeṣaṇam |* (Manorathanandivṛtti on Pramāṇavārttika 4,190-2).

⁷ *caitre dhanurdharatvasandehād viśeṣaṇenā 'yogamātram vyavacchidyate |* (Manorathanandivṛtti on Pramāṇavārttika 4,190-2).

c. The exclusion of absolute non-connection (*atyantāyoga-vyavaccheda*):

Example: *nīlaṃ bhavati eva sarojam*.

‘A lotus is dark blue indeed.’

Except that *eva* here lays emphasis on a verb rather than a non-verbal predicate, this case is, according to Ganeri (2011, 238-40), essentially the same as the exclusion of non-connection.

If Manorathanandin’s interpretation is correct, we can safely claim that Dharmakīrti’s analysis on *eva* is of two-valued logic – by uttering “It is Pārtha who is an archer”, that there are other people who are also archers is automatically denied; by uttering “Caitra is indeed an archer”, the possibility that he is not an archer is tacitly excluded.

Dharmakīrti can be viewed as a key figure in interpreting *eva* through the lens of logic. The discussion of this issue was continued by the later Indian logicians who followed the same analytical paradigm pioneered by Dharmakīrti and only deviated from his conclusion in classification details.⁸

The above raises the main problem with the definition of *eva*. On the one hand, traditional grammarians have only offered us overly concise explanations, which are insufficient for a contemporary learner to understand *eva* in detail. On the other hand, logicians like Dharmakīrti tend to define *eva* purely with regard to two-valued logic, which is not the case on the level of natural language, as I will argue in the next section. In this paper, I will examine language material from classical Sanskrit texts and closely examine four main ways of using this particle. I will argue that *eva* has a clear and coherent meaning, namely, indicating a mental operation to select one member out of a group while being neutral on other members’ qualifications.⁹ The analytical tool employed here is a (greatly) simplified and modified version of Cognitive Grammar developed by Langacker (1987; 1991; 2008), which provides both an intuitive and consistent theory covering semantics, grammar, and pragmatics and is considered by me to be the fittest framework for this issue.

⁸ Ganeri (2011, 237-43) classified the main opinions of traditional logicians on this topic into two categories – the opinion of Buddhists and that of Navya-Nyāya scholars. According to the material Ganeri quoted, the second opinion claims that *eva* only suggests the exclusion of connection with others.

⁹ This paper offers a preliminary proposal rather than an exhaustive analysis; therefore, I will refrain from discussing the applications of *eva* that some linguists classify as exclusively metrical. I propose that even in these so-called ‘metrical’ contexts, *eva* retains certain semantic significance. This viewpoint is reinforced by Coenen (2024, 317-18) in his exploration of the semantics of the Vedic emphatic particle *īd*.

2 Emphatic Expression

Many natural languages have one or more ways to create a specific kind of semantic structure, which I will refer to as a ‘single-out expression’ for now, for the sake of convenience. For example, in English, we can say:¹⁰

- (1) ***It is** Jerry **who** saved Tom.*
- (2) *Tony is the **very** person who has been courting me.*

In the above two sentences, the words highlighted in bold make up the ‘single-out’ structure. Of course, the object is not limited to a thing or a person; it can also belong to the category of place, time, e.g., as in the following examples:

- (3) ***It was** at three o’clock **when** I suddenly woke up.*
- (4) *This is the **exact** place where I first met my wife.*

The semantic difference between these sentences and the corresponding simple declarative ones, like, for instance, the minimal pair of (1) and (5) below, is usually vaguely claimed to be that the former one ‘emphasizes’ the agent. But how could we articulate this concept with greater precision?

- (5) *Jerry saved Tom.*

By reflecting on (1), one thing we can be sure about from intuition is that, somehow, in our mind, we select Jerry out of some ‘group’. I assume it is a set of candidates accessed in the speaker’s mind, with no clear bounds, pulled together according to the speaker’s world knowledge. In the case of example sentence (1), such candidates are the characters having to do with Tom in the well-known cartoon *Tom and Jerry* (e.g., Spike, Butch, Tuffy).

To illustrate this nuance [fig. 1], a dotted-line circle represents the set of candidates, and smaller solid-line circles represent the single characters activated in the speaker’s mind, including Tom. Further, an arrowed double line is used to represent the action ‘save’, and an arrowed single line the mental operation of this selection.

¹⁰ The example sentences (1)-(5) in English are composed by the Author.

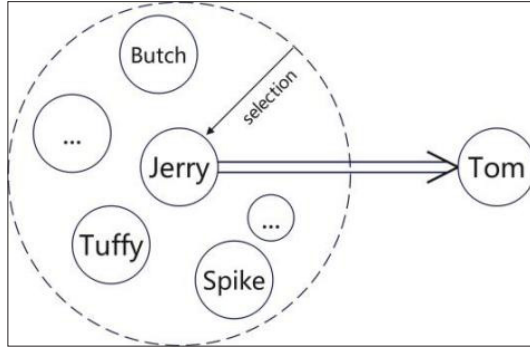


Figure 1 Base of example sentence (1)

This whole schema can be considered diagramming the base of the sentence (1). A base can be thought of as the set of all conceptions and knowledge involved in a linguistic expression, whether explicitly mentioned or not.¹¹ The most straightforward instance is that of quantifiers, for it would be impossible to understand *some students*, *most women*, or *all Americans* without invoking conceptions of a larger scope, i.e., ‘all students’, ‘all women’, and ‘all Americans’ [fig. 2].¹² However, it’s important to note that the base is not the same as the meaning of a linguistic expression. Indeed, within this base, we only pay attention to a particular item and linguistically encode it. This constituent is called a ‘profile’, as represented by the three small circles in bold in figure 2. It is the profile against the base as a whole that forms linguistic meaning.¹³

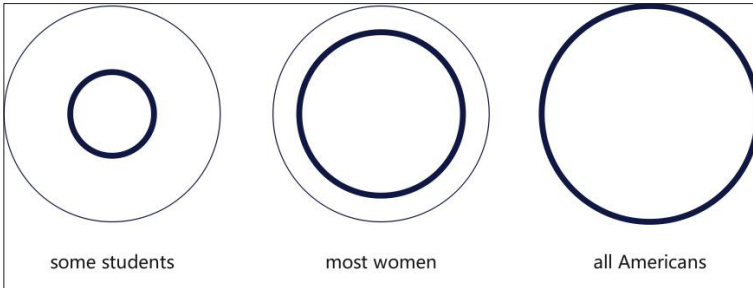


Figure 2 Profiles of quantifiers

¹¹ Cf. Langacker 1987, 183-9 for a detailed definition.

¹² In the case of ‘all Americans’, the invoked concept overlaps with the expression regarding their scopes.

¹³ Again, for a comprehensive discussion with regards to base and profile, cf. Langacker 1987, 183-9.

With this discussed, we are now equipped with the necessary theoretical tool to analyze the English ‘single-out’ pattern *it is... who* in (1). Since it indicates a mental operation of selecting some member out of a big group, its base consists of at least three elements – the group of candidates, the selected member, as well as the operation of selection itself. Compared with the simple declarative one (5), the ‘single-out’ sentence (1) illustrates Jerry’s role against the set of candidates currently accessible in the speaker’s mind. Therefore, I argue that it is this mental selection that constitutes the profile of the ‘single-out’ expression,¹⁴ and together with the selected member and the whole candidate group, they constitute the complete semantics of the ‘single-out’ expression [fig. 3].¹⁵

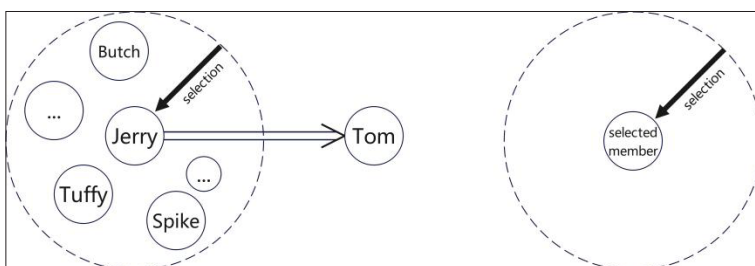


Figure 3 ‘Single-out’ expression in example sentence (1) (left) and semantics of ‘single-out’ expressions in general (right)

Before we move on, it’s important to clarify some terminology. I will refer to the ‘single-out’ expression we have discussed as an ‘emphatic expression’, with the element singled out being ‘emphasized’ or ‘given emphasis’. Some linguistic works categorize emphasis under the more general term ‘focus’ (e.g., Hengeveld, Mackenzie 2012, 89-92), but I believe this may not be the most accurate classification. In cognitive and functional linguistics, ‘focus’ typically refers to the element

¹⁴ This mental selection can be regarded as the result of a more general cognitive ability called ‘comparison’ assumed by Langacker. According to Langacker (1987, 101-5), we subconsciously and automatically compare things we perceive and register their differences. This is why we naturally notice the differences of pitches in a tune and appreciate the flow of music rather than perceiving a sequence of isolated sounds, and why we could not help but notice that a student is more intelligent than others in a class. In the same way, when we access this set of candidates [fig. 3], we automatically compare Jerry with other members with regards to ‘who helped Tom’, through which process Jerry stands out and is selected in our mind.

¹⁵ To some extent, the analysis proposed in this paper resembles a specific reading of the English adverb *just* proposed in Lee 1991, 55-6, which is termed the ‘intensifactory’ reading. According to Lee, in the sentence *The shop is just near the bank*, “just near the bank” is the extreme member among all the spatial spots conceived as “near the bank”.

conveying new and/or contrasting information.¹⁶ While emphasis shares some similarities with this concept, they are different in essence. As evidence, the Japanese language explicitly assigns two separate particles for them – ‘ha’ (は) for focus and ‘koso’ (こそ) for emphasis.¹⁷

- (6) 森さん は 医者; 太田さん は 教師。
Mori-san ha isha Ooda-san ha kyoushi.
 Mr.Mori **FOCUS PARTICLE** doctor Mr.Ooda **FOCUS PARTICLE** teacher
 ‘Mr. Mori is a doctor; Mr. Ooda, **however**, is a teacher.’

- (7) 太田さんこそ 教師 (だ)。
Ooda-san koso kyoushi (da)
 Mr. Ooda **EMPHATIC PARTICLE** teacher be: PRE
 ‘**It is** Mr. Ooda **who** is a teacher.’

In the example sentence (6), *ha* signifies that Mr. Mori and Mr. Ooda are contrastive; in (7), however, *koso* performs the singling-out function and selects Ooda from a potential group in the speaker’s world knowledge.¹⁸

Another thing worth special attention is distinguishing emphatic expressions from limitative ones, such as (6). As we have seen from *Pramāṇavārttika*, Dharmakīrti interprets emphatic *eva* as essentially equivalent to limitative expressions – if you single out one, you automatically exclude others. Although self-evident within the field of logic, this perspective is not always true conceptually and linguistically. We humans have no problem focusing on only one of two interdependent sides of a thing and totally ignoring the other. Thus, *You are not wrong* certainly does not equal *You are right*, as is the case here – by uttering example sentence (1), the speaker merely recognizes the very existence of other candidates and remains neutral about them; while by uttering (6), the speaker explicitly denies their qualification.¹⁹ As we can see in figure 4, although both of them involve the same base and profile, they choose two different

¹⁶ Langacker 1991, 397; 2008, 57; Halliday, Matthiessen 2014, 118-19; Hengeveld, Mackenzie 2012, 89-92.

¹⁷ The example sentences (6), (7), and (9) in Japanese are composed by the Author.

¹⁸ Both focus and emphasis can be considered as part of a broader concept known as ‘prominence’, as proposed by Langacker (1987; 2008). According to Langacker’s theory (2008, 31), entities that are perceived as more ‘prominent’ than others have a higher level of neural activation. In simpler terms, we can say that these entities are more noticeable than others in our thinking.

