ISSN: 0974-3421
INTERDISCIPLINARY JOURNAL OF LINGUISTICS
Volume No: 18, 2025

PATRON
Prof. Nilofer Khan,
Vice-Chancellor, University of Kashmir

ADVISORY BOARD

Anvita Abbi (B.B. Borkar Chair, Goa University) Vaishna Narang (JNU, New Delhi), Nazir Ahmad Dhar (University of Kashmir), Sonal Kulkarni (Deccan College, Pune), Niladri Sekhar Dash (ISI, Kolkata), Pradeep Kumar Das (JNU, New Delhi), M. J Warsi (AMU, Aligarh), Umarani Pappuswamy (CIIL Mysuru) Prasannanshu (National Law University New Delhi), Mufti Mudasir Farooqi (University of Kashmir). Iffat Maqbool (University of Kashmir), Shantanu Ghosh (Amity University Uttar Pradesh)

EDITORIAL BOARD

Musavir Ahmed, Aadil Amin Kak,
Aejaz Mohammed Sheikh, Sajad Hussain Wani,
Sheeba Hassan, Saima Jan, Nazima Mehdi,
Zahid Bashir.

EDITORIAL ASSISTANTS
Sabia Rafiq
Aqib Majeed
Mehreen Amin
Moachiba Jamir
Ifrah Mushtaq

©Head
Department of Linguistics
University of Kashmir, Srinagar
Price: ₹ 600/= $ 6.26

The publisher is not responsible for the views expressed by the authors of the research papers appearing in this publication.

All rights reserved. No part of this publication can be reproduced in any form whatsoever.

IJL VOL (18) 2025 ISSN 0974-3421

INTERDISCIPLINARY JOURNAL OF LINGUISTICS

(A Peer-Reviewed Refereed Journal)

Editor

Musavir Ahmed

Department of Linguistics

University of Kashmir

Srinagar, J&K, India-190006

Contents

	Page No.
Musavir Ahmed Editor's Note	**i-vii**
1.Anita Pandey and Cynthia Chasteen Virtually Left behind during the COVID-19 School Shutdown: A Study of U.S. Based English Language Learners' Learning Loss	**1-44**
2. Kimiaki Tokumaru Syllables and Reflex Circuit: The Missing Link between Animal Calls and Digital Language	**45-68**
3. Niladri Sekhar Dash Issues in Developing a Comparable Corpus for Indian Languages as a Key Linguistic Resource for Cross-Linguistic Research and Applications	**69-82**
4. Nazir Ahmad Dhar Outline of Linguistic Studies of Kashmiri Language	**83-90**
5. Konchok Tashi From Ladakhi to Bhoti: Sociolinguistic and Political Dimensions of a Changing Language Identity in Ladakh	**91-108**
6. Jegmet Sangyas and Sonam Joldan Language and its Politics in Indian Trans-Himalaya: Beyond Bhoti Versus Ladakhi	**109-130**
7. Garima Dalal Creating Interactive English Language Learning Using LLM in Higher Education	**131- 138**
8. Suaad Muhammad, Nisar Ahmad Koka and Sheeba Hassan Learning Outcomes of Saudi Students using Artificial Intelligence (AI) for Foreign Language (FL) Proficiency Development	**139-162**
9. Javed Ahmad, Mohsin Razza Khan and Basim Hamdan Ibrahim Kana'an Efficacy of AI-Powered Adaptive Learning Systems in Enhancing Grammar Acquisition for Arab Learners of English: An Empirical Study	**163- 170**
10. Javaid Aziz Bhat Negotiating Intimacy and Distance: Pronominal Variation across Kinship Domains in Kashmiri	**171-184**
11. Siddharth Sanket and Vadthya Kishore Agreement in Khortha	**185-202**

12. Niloy Chakraborty A Sketch of Tense, Aspect and Modality in Bongcher	**203-218**
13. Aqib Majeed and Sabia Rafiq Nominal Inflection and Case Alignment in Bhadarwahi: A Descriptive Study	**219- 228**
14. Vijay Bernard and Arpana Between Languages, beyond Voices: Examining the Languages, beyond Voices: Examining the Role of Code-Switching as Tool for Silencing Women through Appraisal Theory in "A Thousand Splendid Suns	**229- 240**
15. Saima Jan, Bushra and Aabida Zahoor Romantic Tales in Pakhtoon Tradition: A Comparative study of *Adam Khan-Durkhanai* and *Momin Khan-Sherino*	**241-250**
16. Sayantani Banerjee, Moachiba Jamir, Greeshma Joseph and Zahid Bashir Verum and Exceptive Constructions in some Indian Languages	**251-264**

Editor's Note

It is a pleasure to present the 18th volume of *Interdisciplinary Journal of Linguistics* (IJL) published by the Department of Linguistics, University of Kashmir. The volume has 16 papers by researchers from foreign and Indian universities, discussing a wide range of issues. The contributions reflect upon diverse areas including the description and documentation of under-documented and endangered languages, questions of language and identity, language and politics, language and technology, language and gender and the socio-pragmatics of modes of addresses and politeness. The papers in this volume endorse the continuing relevance of linguistics to some of the most demanding concerns of our times including Artificial Intelligence (AI) and the disruptions brought about by COVID-19 pandemic.

The impact of the pandemic on human life was unprecedented. It laid bare the fragilities and vulnerabilities of people, societies and their institutions. People died, societies went still and institutions crumbled. The hospitals, probably the only hope against the disease, were literally on verge of collapse. Businesses closed, jobs disappeared, and populations lost their financial security. The seemingly robust interdependence of people across the globe turned out to be so fragile that factories shut down in one country, shelves went empty in another. Months of isolation and lockdown took a real toll on social structures: social inequalities around income, healthcare, housing, education, and who gets help and who does not, became more pronounced. The people who had the least lost the most. Governments and institutions were left improvising, trying to respond to a crisis never conceived before. Education was no exception to the onslaught of the pandemic. Schooling and schools came to grinding halt and people had to get accustomed to the conception of remote learning, hitherto unknown to many. The shift to online learning-teaching wasn't an easy one as it didn't concern education only but rather it involved a whole lot of things, not accessible to many: internet, digital tools and above all the educational and linguistic competencies of parents and guardians.

The paper by Anita Pandey and Cynthia Chasteen takes up the question of real accessibility to online education during Covid-19 with specific focus on English Language Learner (ELL) families in the US. Using a comparative mixed-methods approach, the authors look closely at how three very different school districts in Nevada, Missouri, and Maryland communicated with ELL parents and guardians once classrooms closed in spring 2020. The research draws on two kinds of evidence: a three-part pilot survey that captures firsthand accounts from ELL students, their families, and the educators and administrators who work with them, and a close reading of the actual communications like emails, website updates, social media posts which these districts sent out between March and May 2020, analysed through a cultural competency framework developed by one of the authors. The study has come up with some stark revelations. Across all three districts, there was no real evidence of a considered, sustained plan to engage ELL

families specifically, no consistent multilingual outreach, and no personalized resources – nothing beyond messaging designed for a general English-speaking audience. The real concern has been about whether ELL students were able to learn during the lockdown, and whether their families felt informed and supported at all. By highlighting parent and student opinions together with the schools' own communications, this paper offers a substantiated and practical look at the lacunas and the shortcomings and what schools, educators, and policymakers might do in times of crises.

The complex nature of human language allows it to have enormous vocabulary and the ability to combine units into new meanings, unlike any other form of communication. Kimiaki Tokumaru in his detailed paper argues that human language began evolving like a digital system during the Late Pleistocene era. With increasing control over tools, hominins formed accommodative bonds with the plants and animals around them. As they learned to control fire and sought refuge in caves during glacial periods, these groups grew increasingly eusocial, using vocal calls to signal threats. According to the author, a pivotal moment came around Toba super volcano eruption roughly 74,000 years ago, when the resulting winter coincided with the emergence of click consonants and the first acts of naming objects in the world. With the passage of time, as the larynx descended in the throat, humans gained the ability to produce vowels, string together syllables, and eventually began constructing complete sentences which as per the author is the real starting point in language's digital evolution. Syllables in this scheme were significant as they acted discrete as well as divisible units and allowed vocabulary expansion and grammatical structure possible. Writing systems followed around 5,000 years ago, and binary code emerged in the mid-1900s, each step transforming spoken sounds into forms that could persist and be manipulated. Perceived in this manner, the authors assert that language's development fits naturally within the Complex Dynamic Systems theory as an instance of digital-style evolution. The author upholds that language is the end product of roughly four billion years of evolution, during which humans were pushed to move beyond basic reflexes and tribal instinct toward a more thoughtful, truth-seeking form of intelligence.

Comparing large scale data of two languages on criteria like subject matter, genre, text type, and time frame without being prejudicial towards any of the languages involved can be significant and beneficial for a multilingual country like India. Niladri Shekhar Dash in his paper has underscored the potential uses of comparable corpora (CC) in the development of language technology in India. The author begins by introducing a functional description of a CC and moves through the different forms such corpora can take, reviews its current status in India and presents the case for why Indian languages need this significant resource. The paper provides a workable framework for developing a CC and concludes by proposing tangible ways such corpora could serve linguistics and language technology in India. These provide the foundation for development of multilingual

resources like bilingual dictionaries, fixed expressions, matching grammatical structures, databases of translation equivalents, and comparative grammar, etc. The paper suggests that CC holds particular promise when built around related language pairs, since shared ancestry, structural similarity, and common writing systems among many Indian languages would make alignment, processing, and analysis considerably more manageable. With deepening impact of AI, large language models, and machine learning worldwide, CC in India could become a significant resource, both for building multilingual assistive technologies and for enabling comparative studies among languages that are structurally, historically, or geographically related. To realize this potential, the author calls for a harmonized, nationwide effort to build these corpora and put them to use across linguistic and technological research, with the aim of filing the gaps that separate India's many languages.

Language has never been an apolitical entity. Language not only holds social and cultural distinctiveness but also has political significance for most societies. The paper by Konchok Tashi and another by Jegmet Sangyas and Sonam Joldan have deliberated on politics over the nomenclature of the language spoken in the Union Territory (UT) of Ladakh. Ladakhi is spoken by commoners in several districts of Ladakh UT and the name reflects the geography, ethnic identity and regional belonging of people inhabiting the former Ladakh kingdom. Both the studies have delved deep into the social, cultural, educational and political consequences of renaming Ladakhi as Bhoti in the Union Territory of Ladakh. For many communities, language labels act as symbols of identity, and legitimacy and potentially shape decisions on lower order evaluations like schooling of children to broader assessments like political mobilization. According to the author, Bourdieu's notion of symbolic capital can help explain the dynamics at play in Ladakh; calling the language Bhoti draws prestige from Tibetan religious and literary tradition, while Ladakhi derives its authority from a sense of place and historical distinctiveness rooted in the region itself. The term Bhoti, also referred to as Bodhi, began receiving attention in cultural and political conversations in Ladakh during 1970s. The word is a derivative of Sanskrit *"Bhot-desha"*, loosely meaning Tibet, and refers to the family of Tibetic languages spoken across regions which have received Tibetan religious and cultural influence. Its adoption by a section of populace in Ladakh is thought to be motivated by the objective of having constitutional recognition for Tibetic languages in the Eighth Schedule. Supporters of Bhoti maintain that the term offers a unifying linguistic identity for several culturally connected communities spread across a wide geographic area and thereby consolidating the case for both cultural protection and government recognition. However, there are opposing voices who view the move aimed at robbing Ladakhi of its local and regional colour as well as reducing the diversity within it. Besides, there are speakers of languages like Purgi and Balti who do not want their languages to be identified under the label Bhoti. The rising

popularity of Bhoti hasn't, however, resolved the issue. While Bhoti could serve as a suitable umbrella term for region's political profile, it overshadows the distinct local status of Ladakhi. The strain and the associated concern between stressing a collective regional identity and protecting something local and distinctive continues to be at the core of the ongoing debate over how language spoken by the people of Ladakh should be named and comprehended.

The penetration of AI in human life has just begun and its potential, without exaggeration, can't be less than revolutionary. In higher education, AI is already being treated as a significant force which has greatly reduced faculty and administrative load and provides personalized adaptive learning support to students. Surveys undertaken in recent years have confirmed that majority of students as well as faculty in several countries now use AI and AI based teaching-learning resources. In the domain of second and foreign language teaching and learning, AI based tools are increasingly utilized to have personalized instruction, provide real-time problem-solving, and get individualized automated feedback. The volume has three papers which have addressed the role of AI in second and foreign language teaching-learning. The paper by Garima Dalal has reflected upon functioning of a Linguistic Empowerment Cell (LEC), established in Jawaharlal Nehru University (JNU), New Delhi in 2011 and how AI driven Large Language Models (LLM) could be utilized to enhance the efficiency of the cell. The cell was established for students who had to shift to English from their previous education received in any of the vernaculars with the aim that such students are not left behind because of little or no proficiency in English language. Garima Dalal has made a case for adoption of AI in the cell for creating customizable listening exercises with multiple accents, simulated conversation practice, adaptive reading materials commensurate to student proficiency, and brainstorming and paraphrasing support for writing development. On a positive side, AI can facilitate learning, however preventing its misuse should be given equal attention and as such a multi-pronged approach needs to be developed for ethical use of AI and preventing its misuse in learning environments. The author concludes with the suggestion that the institutions of higher education in India should develop language centres like LEC in tune with the recommendations of New Education Policy 2020, which emphasizes multilingual education and AI should be introduced in these centres through a well thought-out and workable plan suited to Indian context.

The other two studies have investigated how AI-based tools affect English language learning among Saudi Arabian students, focusing on their impact on vocabulary building, speaking fluency, listening and grammatical skills. The one by Suaad Muhammad Saeed Alqahtani, Nisar Ahmad Koka and Sheeba Hassan has involved a methodological approach involving numerical and descriptive analysis. The findings reveal that AI technologies boost vocabulary retention and listening comprehension. The paper suggests that while AI can meaningfully assist language acquisition, the best results come from combining these

tools with conventional teaching methods. The study concludes that although AI can handle varied linguistic input effectively, human involvement remains essential to successful language teaching. The results align with broader research in the field, endorsing that personalized, tech-driven instruction strengthens particular language skills. The study stresses that since AI cannot replicate the interactive, culturally responsive nature of in-person instruction, as such AI use in language classrooms should be carefully constrained and paired with human teaching and monitoring.

On a similar note, the other study by Javed Ahmad, Mohsin Razza Khan and Basim Hamdan Ibrahim Kana'an has investigated the impact of AI-based adaptive learning platform on grammatical accuracy in comparison to standard classroom teaching. Using a quasi-experimental methodology, researchers worked with 60 intermediate-level Arab ESL students split into two groups; the experimental group used AI platform, while the control received conventional instruction. Both groups took pre- and post-tests covering English tense structures. The study concludes that AI-enabled personalized learning can meaningfully help address the grammar difficulties Arab learners commonly face and recommends using such tools alongside traditional classroom instruction.

Pronominals in Indian language serve a variety of functions, besides their grammatical roles, unlike many European languages. Kashmiri is one such language where pronominals are employed to demonstrate respect, closeness and remoteness while conversing with various kind of kin and this is the subject matter of the paper by Javaid Aziz Bhat. The study is based on data collected from 240 participants of varying ages, genders, education levels, and rural/urban backgrounds. The strength of the study lies in its use of quantitative study of usage patterns together with a qualitative understanding informed by theories of politeness and language in social context. The results indicate a clear division between casual mundane pronominal forms and more formal and courteous ones. Speakers select a pronoun consciously for managing social relationships. The results indicate that blood relatives are addressed with informal pronouns most of the time signalling closeness and emotional familiarity, while in-laws and other marriage-based relations demand more formal address reflecting a greater emphasis on courtesy and social consideration. These patterns also shift depending on who is speaking; younger people and those from rural areas tend to use more casual forms, while speakers who are middle-aged, more educated or based in urban areas prioritize more formal options.

A short paper by Siddharth Sanket and Vadthya Kishore examines grammatical agreement in Kortha specifically focusing on the variety called Ramghariya Kortha. Kortha is a language spoken in the Jharkhand state of India and Ramghariya Kortha is spoken in Ramgarh district of the state. The linguistic census of India classifies Kortha, with more than 8 million speakers, as a dialect of Hindi. Despite such a huge number of speakers, some scholars consider it endangered. One

structural difference of Khortha with Hindi is that the former has been reported to rely on a single agreement system, whereas Hindi operates with a dual one. Secondly, unlike Hindi, Khortha has no grammatical gender system at all. Agreement in the language can surface either on the main verb itself or on auxiliary elements. By portraying the structural dissimilarities between Khortha and Hindi, the authors have presented a case for its status as independent language, rather than being a dialect of Hindi.

The paper by Niloy Chakraborty offers a description of the functioning of tense, aspect, mood, and modality in Bongcher, an under-documented Tibeto-Burman language from the Kuki-Chin branch, spoken by the Bongcher community in Tripura. Drawing on field data, the author has reported that Bongcher has three tenses, present, past, and future, which however often blur in practice, and speakers must rely more on contextual time words (like "yesterday" or "tomorrow") than on verb endings. Among Tibeto-Burman languages, strict tense marking tends to matter less than the real/unreal (realis/irrealis) distinction, and consequently the author argues that aspect, not tense, better explains how Bongcher expresses time. Besides, the author suggests that Bongcher has five aspects, and three core moods of indicative, imperative, and hortative along with a range of modal constructions that function like a subjunctive. The study indicates that Bongcher has a rich system for expressing speaker attitudes and degrees of likelihood. Beyond its linguistic findings, the paper, according to the author, has a documentation function, helping preserve knowledge of an endangered language and contributing new data to comparative and typological studies.

The paper by Aqib Majeed and Sabiya Rafiq has presented grammatical description of nominal inflection and case alignment in Bhadarwahi, a Western Pahari language spoken by around 1,20,000 people in the Bhadarwah and Chinta valleys of Jammu and Kashmir's Doda district. The study finds that the two genders (masculine and feminine) and two number categories (singular and plural) in Bhadarwahi are expressed through vowel changes, suffixes, and in certain cases, by different root words. Feminine forms are typically derived from masculine ones by shifting a final vowel, changing honorific suffixes, or adding occupation-related endings. Pronouns in the language are sensitive not just to person and number but also to social honorificity. Speakers choose different pronoun forms depending on the respect level appropriate to the person being addressed. Bhadarwahi uses a split-ergative system; subjects are marked depending on both the verb's transitivity and its tense/aspect. In past or completed actions, the subjects of transitive verbs get special ergative marking, while subjects of intransitive verbs and objects remain unmarked. Despite extensive borrowing of vocabulary from neighbouring languages like Hindi, Dogri, and Kashmiri due to prolonged contact, the study concludes that Bhadarwahi has a distinct character, mutually unintelligible with these larger regional languages.

The paper by Vijay Bernard and Arpana explains how code-switching across English and Persian, Dari and Pashto in Khaled Hosseini's "A Thousand Splendid Suns" serves as a tool through which male characters express judgment, emotion, and control over women. The analysis uses Appraisal Theory to explain how language expresses attitudes, feelings and evaluations. The paper argues that code-switching in the novel actively participates in silencing and subordinating female characters, particularly Mariam and Laila. The term *"harami"* (illegitimate) is repeatedly used against a female character, branding her as socially deviant and justifying her exclusion. Another character is labelled as *"dehati"* (rural, unsophisticated) girl, invoking class-based contempt. Terms like *"nang – namoos"* (pride and honour) reflect how women's chastity is tied to family honour and moral policing. Exclamations like *"bas, chup ko"* (shut up), and *"wallah o billah"* reinforce male dominance through anger, dismissal, and implied threats of violence. The paper concludes that the consistent pattern of negative evaluation directed at women establishes how code-switching in the novel operates as a linguistic mechanism of patriarchal control.

The paper by Saima Jan, Bushra and Aabida Zahoor presents a comparative study of two significant Pashtoon romantic folktales, *Adam Khan and Durkhanai* and *Momin Khan and Sherino*. It examines their plot development, narrative structures, storytelling techniques, and major themes of love, honor, social obligation, sacrifice, and tragedy. Through comparative textual analysis, the study identifies similarities and differences in the two narratives. It highlights their shared cultural values alongside distinctive narrative strategies, demonstrating the richness, diversity, and continuity of the Pashtoon storytelling tradition. The paper also underscores the linguistic and cultural significance of these oral narratives as repositories of Pashtoon literary tradition, identity, and collective memory.

The paper by Sayantani Banerjee, Moachiba Jamir, Greeshma Joseph and Zahid Bashir Lone, offers a cross-linguistic survey of verum and exceptive marking in five Indian languages belonging to different language families. The authors underline that verum and exceptive marking should be considered from a function-to-form perspective rather than to take for granted that categories developed for English will have a one-to-one correspondence across languages. The discussion brings together data on particles, prosody, emphatic constructions, copular strategies and negative clauses and highlights the great diversity across the languages studied. The comparison is especially helpful when showing how discourse context affects the interpretation of both truth-related and exception-related expressions. The paper has explained a hitherto unexplored area, particularly in Indian languages. The results are preliminary, but they offer a useful empirical baseline for broader comparative work on discourse-semantic marking and the interplay of truth, exclusion, alternatives and context in genetically diverse languages.

Musavir Ahmed
Editor

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 1-44

Virtually Left behind during the COVID-19 School Shutdown: A Study of U.S. Based English Language Learners' Learning Loss

Anita Pandey[*]
Cynthia Chasteen[**]

Abstract

The study investigated the effects of the COVID-19 global pandemic on the linguistic, cognitive, motivational, and self-regulatory behavior of ELLs and their parents. It identified challenges that middle and high school ELLs, in particular, faced with virtual instruction, as well as the cognitive, social, educational and psychological impact of the pandemic and pandemic-based instruction on this group. More specifically, this study examines whether and to what extent three U.S. public school districts that served--and continue to serve--a large number of ELLs attempted to not merely include, but also engage and empower their ESL students and families in the months immediately following the COVID-19 school closures in the spring of 2020. Dual-language theory and praxis (Collier & Thomas, 2004) undergirded the project design and methodology employed. Pandey's (2012) language-building-blocks framework was used to identify comprehensible input, interactional opportunities, and feedback provided in the data examined.

Findings point at unique challenges ELLs and their parents encountered in the school districts under focus in the United States, due in large part to the use of wholly-virtual and primarily written language exchanges between schools and students and their families. Most school administrators and instructors employed solely written communication with their students, including with ELLs resulting in limited and relatively fewer opportunities for ESL students to verbally interact and engage in more meaningful exchanges with their peers and instructional personnel. In addition, these students also received relatively limited or insufficient scaffolding or feedback, further delaying their language and literacy gains through formal schooling. This study highlights the need for verbally rich and linguistically inclusive communication between emergent language learners and schools via the use of overtly comprehensible and participatory language with ELLs and their care providers. On a positive note, teen participants reported non-schooling-based (verbal) language learning gains.

Keywords: Covid-19, ELL, ESL, Inclusive Communication, Building Block Framework, Online learning

[*] Morgan State University, Baltimore; anita.pandey@morgan.edu
[**] Migrant Education and English Language Learning, Missouri; chasteenc@missouri.edu

1. Introduction: Overview and Research Gap

The Institute of Education Sciences, a U.S. Department of Education subdivision shares that "an estimated three million students are missing from school as a result of the COVID-19 pandemic." Research suggests that ELLs constitute the largest group of students negatively impacted by the pandemic (COVID, 2021). Indeed, "learning loss" (Foust, 2021) has been exacerbated for many more students for a protracted period due to the lingering and presumably fairly long-lasting COVID-pandemic. While COVID-19-prompted learning loss has been widely reported, the full scale, breadth, and national impact of this loss of learning is still unclear. Through survey-based interviews, personal experiences of ELLs, parents and teachers are shared, in an attempt to shed light on the extent of this "loss," so that mitigation and accelerated learning strategies could be proposed, implemented, and tracked. As such, "learning loss" is arguably most starkly visible for low-income students, specifically Black and Hispanic students.

Noteworthy and inspiring examples of learning loss and resilience or learning gains (Baume, 2018) during the COVID-19 lockdown from linguistically challenged communities in the United States are also shared in this study. The language learning gains were, it should be noted, not the result of schooling or school-instituted initiatives, as will be discussed in the latter part of this paper.

2. Purpose & Scope

Given the COVID-virus-prompted school closures, we were presented with a perfect opportunity to evaluate our schools' success at engaging parents, including the parents and families of our most vulnerable students, namely, young and low-income ELLs. How so? A focus on school-related communication, that is, communication on school websites and/or district sites, as well as on social media outlets employed by area schools, as well as communication sent to families via electronic and other means, provided us with a picture of school-specific efforts to reach out to all families and students, and ELL families and students, in particular. Arguably, one of the most effective measures of an educational institution's efforts to continue instructing and assessing students is the technology mechanism(s) and messaging the school and its agents (i.e., teachers, staff, and others) employ to not merely reach families, but to involve and empower them. "Do the messages communicated and the tools employed inform, engage, and empower parents to stay engaged and help ensure continued language and content learning for the students in their care?" That is the central question investigated in this paper.

This paper examines how schools in three geographically different areas that serve high and rapidly-rising numbers of ELLs have been communicating with these students' parents/guardians following COVID-19-prompted school closures. The content (i.e., message(s), timeliness, customization, and/or relevance of the message), the organization (accessibility/ease of understanding), and language and/or

languages employed, as well as the communicative outlets or media utilized (mono-sensory or multisensory) are carefully examined to gauge these districts' commitment to EL learners and their families. The primary research questions investigated in the study reported here are listed below:

a. What forms of communication are school districts employing with their students/families?

b. What efforts, if any, are being made to reach "ELL" students/families? Are any communication outlets, communicative media, resources, and/or messages intended solely for ELL families/students? If so, what forms do they take, and why and how are they being utilized?

c. How successful are the communication mechanisms employed and the messages communicated in reaching EL students and families?

d. How successful are these communication mechanisms and messages in meeting the needs of ELL students (i.e., in ensuring their continued learning) and in reassuring students, parents and other family members in the areas under focus?

The primary objective of this paper is to illustrate discrepancies and gaps in efforts to engage and empower ELL families and students. The absence of a clear, targeted, and sustained plan to engage and empower ELL caregivers and (through them) ELL students, during the school closure period under focus at every one of the U.S. school districts under focus is noteworthy. The U.S. school districts under focus individually and collectively serve a relatively large number of ELLs. The present study employed a comparative mixed methods analysis; this paper compares the communication mechanisms and messages employed by schools in three states, specifically with their "ELL" students and families. By way of a pilot survey, the relative success of their efforts at providing their EL families and students access to pertinent information and continued learning is analyzed. The present study attempted to gauge selected schools' and districts' level of interest in involving parents and instructing students during the focal period through the communication mechanisms, modes, and messages they employed. The underlying hypothesis was that schools and/or districts vary in their level of commitment to ELLs, and that those that reach out to their ELLs through multilingual, multiparty, and multimodal communication (vs. mass communication intended for all parents/families) and those that utilize instructional tools and other resources targeted specifically at this group have prioritized the needs of this population.

By examining the communication mechanisms, modes, and messages employed by schools as well as ELL experts (i.e., trained educators and parents') responses, we get a relatively complete picture of an institution's commitment to each student and family; alongside their level of success at continued instruction, and much more.

3. Methodology

Data for the present study were gathered over a six-week period in the spring of 2020 during the national quarantine when virtually every public school was closed to minimize the spread of COVID-19. A mixed-methods approach was employed. Two sets of data were gathered and analyzed in the course of this study. The first (primary) set consisted of survey-based data. Surveys and institutional communication intended for students and families constituted the data for this study. In the survey component of the study reported here, ELL students, parents/caregivers, and trained educators were invited to share their insights and experience with Online classes offered at area schools (see Appendices A, B, and C). Participants were invited to share their views on the communication (i.e., key criteria such as timeliness, quality and relevance), instructional success, and the instructional tools (e.g., technology) and other mechanisms employed for virtual instruction. The primary data consisted of a three-part pilot survey (see Appendices A, B, and C) with three pivotal groups, namely, ELL students, parents, and trained educators. The student survey took the form of semi-structured face-to-face and Zoom-based-interviews, so that the nonverbal language of students could also be captured and considered (unlike in phone interviews). Elementary, middle, and high school ELL students were interviewed in person (with social distancing maintined) and, in the case of one student, via Zoom, and each student was asked the questions that appear in Appendix A. Parental consent was obtained prior to the student exchanges. Students were invited to rate their learning success and to share obstacles they have faced since schools closed due to COVID-19, and to offer recommendations.

The second part of the survey consisted of a parent/family questionnaire (see Appendix B), and the third part took the form of semi-focused interview questions with a judgement sample of trained educators, including ESL instructors, ESL support service personnel and liaisons, program administrators, and others who directly provide and oversee policy and district-level implementation of instruction and assessment for primarily ELLs (see Appendix C).These three audiences provided us with additional quantitative and qualitative data.

School-based communication pertaining to continuity of learning plans and updates for students and families at school districts in three geographically varied parts of the county the West, the Midwest and the East Coast was also researched over a six-week period, and constituted the secondary data for this study. This included communication and communicative media utilized by schools from March 20 through May 30, including email communication authored by public schools and/or authorized personnel, and intended for families and/or students in the geographic areas and corresponding school districts under focus, namely, Clarke County, Nevada, schools in Missouri, and schools in Howard County, Maryland.

The language building-blocks framework of cultural competency (Pandey, 2012) outlined in Figure 1 below was employed to analyze this non-survey-based data.

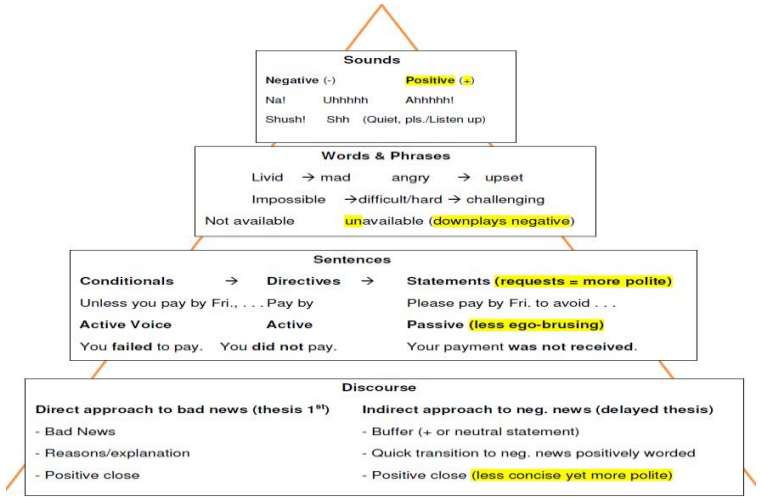

Fig. 1: Language Building Blocks Cultural Competency Model

This critical discourse framework was employed because it examines each unit of language and, as such, offers one of the most comprehensive analyses of language to date. Multimodal language as well as language building blocks that vary in size, starting with the smallest unit of language, sounds, at the pinnacle, to discourse--the functions and associated sequence of language--are analyzed in this framework, level by level and their level of positivity or negativity gauged on a cline. Arguably, on each level of language in this pyramid that charts language units according to their size, the positivity or politeness of each language building block increases from left to right, on a cline. Highlighted segments in the pyramid are meant to emphasize polite language structures. At the top, phonetic (i.e., phoneme or sound and para- or suprasegmental/pitch, tone and intonation) level, for instance, "Na" and no" are neither as positive nor as polite as a relatively non-judgmental "uh" or "ah!" (arguably congratulatory in its message). Similarly, "sh" and "shush" are not as positive as the options presented on the right side of these examples on the sound cline at the top of this pyramid, namely "Quiet please," "Listen up!" or simply maintaining silence until the audience gives the speaker their undivided attention. At the lexical level, this cline is replicated, with "livid" being less positive than "mad" and "angry," with "upset" being the most polite and professional of these choices. At the syntactic level, conditionals have the potential to sound like threats and directives are not as polite as indirect requests communicated via statements. For example, "Please play by Friday to avoid eviction" is less threatening and more polite or face-saving than "Unless you pay by Friday, you will be evicted" and "Pay by Friday or risk eviction." This focus on language that varies in denotation, connotation, and politeness

is consistent across the language spectrum, as outlined in Figure 1. In the present study, positive language use was considered evidence of caring or empathy.

4. Limitations

Phone calls and personal messages to parents and ELL students and educators with Face Book and/or Linked In accounts might have ensured greater participation in the survey segment of the study reported here. Time constraints, including restricted communication by phone with educators and administrators during the crisis when the time-sensitive study reported here was conducted--precluded participation of a representative sample of trained educators, and of a more diverse pool of ELL parents and students. Trained educators were contacted twice via email, two weeks apart, with a second email reminder sent out two to three weeks after the initial request for participation. Of the 150 educator surveys sent out, the survey response rate from this group was a mere nine percent. The International Achievement Liaison for one of the area high schools responded to our initial email (request for their expertise) with the following message: "I have forwarded your email to my supervisor and am awaiting a response." The low response rate of trained educators could be attributed to unforeseen administrative control, suggesting that educators at this school district are likely restricted even as regards choice of instructional resources, course format and delivery, and contact with students and families, among other areas. The sender not only demands that the survey be retracted, she asks that she be copied in the follow-up email requested of the researcher. This message also excludes from the district's approval process any survey that is missing "a university approved IRB." The time-sensitive nature of the study and institutional closure made it virtually impossible for the researchers to obtain the mandated university approval within the allotted time frame.

5. Findings from the Student Survey

The present study affirms our hypothesis that the vast majority of "ELLs" and their families, particularly low-income families, have been struggling with remote instruction (see Appendix A for the raw data and links to the audio files). Given that ten percent of our school-age children (roughly five million) have been negatively impacted by the COVID-closures, this is quite a grim picture.

Of the students that participated in the pilot survey (see Appendix A and Table 1), all but one of them identified Spanish as their home and primary language. Only one of these students, whose home language was Korean, was from a middle-income home; the rest all came from households with a median income of twenty thousand dollars. While students' sentiments varied in the first week following the school closure, most shared that they felt "sad" or "very sad" a month after face-to-face instruction ended, and they all shared that they felt uneasy, stressed, "nervous," or anxious a couple of days prior to the start of Online instruction because they did not know what to expect or

because they did not receive their device. Pseudonyms are employed here to facilitate the discussion for easier identification and analysis of respondents' experiences.

Age + Grade	Gender & home language	# & ages of additional children in household	Item 1	Item 2	Item 6	Item 7	Item 8 (Most challenging classes)
5 Pre-K Dave	M	2 (7, 11)	sad	Very sad	Not enough	No	English
7 3rd Belen	F (Span.)	2 (4 & 11)	sad	Very sad	Not enough	No	English
11 5th Louis	M (Span.)	2 (4 & 7)	happy	sad			English
12 6th Bryan	M (Span.)	3 (4, 17, 19)	V. happy	V. happy ("no homework")	Not enough	"Not really"	Science, math, geog., English, ELA
13 8th Dan	M	2 (3 & 17)	Both happy & sad	Very sad	Not enough	Na	ELA
13.5 8th Trish	F (Span.)	3 (5 months, 2, & 17)		Very sad	Not enough	No	ELA
14 8th Seong	F (Korean)		Sad	V. sad	Not enough	"Not really"	English, ELA
17 10th Jennifer	F (Span.)	2 (3 & 14)	Not happy, some sad	happy	Not a lot; it's not helpful	No	Algebra, APS (advanced physical science), & sometimes U.S. Govt.
17 11th Angel	F (Span.)	3 (4, 12, 19)	Sad	Not happy	Not enough	Na	English Math, Science
17 11th Carla	F (Span.)	1 (12 months old)	"Ni sad, ni happy"	Happy (because she could sleep in)	"Nada" (nonverbal) stopped attending	"No, it's difficult"	English Geometry Biology,
17 11th Roberto	M (Span.)	2 (5 months, 2, 13)	Very happy	More happy "I never liked school"	Yeah	"Yeah—it's easier"	Geometry "stopped attending" yet still passing

Table 1: Student Participants & Responses to Selected Survey Questions

As regards item 4, only two of the 11 students (ages 11 and 17) reported that at least one teacher had contacted them via phone and/or text, namely, their ESL teacher. The 11-year-old's ESL teacher called him at least once during the month-long school closure period to check up on him and to reassure him, and to make sure his father/family had received the email sent to parents outlining the plans for remote instruction. This student shared that this teacher understood and spoke their native language, namely, Spanish, and that by contacting him, she showed that she "really cares" about her students. One of the three high school students who participated in the present study reported that when her ESL teacher called her the weekend after school closed, she felt reassured and "special," and that this teacher has continued to call and text her and has even had food sent to her home. "She is so nice and always worries about me," she declared. When asked how she felt when this teacher reached out to her personally, she responded, "I feel like I'm part of her."

Three students shared that their school counselor had contacted them via text message. Their counselor is a bilingual Korean-English speaker, and while she does not speak their language, she understands many words because so many ELL and special needs students at her school speak Spanish, and she has been increasingly exposed to Spanish, particularly in the last four years as the schools' Hispanic population has grown by 23 percent. These students shared that her text messages were in English and they urged them to contact her if they needed help with their classes. While none of them responded with texts or phone calls, they appreciated receiving her texts and described her texts as evidence of "caring."

5.1 Insufficient Instructional Contact Hours & Relatively Non-Engaging Communication Tools

All but one student reported that they had learned little to nothing since online instruction began (see responses to survey item 6 in Table 1). When asked why, they all noted that the online classes were too few in number per school day and too short in length. They mentioned or hinted at a watered down (Online) curriculum. Jennifer shared that she had expected the Online schedule to look more like the face-to-face schedule. Conversations with these students made it clear that, in the view of 10 out of 11, not enough classes are being offered during the school day, and throughout the week. All students reported having no classes on at least one weekday, unlike when school was in session. Most reported not having classes on Fridays, and the more motivated students reported finding this relatively limited-class-schedule and fewer-contact-hour (Online) delivery to be not challenging enough and, as such, particularly problematic. In general, students found their class schedules and the online instructional contact hours to be insufficient, and they noted that their instructional contact mechanisms were not as effective as they could be.

Not surprisingly, their response to this question (item 6) was matched by their response to item 7; none of them reported having the same opportunities to hear and speak or use English as when face-to-face instruction was offered. The Korean student, Seong, was the only student who reported having access to weekly tutoring in English both before and after school closure. Her interview was, in fact, conducted over Zoom, following a weekly tutoring session she had with an eleventh grader. She shared that her parents had contacted a local church for ESL service offerings and that a high schooler who lived in that area had arranged for her classmate and her to offer tutoring to this student and two other Korean students. Despite this additional weekly (90-minute-long) English language assistance that she was getting, she shared that she would "like more opportunities to hear and speak English, more ESL classes, and more phone calls with my friends" and she noted that the phone calls "should be required by my teacher and also I can do this."

5.2 Language Barriers and Limited English Proficiency during Remote Learning

All the student respondents self-assessed their English skills as relatively limited or weak and identified language (i.e., English) as their biggest hurdle to full participation in the online sessions. In addition to difficulties students noted that they experienced with understanding class sessions (which average 45 minutes in the three locations under focus) and with fully comprehending designated weekly course modules- including readings, and homework materials), all participants identified English and "ELA" or "ESOL" (the names employed at some schools) as their most challenging classes, attesting to the increased language barriers they have experienced since area schools closed. This is hardly surprisingly. "El inglés es mi problema!" shared 17-year-old Carla, who is at risk of dropping out of high school. "No las clases (i.e., not the classes themselves) . . . es que yo no se mucho inglés y yo no entiendo mis clases" she continued. She arrived from Honduras a year ago, and when her father abandoned her mother and her baby sister a few months after they arrived in Maryland, she was emotionally distraught and she had to make taking her of her mother and sister and working as her top priorities. Her participation in online instruction has been filled with obstacles, including technology. She does not consider herself to be computer savvy, and she does not have Wi-Fi access.

5.3 Technology

All three 17-year-olds reported not having access to technology devices for the first week of class. With the help of her ESL teacher, one was able to pick up a computer from her school in time for the second week of class. The other student (Carla) gave up and stopped attending class. Initially she attempted to access class (i.e., Canvas) using her phone, but when her minutes expired, even that became challenging. "Ahora no puedo," she shared when we exchanged ideas on her friend's patio in early June of 2020. "I don't understand (how to use) the computer" she hesitantly reported on her third exchange with the researcher (after she had warmed up to her), and "I don't have Wi-Fi." Almost 90% of the student (and parent) participants in Howard County reported receiving their technology devices late and, as such, being unable to attend remote class sessions for three to five weeks. "My Mom was so worried when our computers didn't arrive," 12-year-old Dan shared.

5.4 Boredom, Lack of Motivation and the Missing Personal Touch

"I'm so bored at home, so I wanna go back school!," shared 13-year-old Dan. All students reported feeling relatively unmotivated in the online classes and once schools closed. "It's hard to feel motivated when the schedule isn't challenging," noted 17-year-old Jennifer. Only student, Roberto, was satisfied with his online classes. He proudly announced "School and I don't get along," and repeatedly shared that he disliked school. "It's easier," he responded when prodded for his thoughts on online instruction. When asked to identify his most challenging online class, he mentioned geometry. "I've stopped

attending," he continued. "Anyway, I'm passing that class." His response suggests that not all students are expected to work as hard as when school was in session in order to pass their Online classes. Missing class does not appear to have any consequences. Whether this is the message schools intend to communicate is however, unclear. 90% of the student and parent participants shared that they preferred spoken and face-to-face interaction. Tech challenges and Carla's limited comprehension of the course materials and assignments that her instructors uploaded on Canvas forced her to stop attending class altogether. She shared that, while some teachers had attempted to assist her via Google Meet after classes ended, at 2 p.m., she didn't understand them because none of her teachers, not even her ESL teacher understands and/or speaks Spanish, so the only way they can really help her is through an interpreter (which has not happened so far) or if they explain things to her visually (i.e., face-to-face--at her home is her preferred learning style, even if it would mean meeting outdoors and maintaining social distance, as we did in the course of our exchange) to replicate as best as possible the instructional-learning mechanisms that has worked when school was in session. Carla was not the only student who reported feeling relatively unmotivated since her school closed. "Right now I just care about my brother. . . I don't know how I can help him," reported 17-year-old Jennifer. She shared that one of the most worrisome things to her was that he had stopped attending his classes. "I think that he just get lazy . . . to get into the computer" and that he preferred to "play on his phone." When asked whether she had shared her concerns with her parents, she responded, "They know that." She later shared that they too do not know what to do to motivate him. So these parents, who work long hours and who are essential workers are themselves uncertain on how best to help their children during this school closure period. "Online, nobody checks to make sure you are listening and paying attention!" she lamented. "You could turn on your computer and walk away, and they (teachers) wouldn't even know, much less care" shared another student.

5.5 Sometimes I don't know something that I don't know

Not knowing what they do not know or understand, and being embarrassed about their (English) language skills are obstacles practically every ESL student mentioned experiencing whether directly or indirectly. "Sometimes I don't know something that I don't know" wrote 12-year-old Bryan in a self-reflection that the researcher asked him to write a week after his mother had contacted her for help with her children's continued learning.

His write-up is eye-opening on multiple levels. How so? His use of phonetic spellings visibly different from conventional spellings and potentially hard-to-decode for many--suggests that he is either accustomed to seeing written script look like it sounds (as is the case in Spanish) or that his literacy skills are limited. Follow-up exchanges with family members (including his parents) suggest that the latter is indeed the case. Brayan was eight when he started school in the U.S.

and he shared that he had had minimal exposure to writing in Spanish in his home country, Honduras. Oral in-class responses are the norm, and students are assigned little to no homework in the early grades. Bryan's low rating of his Online instructional experience at his Maryland-based school is eye-opening; that he would give his school a mere 5 out of a 10 when it comes to instruction and instructional assistance is quite telling. All except the kindergartner reported feeling embarrassed about asking questions even when their instructors asked if anyone had questions. Asking for help with comprehension and/or homework is even harder, according to all student respondents, except the five and seven-year-old.

5.6 Peer Support & Self-Motivation save the Day

Findings from exchanges with ESL students illustrate that students who reside in households with other students have been faring a lot better academically and socio-emotionally than those like Carla and Trish who have no at-home peer support. Findings also showed that when it comes to online instruction and learning, students who share the same home language have been helping classmates and friends who have relatively limited English skills by providing audio and written translations during their class meetings. 13-year-old Dan, originally from El Savador, shared that he has two friends who arrived from El Salvador just recently and since none of their teachers speaks and/or understands Spanish, he has been providing them with audio translations and explanations via WhatsApp and also through Google Translate on Canvas.