¹⁹ Of course, in the case of emphatic expressions, the ‘unqualification’ can also be explicated by other components, like *It is Jerry who saved Tom, not others*.

construals, that is, according to cognitive linguistics, different perspectives from which we interpret the same conception (for the sake of simplicity, candidates other than Jerry is omitted) [fig. 4]:

- (8) **Only** Jerry saved Tom.

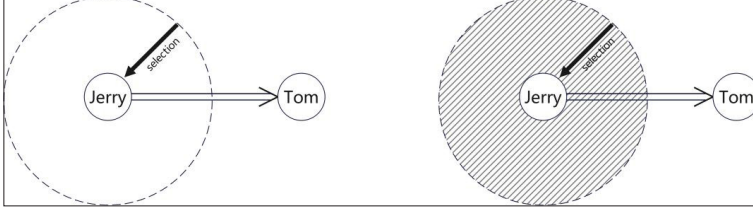


Figure 4 Emphatic expression in example sentence (1) (left) and limitative expression in example sentence (6) (right)

Again, support for this analysis can be found in the Japanese language – Japanese uses two distinct particles, ‘koso’ (こそ) and ‘dake’ (だけ), to convey subtle semantic differences. Compare (7) with (9):

- | | | | |
|-----|--------------------------------------|-----------------|--|
| (9) | 太田さん だけ | 教師 (だ)。 | |
| | Ooda-san dake | kyoushi (da) | |
| | Mr. Ooda LIMITATIVE PARTICLE | teacher be: PRE | |
| | ‘Only Mr. Ooda is a teacher.’ | | |

After discussing the basic conceptions from cognitive linguistics, we can now move on to talk about *eva*, the emphatic particle in Sanskrit.

3 Semantics of *Eva*

Under current analysis, the various usages of *eva* can be classified into three categories. They will be examined one by one and abstracted into one consistent interpretation. A unique usage of *eva* in conjunction with discourse particles will also be discussed. The majority of the example Sanskrit sentences are extracted from classical texts. Nevertheless, I with discretion have also crafted a few *ad hoc* sentences for clearer demonstration, given that Sanskrit is a classical language and finding exact minimal pairs can be challenging (i.e., a pair of linguistic materials differing by only one element) in the extant literature. They are example sentences (18), (20), (23), and (24) without ‘()’ at the end to indicate the sources. All the example sentences in Sanskrit are given without sandhi.

3.1 Emphasizing Things or Properties

Perhaps the most frequent and prototypical use of *eva* is to follow a nominal and emphasize its referent against other candidates:

- (10) *yad* *bhavatā* *abhihitam* *tad*
 RL.N.NOM.SG you: INS.SG to say, to speak: DEM.N.NOM.SG
 P.PT.N.NOM.SG
eva *mayā* *kartavyam...*
EMPHATIC PARTICLE I: INS.SG to do, to perform:
 FUT.P.PT.N.NOM.SG
 ‘Whatever said by you, that **exactly** will be carried out by me...’ (said Saṃjīvakā to Karaṭaka, indicating he would be willing to follow Karaṭaka’s advice after being told a cautionary story).
 Pañcatantra 1,3

- (11) *ananyabhājam* *patim* *āpnuhi* *iti* *sā*
 not dwelling husband: to obtain: IND her: F.NOM.SG
 on another M.ACC.SG PRE.IMP.2.SG
 (woman):
 M.ACC.SG
tathyam **eva** *abhihitā* *bhavana* |
 truth: N.ACC.SG **EMPHATIC** to tell: a name of Śiva: M.INS.SG.
PARTICLE P.PT.F.NOM.SG
na hi īśvaravyāhṛtayaḥ kadācit puṣṇanti loke viparītam artham ||
 “‘May you obtain a husband who does not dwell on another woman!’ – she (Umā) was told the **very** truth by Bhava (Śiva). And never will the words of the Divine foster a reversed meaning in the world.’
 Kumārasambhava 6,63

For instance, by uttering (10), I reckon that Saṃjīvakā indicates there are other choices of actions for him to take which also fall within the group of things that “will be carried out” by him (*tad mayā kartavyam*) while within this category stands out one qualified member, i.e., the things told by Karaṭaka (*yad bhavatā abhihitam*) [fig. 5]. Again, as I argued above, I disagree with defining *eva* here as ‘limitative’, because *eva* merely means to access the group consisting of other members while not commenting on their qualifications. This is why *tad eva* should not be translated as ‘that alone’ or ‘only that’.

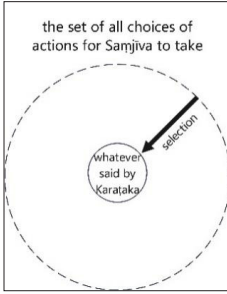


Figure 5
Emphatic expression in example sentence (10)

Of course, to emphasize a nominal does not mean it has to be limited to a nominative or an accusative case:

(12) <i>yasya</i> RL.N.GEN.SG	<i>devasya</i> deity (god or goodness): N.GEN.SG	<i>yaḥ</i> RL.M.NOM.SG	<i>proktaḥ</i> declared: M.NOM.SG	<i>tasya</i> DEM.N.GEN.SG
<i>tena</i> DEM.M.INS.SG	<i>eva</i> EMPHATIC PARTICLE	<i>maṇḍalam</i> maṇḍala: N.ACC.SG		
<i>yasya yasya tu yaḥ mantraḥ yathā dhyānādīpūjanam </i>				
<i>tat</i> DEM.N.NOM.SG	<i>tat</i> DEM.N.NOM.SG	<i>tena</i> DEM.M.INS.SG	<i>eva</i> EMPHATIC PARTICLE	<i>mantraṇa</i> mantra: M.INS.SG
<i>pūjayitvā</i> to worship: ABL	<i>prayatnataḥ </i> zealously: IND			
<i>tasya</i> DEM.N.GEN.SG	<i>eva</i> EMPHATIC PARTICLE	<i>bījamantrābhyām</i> bīja and mantra: M.INS.DU	<i>mūrdhni</i> head: M.LOC.SG	
<i>dadyāt</i> to offer, to place: PRE. OPT.3.SG		<i>pavitrakam </i> a kind of plant: M.ACC.SG		

‘(A man dedicated to devotion) should (perform) such as the maṇḍala and the ritual as meditation, and worship zealously whatever deity (he believes in), **in the exact form** prescribed to the deity **to the chanting of the exact mantra**, and put the Pavitraka on his head to the chanting of **the exact** bīja and mantra **of that deity**.’

Kālikāpurāṇa 59,75b-77a

- (13) *atha tena vyādhena taṇḍulakaṇān vikīrya jālaṃ vistīrṇam |*
saḥ ca pracchannaḥ bhūtvā sthitaḥ |
tasmin **eva** **kāle** *citragrīvanāmā*
DEM.M.LOC.SG **EMPHATIC** **time: M.LOC.SG** named Citragrīva:
 PARTICLE M.NOM.SG
kapotarājaḥ *saparivāraḥ* *viyati*
king of the pigeons: being with a retinue: M.NOM. the sky:
M.NOM.SG SG N.LOC.SG
visarpan *tān* *taṇḍulakanān* *avalokayāmāsa |*
to move along: PRE. DEM.M.ACC.PL rice grain: M.ACC.PL to see, to notice:
PT.M.NOM.SG PF.3.SG
‘Then, the hunter scattered grains of rice about and set a net. He then stayed hidden.
At that very time, the king of the pigeons named Citragrīva, passing through the sky
with his retinue, had noticed those rice grains.’
Hitopadeśa 1,1

A special case arises when emphasis is placed on a universal quantifier, specifically words like *every* and *all*. In English, phrases such as *just every student* or *exactly every student* imply that the suggested group of candidates includes all possible proportions of students – such as ‘some students’, ‘most students’, as well as ‘none of the students’ – and among them, ‘every student’ is selected out. In Sanskrit, we observe a similar usage as follows:

- (14) *na śocanti na vāñchanti na yācante śubhāśubham |*
sarvam **eva** *ca* *kurvanti* *na*
all, every: N.ACC.SG **EMPHATIC** and: to do, to not: IND
 PARTICLE IND perform:
 PRE.3.PL
kurvanti *iha* *kiṃcana |*
to do, to perform: here, whatever: N.ACC.SG
PRE.3.PL now: IND
‘They do not grieve, nor do they desire, or require the pleasant and the unpleasant; in this world, they perform **exactly everything** while also performing nothing.’
Yogavāsiṣṭha 2,13,2

The semantics of *sarvam eva* in this verse is illustrated in [fig. 6]:²⁰

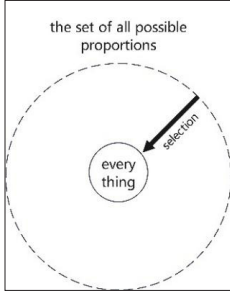


Figure 6
Emphatic expression in example sentence (14)

Further, this use also applies to demonstrative adjectives and adverbials:

- (15) *tad vai yugasahasrāntam brāhmam puṇyam ahar viduḥ |*
rātrim ca tāvatīm eva te
 night: and: IND **such: F.ACC.SG** **EMPHATIC PARTICLE** they:
 F.ACC.SG PRO.M.NOM.PL
ahorātravidaḥ janāḥ |
 knowing days and nights: man, person: M.NOM.PL
 M.NOM.PL
 ‘(Those who) know that the very holy day of Brahma is the end of one thousand ages
 and that the night lasts **just as long**, they are men acquainted with (the length of)
 days and nights.’
 Manusmṛti 1,73

When used this way, the meaning of *eva* is essentially the same as above, except that the object it singles out is not a thing but some specific attribute or state of a thing, both of which I will generally refer to as ‘property’. In (15), the speaker states that among all possible lengths for the night of Brahma, the one that matches the length of the day is chosen. With a square representing the length of

²⁰ Coenen (2024, 263-75) also touches on this topic when he discusses the use of Vedic emphatic particle *id* following universal quantifiers. The perspective from which he looks at this problem, however, is different from mine. Coenen recognized *id* after a universal quantifier as a ‘slack regulator’, a concept borrowed from Lasersohn (1999). A slack regulator serves to reduce the ‘pragmatic slackness’ of a universal quantifier. For instance, let us compare the following two example sentences:

- a. This position is open to everyone except for John.
- b. *This position is open to absolutely everyone except for John.

Without ‘absolutely’, ‘everyone except for John’ is acceptable for the universal quantifier allows for some exceptions. On the other hand, ‘absolutely everyone except for John’ is deemed unacceptable because the slack regulator ‘absolutely’ reduces this slackness and thus does not allow any exceptions like ‘John’.

a day of Brahma's, the meaning of *eva* emphasizing a property can be similarly illustrated as below:

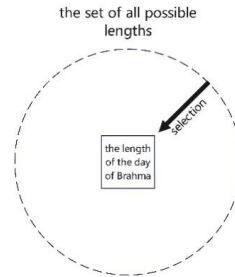


Figure 7

This still applies even when the emphasized component is not explicitly uttered:

- (16) *śrīḥ* *tava* *eva* *me*
prosperity, auspiciousness: you: PRO.GEN.SG **EMPHATIC** I: PRO.GEN.SG
F.NOM.SG **PARTICLE**
*astu,*²¹²²
to be: PRE.IMP.3.SG
'May the prosperity be with me**just** as be with you.'
Gaṇaratnamahodadhi 1

In the example sentence (16), there is an omitted *iva* following *śrīḥ tava* which is the emphasized component here, i.e., *śrīḥ tava iva eva me astu*. Traditional commentators sometimes interchangeably use *iva* and *eva* to explain each other, but as analyzed in Brereton (1982), the two cases are not semantically equivalent, for the latter adds an extra exactness nuance:

- (17) śrīḥ tava **iva** me astu|
 ‘May the prosperity be with me **as** be with you.’

21 The same example sentence given by Apte is *śrīḥ te eva me astu*, which can be explained almost the same way. And Apte also explains it as *tava iva*.

22 Again, Coenen (2024, 253, 256) discusses a similar issue in which the Vedic emphatic particle *íd* comes after the words indicating similarity as *iva* and *ná*. His take is that it indicates a point that is close to full identity on a so-called ‘scale of similarity’.

- (16) *śrīḥ tava eva me astu* |
≈ *śrīḥ tava iva eva me astu* |
or *yathā śrīḥ tava astu tathā eva me astu* |
'May the prosperity be with me **just as** be with you.'

To summarize it, we can reasonably define the first use of *eva* as selecting a thing or a property against a set of alternative candidates accessed in the speaker's mind.

3.2 Emphasizing the Temporal Range

Along with a nonfinite verb, *eva* can emphasize a specific time range of an action. In this case, lexicographers usually interpret it as 'as soon as' or 'while just'. Although these translations are quite accurate in their respective contexts, they lack a coherent and reasonable semantic analysis, especially one that can form a consistent explanation with other uses of *eva*. Here, I classify the temporal use into two subcases, in which *eva* respectively emphasizes the completion time and the progressing time.

Let's take a look at the first one, in which *eva* follows a gerund and underlines the completion time of an action:

- (18) *bhojanam* *bhuktvā* ***eva***
food, dinner: N.ACC.SG to eat: GER **EMPHATIC PARTICLE**
aham *bahiḥ* *agaccham* |
I: PRO.NOM.SG outwards: IND to go: IMPF.1.SG
'It is right after having dinner that I went out.'
- (19) *saḥ* *tām* *sprṣtvā*
He: PRO.M.NOM.SG she: PRO.F.ACC.SG to touch: GER
eva *videhatvam* *prāptaḥ* |
EMPHATIC PARTICLE death: M.ACC.SG to approach: P.PT.M.NOM.SG
'It is as soon as he (Pāṇḍu) touched her (Mādrī) **that** he approached death.'
Mahābhārata 1,90,74

How should we interpret the semantics of *eva* in these sentences? To answer this, we should first take a look at gerunds. Consider the following minimal pair:

- (20) *bhojanam bhuktvā aham bahiḥ agaccham*.
'I went out after having dinner.'

- (18) *bhojanam bhuktvā eva aham bahiḥ agaccham.*
'It is right after having dinner that I went out.'

A gerund is a type of non-finite verb form in Sanskrit, indicating the completed state of an action that happens before another referred to in the main clause. Considering the cognitive domains involved in a verb are highly heterogeneous and complicated, here, only the temporal domain, to which the semantics of gerunds is most related, is taken into account. I use two arrowed lines to represent the actions involved in a gerund sentence, with the left one indicating the action of a gerund and the right one that of a main verb, which is to take place subsequently. Points at both ends represent the beginning and end of the action. A temporal axis is drawn below to illustrate the concept of time flow. Thus, figure 8 shows the base of a gerund in the temporal domain.

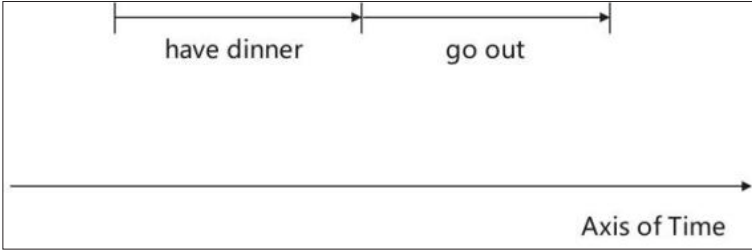


Figure 8 Base of example sentences (18) and (20)

When used without *eva*, as in (18), a gerund highlights the completed state of an action, with no reference to the amount of time that occurs between the completion of such action and the performance of the action encoded in the main clause. A bold point is used to represent it. (20), on the other hand, suggests a very short interval between the end of the former action and the start of the latter. I assume it also implies that other possibilities were available for the addressee as to when the utterer 'went out', and within those possibilities 'right after he had dinner' stands out as the fittest one for the intent situation. Thus, *eva*'s presence here shifts the profile to a specific moment (represented by a point)²³ and selects it within the group of all optional times [figs 9-10].²⁴

²³ This is roughly analogous to the difference between a point on a line and the position of the same point on the coordinate axes.

²⁴ In illustrating a similar use of English *just*, Lee (1991, 55-6) proposes what he calls a 'marginal' reading with which, again, the theory presented here in this paper appears to

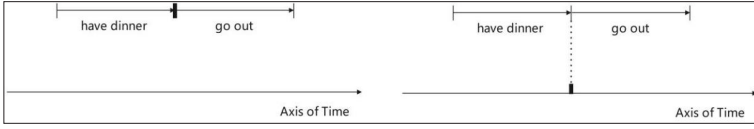


Figure 9 Profiles of the gerund in example sentence (20) (left) and with *eva* in example sentence (18) (right)

A similar but less frequent use should also be mentioned, in which *eva* follows a locative absolute comprising a past participle:

- (21) *adūravartinīm siddhim rājan vigaṇaya ātmanaḥ.*
upasthitā *iyam* *kalyāṇī* *nāmnī*
to appear: this: auspicious: F.NOM.SG name: N.LOC.SG
P.PT.F.NOM.SG DEM.F.NOM.SG
kīrtite **eva** *yat*
to say: **EMPHATIC** because: IND
P.PT.N.LOC.SG **PARTICLE**
‘King! Notice your realization (of the goal) is in no long time. For this blessed one (Nandinī the Cow) has appeared **right after** her name was called.’
Raghuvamśa 1,87

The only difference in this case is that the time of the action is limited to the past, while a gerund is not temporal, i.e. its use is not limited to any specific tense, as illustrated by example sentence (18), which is in past tense, and (22), which, we will be talking about soon, is in present tense. Therefore, in figure 10, I use a point in the middle to indicate ‘now’, the present. What should be noticed is that this ‘now’ point does not necessarily have to overlap with the utterance time. Instead, it can be any time point perceived to be the present depending on the speaker’s perspective, as the use of the historical present in English.

have some commonality. Nevertheless, the framework underlying his argument remains binary logical – the constituents modified by *just* are perceived as either true or false. In one of his original examples *He left just before night*, the action of *leaving* is perceived as being punctual. So, there are only *leave before midnight* and *leave after midnight*. And without *just*, it is not specified how long leaving takes place before this specific point of time; on the other hand, when combined with *just*, the action of *leave* is situated so infinitely close to this point of time that, even if we move it a tiny little bit further, the truth value of *leave before midnight* would not apply anymore. Thus, we say that *leave* is situated at the ‘margin’ on one side of the time point of *midnight*.

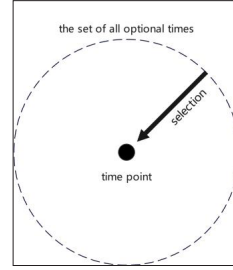


Figure 10
Profile of the locative absolute with *eva* in example sentence (21)

At last, it should be mentioned that the pattern of ‘gerund + *eva*’ also has a slightly different nuance, as in (22):

(22) *yadā vai karoti atha nistiṣṭhati; na ākṛtvā nistiṣṭhati;*

kṛtvā **eva** *nistiṣṭhati;...*

to do: GER EMPHATIC PARTICLE to serve (the teacher): PRE.3.SG²⁵

‘It is when one performs (his duties) that he serves (the teacher); not that he serves without performing; **it is after performing that** one serves...’

Chandogya Upaniṣad 7,21,1

Here, what is emphasized by *eva* is not just the moment right before which the intended action occurs, but all the time after this action. Thereby, example sentence (18) also has a different reading ‘It is after I had dinner that I went out’.²⁶ Nevertheless, this slight difference in nuance should pose no difficulty for the current framework, since it is only a matter of which component of the same base of gerund is selected to be profiled. And this is in fact an important source of polysemes. Think about the English word *across* as in *I ran across the road* and *She lives across the road*. Both involve the same base of the word *across* – a trajectory of moving from one locus through some obstacle to another locus. The only difference is that the former sentence profiles the process (or the obstacle) of *across* while the latter the destination of *across*. Let us name the ‘it is right after/as soon as’ reading as Reading 1 and the ‘it is after... that’ reading as Reading 2, the difference between the two readings and how they derive from the same base can be illustrated below:

²⁵ The interpretation of *niṣṭhā* here is according to Śaṅkara’s commentary (Jha 1942, 401). Also, notice that (22) provides further evidence to our analysis here by explicitly mentioning another possible temporal choice as to when to ‘serve the teacher’.

²⁶ The analysis of this nuance is inspired by Coenen (2024, 256-62). Coenen divides a similar temporal use of *id* into three subcases – specifying the time point right after the event; excluding all earlier time points; and excluding all other time points.

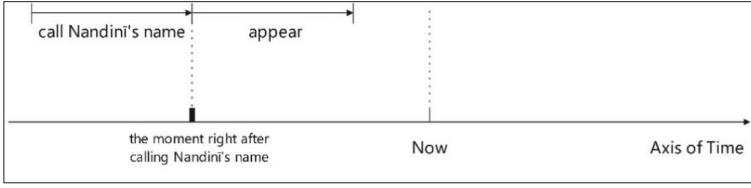


Figure 11 Base of example sentence (18) and its two readings

Then, we move on to the second subcase, in which *eva* emphasizes a present participle. Sanskrit present participle indicates the state of progress during which another action referred to in the main clause takes place:

- (23) *tasya* *viṣaye* *cintayan* *rāmam*
 He: DEM.M.GEN. sphere, territory: thinking: PRE. Rāma: M.ACC.SG
 SG M.LOC.SG PT.M.NOM.SG
dṛṣṭavān
 to see: P.PT.N.NOM.SG
 'I saw Rāma while I was thinking about him.'

The base of (23) is illustrated in figure 12:

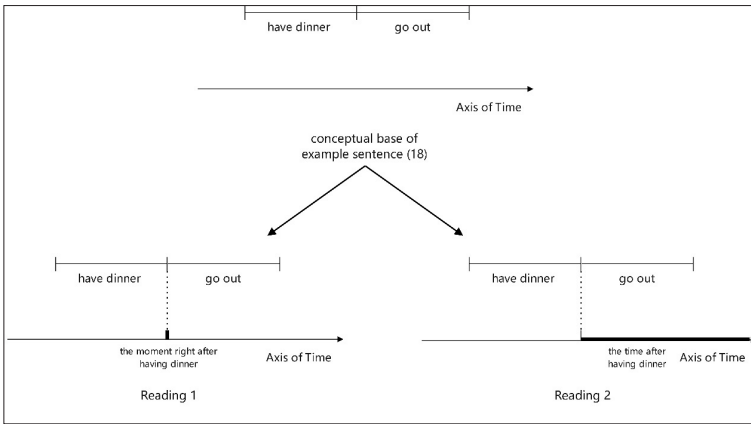


Figure 12 Base of example sentence (23)

Followed by *eva*, this overlapping is further emphasized in the flow of time:

- (24) *tasya viṣaye cintayan eva rāmaṃ dṛṣṭavān.*

‘I saw Rāma **just** as I was thinking about him.’