"Aveces chatiamos" (we sometimes chat) shared 12-year-old Trish, referring to Snapchat exchanges with her friends as the only way she has been able to use "some English" on her own, and of her own accord since her school closed. "But it's not enough and not the same" she shared. Indeed, tech exchanges have not been the same for her as being there in school and talking with and getting help from her classmates, friends and teachers five days a week. "Es difícil" (It's difficult) she repeatedly stated in both English and Spanish throughout every exchange. 13-year-old Bryan shared that if not for his daily exchanges with his best friend Dominic, he would not have had any opportunity to speak or write in English in the seven weeks since schools have closed.

Exchanges with students also revealed that there are clear benefits to multi-student households, and this plus point of peer support is most pronounced in multifamily setups. Students and parents shared that siblings, cousins and/or non-biologically related students in the same household and, in two cases, close friends who shared many classes, periodically shared devices, practiced speaking English and at least one additional language, and knocked heads and co-instructed each otherhowever briefly in other subject areas, as well. These same-grade and cross-grade pairs frequently help each other with access to online modules, and homework, as does Louis with his seven-year-old cousin,

Belen. Belen's device was, in her mother's words "mas de tarde" (more than late), so she used his device until hers arrived.

One special needs ESL student participant shared that she was embarrassed and therefore hesitant to ask her 12-year-old sibling for help with her homework because "I am older," she reported, suggesting that, in her mind, older siblings should not be asking younger siblings for help. For this reason, she preferred to contact one of the researchers for help comprehending and completing classwork and growing her spoken English. She shared that reading and writing were particularly problematic for her so she reached out to the researcher in question via WhatsApp and proposed that they communicate primarily via phone calls and audio messages (i.e., no written messages).

In addition to the frequent lack of understanding reported by every student, lack of motivation was another major challenge many students experienced, including two non-ESL students (14 and 16). Both are the researchers' children who volunteered to share their insights. "It's like talking to a wall; none of my classmates respond when my teacher asks questions. It's like they've given up or have gone back to bed or that they're in the pool or at a picnic we'll never really know 'cause they don't have to turn their cameras and speakers on!" shared the 14-year-old after her e-class ended just 15 minutes after it officially began. "I ended up answering most of the questions and I could tell my teacher really appreciated that I did; he even said "Thank you."

"I don't know what to do to motivate my brother," 17-year-old Jennifer lamented, referring to her 13-year-old brother Dan who had missed at least three classes the week we spoke. She was concerned that he would stop checking in and attending classes. "He tells us that he's done but I think he's just lazy and prefers to be on his phone," she shared. If Dan stops attending class, the continued learning of his two classmates (lower-level "ELLs") that rely on him to make sense of their classes and of their homework will be substantially jeopardized, as well, resulting in a negative chain reaction. In a follow-up exchange we scheduled with Dan to inquire about this matter and about his friend circle and to motivate him to complete the semester to the best of his abilities he explained that Wi-Fi and communication problems had barred him from attending some classes. He shared a message he had received from his ESL instructor titled "A gentle reminder." "We missed you in class today" was how this message in just English opened. Dan revealed that all his friends are "Hispanic" because "at least" he can "communicate with them;" like most of the students we talked to, he is self-conscious about his English and "shy." "I have trouble making friends with anyone who speaks only English," he hesitantly shared, clearly embarrassed about his English.

All the student respondents unanimously shared that the interactivity and peer support they had at school was visibly absent in the online instructional mode employed by their schools. Peer support was found to account for many ELL student respondents' contentment and feelings of academic success. 75 % of the respondents reported

having at least one peer advocate. These included siblings, and other biologically related members who live in the same household or within walking distance, and who are also enrolled at their school. 12-year-old Brian reported that his close "Italian" friend Dominic helped him with homework and even with securing a device. Their families had no idea that remote instruction was to start a month after schools closed because their mother had not seen the emails the school had sent her. This was because she rarely checked email prior to the COVID-induced school closure, even though she had established an account when their families had moved into their apartment complex some five years ago. It wasn't until her two school age children's schools contacted her by phone after they had been absent Online for over two weeks that she learned that classes had resumed. Then began the rush for a device. Her friend, a Dreamer, helped her access her email messages and contacted the schools to request devices, but they didn't arrive until over a week later. Louis' classmate and best friend, Dominic expedited his device receipt. In Louis' words, "Dominic asked my teacher for me and my teacher sent me a device in a few days." Eighty percent of the student participants were self-motivated and this helped them to find ways to ensure that they continued to have a steady and sustained stream of comprehensible input (see Crawford & Krashen, 2015) and interaction in English. For instance, 11-year-old Louis who had had just under a year of classroom-based English instruction by the time this study was conducted proposed and successfully tried out what he termed the "English challenge." His seven-year-old cousin, Belen and him formed a pact to communicate in English for 24-hours, and they initiated and have been following through with this "challenge." It has helped them grow their English skills and, in Belen's case, her Spanish skills, as well. How so? She reported that at her babysitter's she speaks primarily in English, so when her cousin Louis proposed "the English Challenge," she began to "think of the other (Spanish-only-or-mostly Spanish-speech days) as her "Spanish Challenge days." The extent to which their co-facilitated instruction and learning ensured a mix of the third vital language-literacy ingredient, namely, feedback, is unclear. Their reports and the relative ease with which they both responded, in English, suggests that the "English Challenge" they instituted has indeed enhanced their fluency in English.

5.7 Effectiveness of the materials: No More Non- ESL-Friendly Resources

Findings suggest that relevant resources are not always being used to instruct ESL students. For example, 12-year-old Bryan was assigned a weekly word-building program. The narrator in the sample audio clip that he played for the researcher shared that the word "fantastic" is "a fantastic word" because it has "a short a, a short i, and a closed syllable." The researcher had trouble making sense of this and could not understand how this explanation constitutes a viable one, so it isn't surprising that Bryan was clueless about the point of this exercise. For his final reading class assignment, he had to identify any word or words with rising intonation (i.e., higher stress or emphasis) in a sentence by

way of a visual marker, yet he did not understand the instructions and could not explain the objective of the exercise to the researcher. He reported that all students had to do these exercises and shared that some "make no sense." Bryan shared that nobody checked their work or checked in to be sure they understood what needed to be done. All that matters is whether or not students put in the allotted time. "They check to make sure you do 20 or more minutes," he shared. "What happens if you don't?" asked the researcher. "They send you an email." Bryan shared that sometimes his friends or his mother's friends help him complete them but the extent to which helpers' completing these and/or other assigned work constitutes evidence of each ELL student's having learned what was intended is debatable. Could it be that some of these exercises are not catered to ELL students? So if online instruction fails, it's not for lack of technology or apps or programs per se; but for lack of relevant resources and individualized instruction.

5.8 Additional Obstacles to Student Learning

Stress caused by the obstacles shared and by the uncertainty surrounding school re-opening has caused many students to give up. Some students have had to work to help out their families and their online instruction schedules are much more conducive to employment (the only plus point of online instruction shared). Some shared that working has, in fact, been more English-instructional, motivating, and economically advantageous than sitting at home and taking the classes their schools have been offering. As such, at least two teenagers have been drawn to full-time work. Both were quickly promoted (one to head cashier and one to Manager) at their respective places of work (McDonald's and Checker's), and they felt appreciated, needed, and validated when they were recognized through their relatively quick promotion, so not surprisingly, they have found work to be much more rewarding. This new or greater allegiance to her workplace might explain why one of these teenagers, namely, Carla recently stopped attending (remote) classes altogether. She shared that at least she gets to hear English at work, and that since she has to use English, working has helped her hear and practice English daily and much more effectively than Online instruction has, and work has even grown her self-confidence. "En mi trabajo yo trato mas, porque soy cajera!" (at my work I try harder because I'm a cashier) she shared excitedly with a hint of a smile. It is worth noting that the only time she smiled and showed any excitement during our exchanges was when she talked about work.

She shared that additional obstacles to her continued learning include her daily babysitting responsibility and her work schedule. She has to work to help her single mother (who works part-time) and her 11.5-month-old sister to pay their rent and other bills, and she has to be at work by 3p.m. Her second and final class usually ends around 2:15p.m. That has left her with a very small window to get to work; she has had to get her baby sister ready and to drop her off at her aunt's (who gets home from work at 2:40 p.m.). She has to walk about 10

mins to get there on the days that her mother works, and she has to walk to work another 15 minutes. “Maybe next year” she responded when the researcher who interviewed her outlined the advantages of a high school diploma in her efforts to coax Carla to resume her schooling. Yet, if her school continues to offer the same kind of (relatively incomplete, ineffective, and unengaging) instruction as it is presently offering, Carla is unlikely to return, she noted.

In regards to recommendations for any or all of their teachers, the last survey item in the present study, all students offered at least one suggestion.

Their recommendations are summarized below:

I. More classes and more frequent meetings (at least 2 or 3 times a week for each content/subject area) and each should be at least 1.5 hours in length, and preferably two hours long.

II. Longer class sessions (70-90 minutes per meeting recommended vs. 45 minutes)

III. More ESL small group and/or pair meetings, preferably longer than 45 minutes each, and more than once a week to allow for both frequent and sustained access to the English language and support services.

IV. At least one weekly one-on-one meeting with each student (averaging 10 minutes of audio- and/or video interaction to allow for customized, individualized instruction—which each ESL student participant stated would be less intimidating than large group interactions)

As regards the first recommendation, HCPSS student participants shared that 45-minute session over four out of seven days a week is simply insufficient to cover the material for so many classes and to offer in-depth coverage. 45 minutes are also insufficient to ensure that all students comprehend what is being shared, particularly ESL students who frequently need additional time to process information, to formulate questions, and to muster the courage to participate and/or to ask questions.

As regards the third recommendation, currently only one 45-minute session is offered online by ESL instructors and support staff at the largest school district under focus. In general, the schools and districts that have made a concerted effort to reach out to their ESL students and families have done one or more of the following and, as such, have been relatively more successful at engaging and empowering this group of students and their families:

I. Not merely permitted but encouraged their instructors and support staff to reach out to students and families early on and continuously

II. Permitted and encouraged instructors and support staff to reach out to their students and families in their home/heritage languages

III. Permitted and encouraged their instructors and support staff to reach out to their students and families via phone and texts, and not just through email and other (electronic) written media

IV. Have emphasized parent engagement and made their content relatively accessible (using multiple techniques, e.g., a parent portal that is associated with their phone numbers)

V. Have offered free wi-fi

VI. Have provided and emphasized homework completion as an important (practice) segment of learning

VII. Have offered one-on-one and group tutoring

VIII. Have utilized interactive technologies such as Flip grid to ensure that their ESL students are receiving a sustained stream of English, are continually interacting in English with at least one other individual whose English proficiency is higher than theirs and that their students are receiving continued feedback—both direct and indirect, in order to make necessary adjustments to their speech and/or literacy

IX. Have offered printouts or content packets (with activities or homework) at accessible sites, for pick up by families

X. Have conferred or periodically confer with parents/caregivers and students to plan viable and fruitful cross-grade, and/or cross-age homework and conversational and/or literacy partnerships in the same households or residential communities (e.g., Town and Country weekly Conversation Tables could be held in the parking lot of the leasing office or at a playground. With a bit of collaboration between instructors and, in some cases, across schools, (i.e., instructional planning across grades and even across schools), siblings could be enlisted to work with each other, even if they are in different schools.

XI. Have offered additional services to families such a health and nutrition services, and

XII. Have attempted to make these services more accessible to ESL students and families by utilizing more than just English (alone) to publicize them. Multimodal and Bilingual/Multilingual and translingual communication are absolutely essential. Local other-language radio stations could be tapped to assist in getting the word out as well as for soliciting and sharing ideas on what has worked for some families.

6. ELL Parents' Views on Virtual Instruction

Semi-focused exchanges with Spanish-speaking ELL parents (see Appendix B) were conducted almost entirely in Spanish face-to-face

(outdoors) or via WhatsApp video conferencing. Frustration was the most palpable sentiment that all the ESL parents shared in their exchanges with us. All parents all shared that their children's continued learning was their top priority. Most of the low-income parents interviewed shared that they had completed "le basico" (i.e., sixth grade) in their home countries. It is worth noting that, in written texts, the Spanish-speaking parent respondents consistently employ spelling pronunciation; in short, they spell the way they speak. Note, for instance, that "hacer" (to make) is spelled "aser" by one parent and "haser" by another.

All the mothers shared that they had had to run around quite a bit to locate "libros" (books) and/or materials to keep their elementary and middle school children busy and learning for the two to four weeks before virtual instruction was offered by their children's schools. During this time, parents had received no communication from their children's schools. They found this lack of communication to be appalling and most identified it as a challenge. When the Hispanic parent respondents were asked whether their children's education was the primary responsibility of the mothers in their households, all the parents, both males and females shared that this was indeed the case since the families had emigrated to the U.S. The reason for this, the three fathers that participated shared was that the males typically work longer hours and farther away from home (as construction, lawn and other unskilled laborers and their livelihoods require them to work early and late hours (typically before and after school) across county and state lines so that they can pay the rent and other bills and provide as best as they can (their primary objective) for their families, and ensure that are "comodo" (comfortable). All the parents worked full-time in either the food industry or construction.

6.1 Ellos (Schools and Trained Educators) are not doing enough

Parents unanimously shared that schools and trained educators are simply not doing enough. "Challenge them!" one parent announced. "Why aren't they checking on the students by phone, and more often by email and by text, and why haven't they bothered to contact the parents weekly and in a language we understand? It shouldn't be that difficult? Isn't that the responsibility of a school?" Lack of instructor and student motivation was mentioned by all the parent respondents as a major obstacle to successful online instruction and learning. "It's like they have just (all) given up," one parent shared, referring to the schools, teachers, and students. The general consensus among parents was that their children had not learned enough since schools had closed and since Online instruction began. All the parents expressed grave disappointment at what they described as the school's and most teachers' general lack of interest in these (ESL) students and their education. Each shared that they expected a lot more from the schools and from every single teacher and not just a single committed ESL teacher and Counselor. "Quiero que ellos luchar para obtener le mejor educación para ser alguien y no solo exitosos!" (I want them to fight to

get the best education to be someone and not just successful) declared one father. “What are they getting paid for?” one parent asked, referring to teachers and school administrators. “For Forty-five to ninety minutes of contact time?” he continued. “It’s a joke!” he continued, referring to what he described as a half-hearted attempt at continuity of learning.

6.2 Primary Challenges Faced by Low-Income ESL Parents

6.2.1 Technology

One father shared that the cost of Internet was prohibitive in his County. The difficulty of getting their home connected in a timely manner was another obstacle he shared that his family had experienced. He expressed great dissatisfaction at being given the runaround by Internet providers in his area. He shared that his friends that live in Catonsville, under five miles away in neighboring Baltimore County were able to secure Internet access for a reasonable monthly cost of ten dollars ($10), but that Internet companies were charging them over five times as much. “Eso es discrimination,” he declared, visibly angered by this corporate monopoly. He noted that the differences in the areas and schools were not enough to warrant such a huge price difference. His family is paying “mas de cincuenta dollares mensual” (more than fifty dollars monthly), he shared. “Y solo para 45 or 90 minutes por solo cuatro dias?” “Qual es el function?” (i.e., What’s the point of getting and paying for Internet when you have so few class contact hours?).

95 percent of the parent respondents reported that not having Google-Meet-compatible devices and or not receiving their devices from their children’s schools on time posed a major problem because their children missed over a week of classes and this psychological strain, compounded by the parents’ intensive and (Corona-conducive) work made their parents feel helpless and left them and their children emotionally drained. “Tenia mucho stress porque comenzo classes mas tarde porque no recibi su computadora,” reported a 39-year-old parent. One parent stated that she was so “preocupada” about not receiving her seven-year-old’s device that she borrowed money to purchase a device but could not find a computer (she does not have a credit or debit card because she has been unable to secure an ID card since her National ID contains another person’s photo and her attempts to get the problem resolved through her Embassy have been unsuccessful) so she settled for a tablet at Walmart, but the tablet did not work to access Google Meets, even though she successfully shared the hotspot Wi-Fi from her phone. According to her 11-year-old nephew, who also lives with her, “no era competente” (i.e., it wasn’t a compatible device and it didn’t have enough RAM). She shared that unreliable Wi-Fi at their apartment has been a constant source of stress. Most days, she is successful when she attempts to “compartir” (i.e., share) Wi-Fi from her telephone to her daughter’s device, but there have been days when her hotspot quota has been rapidly deleted or the device sharing hasn’t worked. Not being accustomed to checking email, not having or as frequently using email,

and not being as comfortable with written Spanish as with oral Spanish was a problem every Spanish-speaking parent reported.

6.2.2 The i-Phone Craze

Speaking of technology, one parent shared that he was still getting to know his 13-year-old son. He has been trying to build a relationship of trust and mutual respect with his son for the last three years since they were reunited after a long 10-year separation, which has taken a toll (see Author 2104a). His son recently manipulated him into purchasing for him an i-Phone. "Me explica que es mejor y bueno para accessor sus clases" (He told me it's the best and good for accessing his classes), the father shared. He saved up and spent close to five hundred dollars to make his son happy by getting him his phone of choice, and shared that he regretted his decision because thei-Phone turned out to be a great learning distraction. Instead of accessing class in a timely manner and using his phone as the learning tool it was intended to be, his son started playing free video and other games on his phone and got addicted to his device.

6.3 English Language & Literacy

"Solo Inglés utilizan!" 13-year-old Dan's father announced, referring to the language employed by his children's school. He noted that using only or primarily English with ESL parents is inconsiderate and problematic. "Tienes que ser bilingual! (They need to be bilingual") he announced. That his wife would receive messages only in English and that too, never oral or audio messages but only written email and an occasional written text from the school Counselor was indicative of the schools' lack of caring. "Ellos no le importa los estudiantes ni nosotros!" he declared (They don't care about the students and us) If they did, they would, at the very least, communicate through a translator and periodically communicate with the families. He noted that when his wife received written messages in English, she would have to wait until either her 13-year-old or her 17-year-old would help her translate the messages and that this would prolong the communication with the parents. At least if they bothered to send a written message in Spanish, she could decode it faster and on time, he proposed. He noted that fortunately, at least one of the written messages from his children's schools (between late March and June 14) had been in Spanish. He could not recall whether the middle or the high school had sent the message.

"Como yo no conosco (mucho) inglís," a parent shared, echoing a challenge that 99 percent of parent respondents shared. Not understanding English well and not speak it made more parents feel unready to help their children continue to learn as best as they should be, most shared. Many confided that written communication even in Spanish—was challenging for them, and for their children, so the mostly-written communication sent by area schools (via email and occasionally, too via written text messages) was hard to decode. Many schools assume that if parents self-report that their home language is X,

Y, or Z, that they must be literate in these languages. That is hasty on our part. Even to assume that speakers of Latinate languages would have an easier time with the mostly Roman-based script employed in written English is presumptuous.

6.4 Contracting & Compartiring Corona Increased Household Expenses

Two parents reported contracting Corona and how the high fever and body aches and fear of their families and children (both have toddlers and teens) contracting the virus made May a painful month in which they felt minimally helpful to their children and families as regards their breadwinning, parenting, educational motivation/academic and child-raising responsibilities). They both shared with teary voices how their families had being socially isolated and ostracized when word got around as it quickly does in the Latino community that these mothers "tienen Corona" (have Corona). The stigma had a negative impact on their school-age children; they shared that their children acted out more and seemed to resent them during that period. The added economic burden of losing work hours (they both received half a day's pay for three days) for the 14 days that they stayed at home was a challenge they both shared. In addition, two parent respondents mentioned that they had spent an average of $80 purchasing a device and that their expenses had gone up by an average of $54 a month because they have had to pay for access to the Internet.

In our phone exchange the following day, she shared her schooling concerns. What worried her the most was that she had not heard from her child's school over two weeks after schools had closed, and she sought help so that her child could participate in online classes. She had made multiple trips (during her 30-minute lunch break) but "estaba cerado" (it was closed) and she did not see anyone on site that she could ask. I asked if she had checked her email. She said she rarely used email. She also shared that her family had limited Internet access. We shared and translated for her the school's website information. It listed the weekly schedule and class times and modules for each grade. Luckily, her daughter's school also shared a few weekly lessons on a local television station, and this was listed on the school's website. If this mother had not taken the initiative to advocate for her child by reaching out for help, many more weeks might have passed by before her child would have received lessons.

6.5 Not Knowing How to Proceed

Once schools shut their doors in mid-March, most turned their backs on parents; they left behind and left out many academically unprepared educators, namely, parents. Most of the parent respondents come from cultures where the charge of schooling is assigned to the instructor, with little to no interference from the parent. Parent involvement has a different meaning in every (sub) culture. Not knowing how to be an academic educator and not merely a breadwinner and motivator has been a challenge many parents shared that they were unprepared for. A

parent (referred to here as Denora) contacted one of the researchers and offered to pay for weekly English lessons so that she could help her elementary school children. Her third grader had been doing very poorly even prior to the school closure, she shared, largely because he didn't understand) Ingles." Yet when she had requested a meeting in February with his teacher, the wait time for an interpreter was about a month. This is because her child's school does not allow for volunteer interpreters. Parents have to communicate directly with the Office of International Students (not through proxies) and only the services of contracted interpreters hired by the school can be employed, creating a protracted and unduly long wait period. That meeting never took place after schools closed. On April 2, 16 days after schools had closed, she responded to one of the researcher's inquiry about whether or not she had received any communication from her children's school regarding the districts' plans for remote instruction and how to secure a device for her children. She mentions that her husband told her (that the schools would begin teaching virtually and in a follow-up call she stated that he had heard this from a friend at work who had received an email in English from the school) and she asks what she needs to do to get the tablets: "que tengo que haser (sic) para obtenerla."

6.6 A WhatsApp Parent Support Group

One parent of a 13-year-old who arrived from Honduras a little over a year ago shared that many mothers had formed a "grupo" on WhatsApp to support each other and share information regarding their children schooling. Her 8th-grader's device arrived over a week late, so both she and her child were stressed. Being part of this group helped reassure her, she shared.

17-year-old Jennifer who reported getting a great deal of language and emotional support from her ESL instructor proposed a comprehensive three-part plan of action (see her response to the last student survey item): "I would contact the parents first and see that, if my children is (sic) doing the work. My second thing is like checking one-on-one (with) my students to see if they are doing the work and sending a (text) message to see if they are understanding the title(s) or the system that we are trying to teach, so that they can do well in the future."

When prodded for details on her last-mentioned recommendation, she mention one-on-one texting, if calling is burdensome for teachers and/or support staff. She also proposed (in a follow-up exchange) that the teachers collaborate more and ":must talk to the ESL teacher" and that ESL teachers must attend (content) classes or at least have access to the content so that can ask more pointed questions and help each student with language tied to specific lessons.

This student's imprompt response acknowledges the role of parents as educational front liners, and of family support in successful student learning. That a 17-year-old recognizes that parents and siblings (and/or other students in the home) must be more fully engaged in the

educational process in order to ensure that every student is learning as best as they can is noteworthy. We would do well to heed her advice.

7. Educator's Insights

All of the ESL educators who took time to respond to our request for their feedback reported feeling overwhelmed and overworked. All but one responded via email and one opted for a phone interview. Most shared that they are responsible for an average of 45 ESL students, and many have to assist other students, as well, and to attend one to two weekly ESL meetings (averaging 45 minutes) and additional staff trainings and/or meetings. In fact, 90 percent of the educators with whom we dialoged shared that many trained educators, primarily content teachers, had been spending more time training on remote instruction than actually teaching and assessing learning.

A MO-based elementary school ELL Specialist shared that her program serves 275 ESL students and that she had used primarily phone calls and written text messages to reach out to her students during the school closure period. Her objective has been "to inform, reassure, and inquire about technology in home." She noted that her administrators have not dedicated specific communication outlets, resources, and/or messages solely for ELLs. She recommended Talking Points as a useful communication and/or learning tools for ESL students and/or families, particularly K-4 and low-level ELL. Regarding the last question, namely, challenges, encountered in reaching ESL students and/or their families and solutions they would propose, she responded:

> *Calling students on 3-way calls to leave a message with an interpreter and give them a time to I will call back again with the interpreter so they feel comfortable picking up, getting online learning set up with 3-way calls, helping them understand that students should do what they can and not worry about what they can't and parents not knowing how to help them.*

Two Nevada-based ESL educators (one a middle school and the other a high school instructor) shared that they had contacted their ESL students via phone calls, texts, and audio messages through WhatsApp. A third ESL educator, from Maryland, shared that WhatsApp had been the most effective and is, in fact, the preferred and most popular choice of her students and families

Both the Las Vegas-based educators have had considerable success with engaging and empowering their students when they employed Flip Grid, an interaction-engendering app that allows for short 90-minute video clips that can easily be uploaded using a code the teachers send to their students. They highly recommended this. 98% of the educators surveyed reported feeling relatively isolated in their efforts to engage their students and families. All 10 ESL educators shared that while they had been successful at contacting (i.e., reaching) almost every student and/or parent, they had not been as effective as they would like to have

been in meeting the needs of their students and families. The vast majority reported that since schools had closed, they had been unsuccessful in meeting the needs of their students and families. “On a scale of 1-5, with 5 being highly successful, I would say a 2 when it comes to meeting their needs and probably a 4 when it comes to making contact with them,” responded one teacher. The vast majority shared that they felt constrained by lack of administrative support and oversight, and the added obstacles that their students and they themselves faced and still face.

Some educators shared that many teachers, mostly content area teachers had become laissez faire and contented with doing the barest minimum. “As long as they get paid, that’s all that matters to some teachers,” he noted. “Some would prefer not to return to school because then they would actually have to work for their money,” another teacher shared. She was convinced that the ESL and Special needs teachers are the most burdened and, in her view, generally the most committed to their students and families.

A Maryland-based educator shared that she has been more involved with her K-1 ESL students, since they need more help. “Before school closed, we emailed the students and their parents—depending on whose email addresses they had on file. With the older students, they are generally more invested, so they don’t require as much monitoring.” They have her “cell” and know how to reach her. She messages them from time to time. Her school (district) has not been using or encouraging use of any tech tools specifically for this population, and she shared that she found that disconcerting. “They do have “ESL weekly” trainings every Monday and they have been “very helpful,” she shared. The ESL teachers (3 teachers, 2 para teachers) at her school also convene once a week (via Google Meet) to exchange ideas.

Student and parent exchanges from the present study revealed that caring teachers can make all the difference. In every instance in which a teacher went out of their way to help reach and/or empower their “ELLs,” these teachers were consistently ELA teachers. Many have extended more than just language and academic support to their students. This is hardly surprising, given that most ESL and/or language instructors are have at least some training in sociolinguistics. As such, they tend to be culturally sensitive, empathetic, and inclusive. Many are well-traveled, bilingual and/or advocates of bilingualism.

All but one of the educator respondents from Maryland shared that their hands are “tied” due to bureaucratic regulations beyond their control. “We are discouraged from contacting students by phone!” reported one middle school teacher, her voice one of consternation. “They (the administration) told us not to share our number--essentially. We were told it was up to us; we were encouraged to block our number,” she continued. Shortly before the start of remote instruction and until now, this teacher has been calling and messaging her students and the parents with whom she has established a “relationship.” Lately,

she shared, she has been "messaging more and calling less." "I speak Spanish, so I'm drawn to this community. I'm from South America so I understand the culture" and her Spanish speaking ESL students excitedly reported that she called them, suggesting that personal phone calls and texting were more impactful and personally reassuring and that they related to her for this reason and also because she spoke their language. This teacher shared that she likes using phone calls and audio texts for the following reasons:

i. they are "faster"

ii. "I don't have to translate content"

iii. "I have (Spanish-speaking students that speak pretty well but they can't read and/or write, so at least this way" (i.e., via calls and audio messaging). I know this way they understand what is being shared."

iv. My ESL students like the phone calls and messages.

This teacher identified additional challenges that she has faced, many of which were echoed by other respondents and outlined below:

Attendance has dropped dramatically. "It varies from 30-60%," shared 80% of respondents and "it's actually declined since Online instruction started," shared one respondent. One kindergartner "disappeared" altogether and his teacher was unsuccessful in reaching the family. She did call and "leave a message" but she hasn't messaged them because, as she reported "I don't know the family as well. I don't text that family. I don't really have a relationship with this family. My calls went to a voice mail."

8. Teacher Challenges

"It gets tiring. I put a lot of time and effort for week after week" reported a Maryland-based teacher. She has even gone to at least one student's home to show her student how to access Google Meet, and even checks into "some (content) classes" but some students have simply stopped putting in the effort, in her view. She shared how she has called, messaged and continually made every effort to encourage this one student but she still didn't attend class. She would still be sleeping at 10a.m. when she's call her and her first remote class started at 10a.m. "They've got to have some personal responsibility, as well" she shared. So, in the case of this student, this teacher got frustrated and felt that her efforts had been in vain. "So I've backed off now," she shared. (So when students give up, even caring teachers can give up)

"I know of a student whose computer got there four weeks late and even after it was supposed to get there, they had to go pick it up" so that put them "behind by another month," shared one Maryland-based teacher. Devices received late and internet access issues were a "problem" initially, particularly for ELLs in Howard County schools (Maryland) but "they have been resolved for the most part," shared a teacher in a follow-up exchange in early June, 2020.

Tied to online delivery is what we would call "The literacy challenge" that this instructor noted that most of her literacy-using students experienced. "We can't do reading and writing effectively. It's just not the same as being there and helping in person," she shared. "We have identified some apps and other tech programs but not all are effective," she continued.

In one teacher's experience, her ELL students and families have simply not been as forthright about their needs. "I wish they'd be more honest" about what's stopping them from attending every ESOL and/or content class." Another challenge she identified is what we term here "the Family/Community Connection" challenge. Her school has two designated Family Liaisons, but "they are technically the Chinese and Korean liaisons; neither speaks Spanish" so when and if they ever reach out to these students and their families, "they have to do so through an interpretation and that's simply just not the same."

The Uncertainty challenge: this challenge is one of not having sufficient information from the school administrator side regarding plans for these (ELL) students and their families. Not knowing what is going to happen in the summer and/or fall to these students, since the ELL offerings typically offered every summer have been canceled was a challenges most ESL educators shared. Not knowing the consequence on their progress, of not attending or not completing to a satisfactory level their courses for the rest of this semester were additional challenges identified. "I think we are going to repeat the quarter in the fall, but I don't know," shared one teacher.

Whether and to what extent content teachers have been collaborating with ESL teachers to ease comprehension and engagement of their ESL students is unclear. The present study suggests that such collaboration, if any, has been minimally evident.

9. Analysis of Non-Survey-Based Data: Findings & Discussion

As regards the non-survey-based data sample, schools in the Las Vegas area had some of the most ELL-inclusive and parent-engaging communication on their Website and other documents, followed by schools in Missouri. Each of the school districts had varied start dates for remote instruction. While most schools in Nevada started classes just a week following the end of face-to-face delivery, HCPSS in Maryland did not offer e-instruction for a full month following the mandated closure, and began with High School students first, followed by elementary and middle school student instruction a week later, on April 20. Most public schools in Missouri began offering remote instruction two weeks after the mid-March closure. Some might argue that the lag time in instructional delivery could be interpreted as evidence of these districts' varied levels of commitment to their students. The communication these institutions and/or their representatives employed with parents/guardians and students could be analyzed for timeliness, purpose (to inform, to engage, to empower), language (informative, engaging, empowering- i.e., evidencing

compassion, care, and commitment to student success). Of the districts under focus, Clark County Schools were the first to offer remote instruction—just a week following school closures. They offered students free Wi-Fi via their school buses, which would park for at least two hours in residential areas to allow students to access online resources and to complete any practice and mandatory assignments.

A comparison of the language that two districts employ on their web pages (all accessed on May 20, 2020) reveals notable differences that suggest different levels of parent/family engagement which could be categorized into three groups based on their linguistic function: that of informing versus involving versus empowering parents/families.

Take a look at the Maryland-based school page titled "Supporting Instruction at Home" (https://www.hcpss.org/health/coron avirus/supporting-instruction/). The term "families" is only used twice on this parent/family page, so the intended audience is only mentioned indirectly. The singular form "family" is employed twice, and the term "family member" only once. No headings or tabs with the word "family." No tab, heading or content dedicated to the theme of information for families is visible--in the headings and in the body of the message, so parents and/or families are not overtly identified as the intended readers. Theoretically the target readership is parents, as evidenced by the descriptive phase "your child's teachers" employed in the fourth line, as well as in the pronouns "you," "yourself" and "I" that are employed to refer to the reader--presumably to parents (i.e., based on their usage or the context of use). Interestingly, while the word "instruction" is employed and constitutes a section heading, the associated term "assessment" is not used. Arguably this page is intended to help parents assume the roles and responsibilities of teaching sans quantitative assessment. Some might argue that this is contrary to outcomes-oriented instructional practices. In contrast, the home page of Columbia Public Schools in Missouri includes "ParentLink." This parent-first title promptly spells out the importance of parents in this school district's educational mission. Additionally, on this web page, the seventh tab on the home page menu titled "Quick Links" opens by announcing that "Columbia Public Schools' messaging system allows parents and guardians to receive school messages via phone, text, e-mail and mobile app." The multimodal messaging employed with parents/guardians during the COVID-19 school closure is noteworthy. Contrast this with the primarily email-based communication employed with parents/guardians in the Howard County Public School System (HCPSS) in Maryland. Below is an email message (Fig. 13) that school administrators sent out on April 1 to the families of their 59,000 students via email in Howard County, Maryland (HoCo, hereafter):

Please complete the Student Technology Survey in English, using the Chinese, Korean and Spanish language translations as reference. If you need support in filling out the survey, please contact the Call Center:

Chinese: 410–313–5920

Korean: 410–313–1592

Spanish: 410–313–1591

For assistance in all other languages, please contact Monica Ranta at 410–313–7102.
Encuesta de Tecnología Estudiantil

Por favor, complete la encuesta sobre tecnología en inglés [HCPSS Student Technology Survey] utilizando las traducciones como referencia. Si necesita asistencia para completarla, por favor, comuníquese con el centro de llamadas. Español: 410-313-1591

学生技术设个调查

请参照中文翻译件，用英文填写技术问卷调查 [HCPSS Student Technology Survey]。. . .

테크놀로지 이용을 위한 설문조사 [HCPSS Student Technology Survey]는 번역문을 . . .

Fig. 13: HoCo Student Technology Survey (https://www.hcpss. org/health/coronavirus/devices/?fbclid=IwAR03b78u9OOFnv Tp_luyUAWti-ZM-6CeLd9Y4oMEXGnn3Qh4eC1xWiWgQJw)

Note that the survey title and the title of the communication sent out do not clarify the needs-assessment nature of the survey; namely, that the intent is to determine if a family needs a device. Moreover, the survey taker is asked to provide their responses “en inglés.” This use-English-only-to-respond mandate is arguably both baffling and linguistically exclusive to ELL families. How likely are parents to complete this survey when they are required to respond only in English? The accuracy of online translators further muddies the waters.

Furthermore, parents must have e-mail access and an e-mail account to see this note from the ninth largest school district. When schools/teachers send (important) messages to students/families solely or primarily via email, especially during the national quarantine, we exclude a huge segment of our community. We are assuming that parents have access to email and an email account that they necessarily use and that everyone reads one or more languages and/or specific languages that they might have (self)reported as knowing (e.g., Spanish) and, in the case of the Maryland schools under focus, administrators are asking adults who are already self-conscious about their English and/or literacy and educational skills (among other skills) to complete a technology survey "en inglés," and this could be quite anxiety-inducing, not to mention time-consuming, cumbersome, and disconcerting.

Columbia Public Schools’ home page contains both a “Families Portal” tab and a “Families” tab. The former leads to a password-based entry

point for parents/guardians while the latter leads to more information for families. The ninth menu item or tab refers browsers to "Procedures for Identifying and Assessing English Language Learners (ELL) Children." However, whether these procedures apply to qualifying middle and high school students is unclear. The HCPSS does not provide tabs or links using any of these or synonymous keywords, namely, ELL or equivalent terms (e.g., ESL, Emergent Learner(s), Bilingual Students), "Parent" and "Families." The first tab or quick link is titled "Languages" and it leads to a sub-page titled "languages." However, unlike what some parents might expect, this page does not translate the web page and sub-pages into additional languages, despite the following announcement under the heading "English": "We are fortunate to enroll students representing 117 different languages." This pages merely links parents to the district's "International Student and Family Services" division and shares phone numbers for translation services in just four languages, namely, Spanish, Korean, Chinese, and Chinese.

In contrast, the home page of Clark County School District in NV is overtly linguistically, as well as student-and parent-inclusive. Under the "Trending" box that's center stage on this page, the parent portal is the first link and it's titled "Campus Parent/Student Information." This page leads to a Parent/Student Information page. The "Parent" tab lists a variety of resources, including FACES, which stands for "Family and Community Engagement Services" and is also accessible through the "Community" Tab and information is provided in multiple languages.

10. Recommendations

Online instruction needs to be more rigorous and visibly student-and-parent engaging and empowering. Some students, parents, and trained educators have spoken, and it is time to hear them, and to make necessary changes. We could begin by addressing the challenges that respondents in the present study shared. While they might not be applicable to students in every single school district, and a customized needs assessment would be needed to identify challenges at individual schools and/or districts, we should at least give some thought to how to better meet the needs of our "ELLs" and our special needs students and families.

10.1 Provide Genuine Student Family Support and Linguistic Inclusion

Genuine student and family support and linguistic inclusion generally work in tandem. Administrators, instructors and support staff must also make an effort to bilingualize instruction and communication with parents and students not only to ease comprehension and to fortify relationships, but to engage and empower (see Author, 2018, 2017, 2010, 2014). Consistent bilingual messaging was not apparent in the web page of none of the schools under focus in the present study. Arguably, use of even a single word from a student's idiolect (i.e., individual language including mannerisms) could positively impact the

instructional-learning dynamic (Author, 2015b). Indeed, this simple step, of providing input that is genuinely comprehensible to each student, is more likely to motivate each student and family, so that they pay (closer) attention, and actually feel like they matter, and are more likely to want to participate. Translanguaging by students and educators has also been shown to significantly improve learning outcomes (Canagarajah, 2013, Author 2013). This practice is a sure way to empower ESL and special needs students and parents. How so? Students and parents are less likely to tune out when they understand what is being shared. Monitoring each student's receipt and incorporation of feedback (both direct and indirect) is another essential step advised.

Family professional development facilitated by ELL parents and/or other family and community advocates in home languages and or bilingual mode must be prioritized. These would be vital learning opportunities for trained educators and program administrators

We need to remember that one size doesn't fit all. Differentiated instruction requires individualized attention, and use of instructional strategies and accommodations. Educators need to try to check in weekly with each ESL and/or special needs student, and also with each family. This practice will likely engender and yield team support and enhanced learning outcomes. Just because the vast majority of American residents are accustomed to written communication and enjoy the flexibility that e-communication affords them (e.g., of accessing written media their own time, and revisiting e-messages any number of times--some of the plus points of written communication), and writing creates a paper trail that most administrative institutions require does not mean that written communication is as comprehensible and accessible to every student and families.

In the present study, Spanish-speaking participants consistently shared that written communication, including email and Canvas announcements, for example, are arguably harder and more challenging for students and even parents to access and fully comprehend, and not-as-motivating. Many students reported that their Online classes during the spring, 2020 COVID-19 school closure period were relatively formal, distancing, and, in many cases, off-putting. Keeping subtle and not-so-subtle differences in (sub) cultural communication and learning styles in mind, we need to be more linguistically responsive and to utilize more oral and visual (tech) tools to engage and empower parents/families and students. Language researchers have consistently advised exposing language learners to rich spoken language. Oral language is necessary fodder; it constitutes an essential language-literacy ingredient, namely, comprehensible input. Oral language is also a natural precursor to written language, and spoken language is arguably much more engaging than written language.

10.2 Monitor Language Literacy Essentials

Research shows that students need three vital ingredients to grow their language and literacy skills, namely, a sustained stream of language that makes sense to them (i.e., comprehensible input or CI), opportunities to use the language (i.e., lots of interaction), and opportunities to receive and implement feedback both direct and indirect (Pandey, 2014b, 2012). Realistically, 45 minutes with six or more ESL students each arguably with varied skill sets--leaves little time for periodic and comprehensive comprehension checks, and for peer-peer and/or even student-teacher interaction. Is sufficient time available for feedback provision and integration of feedback? Follow-up through one-on-one exchanges or observations of students interacting with assigned partners is essential, as is careful tracking of their receipt and integration of feedback across the language building blocks spectrum sounds (i.e., pronunciation and paralinguistic elements; appropriate use of word and sentence-level stress, intonation, and of proper tone and pitch); words and idioms (technically "words" and particularly confounding for ESL students), phrases, clauses, sentences, and discourse structures. Through carefully planned assignments and strategic pairings and groupings, ESL instructors and support staff could indeed ensure that each and every ESL student receives all three ingredients in a sustained fashion. How so? They could divide the class up to create peer chat partners (for listening and speaking) and arrange for both in-class and external writing and reading mentors who would be required to periodically check in with students via telephone and/or video, and teachers could randomly monitor these sessions (via three-way calls, or camera-off Zoom and/or Google meet participation) and/or by asking students to record and email their exchanges via the Voice Memo feature on most smart phones and/or some other mechanism, such as Flip grid. Both the assigned peer tutors/mentors and the ESL students could be invited to periodically share a written or audio file via email and/or Canvas, outlining what they learned from each session and identifying areas of improvement, so that they get into the habit of summarizing and evaluating (i.e., growing their critical thinking skills). In short, peer tutoring could be used to fruitfully engage all parties and enhance the communication (spoken and written), critical thinking, and literacy (reading and writing) skills of each and every student. ESL students would have the necessary interaction time, plus such mandated sessions would allow ample time for feedback provision and integration of feedback. To empower each student, the mandated exchanges and tutoring could go beyond language. For instance, ESL teachers could work with content area teachers to identify the most suitable math and science (bi) weekly pairs and groups, vs. art and humanities pairs or groups. ESL students with similar or visibly advanced content area skills could be paired with non-ESL students who could offer their enhanced English skills in return, so that each student would be bringing something to the table.

Clearly such pairings and groupings would require more planning on the part of the instructors, but the outcomes would very likely be more in line with instructional learning objectives. Periodically changing the student pairings and groupings would further diversify this experience and allow for students to learn to work with more and different classmates. This in turn, would force them to get out of their comfort zone (visible in student- or self-selected groupings as opposed to instructor-selected pairings and groupings), while in turn enhancing their confidence and communication skills, among others. To guide them with successfully attaining the outcomes desired, students could be provided with checklists. Examples are provided in Appendices D & E.

The linguistic movement or advancement of ELL students is ensured when at least four vital spokes are part of their instructional/learning wheel, namely:

i. CI

ii. interaction

iii. feedback receipt must be timely, meaningful (i.e., must make sense), and sustained

iv. feedback incorporation must be effective and timely for acquisition (unconscious) and/or learning (a conscious process).

Additional spokes include motivation (self- and external), mental state (the lower a student's anxiety, the more conducive to language acquisition or conscious learning), instructional style (i.e., how engaging an instructor is and often, the more animated and interactive a teacher, the more likely the student is to pay attention to what they are trying to convey), approach to teaching (activities and examples-infused, student-oriented, learned-facilitated vs. lecture-heavy).

10.3 Observe & Listen to Students & Cater to Individual Needs

The present study suggests that we need to listen to our students more closely, and this includes inviting their input via weekly audio or other student feedback that we request and encourage. We also need to observe them when we and other instructional aides (e.g., peers, parents, siblings and school support staff) interact with them. The nonverbal language components such as facial expression, posture and body language are essential for impactful exchanges. Then and only then can we cater to their individual needs.

What Bryan shared in his written reflection is very likely what many other ESL students are experiencing around the country. "Sometimes I don't know something that I don't know" (i.e., I sometimes don't know what I don't know), he wrote. His writing is a call to action, and while his spelling, mechanics, grammar, and handwriting reflect the language and literacy challenges he faces, his message is clear, loud, and powerful.

The following are additional recommendations for all teachers and support staff:

Support ELL and special needs instructors and support staff since they are some of the most burdened. Content teachers can assist with ELLs and can be trained as needed, as most have or are likely to have (in the near future) at least one ELL in their classes

Invite ESL teachers or trained ESL support staff and/or volunteers to attend content classes or make class recordings available for them to view individually and with their students, so they can monitor each ESL student's understanding and their level of engagement, and assist each with comprehension via, for instance, the private chat feature on Zoom, and/or through simultaneous phone calls or audio messaging via WhatsApp. Participants consistently shared that WhatsApp was the preferred communication platform, and it only requires Wi-Fi connectivity and does not use up minutes, which is an added plus. These language aides could then offer customized and differentiated one-or-one or pair or small-group follow-up sessions and pay closer attention to each student's receipt of CI, interaction and the time that they get to communicate, as well as feedback received and/or incorporated. For instance, for those Special needs students who have trouble processing written text (i.e., reading and writing), they could employ audio messaging

Incorporate pair or small group work. Assign ELL students buddies from their grade or another and strategically pair students to ensure two-way instruction of specific content

Don't just give students busy work or no work; teach them responsibility, put them in charge of their own education, and provide them with proper guidance (i.e., check in with them individually).

Assign ELL-relevant videos (or other visual aides) and audio material daily and to cover non-instructional days (e.g., Friday to Sunday in HCPSS, Maryland) to ensure that ELL students have a steady stream to CI, and create thoughtful and relevant activities to ensure that they get to interact with the content and apply key concepts under focus as comprehension is generally greatly facilitated and often accelerated through interactive and application-oriented tasks

Similarly, team up with other colleagues (much like IEP teams coordinate and engage to work to enhance the learning of special needs students) to ensure that these students get lots of feedback and on a sustained basis from both content instructors and ESL instructors and/or support staff. One of the challenges is that when homework completion isn't required, ELL students in particular do not uniformly have the opportunity to receive feedback

Follow up with these students to ensure that they understood the feedback you shared in writing and/or through other means or have an ELL specialist assist you with this.

Engage with and empower parents so that this in turn will empower students. How so? When we connect with parents and make a genuine effort (e.g. by getting to know them, learning much more about each student's interests and needs, their learning styles and their and their parent's learning objectives, finding out who resides in our students' homes and communities, what challenges they face (e.g., restricted learning space, physical noise, chores/distractions, social isolation, among others), we can make more informed decisions and can offer more relevant and customized instruction and work with parents to tap the resources available in student's homes, social contact circles, and in their immediate neighborhoods, including siblings, other family members, neighbors, friends, and advocates.=

Enlist the help of parents with monitoring each student's access to these vital language-literacy ingredients. This can be done in a number of ways. Parents or siblings could be invited to monitor each student's language-literacy growth via a tracker or checklist and to send a brief daily audio or email file with their observations to the ESL teacher who can then work with the parents to make necessary changes.

Enlist the help of parents with recording students while they are taking class or while they are interacting with siblings and others (which parents could oversee).

Counselors should routinely check in with each ESL and/or special needs student daily or at a minimum weekly. It would help if they speak or at least understand the student's home languages. If not, they should be encouraged to attend Survival language classes and/or enlist the help of interpreters, particularly other students who know these languages and who attend the same school (i.e., mentors).

Instituting a student mentorship program (by language and grade and/or subject area) at every school would be immensely helpful to all students, particularly ESL students.

One of the student participants in the present study shared that while it was helpful to receive an emailed overview of the modules to be covered at the start of each week, he recommended providing ESL students with an additional list of focus areas before each session, as well. He also recommended that teachers follow up with each student via phone or a video-based medium to check on their understanding and ability to apply what they discussed in each session. "This, to me is how you can tell if a student learns what you talk about in the Google Meets part," this 13-year old shared. Instructors could also spell out learning objectives or outcomes before and after each session, so that students are constantly reminded of key constructs and other content to pay close attention to, and instructors and students are on the same page. In general, while ESL students reported receiving some comprehensible input (CI) in English through classroom materials/ resources and weekly sessions, all of the participants reported, needing many more opportunities to use English, and to receive and incorporate feedback. To reduce their instructional burden, and to enhance and

diversify the CI that ESL students are exposed to, course instructors could plan their courses in such a way as to allow all students multiple opportunities for interaction and feedback. For example, feedback could be invited from multiple sources, including peers, volunteers, and support staff.

10.4 Employ Relevant Resources and Genuinely Interactive Technology

Instructors and support staff must carefully select resources and technology because not all resources are meaningful and relevant, and not all tech platforms are truly interactive and ELL- and/or special needs-student-engaging. Flip grid and other audio- and video-based instructional apps could and should be utilized by instructors to track via brief 90-second video segments that students can submit relatively easily—each student's interactivity, and to assess the nature, quality, and frequency of the feedback they receive from assigned buddies/peer tutors and others. Even the voice memo feature on most smart phones could be used to engage students and peers. All students could be required to engage in and periodically record clips of their peer interactions using voice memos, Google Meets, Zoom, FaceTime, and/or WhatsApp's audio messaging, and to share clips with their instructors. The topics for these interactions (most of which will be in English--with the potential for translanguaging) could be open or assigned. Instructions to students could read as follows: "After reading "No Food is Healthy, Not Even Kale," and/or viewing the TED talk titled "Gut Health," discuss with your assigned partner the author's main point and support points/reasons proffered and video-record your partner sharing what they consider to be strengths and weaknesses in these resources. Next, videotape yourself as you offer them recommendations. Pay close attention to their delivery (including their pronunciation, their tone of voice, pitch and intonation, word choice, sentence structure, organization of ideas and the informational content. Record their reactions to your recommendations, as well and upload at least one video clip each (i.e., 1.5 minutes per participant) onto Flip grid by 5p.m. today."

10.5 Integrate Peer Exchanges

Peer exchanges are essential to student learning. Instructors and support staff should make every effort to incorporate pair or groupwork into every class session and assignment. This is because peer exchanges help most students learn better and faster (Briggs, 2013). They have also been found to motivate most students because students can see that they are not alone, and what their classmates say verbally and nonverbally speaks volumes to them. By routinely assigning each ESL student one or more peer mentors/tutors and/or by creating weekly student groups or partnerships, all students would get to interact with their peers on a sustained basis. Genuine (self-initiated as opposed to brief instructor-prompted) interaction is an essential element on learning at all levels, pre-K through college and beyond. The current set-up does not encourage this. Dan shared that only one teacher had

assigned a group activity and that his classmates and him were invited to record themselves and upload the audio file onto canvas only once in the course of the eight weeks of Online instruction. He shared that these were the only two assignments (the group and individual recording) that he enjoyed and he clearly found them to be more self- and peer-motivating and interactive than the assignments his classmates and him are routinely assigned. While some students and families have taken the initiative to interact with selected classmates and other friends through (group) chats via Snapchat, Face Time, WhatsApp, and Zoom outlets, and a few ESL families (all middle to high income) have reached out to religious organizations to continue to receive Online ESL tutoring for their children once or twice a week, the vast majority of low-income ESL students' language needs have been minimally met in the three school districts under focus. Based on our limited data, Clark County Schools appeared to be quite well positioned to ensure continuity of learning for their ELLs. Of the three school districts under focus in the present study, this particular one serves the largest number of ELLs mostly Hispanic and Spanish-speaking. It also has the largest number of bilingual and Spanish-speaking employees, and designated Parent and Family Engagement specialists at every school. As part of their professional development, employees are offered Survival Spanish and even Certification-oriented Spanish language classes with commensurate salary increases offered as an incentive (Lantz, personal communication). These hallmarks are correlate-able and noteworthy.

10.6 Employ Multimodal & Multilingual Communication

As regards communicating with families, multimodal and multilingual communication are highly recommended, so that parents who are unaccustomed to email or who might not use email have the benefit of constant communication via text messages, and all parents/caregivers are clear about expectations, learning format, delivery mechanisms, subject areas, instructional times, deliverables, timelines, and assessment mechanisms. In the case of the HOCO communication aimed at informing families about plans for "continuity of learning," to hear the content and to determine its accessibility, the researchers called the numbers listed for specific language groups on the evening of April 2 (after hours), and left their number on the Spanish-line voice service in case they need more volunteers to reach out to families. No call for volunteers or a link for language translators is provided

Instructor self-reflection is very helpful for planning and assessment purposes, so instituting an electronic system that would allow for educators to continually and quickly share their reflections on each student with team members involved in ensuring maximal student learning (e.g., content teachers, ESL instructors, para teachers and parents) would be beneficial for all parties and particularly ESL and special needs students. The record feature on most smart phones, for instance, could be helpful in this regard. Files could be saved and instantly saved with all parties via text and/or email and they could be accessed by recipients on their own time and such reflections and sharing would both prompt and enable greater collaboration and more

timely and effective intervention Periodically convening focus groups consisting of parents, students, and educators at individual schools--via Zoom or Google Meets and Online groups, such as WhatsApp chat groups would be immensely beneficial in enhancing services to ESL students and families. Timely resource creation and resource sharing could be one of the outcomes. Given that ESL instructors and support staff reported feeling overwhelmed and overworked in all the school districts under focus (and we can extrapolate that this is a problem nationally) and many are unable to attend every single content class, recording classes would be helpful. Then instructors, para-teachers and other support staff and even parents or assigned mentors and partners could access these lessons, mine key words and other language (e.g., sentence structures) and associated content, and work with ESL students to ensure that they understand key concepts in each and every content class. This approach would allow for them to check on each student's comprehension (i.e., what parts they did or didn't understand) and would be more effective and comprehensive than simply asking a student if they understood the lesson and/or whether they had any questions. Google Hangouts' record feature, while not 100 percent accurate could be employed during class sessions (with chat-based edits shared by instructors when errors in the captions are glaring or when they limit or interfere with comprehension). This way, students would also be steadily exposed to print and would get into the habit of reading and would find themselves having to engage in faster eye movement and could, in fact, acquire speed reading skills to keep up with the 'moving' screen. The ESL educators who participated in the present study identified reading and writing as the biggest learning/language challenges for their ESL students, so a sustained source of comprehensible input (i.e., print) would be helpful, and ESL students could be prompted to type in the chat bar and to steadily increase their literacy habits consciously and unconsciously so that they begin to see and use written English more and more, and teachers can respond by providing direct or indirect feedback in the form of corrections to spelling and sentences (e.g., grammar) and punctuation through private chat (as in Canvas) and/or through Track Changes in Word. Track Changes also allows for audio comments, so these can be inserted, as well.

11. Directions for Future Research

Follow-up studies that examine social media postings by schools and educators, including parents, would be helpful. A look at students' social media postings during the COVID-19 school closure period could also be eye-opening.

Research on the role played by family and community engagement specialists where available during this crisis-induced school closure would be helpful for instructional planning purposes. Clark County Schools in Las Vegas, NV, for instance, has a Family and Community Engagement Services Division. Family and community support staff have a variety of designations and roles. In HoCo, Maryland, the term Hispanic Liaisons is employed, yet many schools with sizeable

Hispanic populations do not have a Hispanic Liaison. Parent and Family representatives must be hired and/or volunteers solicited and obtained in the interim for all "ELLs." On the policy and federal funding front, Title III funding must be increased and certainly not discontinued, as is the case in the present President's proposed educational budget for fiscal year 2020. According to the non-profit UnidosUS, "The English learner (EL) population grew by 11.35% between 2009 and 2016, but Title III funding the only federal funding stream dedicated to the teaching of ELs has decreased by 9.4% since 2009." "Latinos are overrepresented among ELs, making up nearly 4 million English learners in K-12 public schools," she continues. "In order to ensure that millions of Latino students are not denied a high-quality education and inhibited from contributing to the United States economy, Title III funding must increase" (e-mail from Kendall Evans, dated June 9). Income disparities and their impact should be studied, as income does appear to impact students' access to English language and other academic and social support services. For instance, in the present study, only one student, the single respondent from a middle-income home (with a household income of over $90,000) was receiving weekly English tutoring, as well as tutoring in two academic subjects and she had access to these support services even prior to the COVID-induced school closure. In short, by virtue of advocacy from her middle-income parents and her community, she is relatively well-positioned academically, and she arguably has a greater chance of satisfactorily and successfully completing her courses, and will very likely continue to steadily grow her English language and very likely too, additional skills through the summer of 2020 given that she will continue to receive instructional help, unlike the other (low-income) students who participated in the present study.

12. Funding & Policy

Advocacy in the educational policy realm is also needed. If the U.S. House of Representatives were to add an additional one billion dollars in Title II funding to the COVID CARES Act package, we might have adequate funding to cater to this growing segment of our school-age population, namely ELLs. Funding could eb used for PD and to hire Family and Community liaisons and other key educators as opposed to schools and instructors would need to ensure that they collaborate with students and their families to effectively ascertain each family's needs and to be responsive. They could conduct a needs assessment and a follow-up survey to gauge success. Inviting feedback from more area students and families would be helpful and essential. Given that the instructional and learning onus is now more visibly shared by trained educators and parents, trust-building and collaboration between these groups is essential.

References

Baume, D. (2018). Towards a measure of learning gain: A journey. With obstacles. *Higher Education Pedagogies,* 3(1), 51-53. doi:10.1080/23752696.2018.1467213

Briggs, S. (2013). How Peer Teaching Improves Student Learning and 10 ways to Encourage it. *Inform ED*, accessed June 15, 2020 at https://www.opencolleges.edu.au/informed/features/peer-teaching/

Canagarajah, S, Ed. (2013). Literacy as Translingual Practice: Between Communities and Classrooms. New York: Routledge.

Crawford, J. & S. Krashen (2015). *English Learners in American Classrooms: 101 Questions, 101 Answers*. Portland, OR: Diversity learning K 12. 2nd Ed.

Galvin, Gaby (2018). The 10 Richest Counties in America. *U.S. New and World Report*, Dec. 6. Accessed June 12 at https://www.usnews.com/news/healthiest-communities/slideshows/richest-counties-in-america

Seale, Colin. (2020). Parent involvement has always mattered. Will the COVID-19 pandemic make this the new normal in K-12 Education? Forbes, May 19. Accessed June 14 at https://www.forbes.com/sites/colinseale/2020/05/19/parent-involvement-has-always-mattered-will-the-covid-19-pandemic-finally-make-this-the-new-normal-in-k-12-education/#7cebfbe25e46

Pandey, Anita. (2018). Foreword to *Handbook of Research on Engaging Immigrant Families and Promoting Academic Success for English Language Learners,* edited by Onchwari, Grace & Jared Sagini Keengwe, https://www.igi-global.com/book/handbook-research-emerging-immigrant-families/21652 3

Pandey, Anita. (2017a). Foreword. *The Handbook of Research on Pedagogies and Cultural Considerations for Young English Language Learners*. PA: IGI Global. pp. xviii-xx.

Pandey, Anita. (2017b). Conversation Analysis as a Powerful Tool for Enabling Social Justice: Focus on Children's Play and Media. *Issues in Early Childhood Education.* Hauppage, NY: Nova Science Publishers.

Pandey, Anita. (2015a). Beyond Black 'n White: Addressing Cultural Bias in Word Gap Studies. *Teachers College Record*, August 25.

Pandey, Anita. (2015b). Why I Kept My Kids Out of Preschool. *Perspectives.* April-June. Pp. 1-9. http://www.nabe.org/Resources/Documents/NABE%20Perspectives/NP_37n2_Apr_June2015.pdf

Pandey, Anita. (2014a). "Pero Extraňo Mi Mama": Immigrant Children Lost in Transition. Invited piece, *Childhood Explorer.*

Pandey, Anita. (2014b). Chapter 3 In *Young English Language Learners at School: A Guide for Early Childhood and Elementary Leaders,* edited by K. Nemeth, Caslon Publishing. Pp. 60-69. http://caslonpublishing.com/titles/14/young-dual-language-learners/

Pandey, Anita (2014c). Using Mother Tongues as Building Blocks in Childhood Education. *Childhood Education*, Vol 90 (1), 61-67. http://www.tandfonline.com/eprint/rrJSVSqMP5Nzub5HssAn/full#.UtS5stJDsS4

Pandey, Anita. (2013). "When "Second" Comes First--हिंदी to the Eye?: Sociolinguistic Hybridity in Professional Writing. In *Literacy as Translingual Practice: Between Communities and Classrooms.* Edited by Suresh Canagarajah, New York: Routledge. Pp. 215-227.

Pandey, Anita. (2012).*Language Building Blocks: Essential Linguistics for Early Childhood Educators.* New York: Teachers College Press, Columbia University.

Pandey, Anita (2010). *The Child Language Teacher: Intergenerational Language and Literary Enhancement.* Manasagangothri, Mysore: Central Institute of Indian Languages.

Pandey, Anita (2008). *Chat-Pat: A Taste of Hindi*. Washington D.C.: Department of Health & Human Services.

Appendix A: Student Survey (N = 11)

Selected audio files can be accessed at: https://soundcloud.com/pandey-interview/sets/students-interviews

1. Describe your sentiments one week after school closed (ecstatic/very happy, happy, neither happy nor sad, both happy and sad, sad, very sad, other)
2. Describe your sentiments one month after school closed (ecstatic/very happy, happy, neither happy nor sad, both happy and sad, sad, very sad, other)
3. Describe your sentiments 2-3 days before remote classes started
4. Did your school send instructions?
5. Through what media? (email, text, phone calls, other)
6. How much do you believe you have learned through remote classes? [not enough, a little, a lot]
7. Have you had the same or similar opportunities to hear and practice English since school closed? Yes No:___, Why or why not? comments:
8. What have been your most difficult or challenging classes and why?
9. Do you turn your camera on or off during class, and why?
10. If you were your subject areas and your ESL teacher, what would you do differently? Any recommendations for any or all of your teachers?

Appendix B: Parent Survey (N = 10)

Link to selected audio files: https://soundcloud.com/pandey-interview/sets/esl-parent-interviews

1. Gender: ___ male; ___ female
2. Age: ___under 18; ____18-25; ___26-35; ___over 36
3. Total monthly household income during the quarantine: __under $2,000 a month; __$2,000 to 3,000; over $3,000___
4. Family designation or role of person completing this survey: ___parent; ___student; ___other [please specify]

5. Your highest level of schooling: ___ elementary; ___ high school; ___some college; __finished college
6. The language you use most frequently to communicate with your children and other residents:
7. Other languages heard, understood, and/or used in your home:
8. Total number of children living in your household, and their ages:
9. Do additional adults live in your home? __ Yes; ___no. If so, how many? __2; __3; __ more than 3
10. Which of the following best describes the type of work you do: ___services; ___farming;__office work; __ other (please specify type of work: ____)
11. Are you working during the quarantine? __ Yes; ___no
12. Has your child(ren) had opportunities to practice hearing or speaking English during the quarantine? __ Yes; ___no
13. Has your child(ren) learned anything during the quarantine? __ Yes; ___no
14. How have you been trying to help your child continue learning during this quarantine? __ Yes; ___no. If you answered "yes," please share what you have been teaching your child(ren) and/or how you have tried to help them continue to learn: _____
15. Did your child's school or teacher contact you/your family during the quarantine? __ Yes; ___no
16. If "yes," what language did your child's school and/or teacher use to contact you during the quarantine? _____
17. Has your child's school/teacher provided you with information or guidelines to help your child(ren) continue learning during this quarantine? ___ yes; ___ no
18. Please rate the timeliness of the messages you received from your child's school following school closures: Poor (1) Fair (2) Good (3) Excellent (4)
19. Please rate the quality of the communication you have received from your child's school during the quarantine?

    Poor (1) Fair (2) Good (3) Excellent (4)
20. Please rate the value of the information you have received from your child's school or teacher during the quarantine

Not at all useful (1) somewhat useful (2) useful (3) very useful (4)

21. Did you find the communication you received from your child's school or teacher during the quarantine reassuring? __ Yes; ___no
22. Which of the following mechanisms did your child's school use to reach you (check all that apply)?

____Phone calls ____written text messages ___Audio texts on Whatsapp ___Email

___Postal Mail ____home visit(s) ___Other [please specify____]

23. Which of these mechanisms or what additional forms of communication would you like your child's school or teacher to use (please list here in order of preference)? _____
24. Has your child been assigned any homework during this quarantine? ___ yes; ___ no

25. Would you like your child to be assigned homework during the quarantine? ___ yes; ___ no
26. Would you like your child's teacher to check in with or connect with your child during this quarantine? ___ yes; ___ no
27. Using what platform? (check all that apply) ____Phone calls ____written text messages ___Audio texts on Whatsapp ___Email ___Postal Mail ____home visit(s) ___Other [please specify____
28. Is your child ready for remote learning? ___yes; ___no
29. What obstacles or challenges is your child and/or family experiencing with non-face-to-face instruction (check all that apply): ___ no tablet or computer; ___ limited quiet space; ___ other (pls. specify)
30. Did your child's school or teacher offer to provide your child(ren) with a learning device (e.g., tablet or computer) during this quarantine? ___ yes; ___ no
31. Please rate your level of anxiety as regards your child's education during this quarantine?

    Not at all worried/anxious (1) Somewhat anxious (2) Anxious (3) Highly anxious (4)
32. Please rate each child's level of anxiety as regards school work or learning during this quarantine?

    Not at all worried/anxious (1) Somewhat anxious (2) Anxious (3) Highly anxious (4)
33. If you have an elementary school child and a middle and/or high schooler, are you more worried about any one child's continued learning? ___ yes; ___ no. If you responded "yes," please explain why:_____

Please share additional comments or recommendations for your child's teacher or school here:

Appendix C: Educator (Instructor & Administrator) Survey

(N = 13)

Directions: Please take a few minutes to respond to these questions. Your feedback will be used to enhance school communication and student success everywhere. Thank you kindly!

1. What is your job title and/or role and approximately how many ELLs do you serve directly or indirectly?
2. What forms of communication are you currently using or planning to employ with your students/families during this COVID-19 school closure period?
3. Have your ELL students and/or families been contacted? If so, how and what has been your primary reason for reaching them (e.g. to inform them, to inspire them, to . . .), and how soon after schools closed did you contact them?
4. How successful have you been at meeting their needs?
5. Are any communication outlets, communicative media/resources, and/or messages intended solely for your EL families/students? If so, what forms do they take (e.g., web pages, email messages, phone calls, texts, specific apps), and why and how are they being utilized?
6. How successful are these individual communication mechanisms or platforms in reaching and/or empowering EL students and families? Would you recommend one or more tools?

7. In your experience, how valuable are these communication mechanisms, and the messages that they are being used to communicate in meeting the needs of EL students (i.e., in supporting their continued learning) and in reassuring students, parents and other family members in the areas under focus?
8. In your opinion, how successful have you been in reaching and/or meeting the needs of your ELL students during this quarantine-imposed school closure period?
9. What challenges, if any, have you encountered in reaching your ELL students/families, and what are some solutions you would propose?

We would like to thank and acknowledge your contributions in publications and other outcomes (e.g., PD sessions) resulting from this survey. Please share your name if you are willing to be officially recognized.

Appendix D: Pandey's Sample Language Trackers for Teachers, Parents, & Other Care Providers of Infants & Toddlers

This tool can be modified to monitor the language and literacy progress of any language learner.

Date: ____ Child's name & age: _____

Language(s) the child hears &/or uses:_____

Time of day (Circle): morning, lunchtime, afternoon, evening, night

Location (e.g., classroom, playground, home, friend's house, at community event):

1. Is the infant/toddler hearing conversations or spoken language? Yes__No____
2. Does s/he respond to spoken language or visuals (i.e., turns head or eyes) ?: Yes ___; No ___;
3. Does s/he show that they follow the conversation or exchange? (e.g., through responsive actions, like head nods, or through eye movements, and/or other signals?): Yes ___; No ____
4. Is anyone interacting with your infant? Yes__No____ Who?___ Age of person? ____How? _____For how long? (minutes/hours): _____
5. List all individuals that typically interact with the infant, & their estimated ages: e,g., Alex (boy, 9)
6. Does the infant appear to be equally responsive to each person in the exchange? Yes__No____ Explain (e.g., they get more excited and make more noises when they see or hear ___):
7. Who asks the infant the most questions? __a child __one or more adults __nobody
8. Does anyone try to get your infant to make noises and/or respond in other ways? Yes__No____ How so ? Explain (e.g., by providing the correct pronunciation for wawa/agua → Waa'teR!)
9. Which of the following best describes the kind of language (including nonverbal) that one or more interactors uses with your infant? (feel free to select more than one)

Uses mostly nonverbal lang./signals (mime &/or gestures): ___

Uses mostly statements, many of which invite participation:

Uses mostly directives (e.g., stand up, sit on the rug, clap/raise your hands):

Uses mostly music or songs and encourages participation (through speech or gestures):

Allows the child to lead through signals or talk, & responds when appropriate:

Uses mostly sounds and varied intonation (e.g. kissing smacking, farting, "peek-a-boo"):

Asks many yes-no questions___ Asks many Wh- Questions___

Asks many rhetorical Qs (no answer required)___ Asks many tag Qs (e.g., don't you?)_____

Other (e.g., hums, sings, plays music, dances and awaits infant's reactions, etc. pls. specify):

10. Identify the kinds of questions, if any, used by any adult(s) interacting with the child:

Yes-no Questions___Wh-Questions __ Rhetorical Qs ___ Tag Qs ____Indirect Qs ("I wish . . .")

11. Describe the language(s) the infant uses: (e.g., gestures and some Spanish words—no, mio!)
12. Circle all language you observe the infant using:

    ___Nonverbal signals (e.g., , drawing, frowning, smiling, laughing, crying, running to initiate a chase, screaming, nodding, pointing, pulling— to show something, etc.):___ sounds; __words __phrases (2 or more words used for an idea; e.g., No- mio for It's mine!) __ self-initiated turns (e.g., Mama? inviting participation of person identified, for example). Other [pls. specify]: ____
13. Does the infant observe eye movements and/or turn-taking rules (for joining and/or leaving a conversation? (e.g., wait for a suitable opportunity to interject, allow for their peer or another speaker to finish talking, and/or show that they follow)?: Yes ___; No ____ ; Unable to determine:
14. Do you spot any written language in the observation space? (e.g., letters, words, pictures of family and friends, books) Yes__No____ Does anyone draw attention to written language? Yes__No____
15. Does anyone draw the infant's attention to written lang. ? Yes__No____
16. Does the child have an opportunity to USE their primary language (e.g., a language used in their homes) Yes ___; No ____Unclear ___
17. Does the infant have an opportunity to hear and/or USE another lang. (e.g., English)? Yes ___; No ____ Through what means? (e.g., a TV program, a video clip, a song, a video game, etc.)
18. Does anyone correct the infant? Yes ___; No___; NA
19. How does the infant respond to the feedback they receive? (e.g., well/they promptly modify their lang./behavior; not well/they ignore the feedback or implement it incorrectly:___; unclear_____)
20. Who provides the infant with the most constructive feedback, and how do you determine this?____
21. Could you use your observations to plan future pairings and groupings? Yes ___; No___ How?
22. List any new sounds the infant hears (e.g., "s" in "treasure"):
23. Does anyone interacting with the infant emphasize specific sounds? Yes__No____ If so, how? __
24. List sounds and/or other language forms that the infant appears to have difficulty either following or enunciating:

25. List the language (e.g., Spanish) and language forms (e.g. sounds) the infant uses:
26. List any phrases and/or longer language building blocks that the infant uses in this exchange:
27. Does the infant initiate interaction (using nonverbal signals or lang. that invites participation, such as participation-eliciting intonation, a question or a directive)? Yes__No____If so, list examples:

Name, age, and title of person completing this form (e.g., teacher, teaching aide, Mom, cousin, friend, age 32)

Start & end timed of observation or estimated length of observation:

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 45-68

Syllables and Reflex Circuit: The Missing Link between Animal Calls and Digital Language

Kimiaki Tokumaru[*]

Abstract

Human language stands apart from all other animal communication systems through its infinite vocabulary and combinatorial structure. We propose that it evolved as a digital system through two key transitions during the Late Pleistocene.

In the extremely dry and flat Kalahari, upright hominies using tools survived by forming symbiotic relationships with animals and plants. They mastered fire, took shelter in caves during the Ice Age, and became asocial using calls to distinguish friend from foe. During the volcanic winter following the Toba eruption (~74,000 years ago), they developed click consonants and began naming the world. Later, with laryngeal descent, they acquired vowels, formed syllables, and began to weave sentences. This marked the beginning of digital evolution.

Syllables functioned as discrete units in both time and frequency shared digital signals that enabled the growth of vocabulary and the birth of grammar. The invention of writing allowed these signals to persist across time and space, accelerating the development of complex civilizations.

We further propose two biological mechanisms:(1) phoneme imprinting in infancy via cochlear hair cell pruning, and (2) verbal memory supported by B lymphocytes in the cerebrospinal fluid, forming an immune network within the brain's ventricles, one of the most noise-isolated environments in the human body. Together, these mechanisms suggest that language evolved not only through physical and social adaptation, but also through logical and neurobiological structures that preserve and process digital signals.

Keywords: Language Evolution, Digital Signals, Syllables, Phoneme Imprinting, Click Consonants, Eusociality

1. Introduction

1.1 Digital Systems and the Emergence of Complexity

The origin of human language remains one of the most profound unsolved questions in evolutionary biology. Language and humanity are deeply intertwined, yet their definitions are often vague and

[*] Satellite System Engineer, Uenogaoka Higashi, Oita-city, Japan; tokumaru@pp.iij4u.or.jp

interdependent, rendering artificial intelligence-based attempts to define them circular and inconclusive.

This paper proposes that human language represents a digital evolution of the analog phonetic communication systems found in mammals. We hypothesize that the emergence and evolution of digital signal syllables drove the complex development of human intelligence, providing a theoretical framework for understanding increasing complexity through low-noise environments and iterative processing.

Analog symbolic communication, which has no phonemes, uses the waveform of a call to express the meaning of a symbol and the volume and frequency of the call to modulate a message. It can only share a few dozen symbols and cannot be constructed grammatically. On the other hand, digital signals can combine phonemes to produce an infinite number of symbols. Grammar can combine symbols to make complex messages. Specialized signals for long-term storage, such as letters and DNA, are created, persist, then stored, and widely shared. Eukaryotes produce progressively more complex structures, such as organelles, cells, organs, individuals, groups, and species. Languages are also evolving in more complex, specialized and differentiated directions. It is evolution that makes an analog system into a digital system. When it becomes digital, an even greater leap comes from it.

The term "digital" is commonly associated with binary computing in electronic circuits. However, digital systems extend beyond computing to include biological systems such as RNA and DNA, which encode the information of life (Kay, 2000). Information theory, as articulated by von Neumann (1963), posits that any integer-based system (greater than 1) can form a digital system, with the signal-to-noise ratio serving as a critical factor in its dynamicity and evolution. What matters is not the number of signals, but that sender and receiver share an identical set of discrete units. Phoneme sharing, in this sense, underlies digital human language and communication.

We define a digital system, applicable to language, computing, and biological evolution of life, as one that: (i) employs discrete, finite signals arranged in a one-dimensional array to propagate through noisy channels without error; (ii) processes error-free signals via sophisticated and complex circuits in low-noise environments; (iii) (as a result of (i) and (ii), reproduces complex structures without loss of complexity; and (iv) evolves to increase complexity over time. In this sense, digital systems are inherently self-amplifying, aligning with complex systems theories.

While numerous attempts have been made to define phonemes or sets of phonemes, no single definition has been universally accepted. In this paper, we define a set of phonemes as a set of mutually distinctive minimal auditory units shared among members of a language community, which enables the generation of an infinite number of unique word-signs through combinatorial arrangement. This definition emphasizes the discrete and combinatorial nature of phonemes essential

for digital signaling, and highlights their community-level imprinting during infancy, which guarantees finiteness another key requirement of digital systems.

Syllables, as digital signals, exhibit discrete frequency characteristics unique to each language, suggesting that a finite set of phonemes is imprinted in infancy. This imprinting occurs through the reinforcement and pruning of cochlear hair cells in response to native phonemic stimuli during the first year of life (Kuhl, 2004). Furthermore, complex information is conveyed by constructing sentences that combine content and function words according to grammatical rules. In this regard, linguistic sentences, messenger RNA, and data packets in computer networks share a common structure (For detailed comparison, see Table 1): information encoded in one-dimensional sequences of discrete, finite signals (Tanenbaum and Weatherall, 2011). Over time, reactive signals such as syllables, RNA, and voltage bits evolved into long-term storage systems, including writing, DNA, and magneto-optical media.

Attribute	**Syllable**	**RNA**	**Electronic Bit**
System	Language	Life	Computation
Emergence Timing	~66,000 years ago	~3.8 billion years ago	~1950s
Signal Medium	Acoustic Energy	Molecular Bonding	Electrical Signal
Finiteness	Imprinted Native Phonemes	4 Nucleic Acids (A, G, U, C)	2 States (0, 1)
Minimum Unit	Syllable (a vowel with consonants)	Codon (64 combinations)	Byte (256 combinations)
Network	B-cell/Microglia Network in CSF	tRNA in Cytoplasm	Shift Register in CPU
Low-Noise Environment	CSF, Cultivated zones, Monastery, Library	Cytoplasm, Nuclei, CNS, Uterus,	Clean Room, Data Center
Linearity	Grammatically Structured Sentences	Messenger RNA	Data Packets
Modulation	Grammatical Words	Alternative Splicing, Epigenetics	Protocol Switches
Meaning	Memory in Microglia/ B-cell Networks in CSF	Protein Function	Device Output
Persistent Memory	Writing Systems	DNA	Optical/Magnetic Storage
Error Avoidance	Phonemic Redundancy (Sound Symbolism)	Codon Degeneracy (64 codons to 20 amino acids)	Error-Correcting Codes

Table 1: Comparison of Digital Signal Systems in language, biology and computing

To ensure signal fidelity, processing occurs in low-noise environments, such as cultivated zones, academic settings, cellular nuclei, or cooled computational systems. Mechanisms like sound symbolism, codon degeneracy (where 64 codons specify 20 amino acids), and error-correcting codes further enhance signal reliability.

The evolutionary leaps non-literate to literate societies can be understood as digital transitions, marked by the evolution of signals from syllables to written characters. The acquisition of bit-like signals

has propelled humanity into a new evolutionary stage, further fostering intelligence and civilization. Understanding these digital principles is essential for elucidating this process.

1.2 The Dawn of Digital Evolution of Human Language

Mitochondrial DNA analysis has firmly established the African origin of *Homo sapien*s during the Pleistocene, a consensus widely accepted in the scientific community. (Cann, Stoneking, and Wilson, 1987) However, the ecological and cognitive conditions under which language emerged remain poorly understood. The prevailing "pan-African model" posits that human evolution occurred across multiple regions of the continent, supported by archaeological evidence of settlements and lithic technologies. (Scerri and Will, 2023) Yet, this framework has largely overlooked the parallel emergence of language and cognitive capacity.

We propose a novel hypothesis: during the Late Pleistocene (= Ice Age), humans developed digital signal syllables through a two-stage phonemic evolution. First, click consonants emerged, followed by the introduction of vowels, culminating in syllables—discrete, finite, and linearly combinable units. The emergence of these digital signals marked the birth of language and enabled the reorganization of neural circuits specialized for processing structured grammar and abstract concepts mechanisms that remain invisible, enigmatic, and procedurally elusive at the deepest levels of human intelligence.\

As evolution builds on pre-existing structures (Deacon and Deacon, 1999), early hominins upright bipeds using stone and bone tools gradually accumulated ecological and symbiotic knowledge over roughly three million years. These hominins inhabited the harshly arid and exceptionally flat environment of the Kalahari (Dart, 1959, Van der Post 1958, Owens and Owens, 1984), coexisting with diverse animal and plant species. Around 300,000 years ago, the control of fire enhanced nutritional efficiency and energy allocation. By approximately 130,000 years ago, with the onset of an ice age, early humans began occupying coastal caves. These stable, protected environments prolonged infant dependency, increasing exposure to social stimuli and fostering both curiosity and inquisitive behavior in tandem with expanding brain volume (Martin 1990). These transitions are summarized in Table 2 as milestones in the digital evolution of human communication

Period	Innovation	Required Adaptation
3 Ma	Tool Use	Bipedalism, tool use, cranial adaptations.
300 ka	Fire Use	Social gatherings, improved digestion.
130 ka (MIS 5)	Eusociality & Alarm Calls	Cave dwelling, cooperative childcare, sound imprinting, expanded brain volume.
74 ka (MIS 4)	Vocabulary	Click consonants during Toba volcanic winter; Still Bay culture.
66 ka (MIS 3)	Grammar	Mental protuberance, laryngeal descent; vowels and syllables initiated digital evolution; Howiesons Poort culture.

5 ka	Writing	Writing systems in Mesopotamia enabled persistent syllables, fostering civilization.
2.5 ka	Conceptual Thought	Philosophers/scientists in quiet environments (e.g., academies, monasteries) developed abstraction.
1950s	Bidirectionality	Electronic signal bits enabled searchable, interactive communication (e.g., databases).
21st Century	Forward Error Correction (FEC)*	Ultra-low-noise environments (e.g., AI data centers, reclusive settings) support FEC and interdisciplinary integration of concepts.

Table 2. Evolutionary milestones in the development of digital signal systems in language, from early tool use to modern forward error correction

MIS: Marine Isotope Stages

* FEC is a technique to correct errors without contacting information senders.

The involvement of grandparents in childcare likely facilitated the emergence of extended family structures, deepening intra-group communication and knowledge transmission. Cave-dwelling populations began to exhibit eusocial behavior, employing community-specific alarm calls to distinguish between allies and adversaries (Wilson, 2012). Against this evolutionary backdrop, the Toba supereruption (~74,000 years ago) and the ensuing volcanic winter imposed extreme environmental pressures (Ambrose 1998), potentially catalyzing further vocal, cognitive, and social adaptations.

2. Hypotheses

2.1 Emergence of Consonants

Around 74,000 years ago, the Toba super eruption triggered a period of global cooling lasting approximately 6 to 10 years, reducing average temperatures by 3–5°C. This dramatic environmental stress confined early human populations to caves, where food scarcity and social isolation intensified the need for vocal communication. Within these enclosed communities, shared alarm calls originally used to distinguish allies from outsiders evolved into shorter, combinable vocal units.

These mimetic vocalizations, which initially signaled group membership, gradually transformed into discrete sound segments capable of forming referential labels for natural elements, emotional states, and social actions particularly through communal activities such as singing. This transition, possibly rooted in playful naming practices and environmental reverence, marks the hypothesized origin of consonantal phonemes, which helped construct a shared symbolic worldview among group members.

Click consonants, still preserved in Khoisan languages of southern Africa, likely represent remnants of this early phonemic stage, encoding names for landforms, flora, and fauna. Archaeological findings from Blombos Cave in South Africa a representative site of the Still Bay industry (~71–72 ka) support this cultural and cognitive expansion, pointing to the emergence of symbolic artifacts and an enriched vocabulary (Henscilwood et al., 2001).

While some researchers interpret mitochondrial DNA patterns

(e.g., specific SNPs) as evidence of a population bottleneck, we propose an alternative view: this period reflects a linguistic innovation the emergence of proto-languages. Consonants, particularly clicks, relied solely on tongue or lip movements and were independent of the larynx or vocal cords. Their lack of grammatical function further supports the hypothesis that consonants preceded vowels in the evolution of phonemic structure (Westphal, 1971, Traill, 1997).

2.2 Emergence of Vowels

Around 66,000 years ago, as the climate warmed, anatomical adaptations enabled the production of vowels. The development of the mental protuberance observed in fossils such as KRM-41815 (Figure 1) from Klasies River Mouth Caves, representative of the Howiesons Poort industry facilitated laryngeal descent and the formation of a vocal tract optimized for vowel resonance (Singer and Wymer, 1982, Rightmire and Deacon, 1991, Morrissey, Mentzer and Wurz, 2023). Repetitive tongue movements associated with click consonants likely contributed to the development of the uniquely human mandible, shaping a vocal tract with balanced horizontal and vertical dimensions, thereby enhancing formant tuning (Deagling, 2012). The robust and shortened mandible, originally adapted in early humans for consuming their staple food in the Kalahari—herbivore bones—later contributed to enhanced vowel resonance capacity, enabling more distinct and stable syllable production. Despite the increased risk of asphyxiation due to laryngeal descent, the retention of this adaptation suggests strong selective pressure possibly driven by an instinctive urge for linearized expression or the emergence of a second digital system of intelligence, following the first digital system: the genetic code of life.

This transformation may reflect an unconscious recognition embedded in the logic of life that language and intelligence have begun to evolve as a second digital system, following principles similar to the genetic code, in which discrete sequences encode and express information.

X-ray imaging by Chiba and Kajiyama (1941) provides physical evidence of this mandibular adaptation (Fugure-2). The integration of vowels with consonants produced temporally segmented syllables phonological units that support linear structuring through grammatical connections. Since then, all known human languages have relied on syllables as fundamental speech units: discrete, finite, and one-dimensional, fulfilling the criteria of digital signals.

2.3 Digital Systems Dependent on Low-Noise Environment

We communicate using sentences built from syllables. In this process, the sequence of syllables spoken by the sender is transmitted to the listener without a single error. This is possible because syllables are

discrete and finite signals; words have strong phonological representations; and our brains are equipped with native-language phoneme maps that enable the recovery of correct sounds even under variation (see Section 2.4).

When the listener receives a sentence without error, meaning is automatically recalled. However, the mechanisms through which meaning is recalled and what kind of meaning is recalled have rarely been studied. For example, everyday utterances such as “What do you want to eat?” or “Where do you want to go?” may invoke mechanisms very different from those required to understand abstract scientific concepts like post-translational modification in molecular biology or entropy in thermodynamics. How are such invisible phenomena in science correctly understood?

This paper does not delve into such mechanisms, but emphasizes the importance of quiet environments, where both speakers and listeners can carefully, slowly, and repeatedly reflect on each word.

John von Neumann (1963) defined digital systems to reproduce complexity without loss and to be able to evolve complexity over long periods, and to understand them signal-to-noise (S/N) ratio is the most important theory.

This point connects language evolution to a broader biological trend: biological evolution can be seen as a series of breakthroughs in noise reduction, particularly during massive extinction events that gave rise to structures like the nuclear membrane, CNS, cerebrospinal cavities, and the uterus. After environmental crises, these internal spaces functioned as incubators for high S/N signal environments, permitting exponential increases in internal complexity: (Table-3)

Period	Physical Mutation for Noise Reduction	Logical Evolution toward Complexity after Environmental Recovery
2 billion years ago	Nucleus	DNA double helix: restriction enzymes; post-translational modification; endosymbiosis with bacteria
540 million years ago	CNS + Ventricles	Sensory organs: motor control; spinal reflex: symbiosis with multiple cells/organs including neural cells/CNS
66 million years ago	Uterus	Umbilical cord: precise growth control; vocal communication; (eu)social group living symbiosis

Table 3. Biological evolution as a sequence of noise-reduction breakthroughs, with subsequent logical complexity growth after environmental recovery

In humans, a parallel occurred: syllables, letters, and bits were born for specific purposes identifying friends and foes, managing land ownership and taxes, or racing to develop missiles. These digital signals, when used in low-noise environments such as caves, cultivated settlements, monasteries, schools, or seclusion, enabled the emergence

of grammar, abstract concepts, and forward error correction. (Table-4).

Period	Signal Evolution (Property)	Logical Evolution toward complexity in Low-Noise Environments
66 KA	Syllables	In cultivated lands safe from predators, humans switched from binaural to monaural hearing of speech, repurposing directional auditory nuclei for grammatical vector processing.
5KA	Letters = Persistent syllables	In monasteries or academies, isolated from worldly duties, people engaged in deep reading and reasoning, acquiring abstract conceptual operations.
60 YA	Bits = Interactive syllables	With computer networks and AI search functions, scientific concepts can be traced to rejuvenating origins, enabling forward error correction and shared cognitive genome.

Table 4. Language evolution toward complexity in low-noise environments

Remarkably, this adaptation was environmental rather than neural, creating a structural asymmetry akin to an evolutionary "original sin": language became digital, yet our brain circuits remained analog and reflexive. The brain did not evolve new architectures to accommodate the complexity of digital language. Instead, humans reshaped and engineered their environments to reduce noise and improve the signal-to-noise ratio in other words, once a new signal emerged through physical evolution, it became necessary to seek out quieter environments where that signal's potential could be fully realized. Human cognitive evolution may only advance in this way.