- | | | | | | |
|------|-------------------|---------------------------|------------------------------|-----------------------------|-----------------------------|
| (25) | <i>yuṣmābhiḥ</i> | <i>bhasma</i> | <i>bhakṣitavyam</i> | <i>iti</i> | <i>vadan</i> |
| | you: PRO.INS.PL | ashes:
N.NOM.SG | to eat:
FUT.P.PT.N.NOM.SG | IND | to say: PRE.PT.M.NOM.
SG |
| | <i>eva</i> | <i>patitaḥ</i> | <i>taiḥ</i> | <i>vyāpāditaḥ</i> | <i>ca</i> |
| | EMPHATIC | to fall:
P.PT.M.NOM.SG | he: PRO.M.INS.PL | to perish:
P.PT.M.NOM.SG | and: IND |
| | PARTICLE | | | | |

“(I wish) you (to) eat ashes!” -- **just** as he (the tortoise) was saying this, he fell off and was killed by them (cowherds).’

Hitopadeśa 4

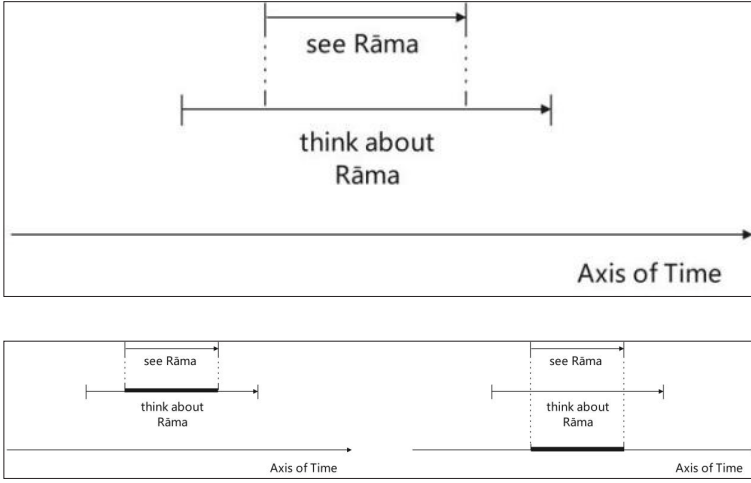
- (26) *yaḥ anadhītya dvijaḥ vedam anyatra kurute śramam.*

<i>sa</i>	<i>jīvan</i>	<i>eva</i>	<i>śūdratvam</i>
he:	to live:	EMPHATIC	the state of being
PRO.M.NOM.SG	PRE.PT.M.NOM.SG	PARTICLE	Śūdra: N.ACC.SG
<i>āśu</i>	<i>gacchati</i>	<i>sānvayaḥ</i>	
soon: IND	to go:	being along with descendants: M.NOM.SG	
	PRE.3.SG		

‘A twice-born man who, not having studied the Veda, engages in worldly business, **it is while** he is still alive **that** he soon falls to the state of a Sudra, and so do his descendants.’

Manusmṛti 2,168

The semantic contrast presented [figs 13a-b] highlights the differing implications of a present participle when used independently compared to when it is followed by *eva*, with an arrowed line below representing the duration of the action referred to by a present participle and one above the action in the main clause. I use a corresponding bold segment to represent this overlapping. In the example sentence (23), what is profiled is the overlap of two actions itself; while in (24), with *eva*, the profile shifts to the time stretch during which this overlap occurs. And this mental selection profiled by *eva* in (24) is illustrated in figure 14, with a small circle inside indicating the stretch of time:



Figures 13a-b Profile of the present participle in example sentence (23) (left) and with *eva* in example sentence (24) (right)

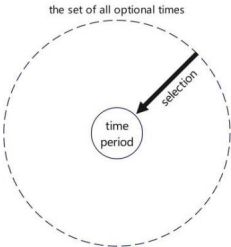


Figure 14
Emphatic expression in example sentence (24)

It should be clear from the above that in the second use, *eva* selects a specific temporal range from a group of optional times. This range can be a moment or a period, depending on which kind of nonfinite verb *eva* follows.

3.3 Emphasizing the Binary Judgment

The last use of *eva* is to emphasize the judgment of a predicate. The judgment is typically a binary one, meaning it is confined to one of two opposing choices.

What does this mean? When we utter an English sentence as *John is indeed Jimmy's husband*, I assume that *indeed* suggests we subconsciously access both the possibilities that 'John is Jimmy's husband' and 'John is not Jimmy's husband'. Within them, 'John is Jimmy's husband' stands out as the one who fits the intent situation.

This is also the case with *eva*:

- (27) *saḥ cet phalasya bhoktvā tarhi ātmā iti hetoḥ |*
ātmavāveśeṣaṇāt iti yāvat | tasmāt kva api ātmantare api tat kim na syāt.
api tu bhavitavyam eva.
rather, instead: IND to be, to exist: FUT.P.PT.N.NOM.SG **EMPHATIC PARTICLE**
'If he is the consumer of the fruit (of sins), it is for the reason of the self. That is, because of the non-distinction of self-being. Thus, would not that (fruit) also exist somewhere else within the self? Rather, it **definitely** would.'
Naiṣadhiyaparakāśavyākhyā 17,52

- (28) *gataḥ eva na te nivartate*
to go: **EMPHATIC PARTICLE** not: IND you: PRO.GEN. to come back:
P.PT.M.NOM.SG **PARTICLE** SG PRE.MID.3.SG
saḥ sakhā dīpaḥ iva anilāhataḥ
that: friend: lamp: as: IND being struck by
DEM.M.NOM.SG M.NOM.SG M.NOM.S wind: M.NOM.SG
aham asya daśā iva paśya mām aviśahavyasanena dhūmitām
'Gone **indeed** is your (Madhu, a name of Śiva) friend (Kāma). He comes not back, as a lamp put out by the wind. I (Rati) am like this lamp wick. Behold me (dispersing) like the smoke under unbearable misfortune.'
Kumārasambhava 4,30

In this case, *eva* can be placed after whatever predicate it lays emphasis on. Again, this case does not deviate from the above two uses. Take (28), for instance, when we claim that somebody is 'gone indeed', we subconsciously access both the possibilities of 'be gone' and 'be not gone' in our mind. Thus, compared to the negative one, 'be gone' stands out as the eligible choice fitting the intent situation.

Here, I define *eva* as selecting one judgment out of two. In figure 15, the bigger circle is employed to represent the set of all possibilities, which contains only two members – affirmation and negation; the square is used to present two of the judgments:

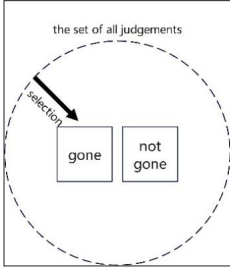


Figure 15
Emphatic expression in example sentence (28)

And the above interpretation applies to negation as well:

- (29) *sarveṣāṃ pṛthivīśānām tvam ajeyaḥ bhaviṣyasi.*
sarvaḥṣayaakaraḥ bhāvī na cirād
holocaust: M.NOM.SG predestined, inevitable: M.NOM.SG not: IND for long: IND
eva bhārgava.
EMPHATIC PARTICLE Bhārgava (a name of Paraśu Rāma): M.VOC.SG
'You (Paraśu Rāma) will become invincible to all the rulers of the earth. A holocaust is due
in **certainly** no long time, Bhārgava.'
Skandapurāṇa 5,2,29,6
- (30) *tasmai datvā tadā siddham sarvam viprāya niścitaḥ.*
niśṣeṣite anne bhagavān
to have nothing left: P.PT.N.LOC.S food: N.LOC.SG the glorious one: M.NOM.SG
abhuktvā **eva** mahātapaḥ
to not eat: GER **EMPHATIC PARTICLE** great ascetic: M.NOM.SG
na kiñcid avadat vipram maunavratam upasthitaḥ
'Then, with the mind determined, having given all the cooked food to the brāhmaṇa and
indeed not eaten, and no food being left, the glorious ascetic (Viśvāmitra) did not speak
a word to the brāhmaṇa and approached his vow of silence (again).'
Rāmāyaṇa 1,65,6-7a

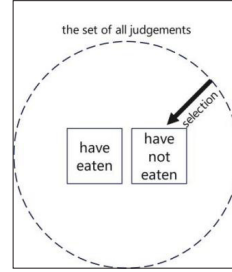


Figure 16
Emphatic expression in example sentence (30)

3.4 Conjunction of *Eva* with Discourse Particles: A Case Study of *Tu*

At last, a specific use of *eva* is yet to be discussed – how should we analyze the combination of *eva* with a discourse particle like *ca*, *tu*, or *hi*?

A discourse is a dynamic process with ‘flowing’ contents and ever-changing topics that develop over time. To illustrate this process, in figure 17, a rounded-corner rectangle is used to represent what is referred to as a unit of content from a particular discourse. A unit of content can consist of expressions of any length, such as a round of conversation, a written paragraph, a sentence, or even several phrases or words, that can be roughly thought of as centering around a specific topic. Since there are no strict standards regarding what counts as ‘centering around a specific topic’, the dotted-line frame represents this ‘roughness’ as we have done before with the set of candidates. In addition, to show the dynamic development of discourse, at least three of these units must be set up separately labeled as ‘last’, ‘current’, and ‘next’.

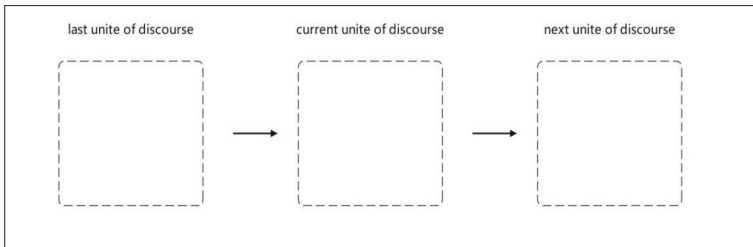


Figure 17 Flow of discourse

On the other hand, change constitutes only one aspect of discourse, and to maintain coherence and clarity in a discourse, there must be connections between units of content.²⁷ These connections can be achieved by using conjunctions in English, which indicate a general connection between two units. However, discourse particles are another part of speech that further specifies the exact element that forms this connection. We can find them in languages like Sanskrit (e.g. *tu*, *ca*, *hi*) and Japanese (e.g. *wa*, *mo*). Here, let us take for example *tu*, one of the most frequently used discourse particles in Sanskrit.²⁸

- (31) *sthūlam* *tarpayate* *Viśvam*;
gross: N.NOM.SG to satisfy: MID.PRE.3.SG Viśva: M.ACC.SG
praviviktam ***tu*** *taijasam*
subtle: N.NOM.SG **however: IND** Taijasa: M.ACC.SG
'The gross (object) satisfies the Viśvam; the subtle (object), **however**, satisfies the Taijasa.'
Māṇḍūkya Upaniṣad 1,6,4

In figure 18, a dotted line in bold shows this connection, which is the profile of *tu*.

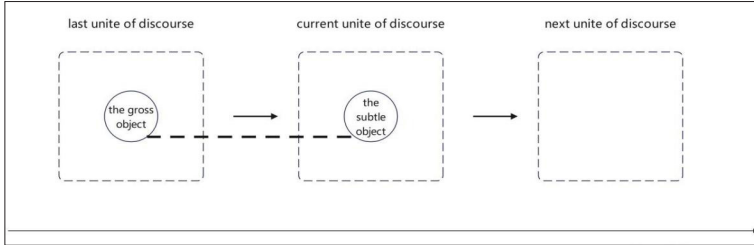


Figure 18 *Tu* in example sentence (31)

Now, it seems pretty straightforward that the combination of *eva* with a discourse particle results in the integration of mental selection and discourse connection. But there is a subtle difference to consider regarding whether *tu eva* and *eva* mean the same. Let's examine the following two sentences:

²⁷ This connection can also be seen as the result of the mental comparison mentioned in section 2.

²⁸ *Tu* is interpreted by some Sanskritists as having both the sense of conjunction ('and', 'moreover') and the sense of disjunction ('but', 'however') (cf. *Monier-Williams Sanskrit-English Dictionary*, *Wilson Sanskrit-English Dictionary*). A simpler explanation of this, I believe, is that *tu* merely signifies discourse connection, while its specific meaning is contextual, which is analogical to Japanese *ga* (が).