2.4 Phonemic Maps and the Immune-Based Network for Meaning Reconstruction

Before exploring the internal mechanisms of phoneme representation and memory formation, it is essential to recognize the limitations of reflex-based language processing. The table below summarizes the weaknesses of the sign-reflex system, which remains embedded in the brainstem circuitry (Table 5):

Aspect	***Weakness***
Involuntary System	Lacks mechanisms to form new word memories, associate them with appropriate meanings, or control reactions.
Recognition System	Easily overlooks new or unfamiliar signs. Tends to disregard the significance of unfamiliar information.
Ego-Centric Tendency	Self-affirming with no provision for self-diagnosis or introspection.
Ignorance of the Mechanism of Meaning	Not knowing that different people have different meanings.

Unsuspecting Nature	No mechanism to verify the authenticity of incoming stimuli.
Reflexive Behavior	Reacts immediately without thoughtful consideration.
Stubbornness	No capability for circuit reconfiguration or error correction.
Difficulty in Reconfiguration	Cannot dynamically update, reconfigure or refine language processing rules based on new experiences.
Inability to Handle Abstract Concepts	Inability to get out of the habit of associating words with episodic memory and build pure abstractions
Empiricism	Inability to calmly observe and judge new situations under the decisive influence of precedents
Formalism	Feel secure when words and meanings are connected, and do not ask whether the meaning is correct
Passivity	Lacks genuine or spontaneous intellectual curiosity.
Survival-First Instinct	Prioritizes self-preservation over truth, making it susceptible to coercion.
Selective Truth-Telling for Ingroup Benefit	Information is filtered based on social boundaries: truth is told to ingroup members, while falsehoods or silence are used toward outgroup individuals.
Single-Sign Limitation	Struggles to process complex systems or multiple signs simultaneously; Requires reference models.

Table 5. Weaknesses of sign-reflex mechanisms when applied to language processing in intelligence systems

(Author's Original)

The ability to withhold or distort truth based on social context particularly in distinguishing friend from foe is not merely a cultural artifact but a biologically ingrained reflex. As eusocial animals, humans are evolutionarily wired to modulate signal fidelity based on group boundaries. This reflex, rooted in ancient survival mechanisms, persists beneath the symbolic surface of modern language, contributing to the asymmetry between digital expression and analog brainstem control.

These limitations highlight the necessity for higher-level systems such as phonemic maps and CSF-based memory circuits—to support meaning mechanisms including grammatical language and abstract thought.

We hypothesize that infants acquire native phonemes through the selective reinforcement and pruning of cochlear hair cells. Auditory stimulation from caregivers activates hair cells tuned to native phoneme frequencies, while non-responsive cells are eliminated. This process creates a phonemic map extending from the cochlear nucleus to the auditory cortex (Figure 3). Unlike traditional models that emphasize cortical plasticity, this framework proposes that foreign accents and dialects are perceived through native phonemic filters, without

requiring higher-level cortical inference. By filtering auditory input through pre-established phonemic templates, the system ensures robust and efficient recognition of digital signals.

We further propose a novel hypothesis: the existence of an intraventricular immune network involved in language processing. Specifically, word representations are mediated by B lymphocytes circulating in the cerebrospinal fluid (CSF), forming a dynamic network of antibodies and microglial antigens within the brain's ventricular system.

This hypothesis was supported by anecdotal reports shared at a regional neuro-rehabilitation study group employing the Bobath approach, in which cases of progressive aphasia and dyslexia were observed following normal pressure hydrocephalus (NPH) shunting. We hypothesize that shunting leads to a depletion of B cells in the CSF, contributing to the gradual onset of aphasia. The simultaneous progression of aphasia and dyslexia suggests that B cells not only store vocabulary memory but also generate and maintain logical memory representations such as binary operators like "O" and "X" for visually processed character strings. In this sense, literacy may depend on B-cell-mediated memory, supporting the principle that "without vocabulary, there is no literacy": unfamiliar words lack corresponding B-cell networks.

These network terminals are shaped by the frequency characteristics of native phonemes and form three-dimensional antigenic structures corresponding to the tonotopic maps established in the cochlear nucleus. These structures engage in localized antigen–antibody interactions, transforming phonological inputs into immune-encoded memory networks.

This mechanism offers a new explanation for why foreign vocabularies are interpreted within the framework of native phonemes: the phonological signal activates immune network terminals aligned with native frequency domains. In contrast to corticocentric models that emphasize Broca's and Wernicke's areas, this hypothesis foregrounds subcortical, immune-mediated circuits for language processing. Functioning as a reflex-like system, this circuitry may distinguish native phonemes as indicators of social affiliation or threat—reflecting an evolutionary legacy rooted in eusocial auditory communication.

2.5 Writing: Syllables that Persist

Evolutionary biologist John Maynard Smith noted the inherently digital nature of language, suggesting that syllables evolved into letters, which in turn enabled electronic storage and transmission systems (Maynard Smith and Szathmáry, 2000). Although human language is thought to have emerged in South Africa around 66,000 years ago and gradually spread across the globe, it remained unwritten for over 60,000 years. Writing first appeared in Mesopotamia approximately 5,000 years ago, in the vast fertile plains formed by sedimentary deposits from

continental collisions. Dynasties governing these expansive territories likely developed writing systems to record ownership and taxation, creating a shared memory infrastructure that extended beyond individual cognitive limits.

It was writing that served as the catalyst: it enabled the intergenerational transmission of accumulated knowledge, thereby giving rise to civilization. Unlike ephemeral speech, written characters persist across time and space, necessitating formal educational institutions for their acquisition. Schools, which also originated in Mesopotamia, were developed to teach reading and writing. Writing gave birth to civilization, not vice versa.

Once initiated by writing, civilization advances through the refinement of orthographic systems, supported by laws and cultural policies that ensure societal continuity. Civilization encompasses educational institutions (schools, academies), teacher training programs, professional educators, literary critics, scholars, and public intellectuals. It also encompasses archives and libraries dedicated to the preservation of documents, including the librarians and clerks who manage them, as well as the publishing industry, bookstores, and secondhand bookshops. Civilizations maintain literary and historical traditions, produce dictionaries to standardize language, and foster translation efforts that integrate foreign literature and philosophy into native linguistic frameworks. In this way, civilization progresses as a collective endeavor sustained by those who preserve tradition and contribute to intellectual discourse.

Writing-driven civilization accumulates intellectual output in low-noise environments such as monasteries or academies, where conceptual thinking first emerged. Unlike reflexive word use based on one-to-one associations, conceptual thought operates through a one-to-all logic, as described in Piaget's (1947) additive and multiplicative operations on groups. This shift from connecting words to individual memories to associating them with sets of memories marks the second digital leap beyond grammar: the emergence of concept formation.

2.6 Bits: Interactive Syllables

Vannevar Bush (1945) proposed the concept of MEMEX a machine designed to help humans organize and connect knowledge across disciplines. This vision inspired Douglas Engelbart, whose work at the Augmentation Research Center laid the foundation for the Alto personal computer, developed at Xerox PARC (Palo Alto Research Center) in 1973. That machine introduced many features we now take for granted: graphical user interfaces, the mouse, text editors, spreadsheets, copy-and-paste functions, email, and even the [PDF Portable Document Format] (Bardini, 2000).

These innovations were not merely for convenience. Their original goal was to augment human thought enabling us to collect, manage, and evaluate knowledge across fields. This vision advanced with the

advent of the World Wide Web (1989), online bookstores like Amazon (1994), and search engines such as Google (1998), all of which brought Bush's idea closer to reality by making linguistic and scholarly knowledge globally accessible.

This digital transformation marks the third major step in the evolution of language: from syllables spoken in real time, to written letters that persist across time, and now to interactive bits. Spoken language connects us in the present moment. Writing allows us to reach across generations but only if we find the right book and take time to read it. Bits, in contrast, enable instantaneous search, retrieval, and interaction. They render language dynamic and computationally searchable.

However, digital information introduces new challenges. In noisy environments where signals are easily distorted we must ensure that an author's original intent is preserved. Just like digital communication systems, we require mechanisms for verifying authenticity. One key principle: a document published while the author is alive is more likely to represent their true intentions. After death, the risk of distortion, misinterpretation, or even forgery increases demanding careful scrutiny by readers.

In 13th-century Japan, the Buddhist scholar-monk Dōgen (1200–1253) took deliberate measures to prevent his disciples from altering or fabricating his writings. In his 75-volume Buddhist text *Shōbōgenzō* (Treasury of the True Dharma Eye), which served as an introductory textbook on Buddhist terminology, he numbered the volumes sequentially and added colophons to each, recording the location and date of both his lectures and the final transcription (Tokumaru, 2018). In the colophon of the first volume, he noted that the editorial work had begun in the year before his death, thereby signaling that the text represented the culmination of his teaching.

Likewise, in his ten-volume *Eihei Kōroku* (Dogen's Extensive Record, 1252/2010), he concluded each volume with identifying marks specifying the number of sermons and verses it contained effectively embedding a checksum-like system to discourage unauthorized additions or deletions. It is noteworthy that Dōgen, in his Chinese poetry, twice alluded to the concept of redundancy in information theory through the metaphor of "Frost is added to the snow**.**" This suggests his intuitive grasp of redundancy based error correction systems (Dharma Hall Discourse 473 and 507). These colophons and identifying markers may represent the earliest known use of error-correction codes in human intellectual history.

When an author's authentic voice can be verified, a regenerative approach to reading becomes possible. This method treats the reading process as a form of source error correction reconstructing the author's experiments and insights through their writing. Even if the author made mistakes, the reader can independently identify and correct them. This

self-corrective method mirrors the principle of Forward Error Correction (FEC) in communication systems.

Bits thus function like persistent, addressable syllables. They can be read and reinterpreted not only by humans, but also by search engines and AI systems. In this sense, every human today gains the unprecedented ability to "converse" with the minds of past scientists, philosophers, and thinkers.

This is not merely a technical shift; it transforms the scale of human cognition. As in signal theory, clarity depends not only on the strength of the message, but also on the surrounding noise. The more precise and less distracted the environment, the better we can understand and refine complex ideas. Reading slowly, returning to passages, and creating quiet spaces these are the timeless techniques for increasing signal clarity.

Interdisciplinary research, too, resembles tuning across frequency bands. Each field speaks in its own codes, with unique assumptions and symbolic systems. To integrate diverse sources of knowledge, we must first define scientific terms clearly, avoiding unnecessary jargon; examine whether they exhibit the properties of a mathematical group; and confirm that they emerged from the recognition of previously unacknowledged phenomena, by revisiting the original works and insights of the scientists who named them.

Paradoxically, the very tools designed to amplify human thought now drown it in noise. We live in a world of constant interruptions, multitasking, and digital distractions. Though our tools are more powerful than ever, they are most effective only in environments of quiet focus. Perhaps the true challenge of the digital age is not technical, but ecological: creating the silence necessary for thought (Carr, 2011)

3. Hypotheses Development

The hypotheses presented above emerged through a long-term integrative process, synthesizing insights from evolutionary biology, field linguistics, anthropology, archaeology, psychology, animal ecology, and digital communication theory. Rather than appearing fully formed, the core hypothesis of digital language evolution developed gradually through iterative reconsideration of empirical data, cross-disciplinary reasoning, and keyword-driven exploration of overlooked biological structures and theoretical models, including Jerne's immune network theory and Vigh's CSF-contacting neurons (Jerne, 1974, 1984, Vigh et al., 2004).

Our initial speculation centered on the possibility that Late Pleistocene cave habitation facilitated vocal communication. Caves offered safe, sound-shielded, dark, and acoustically reverberant environments. These conditions suggested that sound more than gesture could serve as a reliable and far-reaching signaling modality.

A key turning point came with comparative studies of eusocial animals especially the naked mole rat (*Heterocephalus glaber*), which recognizes kin through olfactory cues, coordinates behavior cooperatively, and employs only 17 distinct acoustic signals (Pepper, Braude, Lacey, and Sherman, 1991). This realization was profoundly disorienting: darkness, it seemed, does not foster vocabulary. The long-held notion that dark environment in the cave alone drives linguistic richness was cast into doubt. In this conceptual vacuum, a new idea began to emerge that the syllable itself, as a discrete acoustic unit, may represent the true minimal building block of language. Through the recombination of these units, an open-ended inventory of word-signs becomes possible. A digital breakthrough sparked by a moment of insight.

We propose that syllables evolved from eusocial alarm calls into temporally discrete digital units. In eusocial species, the ability to distinguish between kin and non-kin despite similar appearance is vital for survival. While many species rely on scent, humans evolved to use sound alone. This auditory reliance, likely shaped in cave environments, gave rise to syllables as temporally discrete, combinable units forming the foundation of digital language.

Between 2009 and 2012, we searched extensively for cellular and molecular structures in the cerebral cortex that could account for language memory, but found nothing**.** Drawing upon Jerne's theory of immune networks, we began investigating the possibility that language memory is not confined to cortical circuits but also embedded in neuroimmune structures within the cerebrospinal fluid (CSF)**.** This inquiry led to the formulation of the intraventricular neuroimmune network hypothesis, which posits that word representations are stabilized through interactions among CSF B cells, their antibodies, CSF-contacting neurons, and microglial antigen-presenting networks.

Inspiration also came from earlier thinkers. Von Neumann (1963) foresaw a link between information and thermodynamics, and focused on signal-to-noise ratio in digital systems. Alongside Piaget's (1947) group theory of cognition and Maynard Smith's (1999) signal evolution model (syllables → letters → bits), these ideas helped form the foundation of our hypothesis.

Table 6 outlines the intellectual trajectory of these ideas, shaped through long-term dialogue and reflection. Key developments emerged from feedback at international conferences, workshops, and study groups hosted by IEICE, IPSJ, and JSAI, as well as symposia we organized for the Japan Society for Evolutionary Studies (JSES in 2020 "Evolution of Modern Human: The Localities of Intellectual Mutation" and 2022 "Information Theory of Evolution") and the Japan Society for Cognitive Science (JSCS in 2022 "Refined Human Knowledge Genome"). Each step marked not just conceptual expansion, but also a refinement of our understanding of digital language's saltative evolution and its underlying subcortical neurobiological circuits**.**

Year	Development
2007	Field visit to Klasies River Mouth Caves, sparking interest in language and human origins.
2008	Rejection of cave-based evolution; formulation of ***digital language hypothesis*** inspired by Jerne's immune network theory.
2010	***Two-stage phoneme evolution:*** click consonants (Still Bay) to syllables (Howiesons Poort).
2011	Analysis of digital networks: physical layer (acoustic signals) vs. logical layer (neural circuits) based on the Internet Layer Analysis tool: OSI Reference Model.
2012	Hypothesis of *language processing via **CSF B cells**, microglia, and CSF-contacting neurons in the brain ventricle.* 2nd visit to KRM
2013	Introduction of *Forward Error Correction (FEC) for Scientific Concepts***,** integrating interdisciplinary insights. Two poster presentations at ICL-19 (Geneve)
2014	*Monaural mother-tongue hearing and vector processing of grammatical syllables via reflex circuits.*
2016	*Three-stage digital evolution: syllables → **writing** → **bits***; presented at ICLAP2 (Lahore, Pakistan) as "Logical Linguistics: Bricolage and Breakthrough".
2020	*Concepts as mathematical **groups**; phoneme sharing from **eusocial alarm calls.***
2021	*Ultra-low-noise environments for FEC*; development of *linguistic information reception procedures exploiting bibliographic data and regenerative reception*.
2022	***Literacy** memory in CSF B-cell networks*, inspired by anecdotal reports of progressive aphasia and dyslexia post-NPH shunting
2023	Visit to KRM and Maropeng center for the Cradle of Humankind. Realized the *evolutionary continuity between early hominins and modern humans*
2024	**P**SSA22 Presentations: *Kalahari origin of Bipedalism*, Fossil mandible KRM-41815 is the evidence of digital evolution of human intelligence.(Graff Reinet, S.Africa.)
2025	Phoneme emergence from eusocial alarm calls in Late Pleistocene South Africa; ***phoneme map imprinting** via tonotopic cochlear nucleus pruning with mother tongue stimuli reinforcement during early language exposure.*

Table 6. Timeline tracing the conceptual development of the digital language evolution hypothesis, based on fieldwork, interdisciplinary research, and conference presentations. (*italics*: main hypotheses; **bold**: key hypotheses).

4. Hypothesis Testing

4.1 Evolutionary Sequence: Clicks before Syllables

As part of testing the proposed hypothesis of syllables evolving from alarm-based communication systems, we examine the possible precedence of click consonants over syllables in human evolution. We hypothesize that click consonants emerged before vowels and syllables, as they can be produced without pulmonary airflow or a descended

larynx. Their rarity in grammatical function words and limited integration into coherent phonological systems (Westphal, 1971, Traill, 1998) suggest a pre-grammatical origin. Once syllables evolved, populations who migrated out of Southern Africa relied exclusively on syllables, while Khoisan languages retained clicks in lexical words possibly because the surrounding world continued to be named with clicks. This pattern supports the view that clicks preceded syllables and have been retained primarily in southern African languages.

4.2. Writing and Bits as Indelible and Interactive Syllables

We hypothesize that written letters and digital bits serve as persistent and interactive carriers of syllabic information. Unlike spoken syllables, which depend on immediate shared context, written and digital forms are crafted with unknown readers in mind. These forms are structured to carry meaning reliably across time, allowing it to connect meaningfully with the reader's memory and understanding without relying on arbitrary interpretation.

To test this hypothesis, controlled recall and recognition experiments can be conducted using identical syllabic content presented in spoken form, written on paper, and via digital output devices such as a screen or a speaker. Measures of retention accuracy, recall latency, and interpretive stability over time may reveal whether written and digital syllables function as durable extensions of oral linguistic units.

If confirmed, this would support the view that the evolution of literacy here understood as the formation of visual contact-based memory in individual B cells in response to the visual input of written letters and of digital media represents the second and third digital leaps beyond spoken syllables, transforming language from an evanescent signal into a persistent, computationally searchable, and electrically transmittable form.

4.3 S/N Ratio, Bit-energy and Conceptual Processing

We hypothesize that conceptual learning and forward error correction (FEC) in human cognition depend on a biologically tuned signal-to-noise (S/N) threshold. Comparative experiments conducted under varying environmental conditions (e.g., classrooms, noisy settings, or quiet library spaces) can help determine the critical S/N level required for stable abstraction and error-resistant learning.

However, S/N alone may not fully account for the dynamics of language processing. Drawing on digital communication theory, we also consider the importance of bit energy per noise spectral density (Eb/No), where R (bit rate) reflects the processing speed or syllabic tempo of linguistic input. A rapid influx of information (high R) may dilute the bit energy, reducing interpretive reliability even under high S/N.

Moreover, iteration repeated exposure to the same linguistic content functions as a cognitive amplifier, enabling convergence on intended meaning through successive approximations. This suggests that high S/N, optimal Eb/N_0, and iterative reinforcement together form the physiological basis of robust conceptual processing.

4.4 CSF B Lymphocytes and Verbal Memory

We hypothesize that B lymphocytes floating in the cerebrospinal fluid (CSF) play a central role in verbal memory formation. When new symbol (conditioned stimulus: buzzer, bell, metronome, etc.) is heard and followed by food or poison introduced into the mouth, salivation occurs at the mere hearing of the conditioned stimulus. This is the beginning of a conditioned reflex experiment. At this time, a pair of specific antigen and antibody corresponding to the envelope of the speech waveform of the word are produced. The antigens are expressed on microglia, coding for multimodal sensory memories, and CSF-contacting neurons (CSF-cNs), which become activated by afferent signals at the ventricular wall. Antibodies, in turn, are expressed on the membrane surfaces of B cells floating within the CSF.

Each antibody consists of three amino acid modules, namely Complementary Defining Regions (CDRs), each of which can also function as an antigen. This property allows B cells to interact not only with CSF-cNs and microglia but also with other B cells, thereby forming a decentralized, distributed network. We propose that this dynamic, immune-based memory network underlies conscious access to verbal representations. To experimentally test this hypothesis, we suggest the following three approaches:

(1) Experimental Induction of New Verbal Memory

Using proper nouns (e.g., names of foreign mountains, river or historical figures) unfamiliar to participants, we can examine memory formation. Participants read aloud the name 10–20 times daily over 3–5 consecutive days, with or without writing string of characters, and/or receiving a 1-minute explanation about the associated referent. After one, three and five days, they are asked whether they can spontaneously recall and articulate the name and explain the referent's significance. Long-term retention is then tested at one month, six months, and one year.

A parallel experiment on episodic memory could involve unfamiliar foreign dishes (e.g., haggis, thieboudienne). Participants sample the food over 3–5 consecutive days and are later asked to recall the dish's name and describe it in their own words, with follow-up tests for retention at the same intervals.

(2) Recognition of Literacy Memory

Participants are presented with around 10 or 15 familiar food or beverage names, such as *sandwiches, sushi, omelet, barbecue, spaghetti,* etc. rendered in unfamiliar scripts (e.g., *Chinese characters,*

Cyrillic, Thai, Hebrew, Hangul, etc.). They are explained just once or twice to associate each character string with its meaning in their native language. Recognition is tested a few hours after, the following day, one week later, and again at one month, six months, and one year to assess the stability of literacy memory mediated by B cells.

(3) Clinical Correlation bsetween CSF Loss and Aphasia

(a) Statistical Analysis: Compare incidence rates of aphasia and alexia between neurosurgical patients with and without CSF loss or blood–brain barrier disruption.

(b) Rehabilitation Protocol: For patients who have lost the names of familiar items (e.g., foods), provide repeated exposure to pictures or real objects, paired with verbal repetition of the name (10–20 times per session over 3–5 days). Recovery of naming ability would be used as a measure of verbal memory cell regeneration.

5. Implications

5.1. Missing Links with Pre-Languages

A pivotal evolutionary development in the emergence of modern humans was the transformation of the mandible, which enabled the formation of a vocal tract capable of producing vowel resonance and digital syllables key components of human speech.

Our 2007 visit to the Klasies River Mouth Caves first alerted us to the digital structure of human language (3. Hypotheses Development), prompting investigation into how signals propagate between individuals and are processed by complex neural circuits in the brain.

Hypotheses regarding the origin of phonemes further crystallized following our visit to the Maropeng Centre in South Africa in 2023. There, we were struck by the morphological similarities between *Australopithecus africanus* and modern humans (Figure-4) similarities that had been noted by Raymond Dart and later supported by the taxonomist Ernst Mayr, who classified them within the same genus (Dart 1959).

These observations led us to hypothesize that early bipedal hominins in the extremely flat Kalahari, already using tools and fire, began to inhabit caves during the Ice Age and gradually acquired syllables. Human evolution, we propose, unfolded both in the expansive plains of the Kalahari and in the large coastal caves at the southern tip of Africa. While early hominins in the Kalahari likely communicated through non-verbal or pre-linguistic means for millions of years, the emergence of syllables marked a sudden and unprecedented leap a digital discontinuity in the evolutionary timeline.

5.2. Roadmap for Digital Evolution

Because syllables function as digital signals, the invention of writing enabled linguistic information to be stored, transmitted across generations, and disseminated across space, thereby giving rise to

complex civilizations. This progression from syllables as discrete oral signals, to writing as persistent representations, and finally to digital bits as interactive signals constitutes a three-stage digital evolution of human language.

Today, digital bits interactive, addressable (uniquely locatable), and searchable (logically retrievable) represent the next phase in this signal evolution. If properly mastered, this signaling system may enable yet another leap in human intelligence, potentially surpassing the complexity achieved in previous stages by several orders of magnitude.

We further speculate that genetically guided behaviors such as a child's transition to monaural listening of the mother tongue around age three, or the strong interest around age ten in puns, riddles, and rule-breaking exceptions may represent innate, genetically programmed developmental milestones in acquiring digital language processing for grammar and concepts. These hypotheses warrant systematic investigation through developmental psychology and educational research.

5.3. Language Processing by CSF B Cells

The hypothesis that language processing involves an intraventricular immune network introduces a novel neurobiological framework. It suggests that language memory is mediated not only by cortical glial circuits but also by subcortical structures particularly the cochlear nucleus, which is attuned to the phonemic features of the mother tongue and by immune networks within the cerebrospinal fluid (CSF), including B lymphocytes.

This model suggests that stable phonemic encoding and consistent word recall may depend on B cell interactions at the CSF–brain interface. Accordingly, certain forms of aphasia may result from the depletion of B cells in the CSF, a possibility that warrants empirical investigation.

Furthermore, it is important to examine whether CSF shunting in patients with normal-pressure hydrocephalus (NPH) contributes to immune cell loss, potentially leading to progressive aphasia and dyslexia. If validated, these findings would warrant the development of novel rehabilitation strategies grounded in the CSF immune network framework.

5.4 Archaeological Context and the Need for Reappraisal

The hypothesis of syllable-based language evolution is grounded in South African archaeological sites, such as Blombos and Klasies River Mouth, which provide critical evidence of early symbolic behavior and phonemic development. However, the global recognition of these sites has been limited, partly due to historical isolation during the 1960s–1990s academic boycott of South Africa. Despite recent UNESCO inscriptions (e.g., Pinnacle Point), older sites with richer fossil and cultural records remain underexplored. (ICOMOS,2024) Reintegrating

these findings into global discourse is essential for understanding the digital evolution of language and human cognition.

6. Conclusion

This paper proposes a two-stage model for the evolution of phonemes during the Late Pleistocene. Around 74,000 years ago, in the cold aftermath of the Toba supereruption, click consonants emerged as segmented, reflex-compatible alarm signals. Approximately 66,000 years ago, the development of the mental protuberance enabled vowel production, giving rise to syllables—discrete, finite, and temporally structured digital units. Writing systems appeared around 5,000 years ago, and binary bits in the mid-20th century, marking the transformation of speech signals into persistent and interactive forms. This trajectory is best understood as a case of digital evolution within the broader framework of complex systems theory.

Digital evolution has markedly increased cognitive demands on individuals. The saltative biological transitions from prokaryotes to eukaryotes, from eukaryotes to vertebrates, and from vertebrates to mammals are mirrored by linguistic advances within a mere 70,000 years. This progression is evidenced by archaeological findings from Blombos Cave and Klasies River Mouth Caves, the unique linguistic roles of click consonants and syllables, the pruning of cochlear hair cells, and the proposed intraventricular immune network. Together with the symbolic evolution from syllables to letters to bits, these elements outline a roadmap for the digital transformation of human language.

Language stands as the pinnacle of four billion years of evolutionary refinement. It now calls upon us to transcend mere spinal reflexes and eusocial instincts, and to embrace truth-seeking as intelligent, reflective beings (see Table 2). A hypothetical allele favoring the joy of learning and discovery may be essential for this next step in human evolution.

Acknowledgments

We express our sincere gratitude to the organizers and participants of the research meetings hosted by the Institute of Electronics, Information and Communication Engineers (IEICE), the Information Processing Society of Japan (IPSJ), and the Japanese Society for Artificial Intelligence (JSAI). Between 2009 and 2023, particularly during the intensive periods of 2009–2014 and 2020–2022, we participated in more than 100 meetings, engaging in discussions that significantly accelerated the development of our hypotheses, as outlined in Table 6 (Hypotheses Development). These interactions were instrumental in shaping this work, particularly in refining the digital linguistic framework, the role of signal-to-noise ratios, and the reconsideration of Australopithecus africanus in language evolution.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Use of AI

During the preparation of this work, the author(s) used ChatGPT and Grok in order to ameliorate written English and find appropriate research articles. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Figure 1. Fossil mandible from Klasies River Mouth Caves (*Morrissey et al.*, 2023), excavated from a ~100 ka layer but showing a fully developed mental protuberance. Morphology suggests it may be intrusive, possibly from a later burial.

Figure 2. X-ray images of mandibular and oral cavity morphology (*Chiba and Kajiyama*, 1941), illustrating how anterior oral cavity length and sublingual space affect vowel resonance.

Figure 3. Conceptual diagram of the intraventricular immune cell network for language processing, showing CSF B lymphocytes, CSF-contacting neurons, and microglial antigen-presenting cells (original work by the author).

Figure 4. Comparative skeletal display of *Australopithecus africanus* and *Homo sapiens*, photographed by the author at the Maropeng Centre, Cradle of Humankind, South Africa.

Table 1. Comparative attributes of three digital signal systems syllables in language, RNA in life, and electronic bits in computation highlighting parallels in structure, function, and error control.

Table 2. Evolutionary milestones in the development of digital signal systems in language, from early tool use to modern forward error correction.

Table 3. Biological evolution as a sequence of noise-reduction breakthroughs, with subsequent logical complexity growth after environmental recovery.

Table 4. Language evolution as a progression of signal types in low-noise environments, from syllables to letters to interactive bits.

Table 5. Weaknesses of sign-reflex mechanisms when applied to language processing in intelligence systems.

Table 6. Timeline tracing the conceptual development of the digital language evolution hypothesis, based on fieldwork, interdisciplinary research, and conference presentations.

References

Ambrose, S. (1998). Late Pleistocene human population bottlenecks, volcanic winter, and the differentiation of modern humans. *Journal of Human Evolution*, 34, 623–651. https://doi.org/10.10 06/jhev.1998.0219

Bardini, T. (2000).*Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing*. Stanford University Press.

Bush, V. (1945).As We May Think .*Atlantic Monthly*, 176(July), 101–108. http://www.ps.unisaarland.de/duchies/pub/vbush/vbush/vb ush/vbush-all.shtml

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987).Mitochondrial DNA and human evolution. *Nature*, 325(6099), 31 36.https://doi:10.1038/325031a0

Carr, N. (2010). *The Shallows: What the Internet is doing to our Brains*. W. W. Norton & Co.

Chiba, T., & Kajiyama, M. (1941).*Vowels, its Nature and Structure* Tokyo Kaiseikan Publishing house, Tokyo

Dart, R. A. (1959). *Adventures with the Missing Link*.Harper and brothers, Uk edition.

Deacon, H. J., & Deacon, J. (1999).*Human beginnings in South Africa: uncovering the secrets of the Stone Age*. Altamira Press. Walnut creek, C.A.

Deagling, D. J. (2012).The Human Mandible and the Origins of Speech. JournalofAnthropology.Hindawi publishing corporation https://doi.org/10.1155/2012/20152

Dogen's Extensive Record (2010 English version, Original 1252). Translated by T. D. Leighton & S. Okumura. Wisdom Publications.

Eleanor, M. L. S., & Will, M. (2023).The revolution that still isn't: The origins of behavioral complexity in *Homo sapiens*. *Journal of humanEvolution*,179,103358.https://doi.org/10.1016/j.jhevol.2023.103358

Henscilwood, C. S., et al. (2001). An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: implications for the origins of modern human behaviour, symbolism and language. *Journal of human evolution, volume*,41, issue, 6. 631–678. https://doi.org/10.1006/jhev.2001.0515

ICOMOS, Advisory Body Evaluation (2024).*Pleistocene Occupation Sites of South Africa, republic of south Afarica, WHC ID. No.172*

Jerne, N. K. (1974). Towards a Network Theory of Immune system ann. Immunol (Paris) vol. 125C, issue 1-2, 373–389.

Jerne, N. K. (1984). The Generative Grammar of the Immune System (The Nobel Lecture).Biosci rep jun:5(6): 439-51 https://doi:10.1007/ BF01116941.

Kay, L. E. (2000). *Who Wrote the Book of Life? A History of the Genetic Code*.Stanford University Press.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843. https://doi.org/10.1038/nrn1533

Martin, R. D. (1990). *Primate Origins and Evolution: A Phylogenetic Reconstruction*. Princeton University press, co- published by Chapman and Hall.

Maynard Smith, J., & Szathmáry, E. (1999).*The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford University Press.

Morrissey, P., Mentzer, S. M., & Wurz, S. (2023). The stratigraphy and formation of Middle Stone Age deposits in Cave 1B, Klasies River Main site, South Africa, with implications for the context, age, and cultural association of the KRM 41815/SAM-AP 6222 human mandible.*Journal of Human Evolution*, Vol. 183, 103414.

Owens, M., & Owens, D. (1984). *Cry of the Kalahari*.Houghton Mifflin Boston.

Pepper, W. J., Braude, S. H., Lacey, E. A., & Sherman, P. W. (1991). *Vocalization of the Naked Mole-Rat*. In R. D. Alexander et al. (Eds.), *The Biology of the Naked Mole-Rat*. Princeton University Press.

Piaget, J. (1947). *La psychologie de l'intelligence*. *The psychology of Intelligence*. Armand Colin. Paris:

Rightmire, G. P., & Deacon, H. J. (1991). Comparative studies of Late Pleistocene human remains from Klasies River Mouth, South Africa. *Journal of Human Evolution, vol.* 20, 131–156.

Singer, R., & Wymer, J. (1982). *The Middle Stone Age at Klasies River Mouth in South Africa*. University of Chicago Press.

Tanenbaum, A., & Wetherall, D. J. (2011). *Computer Networks* (5th ed.). Prentice Hall. Pearson.

Tokumaru, K. (2018). Zero error requirement for incoming linguistic information and postscript coding by Dogen (1200–1253). *Vestnik Kurganskogo Gosudarstvennogo Universiteta*, *2*(49), 97–103

Traill, A. (1997). Linguistic phonetic features for clicks in R. K. Herbert (Ed.), *African linguistics at the crossroads: papers from Kwaluseni* (1st World Congress of African Linguistics, Swaziland, 18–22 July, 1994), 99–117.

Van Der Post, L. (1958). *The Lost World of the Kalahari*. Hogarth Press.

Vigh, B., et al. (2004). The system of cerebrospinal fluid-contacting neurons.It's supposed role in the nonsynaptic signal transmission of the brain, *Histol. Histopathol.*,19, 607–628. https://doi.org/10.14670/HH-19.607

Von Neumann, J. (1963).*The General and Logical Theory of Automata*. In *Design of computers, theory of automata and numerical analysis* (Vol. 5, A. H. Taub, Ed.). Pergamon Press.

Westphal, E. O. J. (1971). The click languages of Southern and Eastern Africa. In T. A. Sebeok (Ed.), *Current Trends in Linguistics* (Vol. 7: Linguistics in Sub-Saharan Africa, pp. 367–420). Mouton de Gruyter.

Wilson, E. O. (2012). *The Social Conquest of Earth*. Liveright Publishing Corporation, W. W Norton and Company New York

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 69-82

Issues in Developing a Comparable Corpus for Indian Languages as a Key Linguistic Resource for Cross-Linguistic Research and Applications

Niladri Sekhar Dash[*]

Abstract

A comparable corpus (CC) is compiled by selecting two different languages with a set of predefined dimensions (e.g., topic, genre, text types, time period, etc.), and selecting large amount of texts from respective languages, with equal importance on texts considered for inclusion (Maia, 2003). The multipurpose applications of a CC, in a country like India, are unbound as it develops a new ecology for developing language-based assistive technology to serve the multilingual requirements of Indian societies. In this conceptual position paper, we propose methods and approaches for the generation, composition, content, and application of a CC in the Indian languages with reference to its functional utilities in linguistic and extralinguistic domains. We first proposed an applicable definition of a CC (Laviosa, 1997), then we discuss about its typological diversities (Johansson & Hasselgård, 1999), refer to the present Indian situation of CC generation, argue for generating it for the benefit of the Indian languages, describe the processes one should apply to design a CC, and finally, highlight applications of a CC in linguistics and language technology in the context of Indian languages.

Keywords: Comparable Corpora, Technolinguistics, Lexical Resource, Multilingual, Equivalents, Termbanks

1. Introduction

The multipurpose utilities of a comparable corpus (CC) in linguistics and techno-linguistics confirm the fact that future multilingual works will depend heavily on data and information obtained from it (Teubert, 1996). The modern multilingual technologies need a large amount of processed data from a CC to develop object-oriented and user-centric tools and technologies. A CC with an adequate amount of empirical and verifiable texts and a unique composition becomes conducive for cross-domain, interlingual, and crosslingual applications (Atkins *et al.*, 1992). In the last few years, a CC has been a constant source of linguistic information and data to develop linguistic resources (e.g., *comparable lexicons, termbanks, comparable grammars, translation systems, text miners, and bilingual dialogue systems*) and comparative studies on languages, societies, ecologies, and cultures (Maia, 1997;

[*] Indian Statistical Institute, Kolkata; nsdash@yahoo.com

Laviosa, 2002). Data and information collected from a CC have been useful for developing a better understanding of relational dependency between sister languages, looking into bilingual culture frames, and constructing ethnographic profiles of the sister language communities.

Keeping these advantages in view, in this conceptual position paper, we present some ideas about the form, composition, content, and application of a CC for the Indian languages with reference to its functional utilities within and beyond linguistics. We first define the concept of a CC, address its diversities, refer to the present Indian context, describe the process of designing a CC, and highlight applications of a CC in linguistics and language technology in the context of Indian languages.

2. Defining a Comparable Corpus (CC)

The concept of a digital comparable corpus (CC) is not very clear to many of us working in the area of linguistics and technolinguistics. As a result, we come across many views and definitions of CC, many of which show characteristic differences despite definitional and conceptual proximities (Maia, 2003). To get a clear understanding of the concept, we refer to some definitions to understand how a CC is to be formed, framed, and utilized.

- A CC contains corpora from two different languages that come from the same domains. A news text corpus from English and Bengali, based on this definition, is a CC, although from a critical perspective, they may vary in topic, composition, domain, time, place, size, balance, representation, and others. This definition, in the true sense, does not serve our purpose of designing a CC.

- A CC may select similar texts from more than one language. Although this logic is followed in building a CC, there are a few studies that reveal a different purpose and characteristics of a CC. It seems that it is still in a concept than a reality. In our argument, the basic goal of a CC is to compare different languages in similar circumstances, but avoid the inevitable distortion generated through translations (Borin & Prütz, 2000).

- A CC, in principle, should be made of pairs of texts produced independently of each other. The texts are put into a CC as they are considered comparable in certain aspects. For example, we compile a corpus of written Hindi sports texts similar in form, content, and context to a corpus of written English sports texts.

- A CC should refer to a multilingual corpus containing texts in different languages, where text samples may not be the same in content, genre, or register. It may cover a variety of corpus types, referring to texts designed and generated following similar design criteria. Texts may or may not be translations of each other. If texts (*on the same topic*) are translations of each other, it is a Parallel Translation Corpus (PTC); if texts are aligned, it is an Aligned Parallel Translation Corpus (APTC).

- A CC also consists of two sets of texts: one originally written in language A and one of similar texts translated into language A, from a different language. Its value is mainly theoretical; what is investigated is the linguistic nature of translated text, independently of the source language (Baker, 1995).
- A CC is permitted to contain similar texts from several languages. The parallel translation corpora of the Indian languages that were developed in the ILCI project are a nice example of this (Jha, 2010). Also, the text corpora compiled in *AI4Bharat*, *Vidyaapati (2022-2026)*, and *Ishaan* (2022-2026) of the *Bhashini* project of the Govt. of India are good examples of a multilingual CC. Both *Vidyaapati* and *Ishaan* follow the same policy, model, time frame, genres, and quantity of words from several Indian languages.
- A CC can be made from a variety of sources: texts found in digital form, scanned, or typewritten. The criteria for choosing texts depend on the homogeneity of texts in terms of features such as subject domains, text origin, text constitution (i.e., *single or joint texts*), factuality, technicality, and application (Biber, 1993).
- A CC may contain texts not collected from the same genres, types, and fields. Moreover, texts may not carry the same kind of content and linguistic information. With several disparities in composition and content, a text collection may be treated as a CC if it is made for a specific purpose, such as producing comparable lexicons and dictionaries. The Comparable Reference Corpus (CRC) of the *Commission of the European Community* (CEC) is a nice example of this type.

Our main argument is that the context of defining a CC is that any two natural language texts can be treated as 'comparable texts' provided the two texts adhere to some conditions and features defined beforehand for the purpose of a particular study or investigation. In this perspective, a written text and a spoken text of a language can be treated as a CC provided they match on selection criteria.

Advanced technology makes it easy now to create small or big, single or multidimensional CC in many languages. Technolinguists, corpus linguists, translators, and content developers create and use multilingual CC as a source of information, data, and guidance.

The size of a CC can vary based on how judiciously it is made to meet the criteria of specific requirements. In the case of texts with a high degree of technicality and written by subject specialists for peers, it is fine to collect relevant data and information from a few selected sample texts. In such a situation, the amount of text does not create much problem if samples of texts are of similar content and equal size. On the other hand, in the case of newspaper texts from two different languages, we argue for developing a larger CC taking various linguistic and extralinguistic issues into the design criteria.

We also argue for accepting a CC consisting of texts from similar sources representing different varieties of the same language. For instance, the samples of English texts that have been collected from the newspapers published in different cities of India are a fine example of a CC, the analysis of which provides a clear understanding of how Indian English is represented differently in various states in India. It also provides valuable data and insightful information for developing comparable lexical databases and designing primers in Indian English (Dash, 2020).

3. Types of Digital Comparable Corpus

In our opinion, based on the types of sources of texts, a CC, keeping further subdivision at bay, can be of three types:

(a) Monolingual Comparable Corpus (MCC)

(b) Bilingual Comparable Corpus (BCC)

(c) Multilingual Comparable Corpus (MLCC)

3.1 Monolingual Comparable Corpus (MCC)

An MCC includes texts from one language but incorporates text samples from different genres, domains, and registers. Here, the type of a particular text acts as a pointer to determine the separate identities of the texts included in the corpus. Thus, any comparable collection of texts may be put under an MCC to initiate comparative studies across registers within a single language (Table 1).

Hindi	
Medical texts	**Engineering texts**
Recent publication	Recent publication
Printed texts	Printed texts
Large collection	Large collection
Subdomains inclusive	Subdomains inclusive

Table 1: MCC with texts from a single language

We further classify an MCC using different parameters, such as genre (*written vs. spoken*), time (*synchronic vs. diachronic*), register (*dialectal vs. standard)*, text type (*informative vs. imaginative*), text content (*literary vs. scientific*), and text authorship (*male vs. female*). The basic argument is that an MCC is free to contain texts from two different types of texts, provided the selection criteria are uniform across the text types.

3.2 Bilingual Comparable Corpus (BCC)

It is also possible to create a BCC with texts taken from two different languages, maintaining a kind of uniformity in text selection criteria and composition patterns across the two languages (Church, 2025). Much better investigational advantages are available if it is made with texts taken from typologically (e.g., *Hindi & Punjabi*), genealogically (e.g., *Bengali & Odia*), or geographically (e.g., *Odia & Telugu*) related

languages. One can create a BCC with texts from Hindi and Bengali, as the languages are both typologically and genealogically related. It, however, should be kept in mind that the text samples, which are included from the two languages, should belong to a similar genre, text type, and content (Table 2).

Hindi	**Bengali**
Medical Texts	**Medical Texts**
Recent publication	Recent Publication
Printed textbooks	Printed textbooks
Large collection	Large collection
Subdomains inclusive	Subdomains inclusive

Table 2: BCC with texts from two languages

A BCC may appear similar to a parallel corpus. In the case of a parallel corpus, content, text type, and genre are identical, while texts belong to two separate languages. In the case of a BCC, all the parameters may vary based on the requirements of a study. Like that of a MCC, a BCC is also used in language teaching, translation, and dictionary compilation (Barlow, 1996).

3.3 Multilingual Comparable Corpus (MLCC)

It is not necessary that a CC should always contain texts from two languages or text varieties. The content of a CC may include three or more languages if the texts are both characteristically and compositionally similar in domains and content (Abdulmumin et al., 2024). For instance, the AI4Bharat is engaged in developing parallel comparable corpora in many Indian languages for machine translation purposes (Gawtham et al., 2022).

A MLCC includes texts from three or more languages, while the criteria for considering texts from different languages remain the same. Here, the goal is to collect texts of identical domains and disciplines from different languages in a uniform manner so that comparison among texts becomes organized, systematic, uniform and useful. In a multilingual country like India, a MLCC is indispensable for developing cross-lingual resources, starting cross-lingual studies, and developing multilingual technologies (Table 3).

Hindi	**Bengali**	**Tamil**
Written	Written	Written
Cultural event	Cultural event	Cultural event
New year	New Year	New Year
Celebration	Celebration	Celebration
Activities	Activities	Activities

Table 3: MCC with texts from three languages

The attempt to develop a MCC for the Indian languages began a few years ago, when the project *Indian Languages Corpora Initiative (ILCI)* started with a goal for collecting comparable sentences from

various disciplines and subject domains from several Indian languages (Dash 2009; Jha, 2010). This effort is further strengthened with the initiation of the *Vidyaapati* and *Ishaan* projects under *Bhashini*, as well as with the establishment of the national hub for Indian languages as a part of the AI4Bharat.

4. Designing a Comparable Corpus

There is a demand for CC as it is rich with information to carry out multidimensional intralingual and interlingual studies. Besides, it is an essential requirement for many tools and resources like machine translation, information retrieval, and parallel multilingual lexicons. A CC perhaps does not need further specification of criteria beyond those defined for monolingual corpora (Sabel et al., 2022). Any two monolingual corpora, if put into comparative analysis, can be considered a CC. In such a case, factors controlling the constitution of each monolingual corpus may not be identical or similar. It can be assumed that while one corpus may comprise texts from one language (L1) of a particular discipline, the other one may consist of texts from another language (L2), containing texts from a different discipline. Although both texts theoretically constitute a bilingual corpus, when put to comparative analysis, they show little similarity for comparison. There are a few linguistic criteria that mark their differences (Table 4):

Criteria	**L1**	**L2**
Genre	written	spoken
Register	literature	mass media
Agent	young female	aged male
Domain	story	political speech
Setting	family	market

Table 4: The Bilingual corpus is not a comparable corpus

The differences of the factors stated above prevent texts from being comparable despite their bilingual entity. Any study between the two may produce results and observations hardly comparable between the languages. To overcome this limitation, we argue for developing bilingual parallel corpora (BPC) for interlinguistic comparison and mutual information extraction. It will be developed in such a way that both texts are subjected to alignment at the word and sentence level. It is necessary to define the design criteria for the creation of aligned and bilingual parallel corpora, as predetermination of designing factors relieves investigators from unwanted problems (Váradi and Kiss, 2001).

In the case of a CC, we argue that both texts should follow similar design criteria. This will enable access to the text easily and more simplified manner. Besides, it will help to process texts in an efficient way and utilise results more fruitfully in translation, parallel lexicon generation, and language teaching. Imagine an instance where both texts are parallelly developed following identical design criteria. Suppose two texts are created in Bengali and Tamil on the news of the

election published in two newspapers in 2025 in Tamil and Bengali. If both texts are put in comparison, they can be an ideal example of CC due to their identical design criteria (Table 5).

Criteria	**Bengali**	**Tamil**
Genre	written	written
Register	mass media	mass media
Domain	news report	news report
Sub-domain	politics	politics
Topic	election	election
source	newspaper	newspaper
Time	2025	2025

Table 5: Identical designing criteria for BCC

In some cases, bilingual texts may include original and translated texts of the same domains and genres (Maia, 2021). This is useful in translation and parallel lexicon generation. When we search for the appropriate terms in the source and target corpora, the following things happen.

- A term in the source corpus may not have an appropriate corresponding term in the target corpus.
- Translation 'per se' is probably impossible, as a term in one language may not correspond to a word in another language.
- A translated term may be a hyponym of a term in the source language.
- Language-specific expressions in the source corpus may not have a correspondence in the target corpus.

It implies that a parallel corpus (with translationally equivalent texts) and a CC (made with texts from pairs of languages) can vary in roles and relevance (McEnery & Oakes, 1995). They, however, provide data for those people who are looking for information based on which words and phrases can be translated acceptably (Zanettin, 1994).

When we propose to develop a CC with a collection of scientific texts from two different languages for collecting translational terminology (Kontonatsios, 2015), we suggest considering the following criteria:

- author-reader relationship (expert-expert or expert-naive communication),
- published texts (because they ensure some degree of acceptance of the terminology of a particular field),
- text origin (product of individual or joint authorship),
- constitution of texts (single or composite),
- factuality (text is factual or imaginative),

- technicality (text is technical or semi-technical),
- intended outcome (informative or didactic).

It is known that word-class-based equivalents, which are found in bilingual dictionaries, are not enough for judging the quality of translations. Analysis of CC shows how different innovations largely spread across text-types until they permeate the entire language. We also observe that the terminologies of CC provide a good empirical basis for describing these terms from the perspective of their use in different texts, based on which conceptual equivalent terms are possible to derive from a CC.

What we argue here is that with proper modifications of the existing techniques and rules, we can design a uniform method for developing CC for Indian languages to fulfill the needs of Indian linguistics and language technology (Siripragada, 2020). We can adopt innovative strategies for specific language pairs (e.g., *Hindi & Manipuri*) or develop a new model to include several Indian languages so that existing models of CC are refined to enable a wide range of applications for the Indian languages.

5. Utilities of a Comparable Corpus

Compared to a monolingual corpus, a CC provides much better insight into paired languages. A CC is made to supply linguistic data and information for comparative studies on language-specific, typological, and cultural differences. It presents notable differences between the source texts and their translations or between the two different languages. It shows how domain-centric monolingual texts vary based on various linguistic properties and features. It also reveals striking differences between native and non-native texts, as well as provides structurally equivalent texts and paired lexical databases for multilingual dictionary compilation, language teaching, and translation. Some more applications are defined below.

- Texts of a CC are often used for translation analysis to reconstruct the correspondences between segments of a source text and the segments of its translation (Candel-Mora & Vargas-Sierra, 2013).
- A CC is analysed to search for words, terms, and phrases with special meanings and frequent usages in specific domains, genres, and areas (Dragsted, 2006).
- A CC is used to look for information on ways in which a given word or a phrase can be translated acceptably in a target language (Esselink, 2003).
- A CC may be used to develop the procedural skills of the learners. Instead of dealing with isolated words and rules for combining them

into a text, the learners are trained to use wider syntactic chunks like multiword units, phrases, idioms, and collocations to learn about the formation of larger constructions (Fantinuoli, 2018).

The value of a CC in translation is very high. Technically, a translation is a text production activity in which many factors influence the source and target language situation. The analysis of a CC produces information to understand the respective expectations, experience, and knowledge of the languages involved. After establishing structural and semantic equivalence between comparable texts, we gather information about the way in which discourses are framed in two languages. Evidence collected from CC may guide to creation of new texts in the target language based on the requirements of new purposes.

A CC is a good source for learning about terminology. Made with scientific and technical texts, it is a useful source of terminology and its usage information for compiling bilingual or multilingual terminology databases that have strong applicational roles in Indian linguistic contexts. On the other hand, in a classroom context, a CC can be referred to confirm translation hypotheses and to suggest possible solutions to actual translation problems relating to specific text types. Specialised CCs offer information on concepts and terminology along with attested usages and distribution patterns controlled by contexts.

Another unique application of a CC may be linked with the investigation of particular genres and topic areas, as pre- and post-translational activities. A CC, as a source of potentially endless search processes, creates opportunities for language investigators to find genre-based and topic-based linguistic representations in the comparable texts, depending on the intuition, proficiency, interests, and needs of investigators.

The application of a CC can be envisaged in testing the relationship between those expressions of language that we perceive as holding some kind of semantic similarity or conceptual equivalence. The issue of semantic proximity or referential closeness of words has been an area of serious study in sense disambiguation and machine learning. A CC has a great role in this kind of study as it provides ample data and texts to measure the degree of semantic proximity of synonymous words used in two different texts. We can measure similarities between texts and compare words and phrases by identifying categories that have a strong formal resemblance, such as proper names and cognates, or which are proposed as translation equivalents between the two languages (Kaji and Morimoto, 2002).

A CC may be used as a yardstick to train the translators. In search of embedded constructions in a CC, for instance, a frequent use of embedded sentences can indicate the level of complexity involved in

the composition of the texts. Moreover, contrastive analyses of a CC reveal how similar ideas and concepts are expressed in similar kinds of texts in different languages. The analyses may consider different stylistic preferences such as use of rhetoric, propositions, figurative expressions, collocations, lexicogrammatical features, and others. Also, a CC supplies information about how discourses are realised in texts of different languages, and how linguistic and extralinguistic factors are taken into strategies for constructing different discourse patterns.

A CC does not offer single translationally established correspondence; it actually offers a repertoire of recurring collocational or structural features, which provide a basis for establishing equivalence between stretches of texts in two languages. The frequency of use of various linguistic items in a CC provides an indication of the expectations the prospective researchers may have about the linguistic items and features of the texts of a particular genre.

The translation assessment methods based on a CC can be integrated into a curriculum of trainee translators to guide them about the aspects and issues relevant to the creation of acceptable translations. For instance, a list of concordance retrieved from a CC is critically examined in a translation evaluation task as an aid to the analysis of source and target text features. A paired concordance list can be highly instructive to the translation trainees while engaged in a translation evaluation task.

In sum, the functional value of a CC primarily lies in its strength in opening up unprecedented scopes for comparing similar or different sets of language data as well as referring to different language types and varieties used under similar circumstances of text composition and application. Thus, a CC helps to avoid inevitable distortions noted in a parallel and translation corpora.

6. Why a CC is a Good Option for India?

From a practical perspective, we argue that CCs in Indian languages are far better linguistic resources than others in the present multilingual situation:

(a) **General Linguistics:** A CC made with two or more Indian language texts is a good resource for gathering many new insights, information, and understanding about the Indian languages used, which cannot be found by studying monolingual corpora.

(b) **Comparative Linguistics:** A CC is suitable for a wide range of comparative studies across the Indian languages to expand knowledge and understanding about language-specific, typological, genealogical, and cultural similarities and differences.

(c) **Linguistic Universal:** A CC of Indian languages can be a representative corpus to study, examine, attest, or reject some 'universal features' claimed to be present in the Indian languages.

(d) **Lexical Database:** A CC of Indian languages can be a useful resource for compiling multilingual lists of semantically equivalent words, conceptually similar technical terms, and translational equivalents.

(e) **Translation:** A CC in Indian languages is useful for finding similarities and differences between source texts and translations (Dewangan et al. 2021).

(f) **Practical Application:** A CC of Indian languages is indispensable for producing multilingual lexicons, multilingual idioms, dictionaries, writing comparative grammars, and preparing language teaching materials.

In essence, we find several practical applications of CCs built with Indian languages. Because of many visualized advantages, we propose to develop CCs of various types, forms, and texts between two or more Indian languages. Also, there are many theoretical justifications and cognitive advantages due to which we are in favor of compiling CCs for the Indian languages and argue for generating different kinds of CCs for developing tools and technology for the benefit of the languages.

7. Conclusion

A comparable corpus can be fruitfully used for many works of linguistics and techno linguistics in Indian languages. A CC can be useful if it is built within and across 'language pairs' as suggested above. Due to genealogical, typological, orthographic, and other similarities, a CC made within the paired Indian languages will be easy to align, process, and analyse in the respective languages.

It has been more than three decades now that the Indian languages have been used to develop monolingual corpora. Very recently, there have been efforts to develop CCs in the Indian languages (e.g., *AI4Bharat*), although we have not developed pan-Indian guidelines or protocol for this. In this position paper, we therefore present some ideas about how to develop a CC in Indian languages. We also argue that a CC is indispensable for developing multilingual resources (e.g., *lexicon, dictionaries, frozen expressions, equivalent syntactic constructions, context-based databases for translational equivalents, and ambiguous cross-linguistic lexical sets)* for the Indian languages. Data and examples collected from CCs can be utilized to write comparative grammars and develop multilingual scientific and technical terminologies.

In the present global scenario of Artificial Intelligence, Large Language Models, and Machine Learning, the CCs in the Indian languages can be indispensable resources for developing multilingual assistive tools and techniques, as well as starting comparative synchronic studies among the typologically, genealogically, and geographically related Indian languages. To achieve all the goals stated above, we suggest a collective and pan-Indian initiative to develop CCs in Indian languages and utilize the same in all kinds of linguistic and techno-linguistic works to break the divides existing among the Indian languages.

References

Abdulmumin, I., Mkhwanazi, S., Mbooi, M., Muhammad, S.H., Ahmad, I.S., Putini, N., & Marivate. V. 2024. Correcting the FLORES evaluation dataset for four African languages. In *Proceedings of the 9th Conference on Machine Translation*, Miami, Florida, Association for Computational Linguistics, pp. 570-578.

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing, 7*(1), 1–16.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*(2), 223–243.

Barlow, M. (1996). Parallel texts in language teaching. In S. P. Botley, J. Glass, T. McEnery, & A. Wilson (Eds.), *Proceedings of Teaching and Language Corpora 1996* (UCREL Technical Papers No. 9, pp. 45–56). UCREL.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257.

Borin, L., & Prütz, K. (2000). Translated text. In W. Daelemans, K. Siman, J. Veenstra, & J. Zavrel (Eds.), *Computational linguistics in the Netherlands 2000* (pp. 30–34). Rodopi.

Candel-Mora, M.A. & Vargas-Sierra, C. 2013. An analysis of research production in corpus linguistics applied to translation. *Procedia,* 95: 317-324.

Church, K. 2025. Comparable corpora: Opportunities for new research directions. In *Proceedings of the 18th Building & Using Comparable Corpora Workshop* (pp. 73–82)., ACL, 20 Jan. 2025.

Dash, N. S. (2009). *Generation of translation corpora in Indian languages* [Paper presentation]. 2nd Workshop of the Indian Languages Corpora Initiative (ILCI-I), Dravidian University, Kuppam, India.

Dash, N. S. (2020). Developing a news text corpus of Indian English for language technology, machine translation, and language teaching. *International Journal of Translation, 32*(1–2), 31–57.

Dewangan, S., Joshi, N. & Bhattacharyya P. 2021. Experience of neural machine translation between Indian languages. *Machine Translation*. 35: 71-99.

Dragsted, B. 2006. Computer-aided translation as a distributed cognitive task. In Dror, I. and Harnad, S. (eds.) *How Cognitive Technology Extends Our Minds.* (pp. 237-256) John Benjamins, pp. 237-256.

Esselink, B. (2003). Localisation and translation. In H. Somers (Ed.), *Computers and translation* (pp. 67–86). John Benjamins

Fantinuoli, C. 2018. The use of comparable corpora in interpreting practice and training. *The Interpreters' Newsletter,* 23: 133-149.

Gawtham, R. et al. 2022. Samanantar: the largest publicly available parallel corpora collection for 11 Indic languages. *Transaction Association of Computational Linguistics*. 10: 145-162.

Jha, G. N. (2010). The TDIL program and the Indian Language Corpora Initiative (ILCI). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)* (pp. 982–985). European Language Resources Association.

Johansson, S. & Hasselgård, H. 1999. Corpora and cross-linguistic countries. In Granger et al. (Eds.) *Contrastive Linguistics and Translation*, Benjamins, pp. 145-162.

Kaji, H., & Morimoto, Y. (2002). Unsupervised word sense disambiguation using bilingual comparable corpora. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)* (Vol. 1, pp. 1–7).

Kontonatsios, G. (2015). *Automatic compilation of bilingual terminologies from comparable corpora* [Doctoral dissertation, University of Manchester].

Laviosa, S. (1997). How comparable can comparable corpora be? *Target, 9*(2), 289–319.

Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Rodopi.

Maia, B. (1997). Sentence structure and thematization in comparable and parallel texts. In K. Klaudy & J. Kohn (Eds.), *Proceedings of*

the 2nd International Conference on Current Trends in Studies of Translation and Interpreting (pp. 541–547). Scholastica.

Maia, B. (2003). Training translators in terminology and information retrieval using comparable and parallel corpora. In F. Zanettin, S. Bernardini, & D. Stewart (Eds.), *Corpora in translator education* (pp. 43–54). St. Jerome.

Maia, B. 2021. Corpora, translation, terminology, and beyond: objectives and perspectives. *TradTerm, 37*(1): 10-29.

Mcenery, A., & Oakes, M. P. (1995). Sentence and word alignment in the CRATER project: Methods and assessment. In S. Armstrong-Warwick & E. Tzoukerman (Eds.), *Proceedings of the EACL-SIGDAT Workshop* (pp. 77–86).

Sabel, H. B., Dell'Oro, F., Montrichard, C., & Rossari, C. (2022). Setting up bilingual comparable corpora with non-contemporary languages. In *Proceedings of the 15th Workshop on Building and Using Comparable Corpora (BUCC 2022)* (pp. 56–60).

Siripragada, S., Philip, J., Vinay, P., & Jawahar, C. V. (2020). A multilingual parallel corpora collection effort for Indian languages. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)* (pp. 3743–3751).

Zanettin, F. 1994. Parallel words: designing a bilingual database for translation activities. In A. Wilson & T. McEnery (Eds.), *corpora in language education and research: A selection of papers from TALC 94.* UCREL Technical papers No.4, Lancaster, pp. 99-111. UCREL,

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 83-90

Outline of Linguistic Studies of Kashmiri Language

Nazir Ahmad Dhar*

Abstract

The paper presents an overview of the major linguistic works undertaken on Kashmiri, tracing the development of its systematic study from the nineteenth century to the early decades of the twentieth century. It examines significant contributions to the description and analysis of Kashmiri language in the areas of phonetics and phonology, morphology, syntax, lexicography, and dialectology. The paper also examines Pandit Ishwar Koul's Kashmiri Shabdamritam, the first comprehensive grammar of Kashmiri, G. A. Grierson's extensive documentation of the language and its dialects in the Linguistic Survey of India. The paper highlights the contributions of scholars such as Bailey and J. R. Firth to the study of Kashmiri phonetics and phonology. By bringing together these foundational works, the paper provides a concise account of the evolution of Kashmiri linguistic scholarship and highlights their significance for subsequent research on the language.

Keywords: Kashmiri Language, Linguistic Survey of Kashmiri, Dardic, Indo-Aryan

1. Introduction

Linguistic studies of Kashmiri, documented in the form of brief notes, articles, monographs, grammatical sketches and dictionaries dates back to 19th Century AD. The first comprehensive grammar of Kashmiri language was written by Pandit Ishwar Koul under the title *Kashmiri Shabdamritam* written in Sanskrit. Pandit Ishwar Koul modelled his work on the principles of Panian Sanskrit grammar Ashtadhyayi. Grierson edited *Kashmiri Shabdamritam* with notes and additions and the work was published under the auspices of the Asiatic society in 1877. Ishwar Koul also tried to write Kashmiri-Sanskrit dictionary but could not complete it. During the earlier period of 20th century, various research studies on Kashmiri language appeared elaborating on the subject matters like phonetics and phonology, morphology and syntax, dictionaries and lexical studies. Bailey's work on Kashmiri phonetics appeared in 1937. J R Firth carried out a tentative analysis of Kashmiri sounds in 1939. Grierson's monumental work *Linguistic Survey of India* was carried out during the end of 19th century and early years of 20th century covering most of the languages spoken in the South Asian

* University of Kashmir; proflingdhar@gmail.com

subcontinent. The work was published in 11 volumes and 19 parts. Volume 8 Part 2 focuses on Kashmiri language and its dialects. Outline account of Grierson's work on Kashmiri language is summed up as under:

1. Grammatical description of Kashmiri language explains patterns of speech sounds, Kashmiri Nouns, Kashmiri Verbs, Primary and secondary suffixes of Kashmiri Language. Grierson, however has neglected significant aspects of Kashmiri syntax and has provided a fragmented sketch of syntax (Kachru 1966).
2. With respect to linguistic classification of Kashmiri language, Grierson proposed that it belongs to the Dard group of Dardic sub-family. Grierson used the term Dardic to group the genetically related languages spoken across the geographical terrain called Dardistan. Dardic languages according to Grierson comprise a separate branch of Indo-Iranian sub-family and fall at an intermediate place between Indo-Aryan and Iranian languages. Grierson provides linguistic evidences in support of his view summed up by Kachru (1966) in his article *Kashmiri and other Dardic Languages* are as under:
   I. Lack of voiced aspirates in Kashmiri (For example: bh, dh, gh)

   II. Kashmiri has SVO word while Indo-Aryan languages have SOV word order. (It is worth to note here that Shina and Kohistani and dialects of Kashmiri like Poguli, Siraji and Rambani have SOV word order. Kishtwari follows SVO word order).

   III. Aspiration of stops in word final position.

   IV. The presence of large number of post-positions in Kashmiri which are peculiar to Pisacha.

   V. The numeral system is typical Paisacha.

   VI. Three-way contrast of demonstrative pronouns.

   VII. Three-fold past tense in Kashmiri.

S.K. Chatterjee (1962) supported Dardic notion of linguistic classification of Kashmiri and other related languages. Chatterjee holds the view that Kashmiri, at least in its basic substratum belongs to the Dardic section of Aryan or Indo-Iranian. Chatterjee believes that Kashmiri, in spite of Dardic substratum became a part of the Sanskritic-cultural world of India. This linguistic and cultural contact of Kashmiri resulted in diffusion of very large Indo-Aryan linguistic elements into Kashmiri with Dardic base. Georg Morgenstierne (1961) assumes that Dardic is simply a convenient term to denote a bundle of aberrant Indo-Aryan hill languages which contain absolutely no features which cannot be derived from the old Indo-Aryan. S. S. Toshkhani (1977) did

not find any strong linguistic validity in Grierson's linguistic classification of Kashmiri language. Toshkhani made the following observation in his research entitled *Kashmiri Language: Roots, Evolution and affinity.*

1. The phonetic features of Kashmiri language can be diachronically traced back to old Indo-Aryan.
2. Etymologically Kashmiri shows strong affinity with Indo-Aryan languages.
3. Kashmiri shows certain syntactic similarities with Sanskrit particularly for negative sentences.

Grierson described the regional dialects and social dialects of Kashmiri language. Kishtwari is marked as the only true regional dialect of Kashmiri spoken in Kishtwar. Siraji, Rambani and Poguli are the miscellaneous transitional dialects. Social dialects as hinted by Grierson are Hindu Kashmiri and Muslim Kashmiri showing Sanskrit influence in the former and Persian influence in the latter.

3. Grierson compiled his four volume Kashmiri-English dictionary and published it in 1932. The dictionary is based on the material from Ishwar Koul's incomplete lexicographic work entitled "Kashmiri-Sanskrit Kosha".

The rise of structural linguistics during the beginning of 20th century in Europe revolutionized linguistics in thought and practice. The theory of structuralism led to the postulation of a body of doctrines related to the understanding of the nature and structure of human language. Unlike other languages spoken in the subcontinent, Kashmiri language did not receive any serious linguistic attention on the basis of modern linguistic trends till 1966 except one published work entitled, "Kashmiri Word Phonology" authored by Ashok R Kelkar and Pran Nath Trisal published in Anthropological Linguistics in 1964.

The modern phase of linguistic studies of Kashmiri language was initiated by Braj B Kachru, trained in modern linguistics at USA. Kachru undertook a project funded by USA to write *A Reference Grammar of Kashmiri Language*. The project was completed in 1972. The reference grammar is a pioneering work that has covered phonetic-phonological-morphological-syntactic aspects of Kashmiri. In the reference grammar, Kachru establishes phonetic inventory of Kashmiri speech sounds, and analyzes parts of speech. The rules for phrase structures and sentences structures hitherto unexplored are discussed in the grammar. The grammar has pedagogical as well as research orientation. Kachru later wrote intensively on various linguistic aspects of Kashmiri language on the topics like Kashmiri and other Dardic

languages, Verb to be in Kashmiri, A Manual of Teaching Kashmiri language besides some topics on applied linguistics. Omkar Nath Koul with inquisitive linguistic brain deliberated on diverse themes of Kashmiri linguistics for a period that spreads approximately half a century. Brief sum up of Koul's linguistic study with reference to Kashmiri language is as under:

1. Worked on the core areas of linguistics that includes various types of grammars like Phrase Structure grammar (1977), Descriptive Grammar (1992) Cognitive Grammar, co-authored by Kashi Wali (1998). Research papers on different themes of Kashmiri syntax adopting modern linguistic techniques have been published in national and international journals. Prof Koul and Peter Edwin Hook, University of Pennsylvania, USA, carried out collaborative research on Kashmiri language. The work is published and readily available.
2. Pioneered the research studies of Kashmiri language in the Interdisciplinary context like sociolinguistics, psycholinguistics, and language and communication. Syntactic patterns and lexical categories of Kashmiri language indicate deep rootedness in the cultural categories of the people. Prof Koul studied these aspects of Kashmiri language in the framework of Sapir-Whorf hypothesis.

During the end of 20th century, the linguistic study of Kashmiri language became the focus of inquiry in various academic centers. The thrust areas of research remained core linguistics and interdisciplinary areas. The trained linguist studied Kashmiri phonology, morphology, syntax and semantics within the framework of modern linguistic theories. The interdisciplinary areas studied different aspects of Kashmiri language in the context of sociolinguistic, psycholinguistic and applied linguistic models. Major part of this work is documented in the published form today. Apart from this, the linguistic studies of Kashmiri language also flourished in Europe, America and Russia. Russian Academy of Science published grammars of Kashmiri written in Russian language.

Kashmiri language being the subject matter of central importance of Department of Linguistics, University of Kashmir, wide range of programs have been formulated to explore intensively and extensively different aspects of Kashmiri linguistics in tune with modern technology. The brief summary of these programs is as under:

1. Computational Linguistics received due importance as Ph.D scholars with working knowledge of computer sciences undertook

study in this area using Kashmiri language data for their respective studies.

2. Lexicographic studies have been conducted that resulted in developing online dictionaries of Kashmiri vis-à-vis other languages.

3. Machine translation system of Kashmiri vis-à-vis other languages and vice versa has been developed.

4. Dialect Survey of Kashmiri language has been carried out. The survey revisits Grierson's treatment of Kashmiri dialects. The survey comprises three parts. Part 1 is Linguistic Survey of Kashmiri Dialects in the Kashmir Valley. The survey shows significant linguistic cleavage between North Kashmir, South Kashmir and Srinagar. The use of certain linguistic traits specifies the geographical location of the speaker. For example two variables, the progressive marker /an/ in spite of /aːn/ elsewhere and palatalization for the words which are non-palatalized elsewhere are genetically inherent linguistic traits earmarked for speakers of south of Kashmir. Further isoglosses have been drawn between major regions of south Kashmir. The occurrences of round vowels which are not round elsewhere are significant linguistic trait of North Kashmir. Part 2 specifies regional dialects of Kashmiri language spoken outside Kashmir valley. The dialect description in this volume follows Uriel Weinreich's model of Structural dialectology. With regard to phonological variation, the Kashmiri dialects/variants spoken across Pir Panjal present cases where palatalization ceases to exist in many sounds which otherwise are palatalized in Kashmiri of the valley. For example, /kənʲ/ 'stone' is pronounced as /kə̃y/. Similarly, Kashmiri words starting with the dental aspirated affricate [tsʰ] in valley are pronounced by speakers in Banihal, Ramban, Gool, Maho with a dental aspirated affricate [s] which is however different from the dental fricative [s]. The survey volume has used waveform and spectrogram software to show the distinctness of the two sounds in question.

Schmidt and Koul (1984) report that Kishtwari and Poguli both have developed the central vowels peculiar to Kashmiri. One of these central vowels is also found in Brokskat- a variety of Shina. However, this could be just an aerial feature and not genetic (Ramaswami: 1975).

In Kashmiri, the number of vowels posited by Kelkar and Trisal (1964) is fifteen. They have not mentioned the long back rounded vowel [ɔː] because there is a lone example /sɔːd/ 'one and a quarter' for

it in Kashmiri. However, the vowel was frequently perceived in the speech of Poguli and Kishtwari people while a survey was conducted (Ahmad et al: 2011). Some examples are given below.

Poguli

i./gɔ:š/ light

ii./gɔ:l/ shepherd

iii./sɔ:s/ ashes

Kishtwari

i./gɔ:ray/ will tell you
ii./jɔ:n/ good
iii./gɔ:D/ cowshed

The word /sɔ:s/ finds mention in Sheikh ul Alam's poetry as well and it has the same meaning as above (ashes). However, it lacks the inherent roundedness, which may be a deliberate attempt to maintain the rhyming of the poem.

With regard to Poguli, a complete paradigm of personal pronouns is seen which are marked for number, gender and case. The same have been recorded in the part 2 of the survey project that stands published. A few examples are:

i. [ãv]I -nominative, neuter	[asõ:]	our-neuter
ii.[mi] I-ergative, neuter	[tu:n]	your-neuter-sg
iii.[tes] he-dative, masc.	[ti:n]	your-neuter-pl
iv.[su] he-nominative, masc.	[tesõ:]	his-sg
v. [mʲo:n] my-masc	[tesay]	her-sg
vi. [mʲe:n] my-fem		

Kishtwari morphology is linguists' paradise. Apart from *h-lessness* and *devoicing* with the latter also foregrounding the presence of an incipient tone, the copula in Kishtwari presents a curious case. Instead of /chu/ of Kashmiri spoken in the valley, there is a /thu/ which is marked for number and gender as well. Few examples are:

i./unin thu u:n/	There is a dog. (There is a deictic pronoun and not a pleonastic subject)
ii./yi thi akh kita:b/	This is a book.
iii./bɨ thus gevan/	I am singing.
iv./tu kati thukh/	Where are you?

Both these volumes are published by CIIL Mysore. Volume 3 has taken the account of some syntactic features in correlation with regional variation. However. it is yet to be published.

5. Kashmiri language is one of the Indian languages in the language list of C- DAC Pune to develop software in Kashmiri language'

This summary gives a glimpse into Kashmir language research, but is by no means comprehensive. The past two decades have seen significant advancements in core and interdisciplinary areas, leveraging modern technology.

References

Bhat, R. (1987). *A descriptive study of Kashmiri*. Amar Prakashan.

Chatterji, S. K. (1963). *Languages and literatures of modern India*. Bengal Publishers.

Dhar, N. A., & Ahmed, Z. A. (2018). *Linguistic description of Kohistani language spoken in Ganderbal district of Kashmir Valley*. Rajat Publications.

Firth, J. R. (1939). Kashmiri (Specimen). *Le Maître Phonétique*, vol, *68*, Pp, 67–68.

Grierson, G. A. (1919). *Linguistic survey of India: Vol, VIII, Part 11. Royal Asiatic Society, New Delhi*

Kachru, B. B. (1966). *Kashmiri and other Dardic languages*. In T. A. Sebeok (Ed.), *Current trends in linguistics: Linguistics in South Asia* (Vol. 5, pp. 284–306). Mouton.

Kachru, B. B. (1969). *A Reference Grammar of Kashmiri* [Unpublished manuscript]. Department of Linguistics, University of Illinois.

Kaul, M. K., & Dhar, N. A. (2012). *A linguistic survey of Kashmiri dialects*, (Vol. 2). Central Institute of Indian Languages. Mysuru.

Kelkar, A. R. (1984). *Kashmiri: A descriptive sketch*. In O. N. Koul & P. E. Hook (Eds.), *Aspects of Kashmiri linguistics* (pp. 1–13). Bahri Publications.

Kelkar, A. R., & Trisal, P. N. (1964). Kashmiri word phonology: A first sketch. *Anthropological Linguistics*, *6*(1), 13–22.

Koul, O. N. (2005). *Studies in Kashmiri linguistics*. Indian Institute of Language Studies.

Masica, C. P. (1991). *The Indo-Aryan languages*. Cambridge University Press.

Morgenstierne, G. (1961). Dardic and Kāfir languages. In B. Lewis, C. Pellat, & J Schacht (Eds.), *The Encyclopaedia of Islam* (New ed., Vol. 2, Fasc. 25, pp. 138–139). E. J. Brill.

Sachdeva, R., Dhar, N. A., & Kaul, M. K. (2010). *A linguistic survey of Kashmiri dialects* (Vol. 1). Central Institute of Indian Languages.

Schmidt, R. L., & Koul, O. N. (1983). *Kohistani and Kashmiri: An annotated bibliography of Dardic languages*. Patiala Indian Institute of Language Studies.

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 91-108

From Ladakhi to Bhoti: Sociolinguistic and Political Dimensions of a Changing Language Identity in Ladakh

Konchok Tashi[*]

Abstract

The linguistic landscape of Ladakh has undergone significant transformation in recent years, shaped by evolving government policies and shifting cultural narratives. The shift from designating Ladakhi as the region's principal language to the more inclusive word Bhoti, which places the language within the larger Tibetic linguistic family, is crucial to this transformation. This article explores the sociolinguistic and political aspects of this change, examining the ways in which renaming practices relate to governmental acknowledgment, cultural legacy, and identity creation. Drawing on policy documents, community perspectives, and contemporary debates, the study explores the motivations behind adopting Bhoti as an official category, including aspirations for constitutional recognition, pan-Himalayan solidarity, and the pursuit of linguistic unity across Tibetic-speaking regions. At the same time, the research addresses challenges stemming from this change, specifically concerns about eliminating local language diversity and the ramifications for educational planning, resource allocation, and cultural autonomy. The paper offers a detailed explanation of how nomenclature functions as a symbolic and strategic weapon in influencing the future of linguistic identity in Ladakh by placing Ladakh's experience within larger conversations on language politics and minority rights.

Keywords: Ladakhi, Bhoti, Language and Politics, Linguistic Identity, Language Policy, Sociolinguistic Change

1. Introduction

Located in northern India's trans-Himalayan area, Ladakh boasts a complex linguistic environment that has been molded over centuries by trade, religion, cultural exchange, as well its geographic seclusion. The primary local language, known as Ladakhi, belongs to the Western Tibetic branch of the Tibeto-Burman language family. The language has several regional variants spoken across Leh, Kargil, Nubra, Changthang, and Zanskar regions. Ladakhi displays significant dialectal variations, reflecting the historical, social, and ecological peculiarities of the area, despite being frequently regarded as a single language group. For decades, “Ladakhi” served not only as a linguistic

[*] Central University of Jharkhand; konchok.tashi@cuj.ac.in.

identifier but also as a marker of regional identity and cultural belonging.

However, there has been a notable change in the nomenclature of the language in recent years. Increasingly, the name Bhoti—a larger descriptor used among numerous Tibetic-speaking populations in the Indian Himalayas has acquired awareness in political debate, cultural activism, and institutional discussions. Advocates of the name claim that "Bhoti" better expresses the linguistic and cultural links between Ladakh and the wider Tibetic world, offering support to pan-Himalayan groups seeking recognition under the Eighth Schedule of the Indian constitution. At the same time, the move from "Ladakhi" to "Bhoti" has aroused debate within the region, raising problems about identity, representation, and the potential erasure of local linguistic diversity and linguistic identity. The symbolic importance of language naming how labels may both reflect and actively create political objectives, cultural narratives, and social cohesion is highlighted by this developing dispute.

Studying this shift in nomenclature is also vital for understanding the broader sociopolitical transformations emerging in Ladakh, especially in the aftermath of its transition to a Union Territory in 2019. Changes in language labels overlap with concerns of cultural heritage, minority rights, educational planning, and regional politics, making them an important site for scholarly investigation. In order to better understand how renaming procedures affect identity formation, policy-making, and community dynamics in modern Ladakh, this article will look at the sociolinguistic and political aspects of the shift from "Ladakhi" to "Bhoti."

Guided by this aim, the article addresses the following research questions: (1) What historical, cultural, and political circumstances have led to the creation of "Bhoti" as an alternate term for the Ladakhi language? (2) How is this change interpreted and handled by various stakeholders, including members of the community, cultural institutions, religious organizations, and legislators? (3) What are the sociolinguistic implications of using a more inclusive nomenclature in Ladakh for identity creation, education, and language documentation? Therefore, the objectives of the study are to track the evolution of linguistic labels in the region, examine the reasons and conflicts underlying the shift toward "Bhoti," and estimate its potential impact on cultural and linguistic prospects in Ladakh.

2. Methodology

This study adopts a qualitative interpretive approach to examine the ongoing debate surrounding the nomenclature of the Ladakhi language and its implications for linguistic identity, language policy, and cultural representation in Ladakh. Rather than investigating linguistic structure, the study focuses on the sociopolitical dimensions of language naming and official recognition through the analysis of documentary and publicly available sources. The analysis is primarily based on

qualitative document analysis (Bowen, 2009), supplemented by thematic analysis (Braun & Clarke, 2006).

The study draws on a qualitative analysis of official policy documents, institutional statements, public representations, and media reports relevant to language policy and identity in Ladakh. The primary documentary sources include The Jammu and Kashmir Reorganisation Act, 2019 and The Ladakh Official Languages Regulation, 2025. The analysis also incorporates publicly available memoranda and representations submitted by the Ladakh Buddhist Association (LBA), statements issued by the Ladakh Academy of Art, Culture and Languages (LAACL), and public resolutions of the Apex Body, Leh, and the Kargil Democratic Alliance (KDA) concerning constitutional safeguards and language recognition. To contextualise public discourse, media reports from *The Hindu*, *The Indian Express*, *The Tribune*, *Greater Kashmir*, *Daily Excelsior*, and *Scoop News* were also examined. These materials were analysed using qualitative thematic analysis to identify recurring themes related to language policy, linguistic identity, constitutional recognition, and multilingual governance.

The collected materials were analyzed using thematic analysis, whereby recurring themes relating to language naming, linguistic identity, cultural heritage, official recognition, and language policy were identified, coded, and interpreted. Particular attention was paid to the ways in which different stakeholders including government institutions, cultural organizations, civil society groups, scholars, and community members construct and negotiate linguistic identity through the use of the terms *Ladakhi* and *Bhoti*. The analytical framework draws on sociolinguistic perspectives that view language names not merely as linguistic labels but as social and political markers closely associated with identity, community affiliation, and language planning (Fishman, 1991; Edwards, 2009; Spolsky, 2009). To enhance the credibility of the findings, information obtained from official documents was compared with statements issued by community organizations and publicly available media reports. This process of triangulation enabled the study to identify areas of convergence and divergence among official policy, institutional discourse, and community perspectives. The analysis remains interpretive and descriptive, aiming to understand how language naming has emerged as a contested issue within the broader context of identity, cultural preservation, and language policy in contemporary Ladakh, rather than advocating any particular policy position.

3. Historical and Linguistic Background

3.1 The Ladakhi Language: Classification and Features

The language spoken in the Union Territory of Ladakh often referred to as “Ladakhi” belongs to the Western Tibetic subgroup of the Tibetic branch within the Tibeto-Burman language family. Ladakhi is typically positioned in linguistic studies with similar dialects spoken in nearby

areas like Spiti, Purig, and Balti, which collectively constitute a continuum of Tibetic languages molded by millennia of cross-Himalayan cultural interchange. Ladakhi retains some key traits characteristic of Tibetic languages, including a primarily monosyllabic lexicon, a tonal or quasi-tonal prosodic system in some dialects, and grammatical structures and phonetic sounds that show significant similarities with Classical Tibetan. However, Ladakhi stands apart from other Tibetic dialects due to regional modifications in phonology, lexicon, and syntax, highlighting its status as a linguistically autonomous regional language.

Within Ladakh alone, Ladakhi demonstrates remarkable dialectal variations (Koshal, 1979). The Leh dialect, spoken in the central region, is widely regarded as the prestige or standard and serves as the linguistic basis for most written works and educational content. The Nubra dialect has variances in phonetic realization and lexicon, shaped by the region's historical commerce and migration trends. In Sham dialect (spoken in the Sham district of Union Territory of Ladakh), a more conservative language profile has been noted, with certain archaic characteristics preserved due to the area's geographic isolation. The language situation becomes much more convoluted in Kargil, where the Purig variety, closely related yet distinct, is spoken, reflecting a separate historical trajectory and Islamic cultural influences. These dialectal variations underscore the significance of regional linguistic identities within the larger Tibetic heritage by illuminating the diversity within what is often referred to as "Ladakhi."

Ladakhi has historically been written in the Tibetan alphabet (Uchen), a literary tradition that links the area to the larger Tibetan Buddhist cultural realm. Monastic institutions have long been the curators of this script culture, maintaining classical texts, holy literature, and pedagogical manuals. While the Tibetan script continues to enjoy strong cultural and religious prestige, modern efforts to standardize Ladakhi for use in schools and digital media have led to the exploration of alternative orthographic practices, including phonetic transcriptions and adaptations of Roman and Devanagari scripts. These initiatives are a reflection of larger discussions concerning modernization, language accessibility, and the preservation of classical traditions.

3.2 Evolution of Language Labels in Ladakh

The language used in the area has historically been referred to both internally and internationally as "Ladakhi", a name strongly associated with geographic location and regional identity. For centuries, the term "Ladakhi" served as a linguistic and ethnic marker, indicating the uniqueness of the people living in the former kingdom of Ladakh. Colonial-era language surveys, early anthropological research, and Himalayan trip literature all adopted "Ladakhi" as the principal term, so strengthening its prominence in academic and administrative discourse.

Over time, however, the term Bhoti (alternatively spelled Bodhi) began to appear in cultural and political discourse across the Indian Himalayas. Derived from “Bhot-desha,” the Sanskrit word for Tibet country, “Bhoti” refers broadly to the family of Tibetic languages spoken in territories historically affected by Tibetan culture, religion, and scriptural traditions (Tournadre & Suzuki, 2023). The use of "Bhoti" became popular in Ladakh, especially in the early 1970s, as a result of campaigns to have Tibetic languages recognized under the Indian Constitution's Eighth Schedule (Tashi, 2023). Proponents of the word say that “Bhoti” provides a cohesive linguistic identity for a geographically dispersed and culturally connected group, boosting claims for cultural preservation and governmental recognition.

Buddhist monastic establishments have played a significant role in popularizing the “Bhoti” term. Through their intimate linkages with Tibetan liturgy, scriptural learning, and trans-Himalayan networks, monastic communities have long emphasized linguistic continuity with ancient Tibetan traditions. Ladakhi has been framed as part of a broader Tibetic linguistic and spiritual history rather than just as a regional language because of this cultural focus. Consequently, the term Bhoti has developed symbolic meaning, reflecting not just language affinity but also cultural solidarity, religious identification, and pan-Himalayan unification.

However, discussions concerning the preservation of Ladakh's unique language and cultural identity continue to coexist with the growing popularity of "Bhoti." While some perceive Bhoti as a bigger umbrella term that improves political exposure, others express concern that it may conceal Ladakhi’s particular regional character. Thus, the dynamic interaction between local identity, regional solidarity, and national politics is reflected in the creation of language labels in Ladakh; this interaction continues to influence discussions concerning linguistic identity in the Union Territory today.

4. Nomenclature and Identity

4.1 Symbolic Power of Language Naming

The name of a language is never a neutral act; it is heavily steeped in social, cultural, and political meanings. Linguistic labels are potent symbols that communities use to express their identities, assert their validity, and negotiate their position in larger social hierarchies, according to sociolinguistics research. Names shape how speakers understand their language and how it is positioned in relation to other linguistic forms. They influence everything from educational policy to cultural pride, political mobilization, and intergroup relations. As thus, language naming itself becomes a form of identity-making, an act that reflects and promotes particular ideological orientations.

A helpful theoretical framework for comprehending the dynamics at work in Ladakh is provided by Bourdieu's concept of symbolic capital. Bourdieu defined symbolic capital as the legitimacy, prestige,

and recognition that an identity or cultural practice gains within a particular social sector (Bourdieu, 1984). When applied to language, the notion highlights how particular labels have cultural authority or political weight, enabling speakers to align themselves with powerful organizations or historical traditions. In the instance of Ladakh, adopting the name "Bhoti" invokes symbolic capital derived from Tibetan literary and religious tradition, whereas "Ladakhi" takes its symbolic strength from regional rootedness and historical individuality. As a result, selection of labels reflects conflicting ideas of community, legacy, and political ambition and is not only linguistic phenomenon.

4.2 Local Understandings of "Ladakhi" vs. "Bhoti"

At the community level, the vocabulary employed to characterize the local language generates different and frequently competing meanings. Many Ladakhis associate the term "Ladakhi" with a sense of cultural uniqueness, local history, and regional identity. It represents geographical identity, resonating with the social and historical experience of living in the unique environment of the high-altitude Western Indian Himalayas. The phrase frequently holds links with regional customs, oral traditions, and everyday language practices that separate Ladakh from other Tibetic-speaking places.

The rising use of Bhoti, however, generates a different set of connections. The term's proponents frequently see it as a symbol of pan-Himalayan cultural cohesion, based on common scriptural practices and Buddhist ancestry. Bhoti represents inclusion in a larger Tibetic cultural realm that cuts across geographical limits for its speakers. It also functions as a political vehicle, notably for those seeking constitutional recognition of Tibetic languages in India, as it combines different people and communities under a common linguistic name.