- (32) *yaḥ vai bhūmā tad sukham; na alpe sukham asti; bhūmā eva sukham;*
bhūmā tu eva vijijñāsitavyaḥ
 the Infinite: M.NOM.SG **and: IND EMPHATIC PARTICLE** to explore:
 FUT.P.PT.M.NOM.SG

Iti;

IND

bhūmānam, bhagavaḥ, vijijñāsa iti

“‘What exactly is the Infinite, that is bliss; there is no bliss in the Finite; it is the Infinite that is bliss; **and it is** the Infinite **that** should be explored”. “Honorable Sir, I desire to learn the Infinite”.

Chāndogya Upaniṣad 7,23,1

- (33) *adhyāpanam adhyayanam yajanam yājanam tathā|*
dānam pratigrahaṃ ca eva brāhmaṇānām akalpayat|
prajānām rakṣaṇam dānam ijjādhyayanam eva ca|
viṣayeṣu aprasaktiḥ ca kṣatriyasya samāsataḥ|
paśūnām rakṣaṇam dānam ijjā adhyayanam eva ca|
vaṇikpatham kuśīdam ca vaiśyasya kṛṣim eva ca|
ekam eva tu śūdrasya
 one: N.ACC.SG **EMPHATIC PARTICLE** **however: IND** Śūdra: M.GEN.SG
prabhuḥ karma samādiśat
 the lord: duty: N.ACC.SG to assign:
 M.NOM.SG IMPF.3.SG

eteṣāṃ eva varṇānāṃ śuśrūṣāṃ anasūyayā

‘(Brahmā) assigned to the Brāhmaṇas teaching, studying (Vedas), performing sacrifice for their own as well as for others, and giving and accepting (alms). It is protecting the people, giving (alms), sacrificing, and studying (Vedas), as well as detachment from the mundane pleasures that (are assigned) to the caste of Kshatriya together. It is tending the livestock, giving (alms), sacrificing, and studying (Vedas), and trading, lending money, as well as agriculture, that (Brahmā) (assigned) to the caste of Vaiśya. **It is** one duty, **however, that** the lord (Brahmā) assigned to the caste of Śūdra - to be at the service of these (three) castes without evil intention.’

Manusmṛiti 1,88-91

According to cognitive linguistics, the difference in word order also reflects different construals of the same linguistic expression (i.e., it implies different interpretations). Take, for instance, the sentence *He sent a letter to Susan* does not equal *He sent Susan a letter* (cf. Langacker 1987, 39). Here, I argue that in the case of *tu eva*, the connection between the two discourse units lies in the noun marked by *tu*, which, in the example sentence (30), is ‘the infinite’ (*bhūman*). Differently, in the case of *eva tu*, the connection is in the group of candidates suggested by *eva*. Thus, in (31), it is a set of all duties assigned to four castes, and what constitutes connection with this in

the context, is the duties ascribed to three castes other than *śūdra*.²⁹ Figure 19 explicitly demonstrates this difference.



Figure 19 *Tu eva* in example sentence (32) (Up) and *eva tu* in example sentence (33) (Down)

4 Conclusion

While traditional Indian linguists have provided various interpretations of the emphatic particle *eva*, a more nuanced and detailed explanation using contemporary linguistic frameworks is essential for a comprehensive understanding of its usage. This paper has systematically examined the primary functions of *eva* and demonstrated that a thorough semantic analysis is feasible within Cognitive Grammar.

This approach posits that *eva* reflects the speaker's cognitive process of selecting one option from a pool of candidates accessed in their mind while remaining neutral toward the others. This fundamental meaning is evident in the three main usages of *eva*

²⁹ Cf. *Manusmṛiti* 1,87-1,91.

identified in this paper: (a) when it emphasizes a noun, adjective, or adverb, it selects a specific item or quality; (b) when it emphasizes a gerund or a present participle, it selects a particular temporal range; and (c) when it emphasizes a predicate more generally, it indicates a two-valued judgment and selects one of the branches. At last, this paper also explores the conjunction of *eva* with the discourse particle *tu*, asserting that *eva tu* and *tu eva* convey distinct nuances. In the former, the discourse connection is on the component marked by *tu*, whereas in the latter, it is on the group of candidates indicated by *eva*.

Additionally, I believe this work represents the first step – albeit humble and not yet fully developed – toward showing that employing cognitive linguistics to study words typically considered difficult to define is a promising research direction.

Acknowledgment

I sincerely thank Prof. Paolo Visigalli from Shanghai Normal University for his thorough review and detailed comments on this paper's first and second drafts.

Abbreviations

ABL	ablative
ACC	accusative
DAT	dative
DEM	demonstrative
F	feminine
FUT	future tense
FUT.P.PT	future passive participle
GEN	genitive
GER	gerund
IMP	imperative
IMPF	imperfect tense
IND	indeclinable
INS	instrumental
LOC	locative
M	masculine
MID	middle voice
N	neuter
NOM	nominative
P.PT	past participle
PF	perfect tense
PL	plural
PRE	present tense
PRE.PT	present participle
PRO	pronoun
RL	relative
SG	singular
VOC	vocative
1	first person
2	second person
3	third person

Bibliography

Research papers and monographs

- Brereton, J.P. (1982). "The Nonfinite Verb *Iva* in Vedic Prose". *Journal of the American Oriental Society*, 102(3), 443-50.
Available at <https://doi.org/10.2307/602296>
- Brereton, J.P. (1988). "Unsounded Speech: Problems in the Interpretation of BU(M)1.5.10 = BU(K)1.5.3". *Indo-Iranian Journal*, 31(1), 1-10.
Available at <https://doi.org/10.1163/000000088791615339>
- Coenen, P. (2024). *The Functions of Vedic cid, id, im, ī and simk* [PhD Dissertation]. Köln: Universität zu Köln.
- Ganeri, J. (2011). *The Lost Age of Reason: Philosophy in Early Modern India 1450-1700*. New York: Oxford University Press.
- Gonda, J. (1968). "The Sanskrit Nonfinite Verb *Api*". *Lingua*, 21, 183-96.
Available at [https://doi.org/10.1016/0024-3841\(68\)90047-8](https://doi.org/10.1016/0024-3841(68)90047-8)
- Halliday, M.A.K.; Matthiessen, C.M.I.M. (2013). *Halliday's Introduction to Functional Grammar*. London: Routledge.
Available at <https://doi.org/10.4324/9780203431269>
- Hengeveld, K.; Mackenzie, J.L. (2009). *Functional Discourse Grammar*. New York: Oxford University Press.
Available at <https://doi.org/10.1093/oxfordhb/9780199544004.013.0015>
- Langacker, R.W. (1987). *Foundations of Cognitive Grammar*, vol. 1. Stanford: Stanford University Press.
- Langacker, R.W. (1991). *Foundations of Cognitive Grammar*, vol. 2. Stanford: Stanford University Press.
- Langacker, R.W. (2008). *Cognitive Grammar a Basic Introduction*. New York: Oxford University Press.
- Laserson, P. (1999). "Pragmatic Halos". *Language*, 75(3), 522-51.
Available at <https://doi.org/10.2307/417059>
- Lee, D. (1987). "The Semantics of Just". *Journal of Pragmatics*, 11(3), 377-98.
Available at [https://doi.org/10.1016/0378-2166\(87\)90138-x](https://doi.org/10.1016/0378-2166(87)90138-x)
- Lee, D. (1991). "Categories in the Description of Just". *Lingua*, 83(1), 43-66.
Available at [https://doi.org/10.1016/0024-3841\(91\)90051-6](https://doi.org/10.1016/0024-3841(91)90051-6)
- Lühr, R. (2016). "Stressed and Unstressed Nonfinite Verbs in Old Indic". *Discourse nonfinite verbs*, 281-303.
Available at <https://doi.org/10.1515/9783110497151-011>
- Zhou Ranfei 周然飛 (2019). *Gendai nihongo toritatejyoshi 'koso' saikou* 現代日本語と「こそ」再考 (Rediscovery on Modern Japanese Focus Particle 'koso'). *Papers in Linguistic Science*, 23, 51-63.
Available at <http://hdl.handle.net/10097/00127172>

Sanskrit texts and translations

- Bühler, G. (2006). *The Laws of Manu*. New Delhi: Motilal Banarsidass.
- Eggeling, J. (1963). *Vardhamāna's Ganaratnamahodadhi*. New Delhi: Motilal Banarsidass.
- Huang Baosheng 黄宝生 (transl.) (2017). *Luohu Shixi* 罗怙世系 (Raghuvamśa). Beijing: China Social Sciences Press.
- Jhalakikar, V. (1907). *Amarakośa, with the Commentary of Maheśvara*. Bombay: Nirgaya-sagar Press.

- Kāle, M.R. (1917). *Kālidāsa's Kumārasambhava Cantos I-VII*. Bombay: The Standard Publishing Company.
- Kāle, M.R. (1999). *Pañcatantra of Viṣṇuśarma; Edited with a Short Sanskrit Commentary*. New Delhi: Motilal Banarsidass.
- Krishnacharya, T.R.; Vyasaacharya, T.R. (1911). *Srimad Vālmiki Rāmāyana a Critical Edition*. 7 vols. Bombay: Nirgaya-sagar Press.
- Mādhavānanda, S. (1950). *The Bṛhadāraṇyaka Upaniṣad with the Commentary of Śaṅkarācārya*. Mayavati, Almora, Himalayas: Advaita Ashrama.
- Manusukharāya Mora मनसुखराय मोर (1967). *Manusmṛtiḥ Pūrvārddham* मनुस्मृतिः पूर्वार्द्धम् (Manusmṛti, the First Half). Howrah: Udayācala Press.
- Mahāprabhulāla Gosvāmī महाप्रभुलाल गोस्वामी (1988). *Mahākāvī Śrīmadvālmīkīpraṇītaḥ Yogavāsiṣṭhaḥ Hindībhāṣānuvādena Sahitaḥ Prathamō Bhāgaḥ* महाकावि श्रीमद्वाल्मीकिप्रणीतः योगवासिष्ठः हिन्दीभाषानुवादेन सहितः प्रथमो भागः (Yogavāsiṣṭha of Mahākāvī Vālmīki with Hindi Translation Kusumalāta, vol. 1). Varanasi: Tara Book Agency.
- Nikhilānanda, S. (1949). *The Māndūkyaopaniṣad with Gauḍapāda's Kārikā and Śankara's Commentary*. Bangalore: Bangalore Press.
- Pratap, S. (ed.) (2008). *The Kālikāpurāṇa*. Transl. by B.N. Shastri. Jawahar Nagar: Nag Publishers.
- Shastri, G. (1999). *Patañjali's Vyākaraṇa Mahābhāṣya Vth Adhyāya*. Delhi: Pratibha Prakashan.
- Shastri, S.D. (1968). *Pramāṇavārttika of Acharya Dharmakīrti*. Varanasi: Bauddha Bharati.
- Śivadatta, P. (ed.) (1912). *Śrīharsha's Naishadhīyacharita With the commentary (Naishadhīyaprakāśa)*. Bombay: Nirṇaya Sāgar Press.
- Śrīkṛṣṇadāsātmaajhemarājagupta श्रीकृष्णदासात्मजसेमराजगुप्त (1951). *Raghuvamśamahākāvyaṃ* रघुवंशमहाकाव्यम् (Raghuvamśa the Epic). Bombay: Śrīveṅkaṭeśvara Press.
- Śrīmad Vālmiki-Rāmāyaṇa (With Sanskrit Text and English Translation) (1976). Gorakhpur: Gita Press.
- Śrī Nārāyaṇa Rāma Ācārya 'Kāvyatīrtha' श्री नारायण राम आचार्य 'काव्यतीर्थ' (1951). *Nārāyaṇapaṇḍitasaṅgrhītaḥ Hitopadeśaḥ* नारायणपण्डितसङ्गृहीतः हितोपदेशः (Hitopadeśa, Compiled by Nārāyaṇa Pandita) Delhi: Chaukhamba Sanskrit Pratishthan.
- Śrīpāda Dāmodara Sātavalekara श्रीपाद दामोदर सातवलेकर (1968). *Mahābhārata Ādiparva* महाभारत आदिपर्व (Mahābhārata, The Book of the Beginning). Gujarat: Swadhyay Mandal.
- Tagare, G.V. (1992). *The Skanda-Purāṇa*. New Delhi: Motilal Banarsidass.
- Vasu, S.C. (1898). *The Ashtādhyāyī of Pāṇini Book VIII*. Benares: Panini Office.
- Vasu, S.C. (1907). *The Siddhānta Kaumudī of Bhaṭṭoji Dīkṣita*. 3 vols. New Delhi: Motilal Banarsidass.
- Vidyasagara, J. (1873). *The Chandogya Upaniṣad of the Samaveda*. Calcutta: Sucharoo Press.