However, there has been discussion in Ladakh regarding the use of Bhoti. Using an umbrella name, according to some community members, runs the risk of eroding the distinctive character contained in Ladakhi, and homogenizing the linguistic diversity of the area. These issues are especially prominent in locations where language or cultural practices diverge from the Buddhist-majority regions, such as in Kargil, where Purig speakers may not fully identify with the name Bhoti. Thus, the discussion highlights a conflict between the need to maintain regional uniqueness and the aim for cultural unification. While Bhoti may improve political visibility and cultural togetherness, it can simultaneously raise worries of losing the distinctive language heritage that differentiates Ladakh within the greater Tibetic constellation.

The nomenclature of a language is strongly related to the identity, history, and cultural continuity of its speakers and for the people of Ladakh, the name Ladakhi signifies their distinct legacy within India. Changing this established linguistic label to Bhoti risks creating confusion and blurring the distinct identity that Ladakhis have preserved for generations, especially given that Bhoti is widely

understood to mean “Tibetan language,” a classification that does not accurately represent the Indian identity of Ladakhis. While linguistic and cultural similarities among Himalayan regions are apparent, establishing a new nomenclature without broad consent can unintentionally weaken Ladakh’s own rich language legacy. The continued use of the nomenclature *Ladakhi* reflects the linguistic identity and cultural heritage of its speech community and contributes to its recognition as one of the many languages that constitute India's diverse linguistic landscape.

According to Stobdan (2019), “Most of the issues are seemingly steered by a Tibet-oriented agenda, be it sectarian, linguistic or political. For example, the agenda of getting the Bhoti (Tibetan) language included in the Schedule VIII of the Indian Constitution was initially set by Dharamshala through its proxies from the Himalayan region. The move is gaining momentum and being embedded in the agenda of local political parties.”

4.3 The Pan-Himalayan Context

The altering linguistic nomenclature in Ladakh must be viewed within a broader pan-Himalayan framework in which various Tibetic-speaking communities have chosen labels such as Bhoti, Bodhi, or Bodic to characterize their languages. These phrases stress a shared cultural lineage rooted in the Tibetan alphabet, Buddhist textual traditions, and historical exchanges across the trans-Himalayan region. To express a sense of linguistic and cultural continuity, communities in Ladakh, Himachal Pradesh, Sikkim, Arunachal Pradesh, and portions of Nepal also use "Bhoti" or similar terminology. However, the regions host a rich but vulnerable linguistic diversity, with endangered languages such as Ladakhi, Purgi, Balti, Tinani, Gahri, Spitian, Kinnauri, Sherpa, Tamang, Gurung, Limbu, Lepcha, Bhutia, Monpa, and Sherdukpen spoken by the local communities (UNESCO, 2010; Moseley, 2010; Devy, 2013).

A number of movements have surfaced throughout the Indian Himalayas calling for "Bhoti" to be officially recognized as a separate language in the Eighth Schedule of the Indian constitution. These initiatives maintain that a uniform nomenclature gives a stronger communal identity and increases the probability of winning state funding for language preservation, education, and cultural development. By bringing together multiple Tibetic-speaking populations under a unified banner, proponents hope to magnify their political voice and bolster claims to cultural autonomy.

At the same time, the pan-Himalayan use of “Bhoti” emphasizes difficult problems about linguistic taxonomy and identity politics. While the phrase encourages solidarity, it also embraces a vast array of separate but related languages, each with its own sociocultural context and historical history. Engaging with this more comprehensive framework presents Ladakh with both opportunities and challenges: it offers a forum for mutual cultural acknowledgment, but it also poses

important queries regarding the preservation of local identities within wider collectivities.

5. Political Dimensions of the Shift

5.1 Language Policy in Ladakh Post Union Territory Status

Language policy in the area was directly impacted by the major administrative and political changes brought about by the reorganization of Jammu and Kashmir State in 2019, which resulted in the establishment of Ladakh as a distinct Union Territory (UT). As Ladakh embraced a new governance framework under central administration, problems of language identity, cultural preservation, and formal recognition gained prominence. The move to UT status reduced the region's dependence on the previous state government and created new opportunities for local leaders, cultural organizations, and civil society groups to communicate their linguistic goals more aggressively.

The UT administration of Ladakh has consistently underlined the importance of protecting the region's cultural legacy, which is intimately linked to language, religion, and traditional knowledge systems, in the post-2019 administrative landscape. Policy discussions have underlined the necessity of promoting indigenous languages in education, public communication, and cultural initiatives. However, the absence of a constitutionally recognized official language for the UT has caused a vacuum, creating opportunity for competing claims about which linguistic identity best symbolizes Ladakh's diversity and legacy. In this context, the rising institutional reference to the term Bhoti represents an emerging trend among policymakers and powerful cultural actors to locate Ladakh's linguistic identity within a broader Tibetic framework.

The Constitution of the erstwhile State of Jammu and Kashmir recognized "Ladakhi language" as one of the regional languages of the State under Section 146 (Academy for the Development of Art, Culture, and Languages) of the Sixth Schedule (Regional Languages). In addition to Ladakhi, the other regional languages recognized under this provision included Dogri, Kashmiri, Balti, Shina, and Purgi, reflecting the State's rich linguistic and cultural diversity.[1]

The Ladakh Official Languages Regulation, 2025, a landmark decision recognizing the region's distinctive linguistic and cultural heritage designates Purgi, Bhoti, Urdu, Hindi, and English as official languages of the UT. However, despite their long-standing demands being fulfilled, the Bhoti proponents did not celebrate when the announcement recognizing Bhoti as an official language was made. This is quite unusual and suggests that the nomenclature 'Bhoti' may not have widespread support. Further, given the importance of linguistic identity in language planning, the question of whether Ladakhi should also receive explicit official recognition merits further consideration by policymakers.

The regulation has been promulgated pursuant to the powers conferred by Article 240 of the Constitution of India and Section 58(2) of the Jammu and Kashmir Reorganisation Act, 2019 (as notified in the Gazette on June 2, 2025). This significant measure acknowledges and celebrates the multilingual character of Ladakh's population, representing a decisive step towards inclusive governance.

The Gazette notification specifies:

With effect from such date as the Administrator may, by notification in the Official Gazette, appoint in this behalf, the English, Hindi, Urdu, Bhoti, and Purgi languages shall be the official languages to be used for all or any of the official purposes of the Union Territory: Provided that the English language shall continue to be used for all the official purposes of the Union Territory for which it was being used before the date of commencement of the Regulation."

Furthermore, the notification states:

The Administrator may take such necessary steps, as may be provided by rules to be made by him, to strengthen the institutional mechanisms for the promotion and development of other languages in the Union Territory and for the establishment of the Academy of Art, Culture, and Languages. (2) The institutional mechanisms referred to in sub-section (1) shall make special efforts for the promotion and development of other native languages of Ladakh, namely, Shina (Dardic), Brokskat (Dardic), Balti, and Ladakhi."

5.2 The Push for Constitutional Recognition

Demands to include Bhoti in the Eighth Schedule of the Indian Constitution comprise a fundamental political motivation behind the shift in language nomenclature. The movement for constitutional recognition extends back several decades, gaining pace via combined efforts by diverse Himalayan populations who speak Tibetic languages. These communities contend that being included in the Eighth Schedule would guarantee more official representation in national institutions dedicated to language promotion, increased government support for language preservation, the creation of educational resources, and chances for literary advancement.

A network of political, religious, and cultural organizations—including the Ladakh Buddhist Association (LBA), the Ladakh Gonpa Association (LGA), and various civil society organizations under the Apex Body, Leh has played an important role in advocating for the inclusion of Bhoti in the Eighth Schedule of the Constitution of India. Buddhist monastic organizations, with their profound historical linkages to Tibetan script and literature, have played a key role in expressing the cultural reason for adopting "Bhoti" as the language's official title. This initiative has been joined by social groups and political leaders from all throughout the Himalayan region, who have framed constitutional recognition as crucial to protecting their linguistic

legacy and increasing their political prominence inside the Indian Union.

A crucial strategic rationale for accepting the broader moniker "Bhoti" is its perceived inclusivity. Whereas "Ladakhi" is particular to the region, "Bhoti" operates as a pan-Himalayan term that combines numerous Tibetic-speaking communities under a single linguistic flag. This common identity strengthens the struggle for constitutional recognition by presenting a unified front that transcends regional borders. Thus, the use of "Bhoti" as a strategic decision indicates both cultural affinity and political pragmatism.

5.3 Implications for Governance and Representation

The increasing use of the name "Bhoti" has important ramifications for Ladakh's cultural image, education, and administration. If the term were to become institutionalized, it might have an impact on how resources are allocated for language promotion, such as curriculum development, teacher training, textbook creation, and orthographic practice standards. Under a recognized linguistic framework, cultural institutions like monasteries, municipal archives, literary groups, and arts councils stand to gain greater prominence and funding.

In educational settings, adopting "Bhoti" may facilitate the construction of unified instructional resources that fit with pan-Himalayan language traditions. However, this may provide issues for expressing local dialectal variations and potentially marginalize varieties of speech that do not closely fit with the conventional Tibetic paradigm. Similar to this, the emphasis on classical or pan-regional forms in literary composition may cause vernacular idioms that characterize Ladakh's own linguistic legacy to be marginalized.

Beyond the internal dynamics of Ladakh, the move has broader political ramifications. The adoption of "Bhoti" aligns Ladakh more closely with national discourses on cultural preservation, minority rights, and regional integration. It also connects with identity politics inside India, particularly arguments involving indigeneity, the politics of recognition, and the link between regional identities and pan-Himalayan solidarity. At the geopolitical level, the adoption of a Tibetic-associated label carries symbolic significance, gently recalling historical and cultural linkages that reach beyond national borders.

Thus, the political components of the nomenclature shift reflect a complicated interplay between local priorities, regional alliances, and national considerations. The decision between "Ladakhi" and "Bhoti" is not only a linguistic process but is profoundly ingrained in conflicting ideas for the future of governance, identity, and representation in Ladakh.

6. Sociolinguistic Implications

6.1 Language Vitality and Documentation

The move from "Ladakhi" to "Bhoti" carries substantial consequences for the viability and documentation of the region's linguistic landscape.

For successful linguistic description, revitalization, and preservation, researchers studying Tibetic languages frequently stress the importance of precise classification and unambiguous nomenclature. Adopting "Bhoti" as a general term could increase awareness and draw more institutional backing, especially for initiatives pertaining to the preservation of Indian Himalayan languages. But it also runs the risk of masking Ladakh's linguistic communities' inherent variety.

One potential result is the flattening of local language diversity. Because "Bhoti" refers to a broad variety of Tibetic languages, it may cause scholars or decision-makers to treat diverse dialects as a single, homogeneous group. This may lessen the motivation to record smaller or more endangered languages or dialects, especially those that are spoken in remote or culturally unique places like Kargil's Purig-speaking regions, Zanskar, or Changthang. Micro-level language differences that are crucial to comprehending Ladakh's social history and cultural diversity may eventually be overlooked due to the focus on a single linguistic identity. Therefore, although the new nomenclature might increase exposure at the macro level, it might also make linguistic subtleties less noticeable at the local level.

6.2 Education and Curriculum Development

The term "Bhoti" is becoming more and more popular, which poses a number of difficulties for curriculum creation and educational policy. The variety of spoken forms in Ladakh and variances within the larger Himalayan region must be taken into consideration while attempting to standardize the language. Standardization becomes more problematic when a label contains numerous related but separate languages, prompting problems regarding which dialects or varieties should serve as the normative basis for textbooks, grammar guides, and school instruction.

The choice of script is a major point of contention. The Tibetan alphabet has long been connected with classical literature, monastic education, and cultural identity in Ladakh; hence, its usage in formal education promotes ties to the region’s Buddhist legacy. However, for populations whose linguistic practices diverge from classical Tibetan norms, or for regions where the alphabet is not traditionally used (such as Purig-speaking Muslim-majority areas), a Tibetan-based orthography may bring educational issues. Furthermore, some educators support the use of Romanized or Devanagari-based orthographies in national curricula or to increase accessibility for younger students. These varying views are a reflection of larger conflicts between language inclusion, educational pragmatism, and cultural preservation.

The nomenclature change also has an impact on educational resources and teacher preparation. If “Bhoti” becomes the dominant institutional word, teaching materials may increasingly emphasize pan-Himalayan linguistic traits rather than region-specific ones. Teachers may need to adjust to standardized curriculum models, complete new

training programs, and reconcile local spoken forms with traditional Tibetan customs. Additionally, it can put strain on the already fragile educational infrastructure, which has to strike a balance between the requirements of curriculum uniformity and the necessity for culturally specific content.

6.3 Community Identity and Intergroup Relations

In Ladakh, the transition from "Ladakhi" to "Bhoti" has significant effects on intergroup relations and communal identity. While, generally portrayed as a language or cultural issue, the nomenclature dispute links with regional, religious, and ethnic identities in complex ways. The adoption of "Bhoti" resonates powerfully with monastic traditions, scriptural legacy, and trans-Himalayan cultural networks for areas with a majority of Buddhists. It strengthens a sense of belonging to a greater Tibetic world and corresponds with longstanding cultural affinities.

In contrast, populations in Kargil, where Purig and Balti variations are spoken and where Islamic cultural history predominates, may not identify as readily with the name “Bhoti,” which they may consider as intimately tied to Buddhist identity and Tibetan customs. "Ladakhi" or region-specific labels might more accurately represent local histories and sociocultural realities for these communities. These varied opinions underline the prospect that adopting a pan-Himalayan name could alienate specific groups, thereby aggravating interregional differences or reinforcing existing identity boundaries.

Pluralism and inclusivity therefore appear as major themes. As “Bhoti” has already been recognized as one of the official languages in the UT Ladakh, it must now embrace the different linguistic and cultural experiences of all Ladakhis, not primarily those linked with the Buddhist monastic tradition. Fostering a sense of shared belonging requires, making sure that the educational resources, cultural events, and policy efforts represent this variety. Without such initiatives, the nomenclature change could exacerbate social and cultural disparities, influencing intergroup relations in ways that have an impact on regional harmony and political stability.

7. Comparative Perspectives

When considered in larger regional and global contexts where language nomenclature has been crucial in forming identity, politics, and cultural representation, the discussions surrounding the transition from "Ladakhi" to "Bhoti" take on greater depth. Examining analogous cases among other Tibetic-speaking groups as well as global linguistic experiences reveals shared patterns of negotiation between local, particular and collective identity, affording vital comparative insights for understanding Ladakh’s contemporary trajectory.

Across the Indian Himalayan regions and beyond, nomenclature politics have emerged prominently in locations such as Sikkim, West Bengal (Darjeeling, Kalimpong, Siliguri), Arunachal Pradesh, Lahaul-Spiti, Nepal and Baltistan (CIIL, 2008) where language identities

connect with complicated cultural and geopolitical history. In Sikkim, for instance, the language typically referred to locally as "Bhutia" is intimately tied to Tibetan linguistic traditions, while its naming reflects the state's unique sociopolitical situation and cultural diversity. Codifying and promoting the language has required striking a balance between ties to the larger Tibetic tradition and localized identity. Similarly, in Arunachal Pradesh, numerous Tibetic-speaking groups employ designations such as Bhoti to show cultural identification with the Tibetan script and monastic traditions, while concurrently navigating the multicultural and multi-ethnic context of the state.

Communities in Spiti (Himachal Pradesh) speak a language that is closely related to Ladakhi, but they have traditionally preferred labels that are distinctive to the region, like Spiti Bhoti or Spitian (CIIL, 2008). While recognizing linguistic connection with the larger Tibetic family, these labels highlight local uniqueness. Baltistan, now entirely in Pakistan-administered territory, offers yet another instructive example. Baltis speak a Tibetic language formerly written in the Tibetan alphabet, although over time the usage of Perso-Arabic scripts and Islamic influence has transformed linguistic and cultural identification. Discussions about nomenclature in Baltistan show how language identities are shaped by script, religion, and geopolitics in ways that are similar to conflicts in Ladakh (Zeisler, B. (2011).

Comparative evidence from other multilingual contexts demonstrates that language naming and classification are often closely linked to questions of political identity, nation-building, and official recognition. Following the dissolution of Yugoslavia, the formerly unified designation *Serbo-Croatian* gradually gave way to separate standard languages Serbian, Croatian, Bosnian, and later Montenegrin reflecting the emergence of distinct national identities despite their high degree of mutual intelligibility. As a result of attempts to maintain indigenous identity and local linguistic variation, Quechua, which was formerly thought of as a single language across the Andes, is now acknowledged as a collection of different but related languages (Greenberg, 2004; Bugarski, 2004). Even within the Tibetan linguistic community, disagreements exist about whether regional languages such as Amdo and Khams should be recognized as dialects of Tibetan or as distinct languages. These examples mirror the tensions in Ladakh between linguistic unification (as symbolized by "Bhoti") and the preservation of regional linguistic particularities (as represented by "Ladakhi").

Comparative analysis suggests various lessons for Ladakh. First, linguistic labels work best when they strike a balance between local identity sensitivity and political prominence. Although requests for recognition may be strengthened by overly general nomenclature, it runs the risk of obscuring internal variety or marginalizing particular populations. Adopting a unified script or curriculum must be complemented by procedures that take into account regional linguistic realities. Second, script choice and educational policy constantly

emerge as crucial arenas where identity negotiations take place. Third, worldwide experiences suggest that language renaming can have unexpected repercussions, such the rigidification of identities or the loss of linguistic nuance if standardization attempts prioritize governmental objectives over community traditions.

Comparative viewpoints generally show that Ladakh's transition from "Ladakhi" to "Bhoti" is neither singular nor isolated. It is a part of a larger trend where language nomenclature is used to navigate cultural plurality, negotiate governmental acceptance, and proclaim identity. Understanding these connections can help Ladakh create tactics that foster cultural unity without jeopardizing the rich linguistic diversity that distinguishes the region.

8. Discussion

The previous analysis emphasizes the complex interaction between political factors and sociolinguistic dynamics in the transition from Ladakhi to Bhoti. On the one hand, the word "Bhoti" was used as a result of intentional political and cultural tactics meant to increase visibility, obtain constitutional recognition, and claim a common Tibetic ancestry. By connecting Ladakh's language with a broader Himalayan linguistic framework, proponents of Bhoti want to enhance the region's claims for institutional support, educational resources, and cultural preservation. The terminology transition consequently bears concrete benefits in terms of political leverage, pan-Himalayan solidarity, and the symbolic value involved with tying local identity to classical Tibetan legacy.

A number of possible disadvantages are also shown by the sociolinguistic implications. Stressing Bhoti runs the risk of homogenizing a linguistically varied region, possibly marginalizing regional dialects and diminishing the distinctive qualities of Purig, Leh, Sham, Zanskar, Nubra and Changthang variants. These conflicts may be exacerbated by curriculum development, script disputes, and educational standardization, which favour pan-regional forms over regionally particular linguistic practices. Furthermore, some communities especially those in Kargil and other minority linguistic regions may feel marginalized or alienated from a single linguistic category as a result of the transition, which poses social and cultural concerns.

The discussion revolves around this conflict between political unity and linguistic diversity. Internal distinctions that are essential to community identification, oral history, and cultural expression may be obscured by a broad name like Bhoti, even though it improves institutional recognition and promotes collective identity. Policymakers, educators, and community leaders must balance these conflicting needs in order to preserve the rich internal variety of Ladakh's linguistic environment while utilizing the benefits of a single nomenclature.

In a broader sense, Ladakh serves as an example of how language naming can influence cultural futures. Beyond just categorizing speech, linguistic labels actively shape social reality, shape educational agendas, and mediate access to cultural and political resources. A name can become a place of debate over identity, legacy, and regional affiliation, as demonstrated by the decision between Ladakhi and Bhoti. Decisions about nomenclature have long-term consequences: they impact which kinds of language are preserved, which narratives of culture obtain authority, and which populations feel recognized or ostracized. In Ladakh, as elsewhere, language naming operates simultaneously as a symbolic, political, and practical tool that can either support cultural cohesiveness or worsen internal differences.

When these results are combined, it is evident that the transition from Ladakhi to Bhoti is not just linguistic matter. It represents a difficult balancing act between policy, heritage, identity, and community goals. Recognizing the different features of this change is vital for designing inclusive methods that safeguard linguistic variety while supporting the strategic advantages of greater cultural and political recognition.

9. Conclusion

The sociolinguistic and political aspects of the transition from Ladakhi to Bhoti have been explored in this article, emphasizing how language naming serves as a platform for political strategy, cultural expression, and identity negotiation in Ladakh. The study shows that nomenclature is both symbolic and practical: it shapes how communities view themselves, how governments and institutions interact with them, and how linguistic resources are preserved and transmitted. It does this by tracing the historical and linguistic background, investigating community perspectives, analysing post-2019 policy developments, and placing Ladakh within regional and global comparative frameworks.

The article's main contributions include highlighting the significance of language naming in determining long-term cultural trajectories, identifying the advantages and disadvantages of adopting the Bhoti label, and articulating the intricate relationship between regional linguistic diversity and pan-Himalayan political unity. The discussion underlines that while Bhoti may boost political recognition and institutional backing, it also risks homogenizing diverse dialects and marginalizing distinctive community voices, underlining the necessity for careful balancing between inclusivity and uniformity.

In the future, Ladakh's language identity will probably be influenced by the continuous balancing act between local uniqueness and more general Tibetic ties. Policymakers must establish language policies that acknowledge both the aggregate importance of Bhoti and the micro-level diversity of Ladakhi dialects. Educators, cultural organizations, and community leaders play a vital role in ensuring that standardization initiatives do not swamp local linguistic patterns, while

simultaneously exploiting the symbolic and practical advantages of a single label.

For researchers, this work suggests to various avenues for additional inquiry. Ethnographic studies could provide further insight into how different cultures interpret and engage with the Bhoti term in daily life. In addition to protecting linguistic diversity, systematic documentation of dialectal variances throughout Leh, Zanskar, Sham, Nubra, Changthang, and Kargil would guide governmental and educational endeavours. The socio-political ramifications of nomenclature changes throughout time may also be made clearer by longitudinal studies of the effects of language policy and standardization initiatives.

Ultimately, the story of Ladakh reveals the enormous relevance of language naming in constructing cultural futures. It serves as a reminder that language labels are never purely descriptive; they are tools of identity building, instruments of political negotiation, and frameworks for preserving or transforming the heritage of communities in swiftly shifting social and political contexts.

References

Annamalai, E. (2001). *Managing multilingualism in India: Political and Linguistic manifestations*. Sage

Beek, M., & Bray, J. (Eds.). (2012). *Ladakh: culture, history, and development between Himalaya and Karakoram*. Brill.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste* (R. Nice, Trans.). Harvard University Press.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education*. Greenwood Press.

Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press.

Brass, P. (1974). *Language, religion, and politics in North India*. Cambridge University Press.

Bray, J. (Ed.). (2005). *Ladakhi histories: Local and regional perspectives*. Brill.

Cooper, R. L. (1989). *Language planning and social change*. Cambridge University Press.

Driem, G. (2001). *Languages of the Himalayas*. Brill.

Goldstein, M. C. (2004). *A Tibetan revolutionary: The political life and times of Bapa Phüntso Wangye*. University of California Press.

Koshal, S. (1979). *Ladakhi grammar*. Motilal Banarsidass.

Koshal, S. (1990). The Ladakhi language and its regional perspectives. *Acta Orientalia Academiae Scientiarum Hungaricae, 44*, 13–22.

Koul, O. N. (Ed.). (2014). *People's Linguistic Survey of India: Languages of Jammu & Kashmir* (Vol. 9, pp. 233–243). Orient Blackswan.

Roche, G. (2016). The Tibetan language(s): Politics, pluralism, and identity. *Himalaya: The Journal of the Association for Nepal and Himalayan Studies.*

Roche, G., Dede, K., & Pirie, F. (Eds.). (2008). *Modern Tibetan literature and social change*. Duke University Press.

Schiffman, H. F. (1996). *Linguistic culture and language policy*. Routledge.

Tashi, K. (2016). Bhoti: The classical language of the Himalayas. In T. Namgyal (Ed.), *A Cooper approach: The Trans-Himalayan Tibet: History, language & literature (traditional & contemporary)*. Mankin Press.

Tashi, K. (2019). Preservation and development of Tibetan/Bhoti language in Indian Himalayas. In J. Samten, L. Dorjee, P. Dorjee, et al. (Eds.), *In search of truth* (Part II). Alumni Association of CIHTS. Varanasi.

Tashi, K. (2020). Classical Tibetan language and culture in Indian Himalayas. *Kalakalpa (IGNCA Journal of Arts), 5*(1), 201–216.

Tashi, K., Tournadre, N., & Norman, R. (2024). Written Ladakhi and the future of Ladakh's culture. *Forum for Linguistic Studies.*

Tashi, K. (Ed.). (2021). *Languages of Indian Himalayas*. Indu Book Services.

Tashi, K. (2023). *Understanding language attitude and linguistic ecology in Indian Himalayas: A perspective from an endangered language "Ladakhi"*. Indu Book Services.

Tashi, K. (2025). *Tibetan language for non-Tibetans: A beginner's guide to writing and speaking Tibetan*. Springer Nature. https://doi.org/10.1007/978-981-96-2879-7

Tournadre, N. (2009). The Tibetic languages and their classification. *Linguistics of the Tibeto-Burman Area*, *32*(1), 105–182.

Tournadre, N., & Suzuki, H. (2023). *The Tibetic languages: An introduction to the family of languages derived from Old Tibetan*. Lacito, Cnrs.

Policy / Institutional / Community-level Sources

Bazm-e-Adab. (2022). *Letter to the Lieutenant Governor of Ladakh arguing that "Bhoti" is neither a script nor a language of Ladakh*. Voice of Ladakh

District Institute of Education and Training (DIET), Leh. (2023). *Bhoti grammar training/workshop by DIET Leh concludes* [Press release]. Peaceful Societies+2Reachladakh+2

Ladakh Administration. (2024). *Inauguration of Bhoti Cell and Language Lab at DIET Leh* [Press release].

Ladakh Buddhist Association. (2023). *Seminar on the preservation and promotion of the Bhoti language at Chokhang Vihara, Leh* [Press release].

Purig Cultural and Literary Society. (2024). *Press release demanding exclusive use of "Ladakhi" as the language of Ladakh*.

Analyses, Critiques, and Media Discussions

News report: "*Bhoti is not a language: Why parts of Ladakh have an issue with the Education Department's language plan." EdexLive*. (on objections to imposing "Bhoti" across Ladakh) EdexLive

Local media discussion: *"Ladakh Ethnic Groups Unite to Demand Multi-Language Official Status."* Voice of Ladakh, 2024. (on demands to recognise Purgi, Balti, Shina, and other languages along with or instead of Bhoti) Voice of Ladakh

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 109-130

Language and its Politics in Indian Trans-Himalaya: Beyond Bhoti Versus Ladakhi

Jegmet Sangyas*
Sonam Joldan**

Abstract

Language politics in Ladakh involves the twin issues of recognition and preservation. Contestation on nomenclature is concerned with the recognition and the debate on writing system is about preservation. There are two broad views regarding both, one adopts a conservative approach and another liberal. This paper is an attempt to delve into the complexity of the language issue in Ladakh with primary focus on Bhoti versus Ladakhi contestation. The study adopts a qualitative method relying on secondary data based on survey of literature as well as primary data based on government documents and personal interviews with language experts in Ladakh. The study suggests that Bhoti as a classical language with rich literature, geographical expansion and mother language of many Himalayan languages can be considered high standard language appropriate for official recognition in the eighth schedule of the constitution of India. Whereas, Ladakhi being a variant of Bhoti can be provided scope for deviation from classical Bhoti in writing and accord official recognition at the state level. Language education shall move from familiar to unfamiliar by introducing textbooks in local Ladakhi language at primary level and in Bhoti at higher level. While religious and literary texts can be in classical Bhoti, popular literatures like magazines, newspapers etc. can be considered more appropriate in Ladakhi.

Key Words: Language Politics, Ladakhi, Bhoti, Purgi, Balti

1. Introduction

Emergence of regional or social varieties and dialects from a mother language is significant in the history of a language. There are instances where certain variety of a language becomes high variety used by the community for religious and literary purposes, whereas, a regional colloquial variety become low variety used by the community as lingua franca. The Bhoti versus Ladakhi is a similar case in point, where, the former has become high variety and the latter a low variety. Since Ladakhi is one of the Himalayan languages belonging to Tibeto-Burman language family, it has retained classical Tibetan elements,

* University of Ladakh, sjoldan@yahoo.co.in
** University of Ladakh, info@uol.ac.in

however there are elements which are specific to it. People who understand Ladakhi well generally finds it difficult to read and understand written classical Tibetan. It is because, in spoken Ladakhi language, people use many indigenous as well as loan words.

The standard classical Tibetan language and script was developed by the Tibetan scholar, Thonmi Sambhota, in the 7th century. The religious texts used by Tibetans as well as Buddhists of Indian Himalayas are typical examples of written standard classical Tibetan language. Whereas, Ladakhi is a colloquial language used in All India Radio Leh, *Ladakhi Sheraza*[1] and as a lingua franca across Ladakh. Although it is written in the same Tibetan script, there are scope for deviations in the words and grammar from standard classical Tibetan. Meanwhile, Bhoti is an umbrella term developed in 1970s to bring all Himalayan languages, which belongs to Tibeto-Burman language family, under a common nomenclature and project it for constitutional recognition under 8th schedule of the constitution of India. Although, the rationale is to base 'Bhoti' on the script and the grammar of standard classical Tibetan language but consensus is yet to achieve on its standardisation and nomenclature.

It would not be wrong to assume that Bhoti will be closer to classical Tibetan language but Ladakhi on the other hand would be slightly different and localised. It is due to these deviations, particularly in grammar, that there were cases of resistance to written Ladakhi colloquial language by the proponents of Bhoti in Ladakh. Therefore, there has been and still is a debate on writing system between the proponents of Bhoti and Ladakhi. The baseline of Bhoti versus Ladakhi contestation is the debate on maintaining standard and scope for deviation in writings. In this context, it can be argued that given the reality of the language fluidity there must be scope for deviation and modifications in the colloquial language (known as "*Phal-skat*" in Ladakhi) but at the same time standard is also necessary to be maintained in the classical language (Known as "*Chos-skat*" in Ladakhi) to ensure the preservation of the authenticity of the mother language.

Language is also an important basis of social identity and culture. A community acquire and preserve its identity through language and as a cultural marker of group identity language acquires more significance than religion (Warikoo, 1996). Therefore, language assertion and the case for its official recognition become a complex and a contested issue. The case of language issue in Ladakh shows how communities tend to politicise and communalise language in the process of seeking official recognition to it. The movement for the inclusion of Bhoti language in the eighth schedule of the constitution of India and the recent declaration of official languages of the Union Territory (UT) of

Ladakh are main issues around which the language debate in the region revolves.

2. Methodology

The study relies on qualitative method using both primary and secondary data. While the study uses government documents, study report, organisational publications and personal interviews for primary data, its secondary data is based on books, research articles, periodicals and newspaper articles. The case study report of Central Institute of Indian Languages (CIIL) as well as the pamphlet published by Himalayan Buddhist Cultural Association (HBCA) are significant sources of primary data for understanding the case of the movement for constitutional recognition of Bhoti language. Personal interviews were conducted with range of language experts to have better understanding of Bhoti versus Ladakhi debate in particular and the dynamics of policy issue, identity contestation and language politics of Ladakh in general.

Purposive and snowball sampling methods were used to collect data by employing semi- structured interviews. Since the language contestation and politics in Ladakh mainly revolves around Bhoti versus Ladakhi debate, attempt has been made to explore and analyse perspectives of both the proponents of Bhoti and Ladakhi. Therefore, language experts were chosen to ensure proportionate representation of both the groups. Whereas, experts of other languages were also interviewed to know perspectives on language issue in Ladakh beyond Bhoti versus Ladakhi. In doing so, attempt has been made to create a corpus construction of discourses on the issues of pressing language issues in Ladakh regarding the preservation, promotion, recognition and identity.

2.1 Language Diversity of Ladakh

Languages spoken in Union Territory of Ladakh can be classified into two major groups belonging to Tibeto-Burman and Indo-Aryan family respectively. Ladakhi, Balti and Purgi belongs to the first group and Shina and Brokskat belongs to the second group. While Ladakhi speakers predominate Leh district and Zanskar tehsil in Kargil district, Balti speakers are confined to Turtuk tehsil in Leh district and Hardas as well as Lato village in Kargil district. Whereas, Purgi speakers predominate Kargil district excluding Zanskar and Drass. Meanwhile, Shina (Dardic) speakers are predominant in Drass belt[2] and Brokskat (Dardic) speakers are confined to Dah-Hanu belt[3]. The UNESCO placed languages spoken in the Union Territory of Ladakh under the category of endangered languages with Ladakhi, Balti and Purgi as vulnerable and Dardi-Arya[4] language as definitely endangered (Tashi, 2023). However, these languages except Purgi were recognised regional languages in the sixth schedule of the constitution of erstwhile state Jammu and Kashmir. The section 146 of the constitution also envisaged the establishment of an Academy of Art, Culture and Languages for the development of official languages as well as recognised regional languages of the sixth schedule (Noorani, 2014).

Ladakhi spoken in and around Leh and Purgi spoken in and around Kargil has gained dominant position in Leh and Kargil districts respectively (Tashi et al., 2024). Although Ladakhi, Balti and Purgi belong to the same Tibeto-Burman family, there are slight variations among them but mutually intelligible. Meanwhile, Shina and Brokskat though belong to the same Indo-Aryan language family are largely unintelligible to the speakers of each other. However, the differences among them are further being complicated by census classifications, official language policies and demand for recognition.

The union government of India recently declared English, Hindi, Urdu, Bhoti and Purgi as official languages of UT Ladakh. The president of India promulgated 'The Ladakh Official Languages Regulation 2025' on 2 June 2025. The clause 4 (1) of the regulation also envisage establishment of an Academy of Art, Culture and Languages and strengthening of institutional mechanism for the "promotion and development" of other languages in UT Ladakh (MLJ, 2025). Further, the regulation emphasises on "promotion and development" of native languages which include Ladakhi, Balti, Shina (Dardic) and Brokskat (Dardic). However, public debate before and after the official language declaration signifies an ongoing debate within language assertion in Ladakh and the dominant one has been the contestation between Bhoti versus Ladakhi.

2.2 Language Assertions of Ladakh and Bhoti Movement.

Though Bhoti versus Ladakhi contestation is not very old, the trajectory of language assertion and official policies adopted from time to time are pertinent to discuss for better understanding of the contestation. Earliest assertion of a language was made by Kashmir Raj Bodhi Mahasabha[5] in its memorandum submitted to the Glancy Commission, a commission of enquiry appointed by British, in early 1930s. The memorandum contained, among others, the demand for "a change in the medium of instruction from Urdu to Tibetan (Ladakhi/Bhoti)[6]" (Bertelsen, 1996). Having Urdu as compulsory subject in schools instead of Ladakhi was perceived by Buddhists of Ladakh as state imposition of a language that was alien for them. Therefore, the demands that people of Ladakh placed before the state government in 1950s also included making Bodhi[7] the medium of instruction in schools and demanded for the preparation of Bodhi text books for all the classes (Kaul, 2004). Subsequently, the term 'Bodhi' became official name of the academic subject in JKBOSE[8], although 'Ladakhi' was already a recognised regional language under the sixth schedule of the state constitution. This early language assertions made by Buddhists of Ladakh and acquiring the name "Bodhi", a misnomer, have led to the politicisation of language in the region.

Although, Balti, Ladakhi and Shina were recognised as regional languages in the state constitution, none of them became compulsory subject in the government schools in the state of Jammu and Kashmir. In 1969 during their first mass agitation against alleged government negligence and discrimination, Buddhists of Leh demanded recognition

of 'Bodhi' and making it compulsory up to high school level (Kaul, 2004). The agitation compelled the state government to declare that Bodhi will be taught in schools up to high school level. The government also decided to create 32 posts of Bodhi teachers for schools in Leh and Kargil (Shakspo, 2024). However, Muslims of Kargil refused to accept Bodhi teachers and demanded for Arabic teachers instead. According to Gulzar Hussian Kingi[9], Muslims of Kargil demanded Arabic teachers when Buddhist demand for Bhoti teachers was considered by then state government in 1960s (personal communication, April 30, 2025). While, Buddhists of Ladakh criticised the state government of neglecting Bodhi and consistently resisted perceived deliberate imposition of Urdu on them, Muslims of the region did not resist it probably due to their affinity with Urdu. Issa Sabri[10] also observes that since Muslims had to learn Persian and Arabic for the religious purpose, it was easier for them to learn Urdu when introduced in the state during the Dogra rule (personal communication, June 11, 2025). All these factors led to the emergence of a perception of 'Bodhi' as Buddhist language and Urdu as Muslim language in the region.

Buddhists of Ladakh started making language assertion way back in 1950s and 60s in the name of Tibetan or Bodhi but Muslims of the region seldom associated themselves with it. While, Buddhists owned Tibetan script and preferred their language to be written in it and taught in schools probably because all Tibetan Buddhist religious scriptures are written in it. Meanwhile, Muslims of the region often tend to dissociated themselves from Tibetan script because it has no connection with their religious scriptures. Issa Sabri's observation also resonates the same. He opined that with the increasing influence of Islam, the learning of Arabic and Persian scripts became more necessary than Tibetan script for the Muslims of the region (personal communication, June 11, 2025). So, the Muslims in Ladakh preferred learning Arabic and Persian over Tibetan script. Although Purgi and Balti belong to Tibetan-Burman language family and come closer to Tibetan classical language, their literatures are mostly in Perso-Arabic script. Therefore, the popularly used script among Muslims of Ladakh for generations is Perso-Arabic and in the course of time Tibetan script became alien for them. However, the efforts are being made by Balti scholars of Baltistan and Purgi scholars of Kargil of late to revive Tibetan script among the Muslim communities. It reflects growing realisation among the Muslims of Kargil that Tibetan script (Locally called 'Yige') is the actual script with which their language, Purgi and Balti, can be written comprehensively.

According to the CIIL report of a study on Bhoti language which was carried out in 2008, the movement was more then 35 years old. Hence, it can be assumed that the movement for the inclusion of Bhoti language in the eighth schedule of Indian Constitution was started in 1970s. However, the movement got the momentum and has become a popular demand in early 1980s. It was on 10th December 1980, then member of Parliament from Ladakh has raised the issue in the

parliament for the first time (Lok Sabha, 1980). Over the course of time, Members of Parliament from Himalayan regions have submitted number of memorandums to the government and raised matters and questions in the parliament, regarding the inclusion of Bhoti in the eighth schedule of the constitution of India. Then, the movement got intensified with the appointment of Lama Chosphel Zotpa[11] as vice-chairman of National Commission for Schedule Castes and Schedule Tribes in 2002. In 2008 the commission approached CIIL to conduct a study "to ascertain the fitness of Bhoti language for inclusion in the eighth schedule of the constitution of India" (CIIL, n.d.). Accordingly, CIIL carried an extensive study in 2008 and recommended the inclusion of Bhoti language in the eighth schedule of the constitution.

In 2011, Tarun Vijay, Member of Rajya Sabha from Uttarakhand, introduced a private member bill for the inclusion of Bhoti in the eighth schedule of the constitution (Rajya Sabha. VII, 2011). However, the bill was lapsed and could not get passed. Subsequently, MP Ladakh, moved another private member bill in the parliament on 1st April, 2022 proposing necessary constitutional amendment in the eighth schedule to include Bhoti language in it (Lok Sabha. 31, 2022). Considering the symbolic significance the day of introducing the bill in the parliament, the Executive Council of Ladakh Autonomous Hill Development Council[12] (LAHDC), Leh declared 1st April to be celebrated as "Bhoti Day" (News UT Ladakh, 2022). Henceforth, Bhoti day is celebrated every year by the Department of Education, Ladakh Academy of Art, Culture and Languages and the administration of union territory of Ladakh.

Ladakh has been the epicentre of the movement for the inclusion of Bhoti in the eighth schedule of the constitution. Although, other regions of Himalaya resonated the movement in different forms and asserted their collective concerns at various platforms. Yet, Ladakh has maintained the momentum through the support of various socio-religious organisations particularly Buddhist and sustained the movement to date. The main organization which spearheads the movement is HBCA, a Ladakh based organisation. It claims that there are more than 1.6 million Bhoti speakers throughout Himalayan region (HBCA, n.d.). Various Buddhist organizations across Himalayan states have conceived a collective identity and succeeded in mobilising political supports from chief ministers of Jammu and Kashmir, Himachal Pradesh, Sikkim and Arunachal Pradesh as well as Members of Parliament from these states (CIIL, n.d.). Beside HBCA, socio-religious organization like Ladakh Buddhist Association (LBA) has also been advocating the inclusion of Bhoti in the eighth schedule of the constitution. HBCA and LBA have been conducting multiple seminars and peaceful demonstrations in support of the movement. Apart from this, the movement also got support from across religious organisations in Ladakh such as Anjuman-e-Imamia, Anjuman Moin-Ul-Islam and Christian Association (Angmo, 2014). Likewise, LAHDC, Leh has also passed resolutions seeking inclusion of Bhoti language in the eighth schedule of the Indian constitution in 2011 and

2022 (Indus Dispatch, 2022; Reach Ladakh Bulletin, 2022). Therefore, political and social organisations in Leh district have been pressing the government for the declaration of Bhoti as scheduled language.

The government also acknowledged that Bhoti is among 35 non-scheduled languages and demands for their inclusion in the eighth schedule is pending with the government (MHA, 2004). Mentioning the absence of any fixed criteria for the inclusion of languages in the eighth schedule, government stated that the inclusion of more languages in the Eighth Schedule can only be considered after criteria being set and approved (Lok Sabha, 2014). The government constituted committees in 1996 and 2003 to set fix criteria for the same but could not materialised (MHA, 2016). However, the Ministry of Home Affairs in a press release mentioned:

The Government is conscious of the sentiments and requirements for inclusion of other languages in the Eighth Schedule, and will examine the requests keeping in mind the sentiments, and other considerations such as evolution of dialects into language, widespread use of a language etc. (MHA, 2016)

Therefore, at present it seems that the inclusion of languages in eighth schedule is dependent on the political will and the significance of a demand. However, questions, call attention motions and private member bills in the parliament regarding the inclusion of Bhoti language in the eighth schedule, brought by members from Ladakh as well as other concerned states, yielded no result.

3. Bhoti versus Ladakhi Debate

The emergence of the nomenclature 'Bhoti' and subsequent demand for its recognition can be traced back to 1970s. In 1970s the Institute for Higher Tibetan Studies, Sarnath, held a conference with the idea of seeking recognition for Tibetan language in India under the nomenclature of 'Bhot Basha' to integrate it into the larger family of Indian languages (Shakspo, 2024). The purpose of using the nomenclature 'Bhot Bhasha' was to establish the linkage between Tibetan language and Indian culture. Because, Tibetan script is believed to have developed in 7th century by Thonmi Sambhota[13] deriving from Indian Bhrami script (Tashi, 2023). Subsequently, Himalayan Buddhist Cultural Association (HBCA), an NGO, started the movement for the inclusion of Bhoti language in the eighth schedule of the constitution of India in 1970s-80s. The idea of using the term 'Bhoti' seems to have conceived from the 'Bhot Bhasha' term used in the mentioned conference. The term 'Bhoti' is claimed to have broader and inclusive meaning that can be applied to the language and culture of entire Himalayan region in India. Therefore, 'Bhoti' became an umbrella term which seeks to integrate all the languages of Himalayan region from Arunachal Pradesh to Ladakh, belonging to Tibetan-Burman language family, and demand for its recognition under eighth schedule of the constitution of India. The movement received extensive support from the majority of the scholars and monks in Ladakh who are well versed

in classical Tibetan language. However, there is another group of intellectuals in Ladakh, some of them are also well versed in classical Tibetan language, who opposes 'Bhoti' as nomenclature and resisted the inclusion of Ladakhi under the umbrella of Bhoti. Furthermore, this group of intellectuals proposed to maintain Ladakhi as the nomenclature. As a result, two schools of thought have emerged and formed perennial discourses on the debate of Bhoti versus Ladakhi language issue in Ladakh. One is represented by the proponents of Bhoti and another is represented by the proponents of Ladakhi.

After analysing the interviews of language experts, it can be said that the scholarly debate on Bhoti versus Ladakhi revolves around two major issues; the nomenclature and the writing system. On the nomenclature, the general argument of those who are in favour of 'Bhoti' is that the term is more inclusive than the term 'Ladakhi'. Whereas, those who are not in favour of 'Bhoti' argues that the language spoken in Ladakh should rather be called 'Ladakhi'. They also claim that 'Bhoti' term does not enjoy consensus and it also tends to create communal divide in Ladakh. The more serious debate on Bhoti versus Ladakhi is regarding the writing system. The proponents of Bhoti prefer their language to be written in standard classical Tibetan script and strictly according to Tibetan grammar developed by Thonmi Sambhota. Whereas, the proponents of Ladakhi prefer their language to be written in Tibetan script but in colloquial form with slight modifications in classical Tibetan grammar wherever necessary. Therefore, regarding the writing system the approach adopted by the former group is largely conservative, whereas, the approach adopted by the latter group is liberal. Konchok Tashi categorises the prevalent perspectives on language issue in Ladakh into three categories (Tashi, 2023). He observes that there are three group of scholars with starkly diverse views. The first category of intellectuals is the advocator of preservation, development and promotion of 'Ladakhi' and promotes writing colloquial Ladakhi in Tibetan script. The second category of intellectuals are the proponent of 'Bhoti' and they oppose using colloquial Ladakhi in books and other forms of writings. The third category object 'Bhoti' nomenclature but prefer using classical Tibetan in literatures.

One of the arguments in favour of 'Bhoti' nomenclature is based on its historical significance. Urgain Dadul[14], argues that the term 'Bhoti' was derived from the term 'Bhot', which was historically used by Indian scholars to refer to Tibet and in fact 'Greater Tibet' including Baltistan and Ladakh (personal communication, May 20, 2025). The position that the Tibet was historically referred to as 'Bhot' by Indians is also corroborated by the claim that 'Sambhot' was the name acquired by Thonmi Sambhota in India. While, Thupstan Norboo[15] observes that 'Bhoti' term was adopted after prolong deliberations among scholars from different parts of Himalayan region with consensus (Personal communication, May 21, 2025). His observation implies that the term being adopted after prolong scholarly discussion is not justified to be questioned and debated. Another argument is in the line of expanding

the geographical reach of the language. Konchok Thupstan[16] argues that the term 'Bhoti' is more inclusive whereas 'Ladakhi' will limit the expansion of the language and confine it to Ladakh region only (personal communication, June 3, 2025). The underlying concern of all these arguments is the support base and the population of the speaker of the language that is deemed necessary to strengthen the case of its inclusion in the eighth schedule of the constitution of India.

The torch bearer of Bhoti movement seems to have adopted a balance strategy of establishing the Indian root of the Tibetan language, to which all the Himalayan languages trace their origin, and maintaining its Indian identity at the same time. The Centre Institute of Indian Languages (CIIL), Government of India, conducted a case study on Bhoti language in the first decade of 21st century. In the report, it observed:

> The choice of the term Bhoti over Tibetan is a conscious part of the strategy of the leaders of the movement belonging to diverse tribes to affirm their status as a part and parcel of the Indian Identity. They would not want their indigenous speech communities to be confused with their better-known Tibetan refugee cousins, but remind instead how Bhoti has drawn the script and cultural content from the fertile and liberal soil of India with its great intellectual tradition. (CIIL, n.d.)

Moreover, HBCA in its pamphlet entitled as "Memorandum: Importance of Bhoti Language & Recognition of Bhoti Language in The Eight Schedule in Context to Recognise" argued:

> Many studies proved that the primary form of this language was dialectics of people of entire Himalayan region of India. Therefore, Bhoti is the mother tongue of Himalayan people as well as source of Indian culture. (HBCA, n.d.)

Therefore, 'Bhoti' term though consciously coined by some Tibetologist (Sherab, 2023) seeks to develop a larger linguistic identity of Himalayan belt so as to strengthen the case for constitutional recognition of a language.

On the contrary, the arguments against 'Bhoti' nomenclature are founded mainly on identity, consensus and communal division. Scholars who favour the term 'Ladakhi' argue that there is lack of consensus on the term 'Bhoti' at larger Himalayan region level as well as among communities within Ladakh. They also argue that their language should reflect Ladakhi identity. Therefore, they find more logic in calling their language 'Ladakhi' than 'Bhoti'. Khenpo Konchok Sherab[17], writes "the name 'Ladakhi' provides self-identification and a sense of proprietorship over the language that we speak in Ladakh". Further, he argues:

> Bhoti might provide an overarching framework to bring together the Indian Himalayan regions. However, if Ladakhi is recognised under 'Bhoti' it will lead to several challenges.

> Since 'Bhoti' refers to the cultural, religious and literary language of Classical Tibetan, most Ladakhis hesitate to use it to refer to their language. Furthermore, Ladakh's Muslim community would have reservations about the language if it is called 'Bhoti' as it seems associated with Buddhism (Sherab, 2023).

His argument suggests that 'Ladakhi' is more preferable nomenclature than 'Bhoti' because the term 'Bhoti' indicates linkage of the language with culture and religion of Tibetan Buddhism but 'Ladakhi' term confers an indigenous language identity. Bakula Randol Nima Rinpoche[18], also says that 'Bhoti' term being viewed as something related to Buddhist religion would not get support from Muslims of Kargil district and will lead to divide people on communal ground. He thinks that the nomenclature 'Ladakhi' is more likely to get consensus of the people of all the communities in Ladakh and unite them on the language front (personal communication, June 10, 2025). Similarly, Purgi scholars of Kargil argue that technically and historically 'Bhoti' cannot be the name of their language and they perceive it as religion specific in nature. Dissociating from Bhoti movement for inclusion in eighth schedule, Nasir Khafi[19] says that the Bhoti movement is not an assertion of Kargili Muslims, rather it is Leh centric. He thinks that Kargil Muslims would rather agree on 'Ladakhi' than 'Bhoti' as nomenclature (personal communication, April 28, 2025). Whereas, Konchok Phanday[20], argues that 'Bhoti' cannot be called inclusive because it is resisted not only by Purgi and Balti speakers of Ladakh but also by communities like Bhutia and Monpa of Himalayan region (personal communication, May 18, 2025). His argument seems valid given the fact that Lepcha, Bhutia and Limbu are separately demanding for inclusion in the eighth schedule. Moreover, people in Ladakh are also divided on the issue of their language be named as Bhoti. Therefore, the analysis of anti-Bhoti perspectives suggest that the 'Bhoti' term not only lacks consensus but it also tends to foster a communal perception particularly among the Muslims of Ladakh.

As far as the writing system is concerned, proponents of Bhoti argue that writing in colloquial form without following the standard classical Tibetan grammar developed by Thonmi Sambhota will lead to distortion of the language. Their position on writing system contends that it will be a disservice to the language if it is not written according to the classical Tibetan grammar and deviates from its standard. Because they believe that it will change the language beyond recognition in the long run. Konchok Thupstan says that the writing should be according to a standard grammar and he claims that it is lacking in the books written in colloquial Ladakhi language (personal communication, June 3, 2025). Similarly, Thupstan Norboo, observe that if a standard grammar is not followed Bhoti/Ladakhi language will lose its structure and become baseless which he thinks is a more serious issue than nomenclature. Further, he argues that the writing in colloquial language will also encounter with problem of regional dialectical variation (personal communication, May 21, 2025). Their

statements imply that the colloquial Ladakhi language has no grammar. Whereas, according to Urgain Dadul, writing in colloquial language will limit the literary works to Ladakh region only because it would not be understandable to people outside Ladakh who also associate themselves with Bhoti language (personal communication, May 20, 2025). This view seems to assume that written Bhoti would be understandable throughout Himalayan region. Therefore, intellectuals who are in favour of Bhoti leaves no scope for modification in or deviations from the classical Tibetan grammar and they believe that the colloquial language has no grammar.

On the contrary, proponents of Ladakhi group argue that there are considerable differences between classical Tibetan language and Ladakhi colloquial language in both spoken and written forms. Therefore, writing colloquial Ladakhi strictly according to classical Tibetan grammar without even a slight modification is not possible. They believe that popular literature barring the religious ones should be written in colloquial language to make it easier for the public to understand. Therefore, they prefer writing in colloquial language with slight modifications in the Tibetan classical grammar to serve the purpose. Konchok Phanday remarks:

> Human Language cannot exist without grammar, whether it is written in the shape of book or not, therefore to say that the Ladakhi language has no grammar in not true. Language written with grammar that doesn't match the grammar of speech may be misunderstood by the readers, whereas any language which matches the written grammar to the spoken language will be clearly understood. (Phanday, 2017)

Similarly, Sonam Wangchuk[21], observes that there is no language without a grammar and if two persons can communicate with each other without any confusion then there is a grammar (personal communication, June 9, 2025). His argument also implies that Ladakhi colloquial language has a grammar and can be written like any other language. Whereas, Bakula Rangdol Nima Rinpoche, who has written *A Ladakhi Grammar* in colloquial language, says that a writing meant for common masses in Ladakh should be written in colloquial language. However, he agrees that writing in classical Tibetan language following the Tibetan classical grammar rules would also be appropriate if it is meant for those who understands it (personal communication, June 10, 2025). Sonam Wangchuk also observes that there should not be any restriction either on writing in classical Tibetan or colloquial language and both should rather be encouraged (personal communication, June 9, 2025). Therefore, Konchok Phanday has written his book *Comment on Thon-mi's Grammar and Clarification of Language* in classical Tibetan language and *New Dawning Knowledge for Ladakhi Youngsters* in colloquial language. The former book has a section entitled "Clarification of the Language" in English, in which he suggests that literacy in Bodyik (Tibetan script) should be promoted through newspapers, magazines and school text books for 1st and 2nd

class in Ladakhi language (Phanday, 2017). The underlying assumption behind the arguments put forth by proponents of Ladakhi is that a language should not be rigid and there should be a scope deviation and modification if required to match the written form of the language with the spoken one. The proponents of 'Ladakhi' also believe that the elementary text books in colloquial language will eventually help student learn classical Tibetan in higher classes.

There are contesting scholarly perspectives on the issue of promotion and preservation of language and culture as well. The 'Bhoti' proponents hold that the inclusion of it in the 8th schedule would be eventually necessary for the promotion and preservation of Bhoti language and culture. Lama Chosphel Zotpa, chairman HBCA, in an interview argued that the recognition of Bhoti as scheduled language in the constitution is necessary for governmental aid and facilities for its promotion, generating interest among learners and creation of jobs (DD Kashir, 2020). Similarly, Urgain Dadul maintains that since Bhoti has already made a place in the government recognised non-scheduled languages pending inclusion in the eighth schedule, there should not be any issue on nomenclature. He argued that the inclusion of Bhoti in the eighth schedule would encourage learners and enhance the use of the language (personal communication, May 20, 2025). A major concern of Bhoti proponents was the declining interest and use of language among public. This is because the widespread use of language is considered necessary, though not a prescribed criterion for a language to be included in the eighth schedule. The projection of Bhoti as a language spread over entire Himalayan region from Arunachal Pradesh to Ladakh is also done considering the population factor. However, Ladakhi proponents argue that 'Bhoti' being a constructed nomenclature of a language lacks consensus and consequently it has not received constitutional recognition. Konchok Phanday argued that 'Bhoti' literally means Tibetan and the latter being already a recognised language in China[22] may not be considered by the government of India to include in the eighth schedule, moreover, ethnic groups in Sikkim and Balti and Purgi groups in Kargil do not agree with it (Gulistan News Ladakh Official, 2022). Whereas, Nawang Tsering Shakspo[23] in an interview to Reach Ladakh observed that though the demand for Bhoti inclusion in the eighth schedule has been raising since early 1970s, there is confusion and misunderstanding regarding the term Bhoti. He rejects the existence of a language called "Bhoti" and instead suggests "Bodhyik" as nomenclature (Padma, 2024). Therefore, the unresolved language issue in Ladakh is evolving into a more complicated debate among intellectuals.

4. Policy Lacuna and Politicisation of Language

The branch offices of Jammu and Kashmir Academy of Art, Culture and Languages were established in Leh and Kargil in 1969 and 1989 respectively. The academy has been publishing annual books and journals in Ladakhi language since 1976 and 1979 respectively (Shakspo, 2024). According to Balti writer Mohammad Sadiq Hardassi,

Nati Adaab, a journal in Balti language, was published by the academy from 1980s to 2010 (personal communication, June 12, 2025). While *Ladakhi Sharaza*, *Shesrig Melong* and other publications by the academy are written in Tibetan script, *Nati Adaab* was written in Perso-Arabic script[24]. While Bodhi/Bhoti is taught in government schools since 1970s and JKBOSE approved text books are available up to 10th class, Balti and Shina were started teaching in government schools only from 2016 onwards. According to Mohammad Sadiq Hardassi, Balti was taught in government schools up to 5th class with JKBOSE approved text books written in Perso-Arabic script since 2016 (personal communication, June 12, 2025). Whereas, according Mohammad Shafi Sagar[25], Shina was taught in the schools up to 7th class but discontinued after Ladakh becoming UT said (personal communication, June 10 2025). Meanwhile, Centre Institute of Indian Language and National Council of Educational Research and Training (NCERT) published primers for Balti, Ladakhi and Shina. The Ladakhi and Shina primers were published in July 2024 and Balti primer was published in May 2025. Ladakhi primer is written in Tibetan script and Balti as well as Shina primers are written in Perso-Arabic script. All these not only created complications in the language issue in Ladakh but also fostered perceptual differences between Buddhists and Muslims of the region regarding script (*Yige*) and its usages.

The ambiguous and confusing classification of mother tongues of Ladakh in linguistic census reports seems to have further contributed in the complicated language issue of Ladakh. The Linguistic Survey of India (LSI) report 1961 mentions about Balti, Budhi [sic], Ladakhi, Zanskari and Sham [sic] as mother tongues classified under Bhotia group of Tibeto-Himalayan branch of Tibeto-Burman language family and Shina and Brokpa [sic] classified under Dard group of Indo-European language family (Mitra, 1961).While, LSI 2001 report mentions about Balti, Ladakhi, Zanskari and Shina in the list of mother tongues wherein Zanskari is listed as a mother tongue under Bhotia group (ORGI, 2001). The LSI 2001 (Paper 1 2007) has listed Balti, Ladakhi and Shina as languages and Zanskari as a mother tongue under Bhotia language (ORGI, 2007). Whereas, CIIL also observe:

> In the first instance, the complexity stems from the fact that in the census reports of 2001 [or even of earlier decades] tabled by the Home Ministry, the name Bhoti is neither included in the list of 100 non-scheduled languages, nor as one of the Mother tongues listed under them [as census would reason, it is obviously only because less than 10,000 people must have returned it and not out of any bias] (CIIL, n.d.)

Further, LSI 2011 (Paper 1 2018) mentions about Balti, Ladakhi and Shina as languages and Zanskari got no mention (ORGI, 2018). However, Purgi and Bhoti did not figure in all the census reports, while, Brokpa and Zanskari figured in 1961 but missing in 2001 and 2011 census. Therefore, the projection of Bhoti and it missing in census reports have further made the language issue intricated.

The use of different nomenclatures for the same language has also contributed in complication and politicisation the language issue. The use of multiple terms like Bhoti, Bodyik, Bodhi and Choskat to denote written Ladakhi language were ambiguous and have created confusion (Tashi et al., 2024). Muslims of Ladakh perceive these terms as affiliated to Buddhist religion and they tend to distance themselves from associating with any of them. Though Balti and Purgi scholars agree that their mother language is classical Tibetan and the Tibetan script they called "Yige" will be the best to write their language, they explicitly dissociate themselves from the projection of a common 'Bhoti' nomenclature. The general perception among Kargil scholars and leaders is that the 'Bhoti' term is nothing but a changed version of the term 'Bodhi'. Issa Sabri, claims that Bodhi was made the name of the subject on the suggestion of Leh scholars without consulting Purgi and Balti scholars. Further he argued that the Bodhi text books of JKBOSE generally lacked lessons on Balti and Purgi culture and history of Kargil leading them to further alienate from Bodhi (personal communication, April 28, 2025). Similarly, Nasir Khafi argues that since Kargil scholars were never consulted on the adoption of the term 'Bhoti' they not only felt alienated but also perceived it as an attempt to get religious consensus among Buddhists in the garb of language (personal communication, April 28, 2025). The arguments put forth by Balti and Purgi scholars of Kargil also reveals the prevalent communal perception towards Bhoti among Kargil Muslims. Buddhist scholars of Leh also thinks that the general reason behind Kargil's dissociation from Bhoti was religion. Moreover, the apathetic policy adopted by the state government towards regional languages with overarching emphasis on Urdu and the introduction of Arabic alongside Bodhi in government schools further led to polarisation of Ladakhi identities along communal lines (Warikoo, 1996). Then, emphasising Buddhist religious texts as important Bhoti literatures and Bhoti as Himalayan Buddhist language also seems to have encouraged communal perception on the language particularly among Muslims of Ladakh.

It is pertinent to note that Buddhist scholars generally tend to mention important religious scriptures like '*Kagyur*' '*Stangyur*' as important literatures while arguing for inclusion of Bhoti in the eighth schedule. Similarly, CIIL report also observed that just as Sanskrit being sacred language for Hindus, Bhoti is regarded as language of Himalayan Buddhists (CIIL, n.d.). These can also be said to have provided reasons for reinforcement of the perception that associates Bhoti with Buddhism which in turn led to the division of Ladakhi society by its language. John Bray observed:

> All indigenous inhabitants of the region speak a variety of Ladakhi yet, at least as far as the written version is concerned, it is even more true of them than of the British and Americans (in Winston Churchill's well-known phrase) that they are "divided by a common language". In this case the sources of division are explicitly religious. (Bray, 2007)

His observation is pointing to the paradox of the society of Ladakh having a common language but being divided due to the communalisation of that language.

The religious influences or communalisation have made the language matter a sensitive issue. As a result, there were instances of strong resistance to deviation from classical Tibetan grammar in the writing system. For instance, an experiment carried out in 1950s by Eliyah Tsetan Phuntsog, a Christian convert, with modified spelling system in Bhoti/Ladakhi to make it more appropriate to write in colloquial Ladakhi received severe criticism from local Buddhists as "form of linguistic religious subversion" (Bray, 2007). Similarly, a local bilingual magazine *Ladags Melong* published from 1993-2006 was also criticised for writing in colloquial language and had to stop publishing. The magazine was popular because of written colloquial Ladakhi language it used. However, it was accused of diluting the purity of classical Tibetan language, which according to critics may subsequently lead to deadening of Buddhist religious scriptures and teachings if translated into colloquial Ladakhi language (Chosjor, 2019). These resistances show how religious concerns caused the language to become sensitive. Therefore, when language, script and religion are perceived as combined, the confusion prevails and language become a religious sentiment. Consequently, scholars in Ladakh tends to consider Tibetan script as inseparable from its usage for religious scriptures (Tashi, 2023). Such view also seems to have stemmed from deifying a script as sacred and unalterable.

5. Language and Identity

The contestation on language in Ladakh also implies an undercurrent identity issue. The language debate reflects tensions between Ladakhi local identity and its "wider cultural inheritance" (Bray, 2007). While the consideration of cultural inheritance takes preeminent position in the pro-Bhoti perspective, meanwhile, the concern for local identity outstands the former in the pro-Ladakhi perspective. The Bhoti nomenclature is an attempt to construct a larger identity of a language and associated culture. In its pamphlet, HBCA claim that Bhoti is the language and culture of people living in the Himalayan region from Arunachal Pradesh to Ladakh including Sikkim, Darjeeling, Kalimpong, Spiti, Kinnuar, Uttar Kashi etc (HBCA, n.d.). Whereas, proponents of Ladakhi argue that although the root of language and culture of Ladakh lies in Tibet, the language spoken by the people of the region is different from Tibetan and therefore have its own identity. Konchok Phanday opines that Ladakhi language is the identity of the people of Ladakh and it needs to be preserved and promoted (personal communication, May 18, 2025). He prefers the name Ladakhi over Bhoti between these two different identities. Similarly, Khenpo Konchok Sherab observed that the name 'Ladakhi' would provide the people of Ladakh with "self-identification and a sense of proprietorship" over their language (Sherab, 2023). However, Elijah Gergen view Bhoti versus Ladakhi debate a two-prong approach of

preserving classical Tibetan as well as Ladakhi and he thinks it is not bad an approach (personal communication, May 19, 2025). Therefore, the local and wider identity tension is implicit in the contestation over the nomenclature of the language. However, both the perspectives acknowledge that Ladakh owes a great deal of its language and culture to Tibet but at the same time the former has its own distinction.

6. Declaration of Official Languages

After the fierce debate on *Ladags Melong* that took place in early 21st century leading to the cessation of its publications, another public debate on language erupted in Ladakh after it became a Union Territory (UT) in 2019. The declaration of official language of Ladakh was the underlying matter of public discussion aftermath the formation of UT. The debate on language got heightened when Nawang Tsering Shakspo filed a Public Interest Litigation (PIL) in the high court of Jammu Kashmir and Ladakh challenging the term Bhoti being used to refer to "the linguistic and cultural heritage of Himalayan region" (Khajuria, 2024). It led to condemnation by Ladakh Buddhist Association, Ladakh Gonpa Association and Ladakh Cultural Forum as misleading and against the public demand for Bhoti inclusion in eighth schedule of the constitution (Reach Ladakh, 2024). Subsequently, number of conferences and meetings were held by Ladakh Academy of Art, Culture and Languages and the Hill Councils of both the district to deliberate on the languages to be put forth for recognition as official languages of UT Ladakh. Wherein, according to Thupstan Norboo, larger consensus was reached in the conferences on Bhoti to be recognised as one of the official languages (personal communication, May 21, 2025). However, according to Nasir Khafi, scholars of Kargil did not agree with Bhoti but they rather proposed Ladakhi as nomenclature on which they agreed to support (personal communication, April 28, 2025). However, according to Mukhtar Ahmad, Shina writer, scholars and experts were divided on Bhoti and Purgi in the conferences. Therefore, it seems that there was no consensus on Ladakhi and scholars of Leh and Kargil proposed Bhoti and Purgi respectively to be considered for the declaration as official languages. Subsequently, the official languages regulation 2025 declared Bhoti and Purgi among other three as official languages of UT Ladakh. Wherein Shina, Brokskat, Balti and Ladakhi are mentioned as other native languages and the promotion and development of them through proper institutional mechanism like Academy of Art, Culture and Languages is also being envisaged.

7. Conclusion

The language contestation in Ladakh particularly Bhoti versus Ladakhi represents a classic case of conflict between high variety (locally known as *Chos-skat*) and low variety (locally known as *Phal-skat*) language. The case of 'Bhoti' and 'Ladakhi' can be said to be in diglossia[26] situation where the former is standardised and the latter is yet to be standardised. Since the movement for constitutional recognition of 'Bhoti' envisages the language to be standardised

according to the classical Tibetan language used in Buddhist religious texts, 'Bhoti' can be considered as high variety language. Whereas, 'Ladakhi' being a colloquial language used as lingua franca across Ladakh and as medium of communication in All India Radio Leh and popular literatures can be considered as low variety language. While 'Bhoti' is learned through formal education in schools and monasteries, 'Ladakhi' is learned through informal education in families. Therefore, Bhoti and Ladakhi embody an inter-dependent yet complicated positionality.

By analysing the issue at length, it can be posited that the rigid stance taken by both the proponents of Bhoti and Ladakhi has obfuscated the 'Bhoti' versus 'Ladakhi' issue. Therefore, the stances taken by both the groups needs to be slackened because it does not help in strengthening any of the cause. The Bhoti proponents should allow and encourage popular publications like newspapers and magazines be written in the colloquial language because that would help in the promotion and popularisation of both Ladakhi and Bhoti. Moreover, the readership of Ladakhi is far more than that of Bhoti. Whereas, the proponents of Ladakhi should also support the movement for constitutional recognition of Bhoti language as it would ultimately help in preserving the mother language of the land.

The Bhoti versus Ladakhi contestation needs to be dealt with careful consideration of both the arguments adopting moderate approaches without being influenced by mere sentiment. Bhoti is the mother of many languages spread over the entire Himalayan region, and has been projected as a Himalayan language with a strong demand for its inclusion in the eighth schedule of the constitution for last more than five decades. In the past, CIIL in its report submitted in 2008 strongly put a case in favour of inclusion of Bhoti in the mentioned schedule in its recommendation. Bhoti also claims to have majority support in the Himalayan region barring some ethnic groups. However, the debate on writing system and standardisation of Himalayan languages to be included in the larger nomenclature of Bhoti would probably persist even after the inclusion of Bhoti in the eighth schedule. Therefore, dealing with the issue with outmost prudence will be necessary for the preservation and promotion of Himalayan languages. Meanwhile, the objection to Bhoti cannot be overlooked and needs to be addressed prudently by giving due consideration to the arguments of the proponents of Ladakhi. Moreover, the apprehension among Ladakhi proponents that inclusion of Bhoti in the eighth schedule would lead to demotion of Ladakhi, also needs to be removed by ensuring equal promotion of colloquial Ladakhi language.

Since there is no denial of Tibetan script and grammar being the mother language from which Ladakhi, Bhoti, Balti, Purgi etc languages have emerged, slight modification if required may be allowed to make the popular literatures more public oriented. Since there is only a slight difference in the grammar of written classical Tibetan and written Ladakhi, the fear of severing common people from classical Tibetan

language due to use of colloquial language, otherwise claimed by some Ladakhi Buddhist scholars, should not arise. Proponents of Ladakhi argue that promotion of learning and writing colloquial Ladakhi language would not harm Bhoti, rather it will help one to start from mother tongue and progress to learn classical Tibetan language more effectively in higher classes. Similarly, proponents of Bhoti also argue that learning and writing of Bhoti would in a way promote Ladakhi colloquial language as well as classical Tibetan language. In such situations, there should not be any problem on either side to promote learning and writing classical as well as colloquial languages in Ladakh because both are mutually complimentary for the promotion of each other. While, religious and literary texts may be written in classical Tibetan with standard grammar of higher level, the popular literatures may also be allowed to be written in colloquial language with necessary deviation from standard classical Tibetan grammar. Such distinction between popular and literary literature and creation of space for the promotion of both would help Bhoti and Ladakhi coexist and flourish.

The issue of recognition and preservation has been the underlying concerns of language contestation in Ladakh. While the contestation on nomenclature is concerned with the recognition and the debate on writing system is about preservation. There are two schools of thought, one adopts a conservative approach and another a liberal approach. We suggest that Bhoti as a classical language with rich literature, geographical expansion and mother language of many Himalayan languages can be considered high variety language appropriate for official recognition in the eighth schedule of the constitution of India at national level. Whereas, Ladakhi being a variant of Bhoti can be provided scope for deviation from classical Bhoti in writings and given deserving official recognition at the state level. One of the common concerns of both the schools of thought is the declining interest on their language among the people of Ladakh. Therefore, a collaborative effort from both the groups, to find ways to generate interest on the language among the masses and promote language education in Ladakh, should be encouraged. The language education shall move from familiar to unfamiliar by introducing textbooks in colloquial Ladakhi language at primary level and in Bhoti at higher level. While religious and literary texts can be rendered in classical Bhoti, popular literatures like magazines, newspapers etc. can also be considered appropriate to be written in Ladakhi colloquial language. The contestation between Bhoti and Ladakhi in other sense is also about 'maintaining standard' versus 'matching mother tongue' in writings. We maintain that given the fluid nature of language, the scope for deviation and modifications in the colloquial Ladakhi language is required. But at the same time, maintaining the standard of the mother language is also necessary to keep the authenticity of the language intact. Therefore, a balanced approach of preserving the standard of classical language, and the preservation and promotion of Ladakhi colloquial language is recommended.

References

Warikoo, K. (1996). Language and Politics in Jammu and Kashmir: Issues and Perspectives. In P.N. Pushp & K. Warikoo (Eds.), *Jammu, Kashmir and Ladakh: Linguistic Predicament* (pp.183–212). Har-Anand Publications.

Tashi, K. (2023). *Understanding Language Attitude and Linguistic Ecology in Indian Himalayas*. Indu Book Services Pvt. Ltd.

Noorani, A. G. (2014). *Article 370: A Constitutional History of Jammu and Kashmir*. Oxford University Press.

Tashi et al. (2024). Written Ladakhi and the Future of Ladakh's Culture. *Forum for Linguistic Studies*, 6 (5), 764-782. https://journals.bilpubgroup.com/index.php/fls/article/view/7240.

MLJ. (2025). Ministry of Law and Justice. *The Ladakh Official Languages Regulation, 2025*. https://docs.publicnow.com/viewDoc.aspx?filename=32588%5CEXT%5CB54438D4079DE1362B21F368D8015AADAFFE7F6B_EE0BEA6679BBF20F9939D4ADA832378797FE8B79.PDF.

Bertelsen, K. B. (1996). *Our Communalised Future: Sustainable Development, Social Identification and Politics of Representation in Ladakh* [Unpublished doctoral dissertation]. Aarhus University.

Kaul, S. (2004). *Ladakh Through the Ages: Towards a New Identity*. Indus Publishing Company.

Shakspo, N. T. (2024). Language Politics in the Union Territory of Ladakh. *Ladakh Review*, 9–10: 1–9.

Parliament of India. Lok Sabha. (1980, December 10). *Inclusion of Bhoti Script of North Eastern in Eighth Schedule* (Unstarred Question, 3203). Parliament of India. https://sansad.in/ls/search.

CIIL. (n.d.). Central Institute of Indian Languages. *On the Case Study of Bhoti Language: Planning for Recognition Development*.

News UT Ladakh. (2022, April 02). *Decision of Executive Council| LAHDC Leh | for Celebrating Bhoti Language Day in Leh* [Video]. YouTube. https://www.youtube.com/watch?v=HNgV9Z Yc3LU&t=103s.

HBCA. (n.d.). Himalayan Buddhist Cultural Association. *Importance of Bhoti Language & Recognition of Bhoti Language in the Eighth Schedule in Context to Recognise*.

Angmo, P. (2014, December 24). People of Ladakh Unite for Inclusion of Bhoti Language in 8th Schedule. *Reach Ladakh Bulletin*. https://www.reachladakh.com/news/art-culture-and-languages/people-of-ladakh-unite-for-inclusion-of-bhoti-language-in-8th-schedule.

Indus Dispatch. (2022, March 27). Everything You Need to Know About Bhoti Language. *Indus Dispatch.* https://indusdispatch.in/everything-you-to-know-about-bhoti-language/.

Reach Ladakh Bulletin. (2022, April 5). Two-day General Council Meeting of 6th LAHDC, Leh Begins. *Reach Ladakh Bulletin*. https://www.reachladakh.com/news/social-news/two-day-general-council-meeting-of-6th-lahdc-leh-begins.

MHA. Ministry of Home Affairs. (2004, December 22). *Languages in the Eighth Schedule* [Press Release]. https://www.pib.gov.in/newsite/erelcontent.aspx?relid=5928.

Parliament of India. Lok Sabha (2014, August 6). *Inclusion of Bhoti language in Eighth Schedule* (Unstarred Question, 420). Parliament of India. https://sansad.in/ls/search.

MHA. Ministry of Home Affairs. (2016, July 26). *Inclusion of Languages in Eighth Schedule* [Press Release]. https://www.pib.gov.in/newsite/PrintRelease.aspx?relid=147752.

Sherab, K. Khanpo. (2023, September). Bhoti or Ladakhi: What Should We Call Our Language? *Stawa*, 10 (1): 14-15.

Phanday, K. (2017). *Comment on Thon-mi's Grammar and Clarification of Language*. Author.

DD Kashir. (2020, November 16). *Ek-Mulaqat. With Lama Chosphel Zotpa & Dr. Tsewang Yanjor* [Video]. YouTube. https://www.youtube.com/watch?v=j4hzC3a6qlo&t=558s.

Gulistan News Ladakh Official. (2022, October 12). *Exclusive Interview with Ven. Konchok Phandey ||Ladakhi Language and Its Importance* || [Video]. YouTube. https://www.youtube.com/watch?v=Kfh-uohGxSU&t=773s.

Padma, S. (2024, July 19). In Conversation with Nawang Tsering Shakspo, Founder Director of Centre for Research on Ladakh [Interview with N. T. Shakspo]. *Reach Ladakh Bulletin*. https://www.reachladakh.com/news/in-conversation-with-reach-ladakh/in-conversation-with-nawang-tsering-shakspo-founder-director-of-centre-for-research-on-ladakh.

Mitra, A. (1961). *Survey of India 1961: Part II-C (ii) Language Tables*. Office of the Registrar General & Census Commissioner, Government of India.https://language.census.gov.in/eLanguageDivision_VirtualPath/eArchive/pdf/7.pdf.

ORGI. (2001). Office of The Registrar General, India. *Census of Indian 2001: Series 1 India: Language: India, States and Union Territories Table C-16*. 00-147-2001-Cen. Book (E). Office of The Registrar General, Government of India. https://language.census.gov.in/eLanguageDivision_VirtualPath/eArchive/pdf/49.pdf.

ORGI. (2007). Office of the Registrar General, India. *Census of India 2001: Paper 1 of 2007: Language: India, States and Union Territories Table C-16*. OO·l44·2001-Cen-Book (E). Office of The Registrar General, Government of India. https://language.census.gov.in/eLanguageDivision_VirtualPath/eArchive/pdf/C-16_2001.pdf.

ORGI.(2018).Office of the Registrar General & Census Commissioner, India. *Census of India 2011: Paper 1 of 2018: Language: India, States and Union Territories Table C-16.* https://language.census.gov.in/eLanguageDivision_VirtualPath/eArchive/pdf/C-16_2011.pdf.

Bray, J. (2007). Introduction: Locating Ladakhi History. In J. Bray (Eds.), *Ladakhi Histories: Local and Regional Perspectives* (pp. 1-30). Library of Tibetan Works and Archives.

Chosjor, S. (2019). Beyond Kashmir: Understanding Ladakh. In R. Chowdhary (Eds.), *Identity Politics in Jammu and Kashmir* (pp.135-153). Vitasta Publishing Pvt. Ltd.

Khajuria, Y. (2024, December 30). Ladakhi Historian Shakspo Moves HC Against LAHDC's 'Unilateral' Decision Renaming Ladakhi Language, Pot-Yig Script as Bhoti. *The Earth News*. https://earthnews.in/ladakhi-historian-shakspo-moves-hc-against-lahdcs-unilateral-decision/.

Reach Ladakh. (2024, December 14). *PIL Filed in HC for Changing Nomenclature of Ladakhi/Pot-Yig to Bhoti; LBA, LGA & LCF Condemn* [Video]. YouTube. https://www.youtube. com/watch?v=XabK5Ubud54.

The Constitution (Amendment) Bill, 2011, No. VII of 2011 (2011). Rajya Sabha, Parliament of India. https://sansad.in/getFile/BillsTexts/RSBillTexts/Asintroduced/eight%20schd.pdf?source=legislation

The Constitution (Amendment) Bill, 2022, No. 31 of 2022 (2022). Lok Sabha. Parliament of India. https://sansad.in/getFile/Bills_Texts/LSBillTexts/Asintroduced/31%20of%202022%20AS%20Intro.pdf?source=legislation

[1] Ladakhi Sheraza is a periodical and Shesrig Melong is an annual publication, both published by the Jammu and Kashmir Academy of Art, Culture and Languages (now known as Ladakh Academy of Art, Culture and Languages).

[2] It consists of Tehsil Dras and some villages of Kargil tehsil namely, Chhanigon, Badgam, Karkechu and Karsar.

[3] It consists of villages like Darchik, Garkon and Biama villages of Kargil district and Dah and Hanu villages of Leh district.

[4] Languages belonging to Dardi-Arya family are classified as Shina (Dardic) and Brokskat (Dardic) by The Ladakh Official Languages Regulation 2025.

[5] An organization formed by some Kashmiri pandit neo-Buddhists to represent Buddhists of the princely state of Jammu and Kashmir to the Glancy Commission.

[6] Because, during the Dogra rule Urdu became the official language of the state and it was compulsory in schools. Thereafter, Urdu remained an official language of the state of Jammu and Kashmir and was a compulsory subject in Jammu and Kashmir State Borad of School Education (JKBOSE) up to 10th class.

[7] Bodhi, though a misnomer, was nomenclature used for Ladakhi mother-tongue by the Jammu and Kashmir State Board of School Education. According to Tashi et al. (2024), 'Bodhi' term was coined by Pandit teachers then posted in Ladakh mistakenly as a romanticisation of 'Bodyik' (meaning Tibetan script)

[8] Which subsequently got changed into Bhoti.

[9] He is a Purgi language expert and director of Munshi Habibullah School Kargil.

[10] He is a Purgi language expert and a writer, who has initiated teaching 'Yige' (Tibetan alphabets) in Kargil.

[11] Venerable Lama Chosphel Zotpa is from Zanskar Ladakh. He is the head of the Himalayan Buddhist Cultural Association and vice president of Indian Himalayan Council of Nalanda Buddhist Tradition.

[12] One of the regional autonomous political institutions in Ladakh, other being LAHDC Kargil.

[13] It is said that Tibetan king Srongtsan Gampo sent his minister Thonmi Sambhota to India for the purpose of learning scripts of Indian languages in which teachings of the Buddha was available. He is believed to have learned Indian languages and developed Tibetan script to translate Buddha's teachings in to Tibetan language.

[14] Venerable Dr. Urgain Dadul is a retired Associate Professor of Bhoti and the current Chairman of Language and Cultural Committee, Ladakh Buddhist Association.

[15] Dr. Thupstan Norboo is the Editor Ladakh Academy of Art, Culture and Languages.

[16] Venerable Konchok Thusptan is an associate professor of Bhoti at the Central Institute of Buddhist Studies, Ladakh.

[17] Venerable Khenpo Konchok Sherab is a research officer at Srongtsen Library, Centre for Tibetan and Himalayan Studies, Dehradun, India.

[18] Bakula Rangdol Nima Rinpoche is a scholar and he has also written many books in colloquial Ladakhi language including "Ladakhi Grammer".

[19] He is a Purgi writer and an assistant professor of English.

[20] Khenpo Konchok Phanday is well versed in classical Tibetan grammar and has written many books in colloquial Ladakhi language. He was the editor of the section of *Ladags Melong* which contained writeups in colloquial language.

[21] He is an Educator and the founder of 'Ladags Melong' a local bilingual magazine published from 1993 to 2005.

[22] According to Wikipedia, Tibetan language is a recognised official language in the Tibet region under China.

[23] He is a historian, a cultural activist and former head of Jammu and Kashmir Academy of Art Culture and Languages office Leh.

[24] Ladakhi Sheraza is a periodical and Shesrig Melong is an annual publication published by the Jammu and Kashmir Academy of Art, Culture and Languages (Now Ladakh Academy of Art, Culture and Languages)

[25] He is Shina writer from Drass and written books on Shina culture and poetry.

[26] It is a concept propounded by American Linguist, Charle A. Fergusson, that refer to a linguistic situation in which a highly standardised language (not used by the community for ordinary conversation) and one and more less standardised language (popularly used by the community for ordinary conversation) coexist.

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 131-138

Creating Interactive English Language Learning Using LLM in Higher Education

Garima Dalal[*]

Abstract

Large Language Models (LLMs) have influenced higher education, educational institutions, educators and students. Contributions of LLMs have been widely discussed in the literature, which include easing out mechanical tasks for educators leaving them with time to focus on higher skills and providing them with creative alternatives for classrooms. However, concerns over the use of LLMs have also been raised. Reportedly, these are thought to pose a threat to the traditional methods of teaching and learning, bringing in issues concerning human bias, discrimination, ownership, copyrights, consent, and privacy. The Higher Education Institutions (HEIs) are struggling to manoeuvre in this shift. This work discusses the Linguistic Empowerment Cell, one of its kind in the Indian landscape, and how AI and LLM support can aid students in their English language learning and enhance their communication skills.

Keywords: Large Language Models (LLMs), English Language Teaching, Artificial Intelligence, Higher Education Institutions

1. Introduction

With a surge in Artificial Intelligence (AI), the new LLM tools for academic purposes are changing by the hour. This emergence of LLMs marks a critical moment for higher education by introducing a paradigm shift, one that is reshaping pedagogy, research, and the role of academic institutions (Xu et al., 2024). AI assistants are reshaping the higher education landscape by transforming institutions and educators (Schön et al., 2023). LLMs have rapidly expanded tools, aimed at student support, teacher growth, and reimagining of academic assessment.

LLMs are increasingly used as tools for study assistance, offering personalised learning, pathways, real-time problem-solving and automated feedback for students (Xu et al., 2024). According to Wang et al. (2024), the systems can adapt to individual learning styles, provide targeted guidance on complex topics and act as tireless tutors, enhancing student engagement and learning efficiency. AI assistants provide indefinite support to add to the students' existing knowledge by

[*] Jawaharlal Nehru University, New Delhi; garimadalal@jnu.ac.in

building on their prior knowledge. Moreover, Schön et al. (2023) believe that "these systems are tailored to the students' learning patterns and pace. In this sense, AI serves as an Intelligent Tutoring System (ITS)". Few researchers feel that LLMs are boosting the output by easing out administrative and pedagogical tasks of teachers by producing course materials, helping them in creating outlines for lesson plans and most importantly, in achieving automatic grading of assignments, leaving them time to focus on higher-order tasks (Schön et al., 2023; Wang et al., 2024; Xu et al., 2024).

Several scholars have also raised ethical concerns about LLMs (Schön et al., 2023). These include discrimination since LLMs can learn human biases or discriminatory answers, depending on the inputs received. Other concerns include acknowledgement, such as ownership and copyrights, as the original sources are not shared or provided; and lastly, concerns like consent and privacy, since the data may include information shared with usage, restrictions or without explicit consent, leading to a breach in information leakage. LLMs are posing a threat to the traditional assessment methods as well. They generate human-like text but also present an opportunity to move towards a more 'authentic and process-oriented evaluation' (Bobula, 2024). Schön et al. (2023) believe that "AI provides research and writing support by assisting in summarising papers, extracting facts, and helping in understanding the conventions for writing assignments" According to Bobula (2024), this poses a disruption in educational assessment. Traditional plagiarism detection tools are also becoming ineffective against AI-generated content, forcing an unnecessary shift in assessment design. A step towards assessment that emphasises critical thinking, creativity and real-world problem-solving by using the higher-order thinking skills of Bloom's Taxonomy is less susceptible to automation. Bobula (2024) believes that evaluating students on their process, drafts, and reflections, rather than solely on the final output, can be a step towards rethinking and re-looking at the way we look at assessment.

2. LLMs and Higher Education Institutions (HEI)

The integration of LLMs has deep repercussions for the core identity and mission of a Higher Education Institution (HEI). Any HEI needs to navigate through these uncertain, fast-paced environments of AI. They need to adapt to the ever-changing background of AI systems quickly. The technology is forcing a revaluation of how institutions function as creators of knowledge, providers of credentials, centres of innovation, and drivers of social change. Idris et al. (2024) highlight both the possibilities and the significant challenges that LLMs pose for changing the traditional roles of HEIs. According to the authors, HEIs must ensure that 'micro-credentials' driven by LLMs meet established standards, account for new forms of certification, and do not diminish the value of traditional degrees. In addition, Idris et al. (2024) believe that HEIs should tread with rules under human supervision, as it ensures a responsible use of technology and that HEIs must be cautious to cross-check for any biases for underrepresented groups.

Within English language teaching specifically, a growing body of systematic reviews synthesises how generative AI is reshaping classroom practice. Crompton et al. (2024) review the use of AI in English language teaching and learning. They find that most of such research has so far been conducted in Asian higher education contexts, leaving other regions, including India, comparatively under-examined. Du and Daniel (2024) show that AI-powered chatbots are increasingly used to scaffold English-as-a-foreign-language speaking practice, reinforcing the relevance of AI-mediated interaction for the kind of language skill-building LEC offers.