Reviews

Alyssa Ayres, *Speaking Like a State. Language and Nationalism in Pakistan*

Marta Varini

Università Ca' Foscari Venezia, Italia

Review of Ayres, A. (2009). *Speaking Like a State. Language and Nationalism in Pakistan*. New York: Cambridge University Press, 217 pp.

Summary 1 Introduction. – 2 Urdu and the Nation. – 3 Why Can't a Nation be Multi-Ethnic and Multilingual? – 4 The Case of Punjab. – 5 Imagining Pakistan. – 6 Speaking Like a State: The Indonesian Case. – 7 Conclusion.

1 Introduction

With the monograph *Speaking Like a State. Language and Nationalism in Pakistan*, Alyssa Ayres examines the “logic of the nation” in which culture and language are the source of political power through the complex case of Pakistan. The fascinating volume challenges – through several different and complementary lenses – the idea that in the world of modern nation-states, a distinct “national language” is necessary to claim the nationhood itself. Through the paradigmatic case represented by Pakistan, the book places language as a central object of investigation in order to explore various themes that interest the theoretical studies of linguistic anthropology. As noted by the author, in the modern state of Pakistan, the formation of national (and linguistic) ideology goes hand in hand with the religious-based foundation. Exploring how different leaders who ruled through the newly born nation chose Urdu as the symbolic national language of its great – Islamic – cultural and religious past, Ayres suggests right



Submitted 2025-04-29
Published 2025-05-05

Open access

© 2025 Varini | © 4.0



Citation Varini, M. (2025). Review of *Speaking Like a State. Language and Nationalism in Pakistan* by A. Ayres. *Bhasha*, 4(1), [1-15] 157-172.

from the start that the emergence of Urdu was “neither obvious nor natural” (2009, 16). As we shall see, the volume presents how the state of Pakistan has dealt with the challenges of creating national consciousness and cohesion in a multilingual territory: as a postcolonial state who faced the abrupt and dramatic changes – both geographical and cultural – caused by Partition (1947) the book explores the difficulty of implementing cultural and linguistic reforms in such a multifaceted landscape. In this sense, Ayres focuses on the processes of linguistic-cultural (and religious) homogenisation underpinned by different political reforms and different propaganda tools (such as radio, literature and cinema) that in many cases led to instances of language revivalism, violent protests “in the name of language”, and contrasting results in the structuring and deconstruction of group identity boundaries in Pakistan. Indeed, the “Two-Nations Theory”¹ underlying the national ideology, together with the assumption – and imposition – that Urdu should by right be the national language of the territories reorganised into the modern form of the nation led to – often of a violent nature – conflicts: the case of the prejudices against the Bengali language – which participated in the reasons for the splitting with today’s Bangladesh (1971), and the case of the *Punjabiya*t movement and the ‘marginal’ conflicts against Sindhi, Siraiki, Pashto and Balochi – which participated in the fractures in the tissue of the national narrative – will be addressed in more detail in the following paragraphs. Ultimately Ayres offers an interesting insight by comparing the similarities and differences between Pakistan’s experience, the complex case of India and the revealing case of Indonesia, as “the most successful national language project perhaps in human history” (Ayres 2009, 172).

The research developed in *Speaking Like a State* is indeed of immense value and interest to scholars in the fields of history, political theory, South Asian studies and history of culture and nationalism.

The book I am reviewing opens with an introduction and is divided into ten chapters of which the last is the conclusion. Furthermore, this review is organised as an *excursus* of Ayres’ book in order to better understand the contents of the individual chapters. Since Ayres’ monograph was published a few years ago, although it is a comprehensive and relevant text, I have included some suggested reading of more up-to-date recently published articles. In the conclusions, however, having clarified the key points of the monograph, I will briefly offer a personal evaluation of the reading of the work in its entirety.

1 The ‘Two-Nations Theory’, an ideology of religious nationalism that claimed Muslim Indian nationhood, will be explored in more detail in the next paragraph.

2 Urdu and the Nation

Aiming to trace the roots of the ‘national language’ idea for Pakistan, Ayres introduces with the first two chapters the historical, linguistic and political itinerary that characterised the “Evolution of Partition Idea” (Ayres 2009, 134) and finally the articulation of the independent nation. The assumption that Urdu was the “natural” language for Pakistan – to “protect” – (17-18) arose from two interconnected factors: the geographical location of the Muslim League’s primary support, and the “Hindi-Urdu dispute”.

The All-India Muslim League (AIML), which until 1946 was concentrated in areas of northern India (where the Muslim community was a minority) was the political party that advanced the ‘Two-Nations Theory’, a fundamental ideology that reinforced the concepts of cultural and ethnic difference defined on a religious and civil basis, supporting that “Hindus and Muslims belonged to two separate nations which could never satisfactorily live side-by-side” (Ayres 2009, 24). Reasoning about the progressive and complex differentiation between Hindu and Muslim communities, the author emphasises the colonial responsibility in the creation of linguistic and cultural barriers, resulting in the Hindu-Urdu controversy. To this regard, in the first Chapter, Ayres addresses the issue of conflicts over culture and religion, firstly introducing the role of the East India Company – and for instance the Fort William College institution – in the definition of linguistic boundaries and the undermining of the cultural economy connection between Hindi and Urdu.

As pointed out, the concept of a new Muslim state, as supported by Muslim nationalist poet Mohammad Iqbal (1877-1938),² takes its name from student Choudhary Rahmat Ali’s pamphlet “Now or Never” (1933): the adoption of the name Pakistan (an acronym resulting from the amalgamation of Muslim-majority provinces)³ reinforced religious ideology and cultural distinctiveness in order to provide legitimacy nation existence. As suggested: “The nation along with its ideology was already there for centuries but the country came into existence afterwards. Hence Pakistan’s geography is a result of

2 Mohamman Iqbal, poet, politician and academic, is considered the spiritual father of the Pakistani nation, as well as the ideologue and forerunner of the idea of Pakistan identified in the Northwest Indian areas as the final destiny of the Muslim community. In 1930, he gave an important speech to the All-India Muslim League about the concept of the Muslim state, the *millat*. This address – delivered in English – can be read in its entirety: see Pirzada 1970, 153-70.

3 The pamphlets “Now or Never”, published by the young Cambridge University student Choudhary Rahmat Ali, were released periodically until the end of 1946. In these handouts he brought together the names of the provinces for the Muslim community, extrapolating the acronym ‘Pakistan’ from Punjab, Afghanistan, Kashmir, Sindh e BaluchisTAN.

its ideology” (Ayres 2009, 134). Moreover, the word ‘Pakistan’ refers to the ‘land’ (-*stān*) of *pāk*, where the latter adjective denotes purity, noble virtue, or even a sacred life. Yet, both *pāk* and -*stān* morphemes have Sanskrit origins: the fact that the term chosen by Pakistan’s foundational ideology associated to the idea of purification and discernment with the Indian-Hindu counterpart was instead linked to a common linguistic origin appears as a case of etymological irony as well as an element in disharmony with the ‘Two-Nations Theory’, which generates opposition e competition.⁴

Tracing the process of differentiation between Hindi and Urdu, where the British presence has played a fundamental role, Ayres investigates the role of two other languages interfering with the linguistic registers of power, namely English and Arabic.

In the first case, English was indeed a language of prestige: with independence from the British, Pakistan also inherited the institutions of administration and education, structuring an élite of English-educated citizens. Moreover, the national language implementation projects themselves (such as the first language laws, state education reforms and media campaigns) were conveyed or transmitted in English: this symbolic dilemma is represented by many speeches made by the leaders of the new nation were delivered in English, such as the paradigmatic case of Mohammad Ali Jinnah’s notorious public speech stating that “Urdu is the language of Pakistan and no other”. Despite the efforts made by the nation’s various rulers in establishing institutions to propagate Urdu education, English has continued to remain in the linguistic orbit as a symbol of modernity and progress⁵ (i.e. the most recent during the 1990s, is the revolt operated by Aga Khan Rural Support Program in the Northern Pakistan that demanded English as the medium of education).

In the case of Arabic, being the language of the Quran, the prestige of this medium shifts to religious identity formation, with the establishment of *madrassa*, private Islamic schools and the “Iqra Centres”: General Muhammad Ayub Khan (1907-1974) was the first Pakistani leader to support the institution of Arabic language teaching as part of national planning (as a secondary language along with English); later, General Zia ul Haq (1924-1988) promoted a new system of *Iqra Centres*, namely a mass literacy and education campaign that continues to this day.⁶

⁴ Although the Urdu alphabet (*Nasta’liq*, derived from Persian) differs from that of Hindi (the *Devanagari*, Sanskrit script) the grammar of these two languages could be described as almost identical.

⁵ For a more updated reflection on the role of English in Pakistan see Aftab, Willoughby 2023, 87-104.

⁶ For a more recent evaluation of the education system see Tamim 2014, 280-99.

Finally, the second Chapter concludes with the violent case of Bengal, as a paradigm of the concept of “unbridgeable boundaries of difference” (Ayres 2009, 46) in the broader reflection on the politics of culture – in which language functions as cause, solution and ‘muse’ for self-nation – and the problematic need for cultural homogeneity pursued by nationalist agendas. The sub-national conflict related to Bengali is part of the motivation that led to the secession of East Pakistan in 1971. Despite the fact that the first census of the new nation in 1951 revealed that 56% of all Pakistanis claimed Bengali as their first language (meanwhile only the 3% claimed Urdu), the controversy grew over the *Indicness* of the Bengali language. Since it is a language that is not written in an alphabet derived from the Arabic-Persian script and has a vocabulary mainly derived from Sanskrit, the Bengali language became a place of prejudice because it was not representative of the Islamic religious culture that underpinned the founding of Pakistan, unlike Urdu which was considered “primordially intertwined with Muslim consciousness” (Ayres 2009, 44).

3 Why Can’t a Nation be Multi-Ethnic and Multilingual?

As anticipated in paragraph 2, with the politics of culture implemented, hence the decisions taken regarding the unique role of the Urdu language for the nation-state, Pakistan immediately faced conflicts (including violent ones) over the legitimacy of the different languages of power that were present in the new national geography. In the third Chapter of the book “The Nations and Its Margins”, Ayres investigates in more detail the problems related to a multi-ethnic and multilingual landscape such as that represented by the modern Pakistani borders, and thus the dichotomising process that rendered linguistic, literary and cultural traditions – other than Urdu – as ‘regional’ (instead of national), making Sindhi, Siraiki Pashto and Balochi ‘marginalised’. In this Chapter, it emerges how the nation was “insufficiently imagined” (Ayres 2009, 33), or rather it becomes more apparent the idea that Urdu could naturally (or primordially) be the emblematic language of the Pakistani Muslim nation was hardly supportable and thus the cause of revolts and internal conflicts. Similarly, to the case of the Bengali language, perceived as non-Islamic and therefore unsuitable, the province of Sindh has been the theatre of violent clashes related to the discriminatory processes of language and local identity. Sindhi, like Bengali, enjoyed regional eminence during the British era (1843-1947), as it held administrative authority and boasted a long literary and poetic tradition. Due to the territorial and demographic upheavals of Partition, the cities of Karachi and Hyderabad (mostly

populated by Hindus)⁷ experienced a huge flow of migrations that profoundly changed their urban tissue: the city centres were claimed by the Urdu-speaking *mohājir* ('immigrant' or 'settler'), in a context where the channel for expression had almost always – or predominantly – been Sindhi. In this framework that divided the population in half (since Sindhi and Urdu were largely not mutually intelligible) let to the development and institutionalization of a Sindhi “national consciousness”: at the University of Sindh (Hyderabad), which declared Sindhi as the official language of administration in August 1970, nationalism gained popularity and saw the emergence of groups such as Sindhi Adabi Sangat and the MQM group.⁸ In contrast, the Urdu press denounced Sindhi supporters as “leftists, anti-Islamic [...] anti-Pakistan dissidents” (Ayres 2009, 53), subsequently causing extreme violence and mass killings.

The Chapter proceeds with minor examples of conflicts at the margins of national territories: firstly throughout the case of both the *mohājir* and the Siraki (a language that referred to ‘Multani’ or ‘Bhawalpuri’ dialects of southern Punjab) Ayres discusses of a reverse Herderian process, which saw the people adapt to the language of a nation, while minor languages became part of new ethnic categories. Secondly Pashto, the primary language spoken in the Northwest Frontier Province, on both sides of the Durrand Line, as a language suspected of channelling irredentism along the Pakistan-Afghanistan border; finally the case of uprising in the large territory of Baluchistan, considering the role of language and cultural politics in its strong nationalist movement.⁹

4 The Case of Punjab

The *excursus* on language movements considered so far is extended in Chapters 4 and 5 by examining the case of Punjab, and respectively the Punjabinat ‘ethnonationalist’ movement between its elite and popular forms, seeking “rediscovery of the rational basis of the national identity” since “to be a Punjabi is to be as much a Pakistani

⁷ Hindus in fact comprised 64% of the population of Sindh prior to Partition: their migration to India left the cities of Karachi and Hyderabad carved in half, since in 1951 made up 57.55% of the population in Karachi and 66.08% in Hyderabad. See “Sindh, Sindhi, and the Emergence of the Category *Mohajir*” in Chapter 3 (pp. 48-55).

⁸ The Sindhi Adabi Sangat (Sindhi Literary Society) participated in the efforts to declare Sindhi as an official language for Sindh; the MQM group (“Muttahida Qaumi Movement”), founded by the student leader Altaf Hussain (1953-) was a nationalist group that stressed the centrality of Urdu-speaking *mohajirs* as “a mobilizing construct”.

⁹ For a more recent reflection on the nationalist movement in Baluchistan see Shakoor 2016.

as Punjab is an integral part of Pakistan”.¹⁰ The author highlights to this specific movement, localised in Lahore – the heart of Pakistan’s ‘hegemonic’ region – as a model of nationalism that draws on the theorisations of social communication as well as on Benedict Anderson and Ernest Gellner’s texts¹¹ focusing on the terrain of language in the articulation of the modern state.

As a starting point, Ayres presents as emblematic the banning of works in the Punjabi language by Fakhar Zaman (1943), who was accused of obscenity in 1978. F. Zaman, in order to dismiss the charges, went to the Lahore High Court in response to the obscurantist and oppressive petition against the historical freedom of expression and the development of the Punjabi literary language. Punjab is the most populous province of Pakistan, and its demographical and geographical size allowed it to dominate Pakistan’s institutions,¹² providing the widespread of resentment about its cultural legitimacy and symbolic capital – that takes shape in the idea of “Punjabistan”. Following the suggestion that national languages “were very far from simply choosing themselves as the natural expression of majority usage” (Eley, Suny 1996, 7) a census¹³ shows how Urdu represented a very small percentage of the overall population of Pakistan. Becoming a prestige language for a minority elite group, Urdu (supplemented by English) “marginalised Punjabi”, despite it being “the first language of the majority of the country’s population” (Ayres 2009, 73). The Punjabinat movement claimed the restoration of a ‘lost’ identity throughout a campaign of cultural revivalism, developing a historical-literary recovery project – supported by various Punjabi writers.¹⁴ This ethno-literary project progressively developed taking on a more openly stated agenda: through the allegorical literature that was published (as in Fakhar Zaman’s *Bewatna*, 1988) Urdu assumes the role of “oppressor” (or even “murderous” and “man-eating”) against Punjabi who performs as a strong, resistant and iconic ‘hero’ that fights for the lost self (Punjab). Ayres outlines

10 Partial text, Article 18 of Writ Petition no. 3603 of 1978 in the Lahore High Court, Fakhar Zaman son of (Retd) Major Muham-mad Zaman resident of 178-C Model Town, Lahore (Petitioner).

11 Namely: Anderson 1991; Gellner 1983.

12 Indeed, Punjab – as the land of five (*panj*) rivers (*āb*) – emerges as one of the most productive and agriculturally as well as technologically advantaged territories, where the education-literacy dimension is well developed in a demographically predominant scenario compromising the 55.6% of the population of Pakistan (according to the 1998 census). See paragraph 4.

13 The Census of Pakistan (Government of Pakistan, 1951) shows that Urdu was the first language of a very small percentage of the population of Pakistan overall (3.3% in 1951), yet rising to 7.53% in the 1998.

14 For instance, among the most prominent authors: Najm Hosain Syed (1936-), Munnoo Bhai (1933-2018), and Shafqat Tanveer Mirza (1932-2012).

the main authors and their publications (folk songs, poetic forms and modern narrative) that participated in the project to recover historical heroes and the marginal national memory of Punjab and expands the analysis of the establishing the notion of Punjabi heroism in the context of popular culture.

This turn to the 'counter public' in the Punjabi language is particularly interesting because it takes into account the different tools that the Punjabiyyat movement employed to 're-fashion' the so-called lost self: in these two Chapters, Ayres moves from print to 'non-print capitalism' by analysing the medium of cinema and claiming how

the Punjabiyyat movement places a great deal of emphasis on aesthetic rather than economic-instrumental considerations, suggesting that we ought to take seriously the idea that culture or aesthetic values may not be epiphenomenal developments. (Ayres 2009, 14)

A key example of the cultural and linguistic revivalism in the popular culture is represented by the celebrated and hugely successful film *Maula Jāt* (1979), in which a peasant-warrior seeking for revenge is the protagonist of a violent and symbolic story. This intriguing case not only shows an example of visual and literal vernacularization, but the importance of symbolic capital in the formation of a national subjectivity. To close, recalling Gramsci, literature (as well as visual culture mediums) plays a key role in crafting and legitimizing the history that conveys linguistic, cultural and ideological unity in the nation. The case of Punjab, as observed, suggests that in the creation of a national present is necessary to give attention to the 'margins' and their local memory.

5 Imagining Pakistan

As examined in Chapters 6 and 7, even if Pakistan was created and naturalized as the expression of the nation, it lacked it-self of national culture consciousness: the idea of the necessity of a strong national culture led to the astonishing process of reconfiguration, re-fashioning and re-situation of the history of Pakistan, i.e. the legitimization of the nation-state through an antique past and the primacy of territory and locality. As inquired by Dr. Jamil Jalibi (an esteemed scholar of Urdu literary history) in *Pākistānī Kalcar* (1964): "What is our past, and what is our relationship to it? Are we the logical result of the past's historical flow?". In this sense, Ayres examines Pakistan's historiography using as a starting point of reference Choudhary Rahmat Ali's pamphlet *Pakistan: The Fatherland of the Pak Nation* (1935) in which he depicts through a series of maps the

millennia-long political and geographical history that would define the primordial existence that locates present-day Pakistan [fig. 1].

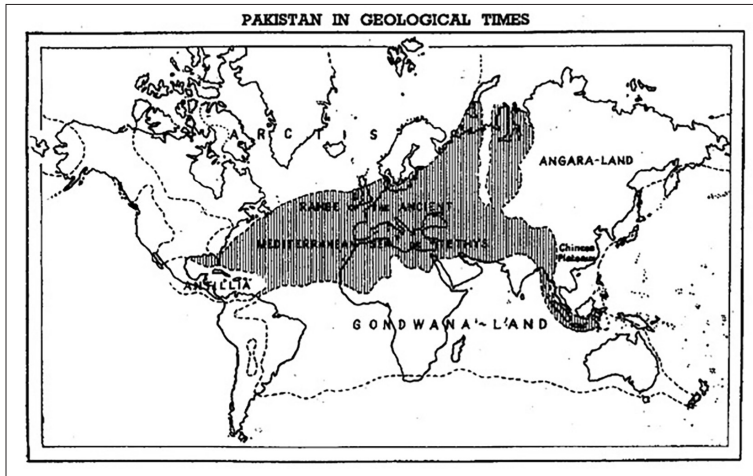


Figure 1 Rahmat Ali's Pakistan in geological times (Ayres 2009, 197)

The visual representation results in an anachronistic attempt to imagine the borders of contemporary Pakistan as a result of the exaggerated historical narrative – a constructed primordialism – placing its past in contrast with the Hindu majority.¹⁵ Moreover, the “Cultural Zones Scheme” – the origins of what is termed “Muslim separatism” – as mapped by Ali Rahman, focussed on the restricted interpretation of the “Islamic” dimensions of the nation, leaving the “margins” contribution to the nation as “provincial”, limited and with no merit (considering, for instance, Punjabi or Sindhi history). As distinctly framed by Gramsci (1985, 256-7):

History was political propaganda, it aimed to create national unity – that is, the nation – from the outside and against tradition, by basing itself on literature. It was a *wish*, not a must based on already existing conditions.

The confusing national epistemology produced a form of historical revisionism entered the educational texts, which “dramatically

¹⁵ As the Muslims community was not equally distributed throughout the territory and, prior to Partition, comprised approximately 20% of the empire's total population. See Chapter 6 and throughout the figures from 9 to 23.

affected the relationship between state education and religion”,¹⁶ as reported by Ayres:

a massive distribution of radio and television sets will be undertaken [...] separate radio and television channels will be established for broadcasting educational programmes to schools and adult literacy centres. On these channels, substantial time will be allocated to the recitation and translation of the Holy Quran so as to saturate the air with the message of God and further forge the bond of national cohesion among the Muslims in different parts of the country. (Ayres 2009, 131)¹⁷

As a continuation on the discourse of the homogenisation project that aimed to form a coherent national culture, the author presents in Chapter 7 a list of works that illustrated the emergence of historical and ideological revisionism.¹⁸ Finally, she offers an interesting insight on the concepts of ‘recognition’ of the nation in conflict with the new historical-geographical concept of the Indus person and region. Based on the important work *The Indus Saga and the Making of Pakistan* (1996) by Aitzaz Ahsan, the geographical significance refers in particular to the line of the “Gurdaspur-Kathiawad Salient”, which demarcates the separation between the Indus region and which would have always had the characteristics of the nation – primordially related to that of Pakistan – and therefore distinct from the Indian area. In this sense, Indus peoples acquire the right to national claim even though they are representative of a continuum that incorporates religious, cultural and linguistic diversity. The author therefore wonders, since

the idea of, the struggle for, and the actualization of Pakistan was manifestly impossible without the participation of Punjabis, Sindhis, and Sarhadis [...] how can the state present a national history that is truly *national* without including the contributions and sacrifices made by people of and in the regions that comprise Pakistan? (Ayres 2009, 146)

To conclude, the profound preoccupation on the part of state political planners with the construction of a superordinate national identity has

16 As suggested by Ayesha Jalal’s essay *Conjuring Pakistan: History as Official Imagining* (1995) “officially approved textbooks display an exasperating degree of confusion as to when and where to begin cataloguing Pakistani history” (Ayres 2009, 125).

17 The quotation obtained from Ayres’ monograph refers to the following text: Government of Pakistan, *The Education Policy*, 1972-80, 29.

18 A complete list opens Chapter 7, where the author considered works related to the emerging historical revisionism (pp. 138-9).

profoundly conditioned the perception of the regional and local areas that now make up present-day Pakistan, conveying (through different communication tools) an imagined, ethnically heterogeneous national past in which “Urdu culture thus became the synecdoche for the larger imaginative leap of anchoring the entire historical narrative” (148).

6 Speaking Like a State: The Indonesian Case

Having discussed language policy, the role of literature and foundational narrative behind independent Pakistan, with Chapters 8 and 9 Ayres explores the potential and actual results of language policy decisions in a comparative study concerning India and Indonesia. More precisely these chapters argue how language policies in South Asia (where various languages coexist) launched nationalist enthusiasm, exploring how the success of a national language “in countries seeking to forge one is intimately related to the symbolic ideologies with which it is invested” (Ayres 2009, 152).

Always bearing in mind how the colonial past affected the foundation of modern states in South Asia,¹⁹ the comparison with India specifically stresses its greater linguistic pluralism. Even in the Indian case, the broad multilingual landscape immediately worried the leaders dedicated to the implementation of cultural and linguistic policies in independent India: although Hindi emerged as one of the most widely spoken languages in the territory, it could never boast more than 40% diffusion among the population, being flanked first and foremost by English and at least eighteen other modern languages spoken by millions of people and with strong literary traditions.²⁰ More precisely, Hindi (in the *Devanagari* script) became an official language on the 14 September 1949, while English (associated with Hindi as the ‘official’ language) was to remain in support of Hindi as an official language in a ‘phase-out’ plan that would meanwhile allow Hindi to develop and take root properly among Indian citizens (over a stipulated total period of fifteen years). Despite the original intent of the Official Language Commission, the

19 To this end, the author reflects on the effects of the wave of decolonisation in the realisation of nation-states in the twentieth century. She then suggests that the new post-colonial epistemologies share the idea of “one language for all” as a necessary condition, along with theories of political development, modernisation, and nation-building held together by language planning.

20 Indeed, the Eighth Schedule of the Indian Constitution establishes – in several mandates from 1949 to 1992 – the legal affiliation of the following languages to the category of “national languages”: Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Malayalam, Marathi, Oriya, Punjabi, Sanskrit, Tamil, Telugu, Urdu, Sindhi (1967), then Konkani, Manipuri and Nepali (1992).

English phase-out arrangements never took place, leaving English with the great prestige of an administrative language and triggering violent protests vehemently objecting the implementation of Hindi as the hegemonic sole official language. The “Anti-Hindi Agitation” in Madras, Tamil Nadu (1949-1964), represents the most violent case of resistance and – sadly – a common thread with Pakistan’s experience: the political class formed in English (in the above-mentioned period) was put at risk by the imposition of Hindi on Tamil speakers.²¹ After the major States reorganization in 1956, a solution was found in the concept of “Unity-in-Diversity”, yet recognizing the new categories of minority and majority languages.

To conclude, as seen from Pakistan’s multi-layered multilingual experience, also in the Indian case literature and mass cultural communication tools have been fundamental in the design of the modern state: for instance, the employment of radio programs such as the successful All-India Radio (for the diffusion of a codified Hindustani vocabulary) and the foundation of Sahitya Akademi (1954) a sort of national literary recognition bureau committed to the “conscious effort to establish a national sensibility of unity-in-diversity through literature” (Ayres 2009, 167).

Moving to the second comparative case study, Ayres introduces the unique case of Indonesia, investigating the intriguing challenges faced in the promotion of a new language, Bahasa Indonesia, – while respecting regional languages and cultures – that literally become “the language of a nation and thus the national language” (Ayres 2009, 182). The author firstly illustrates the similarities between Pakistan and Indonesia’s challenging histories: both modern-states share a past of colonization and have been ruled by highly centralized polities and authoritarian regimes since their independence. Furthermore, both countries are characterized by the presence of a large Muslim community as well as strong ethnolinguistic diversity.²² Yet a key point resulting in the perhaps most successful national language project in human history lays in the explicit aim to create a national language allied with concept of modernization and progress – in discontinuity from the traditional/old. Bahasa Indonesia was in fact literally ‘constructed’ as a unique modern tool of expression, with extensive vocabulary and grammar, that consciously lacked of a remote past. This new lingua franca was

21 Such objections were not limited to Tamil speakers: the states of Bengal, Mysore and Kashmir also had serious objections to the assumption of Hindi as the only official language.

22 To this regard, the author points out a further comparison between the spread of Urdu speakers in Pakistan (no more than 3%) and of Bahasa Indonesia in Indonesia (only 4.9%), meanwhile Punjab constituted 56% of Pakistan and similarly the Javanese 40% of Indonesia.

identified, chosen and sponsored firstly by the Youth Pledge of Indonesia's young anti-colonial nationalists (*Sumpah Pemuda*) that envisioned the unification of a great geographical area into one with a larger sense of cohesion. Although Javanese language, for instance, enjoyed a rich cultural heritage and was widely spoken, the vanguard of the early independence movement of the young nationalists enabled this 'constructed' language to reach 4.9% of diffusion as the first language spoken by the population by the time of independence (1945). The success of this planning, as mentioned, lies in the juxtaposition of the idea of a modern Indonesian state in which the language of unity allowed citizens to participate in the national public sphere, coexisting with the country's other linguistic and cultural traditions in a relationship that did not emphasise a hierarchy among the languages of power. As follows, Indonesian language become a vehicle of expression for the modern nation-state: the briefly introduced engineering project orchestrated involved in the creation of a new (linguistic and ideological) internal vocabulary takes into account the demands of modern life both linguistically (terminologically) and ideologically.

To conclude, as concisely expressed by the author Ayres (2009, 184-5):

Speaking Indonesian never required Indonesian citizens to abjure their cultural history, nor their sense of faith - while speaking Urdu ideologically reminded Bengali, Punjabi, Sindhi, Siraiki, Pashto, and Balochi speakers (among others) that the national language implicitly displaced their cultural and regionally-specific Islamic pasts.

7 Conclusion

As seen through the chapters addressed, Alyssa Ayres' work offers a series of interconnected and interdisciplinary discussions regarding the analysis of the linguistic-cultural logic of the Pakistani nation. The relevance of this volume lies in its incisive ability to present the difficult burden that post-colonial Pakistan had to face - through different historical examples - hence, making evident the different aspects that characterise the dilemma that unfolds between nationhood and choosing single national language. From the very first chapters Ayres addresses a theme that we could call the 'linguistic paradox' by tracing the historical, political and cultural roots that constitute the state project "to forge a Pakistani ethnicity through the cultural heritage of the Urdu language created antipathies where it sought unity" (Ayres 2009, 6). As mentioned, the attempt to reconcile the (Muslim) ideology of the Urdu national language

with the modern nation-state of Pakistan has inevitably provoked regional claims and violent clashes with other linguistic and cultural traditions. The intense process of Urdu legitimisation not only alienates the Bengali language (hence East Pakistan), but also expands to “minority” contexts: this is the case with the process of ethnogenesis involving the Sindhi language and the Mohajir category, the process of dichotomising opposition to the Siraiki movement in southern Punjab and finally the strengthening of nationalist sentiment in Pashtunistan. In “The Nation and Its Margins”, it becomes clear how the cultural policy model adopted in order to reorder a displaced national centre did not really take into account the territorial (hence socio-cultural) dislocation inherent in the Pakistani national project.

Another noteworthy aspect of Ayres’ work is certainly the careful analysis offered on the unusual structural characteristics of the Punjabiyat movement. Through the study of Punjabistan, she not only offers a reflection that differs from the classical concept of nationalism, but also demonstrates the resilience of Punjab culture (both literary and visual) between the finer meshes of Pakistani territory and the relentless desire to restore and protect it. However, I believe that the investigation of local (or marginal) linguistic and cultural processes deserves to be deepened with up-to-date vertical and specific investigations, as suggested by the inclusion of reading recommendations from more recent publications (in the article chapters and bibliography). Regarding the paradox of the ideology of modern Pakistan and the study of processes related to the national language, I believe that Ayres’ monograph offers an excellent understanding of the powerful role of language in shaping nations, thus worthy of being a original textbook. Ayres’s work proves to be a truly valuable contribution for language studies (from regional to global), especially throughout the excellent and intriguing comparative work between the case studies of Pakistan, India and Indonesia.

In conclusion, the author offers a complex linguistic (as well as historical and cultural) account of Pakistan, masterfully disentangling the broad concepts of nationalism, mono- and multilingualism, cultural revivalism, language planning (its models and methods), between regional, national and finally global spheres. For this reason, Pakistan represents the starting point of a reflection much broader than its territorial and ideological boundaries, reaching the interest of scholars interested in gripping readings on linguistic processes for expanding interdisciplinary studies.

Bibliography

- Aftab, Z.; Willoughby, J. (2023). "The Fall of Urdu and the Triumph of English in Pakistan". *The Pakistan Development Review*, 62(1), 87-104.
Available at <https://doi.org/10.30541/v62i1pp.87-104>
- Anderson, A. (1991). *Imagined Communities*. 2nd ed. London; New York: Verso.
- Ayres, A. (2009). *Speaking Like a State. Language and Nationalism in Pakistan*. New York: Cambridge University Press.
Available at <https://doi.org/10.1017/cbo9780511596629>
- Eley, G.; Suny, R.G.; (eds) (1996). *Becoming National*. Oxford: Oxford University Press.
- Gellner, E. (1983). *Nations and Nationalism*. Ithaca: Cornell University Press.
- Gramsci, A. (1985). *Selections from Cultural Writings*. Transl. by W. Boelhower. Ed. by David Forgacs, Geoffrey Nowell-Smith. Cambridge: Harvard University Press.
Available at <https://doi.org/10.2307/2069991>
- Pirzada, S.S. (1970). "All-India Muslim League, Twenty-First Session, Allahabad, December 29-30, 1930". *Foundations of Pakistan*, vol. 2. Karachi; Dacca: National Publishing House, 153-70.
- Qaiser, R. (2015). "Nationalism and Political Identity in Pakistan". *The Johns Hopkins University Press, The SAIS Review of International Affairs*, 35(2), 105-16.
Available at <https://doi.org/10.1353/sais.2015.0027>
- Schiffman, H. (ed.) (2012). *Language Policy and Language Conflict in Afghanistan and Its Neighbors. The Changing Politics of Language Choice*. Leiden: Brill. Brill's Studies in South and Southwest Asian Languages 2.
Available at <https://doi.org/10.1163/9789004217652>
- Shakoor, A.W. (2016). "The Changing Dynamics of the Baloch Nationalist Movement in Pakistan". *University of California Press, Asian Survey*, 56(5), 807-32.
Available at <https://doi.org/10.1525/as.2016.56.5.807>
- Tamim, T. (2014). "The Politics of Languages in Education: Issues of Access, Social Participation and Inequality in the Multilingual Context of Pakistan". *British Educational Research Journal*, 40(2), 280-99.
Available at <https://doi.org/10.1002/berj.3041>

Semestral journal
Department of Asian
and North African Studies



Università
Ca' Foscari
Venezia