3. A Case of Linguistic Empowerment Cell (LEC), JNU, India

To pursue studies in International Universities on a global level, students undertake an International English language proficiency test for non-native speakers, such as the International English Language Testing System (IELTS), which assesses their listening, reading, writing and speaking skills. This test is trusted by universities across the globe. Universities in English-speaking countries use IELTS scores for academic purposes to assess students' proficiency in English and admit them to various courses in their Institutions. Based on the scores, students are required to take basic courses to enhance their English for both academic purposes and communication skills. These courses are usually offered by language support centres in international universities. In the past decade and a half, Indian universities have been robustly hustling with such language support centres for their students as well. India presents a unique educational landscape due to its large student population, diverse linguistic backgrounds and varying access to technology. In a typical higher education classroom, students from different states of India come for studies with English as a medium of Instruction (EMI). The Linguistic Empowerment Cell (LEC) was an initiative in Indian higher education, dedicated to fostering linguistic inclusivity and empowering students from diverse backgrounds to excel in an English-dominated academic environment. Established in 2011, as per the *JNU Newsletter*, LEC "has been committed to ensuring that a student's Linguistic proficiency does not become a barrier to the cognitive and creative potential. This Cell was the first of its kind to be established in India at one of the most premier research universities, the Jawaharlal Nehru University (JNU). The inception of this cell was a direct response to the diverse student population of the university, representing a wide spectrum of linguistic backgrounds, as it draws students from all corners of India" (JNU News Letter, 2012).

Many students, who completed their schooling in vernacular mediums, find the transition to an English-dominated academic setting extremely challenging. The linguistic disadvantage hinders the ability of students to comprehend lectures, participate in academic knowledge, share and articulate their thoughts in well-written forms. Founded on the guiding principle of 'linguistic empowerment as cognitive empowerment', LEC is a language intervention to bridge the gap for students with the requisite skills to excel within their respective

academic disciplines (JNU News Letter, 2012). It was a two-fold move that created a design for other Indian universities. Its success has demonstrated that an institutional commitment to language support is not '*remedial'* but '*essential'* for academic excellence and inclusivity in a country with immense linguistic diversity. This model, conceptualised by JNU, has served as a stepping stone and has been adopted by numerous institutions in Indian higher education.

Since the challenges that led to the creation of LEC are not unique to JNU, universities across India have increasingly recognised the need for structured language and writing support with different names but sharing the LEC core concept. For example, Ashoka University's Centre for Writing and Communication (CWC); Krea University's Centre for Writing and Pedagogy (CWP); National Law School of India University's Writing Centre; and Delhi University's Writing Centre at the Institute of Lifelong Learning, to name a few. This demonstrates a shift in Indian higher education; LEC being at the forefront has provided a model that has proved successful, with dedicated institutional support for essential language teaching, creating a truly inclusive and interactive academic environment in India where universities must actively empower their students to master the language of academic discourse. The Cell has manifested in a practical, unique approach that has resonated with these centres. Over the past decade, it has made a significant contribution to English Language Teaching (ELT), especially English for Academic Purposes (EAP), through its carefully designed certificate programmes and student-centric approach. LEC currently offers capacity-building courses in English for Academic Writing and Basic Communication Skills open to all JNU students. These courses in academic writing and communication skills provide a strong foundation. It also recognises that language for social sciences differs from that of natural sciences; therefore, a discipline-specific approach to English for academic purposes, catering to different disciplines, is vital. Writing is not only about language, but also about structuring arguments and ideas, thereby focusing on critical thinking. Moreover, it also aims to provide a supportive environment where students can overcome linguistic inhibitions without creating a judgmental space.

4. Interactive LLM Learning Environment

LEC offers its two foundational courses, Basic Communication Skills and English for Academic Writing, as two-semester, part-time programmes. The former course is structured around listening, speaking, reading, writing and presentation modules, and the latter moves from academic writing conventions through summarising, paraphrasing, discourse-level coherence, report and abstract writing, to referencing and citation practice. Both courses are now moving toward graded status and, from the coming semester, will be revised to four-credit courses. These courses using LLMs can transform into a dynamic, interactive and personalised journey for each student. The traditional approach to listening skills is primarily reliant on pre-

recorded audio materials; however, LLMs can enhance this experience by providing limitless customised audios for an enhanced experience. The content of these audios can be discipline-specific and tailored to the different levels of difficulty for the student. Moreover, a wide range of global and Indian accents can also be explored to help achieve a neutral accent and be understood. LLMs can also function as interactive tutors where students in groups can engage in a conversation with the AI chatbots to discuss an audio clip. This provides an enhanced and limitless supply of customised and engaging audio content. For instance, a student preparing for a discipline-specific seminar presentation could use an LLM chatbot to rehearse subject vocabulary through a simulated question-and-answer exchange, receiving immediate feedback on pronunciation, coherence and register before the actual presentation.

Moving from listening to speaking is a daunting experience for many students due to the fear of making mistakes. LLMs can create a non-judgmental environment for practice. For reading and comprehension, LLMs can generate text on topics which create interest among the students, especially in a particular discipline, as these classes cater to all disciplines of the university. The level of the text for reading can be challenging; therefore, LLMs can develop readings in which the complexity of the material grows with the student's proficiency. Interacting with the text can also help students enhance their vocabulary in context and build on prior knowledge. Moreover, LLMs can also develop the text according to the levels of Bloom's Taxonomy (Bloom, 1956), starting from basic to higher-order instructional objectives. Students can make mistakes and receive corrections from AI without the fear of being judged by their peers. LLMs can also be used to structure ideas for presentations. This empowers them to take control of their own weaknesses and work on targeted exercises to improve their skills by making them autonomous learners.

Reading and writing skills are closely related, and students often struggle in organising their thoughts and structuring their arguments in a coherent manner. The traditional approach to reading limits the exposure of students to readings, which sometimes can be beyond their level of understanding. This can be replaced by customising their experience with passages tailored to each student's proficiency level, interest and academic context. For writing, LLMs can act as a brainstorming partner, helping students outline their ideas logically for an introduction and moving towards the core content of their writing. Further, AI can assist students in understanding the errors in spelling and grammar; suggest better vocabulary, stylistic improvements and well-structured sentences. Also, AI can curate examples of both good and bad writing to develop critical writing analysis for students.

Establishing a clear and explicit policy in the system, syllabus and classroom, stating what is and is not acceptable use of LLM, is crucial. For example, it is important to understand the difference between using

an LLM to brainstorm ideas, check grammatical errors and suggest alternative paraphrasing for a sentence or explain a difficult concept, which can be acceptable, but not generating entire sentences, paragraphs or essays and submitting them as their own authentic work without acknowledging the original sources. LLMs can be used for creating interactive paraphrasing practice or dynamic summarising exercises, for example, the student takes a paragraph from a scholarly article and reframes the whole set in their own words. LLMs can be provided here with an original paragraph and an attempted paraphrase to compare the paraphrased version with the original and to confirm if the core meaning stays intact. Properly integrating quotes and maintaining an academic tone is also a crucial skill, especially in EAP for students pursuing higher education. Students can use LLM to review their drafts, especially to check the tone and highlight any instances of informal language like slang and contractions. It is important to make students aware of the fact that AI-generated content needs to be reviewed critically. Since LLMs produce incorrect information or fake citation therefore, it is vital to encourage students to evaluate the suggestions made by LLMs. For educators, LLM transparency in the classroom is important; we should be open to criticising its authenticity rather than using LLM as a deceptive device.

5. Conclusion

The rise of LLMs presents both transformative opportunities and significant challenges for higher education. Though these tools can enhance learning, they can also raise concerns about academic integrity and have a high risk of being misused. Although institutions of higher education need to integrate academic policies for the use of generative AI, it is important to encourage students to be transparent about their use of AI tools; therefore, a disclosure of how they have used it in their submissions is necessary. As teachers, educators and students run the potential risk of generating inaccurate and biased information. Using LLMs, students must be made aware that they are responsible for the accuracy and integrity of the work. They are submitting regardless of whether they have used an AI tool or not. Teachers and educators must rethink their assignment designs carefully. LLMs can be effective in generating interactive environments for students. However, the question regarding the misuse of LLM needs to be maintained. LLMs are most useful in breaking down large assignments into smaller steps, such as brainstorming, outlining, drafting or peer reviewing, giving students multiple opportunities to engage in and for the instructor to give feedback, but with caution. Moreover, LLM-generated interactive environments encourage students to engage with material directly and learn from one another. However, we must also not forget that LLM only replicates the real world but cannot replace the real world.

Creativity is a higher-order thinking skill which needs to be worked out by teachers and students alike in order to keep a natural balance. Therefore, an amalgamation of some assignments which can be designed to have students actively engage in LLM-generated

content, relevant for students, is alright instead of banning AI altogether. It is also essential to proactively educate students on how to use LLMs ethically and effectively as a learning aid by not limiting themselves to it entirely. This approach acknowledges the presence of LLM and teaches students to use it responsibly. This also includes an understanding of a potential bias and the importance of fact-checking and verifying sources. Creating a classroom environment for students to feel comfortable in discussing their use of AI should be promoted as an open dialogue. Therefore, preventing unethical LLM use and plagiarism in higher education requires a proactive and multifaceted approach. For the Indian context specifically, this points to concrete next steps: institutions can draw on the National Education Policy 2020's emphasis on multilingualism to fund and formalise language-support cells like LEC across public universities, not only in well-resourced central universities; language centres can develop discipline-specific AI-use guidelines in consultation with faculty, on the model LEC has already piloted for academic writing and communication; and universities can invest in structured training for English language faculty so they are equipped to supervise LLM-assisted learning rather than treat it as an external threat to be policed.

References

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay.

Bobula, M. (2024). Generative artificial intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education.* https://doi.org/10.47408/jldhe.vi30.1137

Choi, W. C., Chang, C. I., Choi, I. C., & Lam, L. C. (2025) A review of large language models (LLMs) development: A cross-country comparison of the US, China, Europe, UK, India, Japan, South Korea, and Canada. *Preprints.org.*https://doi.org /10.20944/preprints202504.2136.v1

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. British Journal of Educational Technology, 55(6), 2503–2529. https://doi.org/10.1111/bjet.13460

De Fine Licht, K. (2023).Integrating Large Language Models into Higher Education: Guidelines for Effective Implement tion. *Computer Sciences & Mathematics Forum, 8*(1), 65. https://doi.org/10.3390/cmsf2023008065

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. Computers and Education: Artificial Intelligence, 6, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Fowler, D. S. (2023). AI in Higher Education. *Journal of Ethics in Higher Education.* https://doi.org/10.26034/fr.jehe.2023.4657

Hu, Z., & McGrath, I. (2011). Innovation in higher education in China: Are teachers ready to integrate ICT in English language teaching? *Technology, Pedagogy and Education, 20,* 41–59. https://doi.org/10.1080/1475939X.2011.554014

Idris, M. D., Feng, X., & Dyo, V. (2024). Revolutionizing Higher Education: Unleashing the Potential of Large Language Models for Strategic Transformation. *IEEE Access, 12*, 67738–67757. https://doi.org/10.1109/ACCESS.2024.3400164

Jawaharlal Nehru University. (2012, November–December). *JNU Newsletter*. Public Relations Office. Retrieved September 3, 2025, from https://www.jnu.ac.in/sites/default/files/jnunews/JNUNewsletter_nov-dec-2012.PDF

Jawaharlal Nehru University. (n.d.). *42nd Annual Report.* Retrieved September 3, 2025, from https://www.jnu.ac.in/site s/default/files /annual_reports/42AnnualReport_Eng8.pdf

Jawaharlal Nehru University. (2025). *LEC: An introduction.* Linguistic Empowerment Cell. Retrieved September 3, 2025, from https://www.jnu.ac.in/LEC/

Rahman, M. M., & Mehar Singh, M. K. (2020). Language ideology of English-medium instruction in higher education: A case study from Bangladesh. *English Today, 36*(4), 40–46. https://doi.org/10. 101 7/S0266078419000294

Schön, E., Neumann, M., Hofmann-Stölting, C., Baeza-yates, R., & Rauschenberger, M. (2023). How are AI assistants changing higher education? *Frontiers in Computer Science, 5*. https://doi.org/10.3 389/fcomp.2023.1208550

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., Yu, P. S., & Wen, Q. (2024). Large Language Models for Education: A Survey and Outlook. *ArXiv, abs/2403.18105*. https://doi.org/10. 48550/ar Xiv.2403.18105

Xu, H., Gan, W., Qi, Z., Wu, J., & Yu, P. S. (2024). Large Language Models for Education: A Survey. *ArXiv, abs/2405.130 01*. https: //doi.org/10.48550/arXiv.2405.13001

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 139-162

Learning Outcomes of Saudi Students using Artificial Intelligence (AI) for Foreign Language (FL) Proficiency Development

Suaad Muhammad Saeed Alqahtani*
Nisar Ahmad Koka**
Sheeba Hassan***

Abstract

This study explores the impact of Artificial Intelligence (AI)-powered tools on the development of foreign language (FL) proficiency among students in Saudi Arabia. It examines the effectiveness of these technologies in enhancing vocabulary acquisition, speech fluency, and listening comprehension. Employing a mixed-methods design that integrates both quantitative and qualitative analyses, the research provides a comprehensive evaluation of AI-assisted language learning practices. Findings indicate that AI tools significantly improve vocabulary retention and listening skills, however, challenges persist in fostering speaking proficiency and cultural interpretation. The results highlight that while AI can effectively support language learning, optimal outcomes require the integration of technological tools with traditional pedagogical methods to bridge cultural gaps and enhance oral communication skills. The study concludes by emphasizing the potential of AI to accommodate diverse linguistic inputs, while underscoring the enduring importance of active human collaboration in effective language education.

Keywords: Artificial Intelligence (AI)-Powered Tools, Foreign Language (FL) Proficiency, Language Learning Tools, Classroom Technology Integration

1. Introduction

Integration of Artificial Intelligence in language education research has gained major scientific attention as researchers seek to maximize educational effectiveness of innovative tools. AI applications including machine learning, natural language processing and speech recognition technology offers personalized adaptive learning experiences that supposedly advance student language abilities (Yamijala et al., 2025). Several countries have been seriously exploring the potential of technological tools to transform their learning environments.

* King Khalid University, KSA; ssaeed@kku.edu.sa
** King Khalid University, KSA; ncoka@kku.edu.sa
*** University of Kashmir; sheeba@uok.edu.in

Many AI-powered language learning applications including Duolingo and Babbel present individualized instruction through their advanced tracking systems that adapt lessons specifically for student requirements (Bernard et al., 2019).Despite these achievements, challenges for universal effectiveness of AI tools for foreign language acquisition persist. Literature analysis of AI applications in language education focuses primarily on Western educational models while leaving major gaps in comprehending how such solutions can address unique learning barriers facing non-Western education systems specifically the Saudi Arabian institutions (Ibrahim & Barnawi, 2022). Teaching of English as second language in Saudi relies on traditional language teaching methods but continuing proficiency gaps indicate that these teaching strategies have insufficiently closed this language divide (Abahussain, 2016). Successes of AI applications elsewhere can be the motivation for considering the potential of AI in solving faced in English language teaching in Saudi albeit with its unique challenges.

In Foreign language (FL) classroom, integration of Artificial Intelligence (AI) makes it possible to offer personalized learning that traditional pedagogical approaches fall short of achieving. Through Artificial Intelligence (AI) tools, students can receive personalized learning content that matches specific learning preferences and receives real-time feedback and specified practice together with interactive exercises matched to their proficiency level (Yang et al., 2010). Individual attention remains essential in language acquisition because different student abilities and learning speeds challenge traditional classroom structures (Schmidt, 2012). AI's ability to improve language learning results across all situations remains an oversimplified view. The potential advantages of AI for language ability development in vocabulary and grammar extend only when considered alongside the training history of users with technology and their cultural roots, and the AI system quality (Zhai & Wibowo, 2023).

In the Saudi context, AI based tools like speech recognition and immediate feedback can enhance learning among learners of English language (Dhouib et al., 2022), yet serious challenges remain in implementing AI technology in the country. Previous research demonstrates that AI tools experience difficulties in providing context-sensitive as well as culturally proper feedback particularly when operating in countries with distinctive cultural personalities like Saudi Arabia. Language learning tools require cultural relevance because students perform better within materials that relate to their experience and local cultural standards (Tomlinson, 2012). Research conducted by Cope et. al., (2021), together with Sajja et al., (2024) shows that generic AI tools supply standardized content which fare insensitive to individual needs and cultural experiences leading to unwarranted outcomes.

Impact of AI on linguistic education depends on three essential components identified by Hassan & Fook (2014). The first requirement is superior tool quality, and secondly solid educational methods

together with efficient implementation within the academic framework are necessary. Research has shown that AI devices receive criticism because they depend on basic memorization instead of humanlike interaction which builds pronunciation and speaking ability. AI systems prioritize individual learning at the cost of limiting students from participating in essential social language acquisition practices which are needed to achieve fluency in real-life linguistic environments (Kingston et. Al., cited in Taguchi, 2008).

Implementation of AI in language teaching faces crucial challenges in Saudi Arabia because issues with high-quality tool access alongside widespread digital gaps reduce its effectiveness (Aldossary, 2023). Variations in students' economic status prevent them from acquiring modern technology tools equally creating learning opportunity gaps between groups. The quick advance of AI tool integration in education system commonly happens faster than educators get proper training to execute AI-based teaching techniques. Teacher preparation at current levels creates substandard results because AI tools reach their full potential only when instructors possess an in-depth understanding of their mechanism and implementation (Chng et al., 2023).

The research examines AI tools' impact on foreign language (FL) education in Saudi Arabia by studying quantitative measuring systems and their effects on classroom education. To properly evaluate language learning tools, the analysis must extend beyond measurement of skill development into an assessment of their relation to Saudi educational elements combined with their capability to solve local student needs. The research investigation generates new insights about AI in foreign language (FL) education by observing its function within the Saudi Arabian educational system which maintains a non-Western cultural setting.

2. The Present Study

Adoption of Artificial Intelligence (AI) in educational practices worldwide shows increasing trend, however, there have been insufficient studies to understand its potential impact on the educational setup of Saudi Arabia especially on English language proficiency. Saudi Arabia has permitted Artificial Intelligence (AI) to rapidly penetrate all sectors including educational fields. The current body of research fails to demonstrate how AI-driven language learning tools genuinely affect Saudi students' achieved levels of proficiency in school and university settings. Despite showing promise for shaping personalized learning experiences, AI tools remain uncertain in their ability to address the proficiency challenges observed within Saudi foreign language education setup. This research studied how AI tool utilization impacts the results of Saudi students' foreign language (FL) proficiency. The study is based on the following questions:

a) To what extent does the use of AI-powered tools impact the foreign language (FL) proficiency of Saudi students?

b) How do AI tools contribute to the improvement of specific language skills (e.g., vocabulary, grammar, speaking, and listening) among Saudi foreign language (FL) learners?

c) What are the challenges and limitations faced by Saudi students in using AI tools for foreign language (FL) learning, and how do these factors affect their learning outcomes?

The present research adds important value to language education theory by examining how Artificial Intelligence (AI) tools benefit non-Western language education institutions. Recent expansion of AI education research includes few longitudinal studies despite most analyses concentrating on Western student populations followed by a notable research gap regarding Saudi Arabian educational contexts. Research in the Saudi context demonstrates the adaptive potential of AI tools for managing key local educational challenges with their language and cultural distinctions. The research findings deliver valuable knowledge to education professionals alongside policymakers and developers who work on Artificial Intelligence (AI)-powered language platforms. The study provides knowledge about factors affecting Saudi schools and universities regarding their capacity to use AI tools while adding to discussions about digital access and educational technology equity. This research aims to establish guidelines for AI language learning tools and their educational frameworks adoption in Saudi Arabian institutions.

3. Literature review

Widespread adoption of Artificial Intelligence (AI) across language learning spaces has become a debatable topic in recent years because of emerging inquiries about its benefits and drawbacks. Language learning apps together with speech recognition systems and intelligent tutoring systems operate through Artificial Intelligence (AI) to create personalized learning settings while providing instant feedback and targeting each student's specific educational requirements. The ability of AI to improve foreign language (FL) skills has extensive support yet researchers fail to conduct detailed studies on its practical effectiveness in Saudi Arabian education system.

AI systems offer flexible learning spaces which allow students to practice individually through personalized activities that adapt to their own learning rate. The systems monitor student advancement by recognizing learning weaknesses to deliver customized lessons and feedback after assessing each student's response. The AI tools on Duolingo among other platforms build personalized lessons utilizing algorithms that match student ability to reach notable gains in vocabulary memorization and language competence. AI tools deliver important feedback immediately to learners during language acquisition which enables them to correct mistakes instantaneously before Piering problematic linguistic patterns according to a study by Betaubun et al. (2023).

AI systems display the capability to deliver education through a non-standard student-centered methodology. Compared to traditional classrooms which force teachers to manage many students at once, AI tools function exclusively toward specific learning requirements (Pedro et al., 2019). AI tools adapt their teaching approach to recognize learner specific strengths and weaknesses with the weaker academic areas receiving proper focus without creating student confusion. AI tools create extensive immersive educational experiences for learners when teaching vocabulary and grammar along with pronunciation fundamentals that lay the foundation of foreign language (FL) acquisition (Song et al., 2023).

Studies have revealed that AI brings many benefits yet researchers caution that AI tools demonstrate limited ability to conduct natural interactions with humans. Farr (2024), believes AI provides solid support for rote learning and vocabulary expansion yet shows constrained ability to teach communicative competency which demands sophisticated understandings of social context. The natural social dimension of language disrupts AI-powered simulations because traditional classrooms remain the optimal environment where students learn language skills through authentic interactions between peers and teachers according to George (2023). Several researchers confirm that AI tools demonstrate excellent performance in managed repetitive tasks yet lack effectiveness when it comes to fostering genuine language practice.

The use of Artificial Intelligence (AI) for language education remains immature throughout Saudi Arabia since procedural research on its effects within such academic settings is meagre. Using the Vision 2030 initiative, Saudi Arabia endorsed major improvements in educational technology adoption with special focus on integrating advanced technologies throughout all sectors including schools. The current research on AI tools in Saudi Arabian language education remains scare with the need for additional quantitative studies that demonstrate the effectiveness of these tools for Saudi learners' specific challenges.

Foreign language (FL) acquisition in Saudi Arabia faces substantial hurdles due to psychological barriers created by the vast differences between Arabic as the native language and English the preferred target language. Linguistic distance between Arabic and English affects the ability of students to acquire English proficiency because it creates major obstacles in pronunciation, alongside syntax and vocabulary acquisition. Language learning tools powered by AI and speech recognition capabilities provide immediate feedback during practice to help students develop their pronunciation skills and listening abilities (Dennis, 2024). AI tool success in Saudi Arabia also depends on the cultural relevance. According to Strielkowski et al. (2024), it is essential to develop AI-driven educational tools that both fit the cultural norms of local areas and effectively match student needs. Cultural sensitivity is essential because absent features lead to a breakdown in

student engagement and a mismatch with Saudi language learning requirements.

The wider influence of AI in language education remains a research topic but most studies focus on Western case studies. Liu & Wang (2024) published results showing AI tools achieving successful English proficiency development for Chinese EFL students. According to Rad et al. (2024), AI tools demonstrate effective results for fostering vocabulary growth and grammar skills during learner-oriented personalized practice. These research studies overlook the special difficulties that non-Western educational structures consisting of Saudi Arabia's educational system face.

Almuhanna (2024), investigated Saudi classroom technology integration showing that AI tools can enhance language acquisition but instructional limitations emerge from inadequate teaching instructions and insufficient guidelines for AI curriculum integration. Research from Almansour (2019) showed that technology tools like Duolingo received acceptance from language learners in Saudi Arabia yet produced inconsistent outcomes since users gained vocabulary knowledge without noticeable progress in speaking or writing skills. The success of AI tools depends on several critical elements including student participative commitment, suitable tool selection and appropriate pipeline integration within current educational programs.

Aljabri & Bhutoria (2020) demonstrated in an analysis that Saudi Arabia's digital differences reduce the reach of AI tools especially when considering students with lower household earnings. When students are not provided equal benefits from AI-driven learning tools, the potential for these technologies to boost language proficiency suffers. AI tools present strong opportunities to change foreign language (FL) instruction, but their capabilities and equity distribution within Saudi learning environments need continued research.

The research application of information technology shows potential for reshaping language education through non-traditional learning spaces yet its implementation within Saudi educational institutions remains understudied. Problematic research exists about how Saudi students use AI tools and their views of AI with concrete findings regarding their learning performance. This research investigates how the usage of AI tools affects educational results in foreign language (FL) performance across Saudi territory. This analysis evaluates both educational frameworks and socio-cultural conditions which impact AI tool performance in non-Western settings for foreign language education.

4. Methodology

The study was conducted between January 2024 and June 2024 on the students enrolled in English language courses at three Saudi Arabian universities. A total of 300 students from both the genders participated in the study. Participants were divided into two groups of 150 students each

between 18 and 25 years of age. The experimental group, which used AI-powered language learning tools, while the control group continued with traditional learning methods. The experimental group used the learning applications, Duolingo and Babbel which provided customized feedback and dynamic learning progressions according to student achievement. Pre- and post-tests were used to assess students' core language competencies, including vocabulary knowledge, grammar skills, and both spoken and listening proficiency. Survey techniques alongside interview methods were combined with pre-interactive and post-interactive assessments to obtain qualitative insights from students regarding their AI tool experiences. The study used mixed-methods research design consisting of quantitative pre- and post-tests coupled with interviews with selected participants.

The primary instruments for data collection were pre- and post-tests designed to measure proficiency of the participants in key areas of foreign language learning. Students completed basic proficiency tests at the beginning of the study to establish their initial competency and received post-test assessment six months later to evaluate skill level modifications.

The experimental group utilized AI tools daily for thirty minutes during five days per week. Students utilized these learning tools through their mobile applications and their computer desktop interfaces based on personal preferences.

The qualitative assessment of student feedback and tool experiences involved semi-structed interviews with selected participants. The participants answered questions about tool accessibility and user motivation while reporting on their interest levels and perceiving language skill growth. Analysis became easier with the Likert-type format used in the survey.

Data collection began by giving the participants the pre-test for measuring their initial proficiency levels. The assessment contained four sections evaluating vocabulary knowledge alongside grammar knowledge and listening comprehension along with speaking proficiency. During six months the test group implemented AI tools as their standard learning practice. The control group engaged in traditional language learning methods involving classroom teaching alongside textbooks and peer discussions throughout the experiment.

A post-investigation proficiency assessment was administered to all students after six months of instruction. The post-test evaluation structure followed exactly the same pattern as the initial assessment to prevent testing measurement inconsistency. Semi-structured interviews were conducted and surveys were distributed after the post-test phase to the entire sample, but the researchers conducted discussions with a reduced subset consisting of 30 students (15 each from the two groups). Recording capabilities were applied during interview sessions due to student preferences for conducting discussions either in English or Arabic.

Variable analysis employed paired-sample t-tests to evaluate statistical differences in language proficiency scores between the experimental and control populations at pre- and post-test intervals. The analysis used t-tests to perform both intra-group comparisons of mean pre- and post-test scores and between-group comparisons of the scores. Analysis of covariance (ANCOVA) supported the examination by controlling prior language proficiency levels together with AI tool usage time to confirm the intervention as the only cause of reported changes.

The study measured AI tool effects on individual language abilities through a multivariate analysis of variance structure that analyzed changes in vocabulary and grammar alongside listening comprehension and speaking tasks. Additional analysis through this method revealed precise areas where students achieved greatest enhancements through their utilization of AI tools.

Researchers analyzed the survey and interview data through thematic approaches. The survey responses received thematic codes that captured recurring ideas about students' views regarding AI tools and their engagement with learning materials and difficulties they encountered. A grounded theory analysis of interview transcripts extracted verbalized patterns that students encountered while using AI technologies. Through qualitative analysis the study obtained expanded insight into AI tool usage patterns among students while uncovering what motivates them to use the tools and what barriers prevent effective use.

5. Results

5.1 Quantitative Analysis

The variance equality test known as Levene's assessment determined uniformity between students in the control and experimental groups before the project began. For t-tests and ANCOVA to produce valid results, the variance between groups requires homogeneity. The analysis verifies if both groups maintain comparable levels of statistical variation.

Group	F-value	p-value
Experimental Group	1.23	0.267
Control Group	1.27	0.254

Table 1. Levene's Test for Homogeneity of Variance

The experimental and control groups demonstrate comparably homogenous variances because their Levene's test p-values exceed 0.05. The homogeneity of variance assumption proves valid making further analyses with t-tests and ANCOVA appropriate.

Both pre-test score distributions were assessed for normality through the Shapiro-Wilk test. The ability to use t-tests and ANCOVA requires that test data exhibits normality characteristics.

Group	W-value	p-value
Experimental Group	0.986	0.144
Control Group	0.981	0.201

Table 2. Shapiro-Wilk Test for Normality of Pre-test Scores

The Shapiro-Wilk test demonstrates that experimental ($p = 0.144$) and control ($p = 0.201$) dataset p-values exceed 0.05. The pre-test scores of both groups exhibit normal distribution patterns which enable subsequent analysis.

To establish baseline proficiency indicators, we began by examining the pre-test descriptive statistics for both groups. The pre-test measured language skills in four key areas; students obtained scores from assessments evaluating their vocabulary alongside grammar use alongside listening comprehension and speaking skills.

Group	Mean Score	Standard Deviation	Minimum Score	Maximum Score	N
Experimental Group	56.5	8.3	42	72	150
Control Group	55.3	7.9	41	71	150

Table 3. Descriptive Statistics of Pre-test Scores

The recorded mean scores for experimental participants (56.5) matched closely with those of the control group (55.3). Standard deviations show that test scores spread moderately throughout both groups (8.3 and 7.9). Pre-intervention testing demonstrated similar language skill strengths for both participant groups.

Students received post-tests after the six-month AI tools implementation within the experimental group combined with traditional teaching methods in the control group. Table 1 below shows the descriptive statistics that analyze post-test results after assessment of both groups.

Group	Mean Score	Standard Deviation	Minimum Score	Maximum Score	N
Experimental Group	78.4	9.2	61	95	150
Control Group	69.2	7.6	54	84	150

Table 4. Descriptive Statistics of Post-test Scores

The experimental group showed superior success in foreign language acquisition following intervention since their mean score reached 78.4, while the control group scored 69.2. Both groups show moderately spread data points although the experimental group displays higher score variability indicating participants improved their language skills with diverse effectiveness.

A paired-sample T-Test compared students' scores from pre-test to post-test periods to identify significant changes between pre- and post-retest scores for each experimental bloc.

Test	Mean Difference	t-value	df	p-value
Pre-test vs Post-test	21.9	15.24	149	0.000

Table 5**.** Paired-Sample T-Test Results for Experimental Group

The paired-sample t-test results for the experimental group show a significant improvement in language proficiency from pre-test (M = 56.5) to post-test (M = 78.4), with a t-value of 15.24 and a p-value of 0.000. This indicates that the use of AI tools significantly improved the students' proficiency in foreign language skills.

Test	Mean Difference	t-value	df	p-value
Pre-test vs Post-test	13.9	10.56	149	0.000

Table 6. Paired-Sample T-Test Results for Control Group

Post-test data from the control group provided evidence of significant language proficiency growth between pre-test (M = 55.3) to post-test (M = 69.2), reflected in the t-value (10.56) and p-value (0.000). The experimental group exceeded the control group in its success rate at acquiring language skills although both groups experienced improvements.

The contrast between experimental and control group post-test results was evaluated with an independent-samples t-test protocol.

Group	Mean Score	t-value	df	p-value
Experimental Group	78.4	8.67	298	0.000
Control Group	69.2	8.67	298	0.000

Table 7. Independent-Samples T-Test for Post-test Scores

A significant difference emerged between experimental group mean scores (M = 78.4) and control group mean scores (M = 69.2) as determined by an independent-samples t-test with t = 8.67 and p = 0.000. The experimental group achieved greater language proficiency gains when using AI tools than the control group without access to AI tools.

Researchers used MANOVA to evaluate how AI tools affected students' language abilities combining vocabulary results with grammar results and listening comprehension scores with speaking performance data.

Language Skill	F-value	p-value
Vocabulary	21.34	0.000
Grammar	18.92	0.000
Listening Comprehension	16.58	0.000
Speaking	15.72	0.000

Table 8. MANOVA Results for Post-test Scores by Language Skill

Tests from MANOVA analysis demonstrate that AI tools produced meaningful consequences across every element of language proficiency. Statistics from the four language skill areas show AI tools produced positive relationships with student performance since all p-values including vocabulary ($p = 0.000$), grammar ($p = 0.000$), listening comprehension ($p = 0.000$) and speaking ($p = 0.000$) fell below 0.05.

Language Skill	Mean Score	Standard Deviation	Minimum Score	Maximum Score	N
Vocabulary	19.4	3.1	14	24	150
Grammar	20.2	2.8	15	25	150
Listening Comprehension	19.6	3.2	13	24	150
Speaking	19.2	3.5	14	23	150

Table 9: Breakdown of Post-test Scores for Experimental Group

Results indicate that experimental group students achieved the most significant gains in grammar ($M = 20.2$) and listening comprehension ($M = 19.6$). The detected levels of standard deviation across four language skills demonstrate that students experienced average variation while improving their performance. Students achieved substantial improvements in grammar proficiency and listening comprehension through the usage of AI tools because these tools adapt to learner needs.

Language Skill	Mean Score	Standard Deviation	Minimum Score	Maximum Score	N
Vocabulary	16.5	2.9	12	21	150
Grammar	17.3	3.1	13	22	150
Listening Comprehension	17.2	3.0	12	22	150
Speaking	18.2	3.3	13	23	150

Table 10. Breakdown of Post-test Scores for Control Group

Students who did not employ AI tools achieved less growth in their language abilities than participants who adopted AI tools. The experimental group achieved the greatest improvement in speaking ($M = 18.2$) yet maintained higher total proficiency scores than the control group. AI-based language learning techniques yielded superior outcomes when compared to traditional methods which otherwise maintained their effectiveness.

Variable	B-value	Standard Error	Beta	t-value	p-value
AI Tool Usage (Experimental vs. Control)	8.5	1.2	0.42	7.08	0.000

Table 11. Regression Analysis Results

The experimental group students achieved improved language proficiency through AI tool use according to regression analysis.

Students who participated in AI tool utilization achieved an average 8.5 point higher improvement compared to the control group students according to the B-value of 8.5. The calculated t-value (7.08) and p-value of 0.000 show significant statistical effects plus the beta coefficient (0.42) confirms moderate strength of relationships between variables. Research evidence demonstrates AI tools produce substantial improvements in language learning achievements.

Gender	Mean Score	Standard Deviation	N
Male	79.2	8.5	75
Female	77.6	9.1	75

Table 12. Gender Differences in Post-test Scores for Experimental Group

Male students in the experimental group scored slightly higher (M = 79.2) than female students (M = 77.6) during testing although their overall performance remained equivalent. The standard deviation results indicate varied levels of achievement in each gender group yet the mean score differences indicate minimal evidence of a gender disparity in AI language training effects.

Gender	Mean Score	Standard Deviation	N
Male	70.1	7.9	75
Female	68.3	7.7	75

Table 13. Gender Differences in Post-test Scores for Control Group

Post-test scoring data demonstrated that male participants in the control section achieved higher overall scores than female participants with M = 70.1 versus M = 68.3. Effects from the intervention showed improvement across both genders but genders remained close to one another through their mean performance points. The analysis points to a minor gender effect on learning achievement which remains below the significant improvements across all participant groups.

Age Range	Mean Score	Standard Deviation	N
18-22	80.3	8.2	90
23-27	77.8	9.3	60
28 and above	75.9	9.6	30

Table 14. Age Differences in Post-test Scores for Experimental Group

Research findings indicate that students between 18-22 years in the experimental group achieved the highest post-test scores at M = 80.3. Students aged 23-27 obtained scores at M = 77.8. The language proficiency results indicated older students (M = 75.9) achieved the lowest scores among all age groups which points to a possible moderate relationship between age and language acquisition improvement. Participants in the younger age bracket demonstrated higher scores because they comfortably use both technology tools and AI-based educational resources.

Age Range	Mean Score	Standard Deviation	N
18-22	71.2	8.0	90
23-27	68.3	7.5	60
28 and above	67.1	7.8	30

Table 15. Age Differences in Post-test Scores for Control Group

The control group students between age 18 and 22 obtained higher scores (M = 71.2) than all other participants while individuals aged 28 or older demonstrated the lowest mean score (M = 67.1). Results show that age impacts language learning performance independently of AI tools utilization due perhaps to students' differing past experience with language methods and their varying design to learn.

5.2 Qualitative analysis

Following are presented statements drawn from the semi-structured interviews with selected participants of the study upholding the significance of AI based tools in supplementing language learning.

At first, I was skeptical about using AI for language learning. But after a few sessions, I saw how quickly I was learning new words and phrases. It felt like I had a personal tutor.

The participant's viewpoint suggests that AI tools which foster student-teacher interaction serve as essential educational tools that create improved learning engagement. This has been observed by Henrie et al., (2015) that students found technology-mediated learning experiences to be more engaging and satisfying.

The AI system seemed to understand my mistakes and corrected them immediately. This was different from how teachers do it in class, where you may not get feedback right away.

The main benefit of AI tools becomes evident because students receive immediate evaluations. Delays in feedback emergence in traditional classrooms occur as a result of either teacher workloads or too large student-to-teacher ratios. Through real-time feedback AI technology shows its ability to deliver corrections which confirms the system's potential to connect gaps between feedback and learning progress. Language learners benefit strongly from fast feedback because it enables them to retain information and correct mistakes with greater speed (Sheen, 2007).

It was interesting how the AI would adjust its difficulty level based on how I was doing. It didn't feel like I was stuck on the same task for too long.

AI tools adjust their instructional approach to a learner's personal speed and educational levels, thereby enhancing personalized learning standards. AI systems enables an educational system through their adaptive capabilities to deliver standardized teaching instruction. A

study by Xu (2024) concludes that AI-driven learning when tailored to students' specific needs produces improved academic results.

Using AI, the lessons feel less intimidating. It was like practicing in a low-pressure environment, so I felt more confident trying new things."

AI tools establish learning environments to support students especially those who experience higher levels of anxiety in speaking. Artificial Intelligence (AI) creates a non-threatening learning space for Saudi students who find themselves in performance-success pressure environments, so they feel more comfortable participating in activities (Doghan, 2020).

I could see how much my vocabulary improved. The AI gave me so many new words, and the quizzes helped me remember them better.

Trackable language improvement shows how AI-driven tools effectively train learners in precise language competencies. Computer-driven interactive exercises that combine learning with memory reinforcement dramatically speed up the acquisition process for newcomers to language vocals. The results connecting vocabulary mastery to language skills quality show AI capability for creating streamlined and captivating vocabulary learning pathways especially effective for the Saudi Arabian education context (Almuhanna, 2024).

I struggled with grammar, but the AI's interactive explanations and examples helped me understand the rules better than I expected."

One of the most difficult language learning areas is grammar but the positive experience with AI-based interactive methods illustrates fundamentally how AI effectively addresses confusing language components. AI-directed personalized explanations guide students through confusing grammatical structures thanks to adaptive instructional approaches while serving English-language learners in Saudi Arabia who face challenges because of linguistic differences between Arabic and English (Al-Nasser, 2015).

The speaking exercises were the hardest for me. But over time, the AI's voice recognition system helped me get more confident with my pronunciation.

The participant's observations about speaking exercises coupled with their experience using AI voice recognition systems underline important aspects of language learning pronunciation and speech development. A significant number of learners in Saudi Arabia avoid speaking classes because they lack confidence and make mistakes when speaking with their classmates. Students benefit from AI solutions which provide uncondemning opportunities to practice speech leading to improvement in their fluency and confidence (Poelling, 1971).

However, participants also reported a series of challenges while using the AI tools, some of which are as below.

Sometimes the AI misunderstood what I said, and that was frustrating. I had to repeat myself several times.

The participant's experience demonstrates how voice recognition systems alongside natural language processing features in AI are currently restrictive. AI technology demonstrates numerous advantages yet it still possesses recognized limitations. Backup issues from technology could interrupt students' educational progression specifically when these issues exceed their ability to address them successfully. The wider adoption of AI tools in language learning faces an important limitation because AI systems require further evolution to correctly understand different accents and speech styles (Dong et al., 2024).

I felt that sometimes the AI focused too much on grammar rules, and I wanted to practice speaking more.

This participant emphasizes the critical equilibrium requirement which exists between language capabilities in Artificial Intelligence (AI) learning systems. Speaking practice gets insufficient attention according to this participant despite the important role of grammar instruction. The Saudi Arabian educational system typically gives limited importance to speaking abilities, so AI tools must develop balanced domains to serve students best (Khosravi et al., 2022).

The system wasn't always easy to use, especially when there were updates or changes. It was confusing to navigate sometimes.

The educational benefits from using AI tools remain inaccessible to students especially in Saudi regions because of varying student proficiency levels with technology. The adoption of new technologies presents a learning curve that creates challenges for learners whose pattern of exposure to frequent technological changes and updates is limited according to Blin & Munro (2008).

I didn't feel like the AI could answer all my questions. Sometimes I needed a human teacher for more in-depth explanations.

The instant data AI provides about language practice needs human educators to supply the essential depth and empathy needed to enrich classroom learning. This learner emphasized a need for human-instructed explanations that address intricate language problems but AI tools currently show limitations in handling these complex situations. Research suggests that Saudi Arabian language learners will benefit most from combining Artificial Intelligence (AI)tools with traditional teaching approaches as found in Alotaibi & Alshehri (2023).

6. Discussion

The study evaluated how AI-powered technologies influence the skills development of foreign language (FL) operations in Saudi Arabia. Both vocabulary acquisition and listening comprehension benefited substantially according to data collected from the study. The research demonstrated key obstacles as well as restrictions for AI in both

improving speech proficiency and recognizing cultural differences. The results established both their educational impact of AI within Saudi language education and their broader international relevance.

The study demonstrated striking vocabulary gains together with improved listening abilities when students utilized AI tools for their language learning. The results indicate that AI-powered learning approaches work best when students receive personalized education which underlines recent academic research on using AI in language education. Research shows that AI-powered learning tools create personalized feedback and adapt educational content to student advancement so learners build stronger vocabulary knowledge alongside better listening abilities (Newton & Nation, 2020). Participants in this research mentioned that AI tools improved their word retention since they came across words in multiple situations that is vital for effective language learning. A student observed that AI presented the words in multiple contextual examples to make meanings more understandable to students. Coșulean (2024), demonstrated a similar finding which showed that customized AI-learning through dynamic educational platforms helps students better maintain their vocabulary.

The research demonstrated valuable progress in vocabularies and listening comprehension but showed important limitations in AI systems when it comes to teaching students conversational fluency. This study confirms an isolated discovery. Literature about language acquisition reveals mounting evidence that Artificial Intelligence (AI) struggles to teach fluent speech skills (Li & Lan, 2022). The investigation showed participants being disappointed about their limited access to perform authentic dialogues since they saw actual conversation practice as vital for better vocal proficiency. One participant told us, "I could remember words and phrases but struggled to communicate naturally." Language learning faces difficulties because actual speaking practice stands as an essential component for developing fluency alongside real-world communication confidence (Richards, 2005). AI tools succeeded in helping students learn pronunciation and word recognition yet these resources did not provide essential conversational context which communication competence needs.

The research found an essential problem with AI language learning systems because of their cultural competence barriers. The study participants noted how AI tools lacked capability to deliver the essential cultural background information tying certain phrases together. According to one interviewee AI systems provided vocabulary support but did not teach cultural context because this made learning difficult at times. Current AI systems demonstrate an important functional limitation when it comes to teaching cultural competence to language learning students. Research has documented how AI struggles to resolve cultural complexities during language educational processes

when cultural and language elements are densely linked (Silverstein, 2004). Teaching true proficiency requires full understanding of cultural nuances because Saudi language frequently follows both cultural and religious directives. The effectiveness of AI tools for vocabulary instruction and grammar teaching is diminished by their lack of capability to deliver cultural context which smooths the learning process.

Students demonstrated different levels of contentment with AI tools depending on their personal engagement with the systems along with their technical complications. The students reported differing experiences with AI tools through positive assessments of their ease of access while some students struggled with the technological challenges especially how effectively the tools recognized speech. A student noted his dissatisfaction after the AI handling system failed to properly understand his speech pattern in answering questions because it misidentified his accent. Recent studies show that AI speech recognition systems frequently encounter difficulties understanding various accents in multicultural settings particularly Saudi Arabia (El Zahraa, 2025). Serious problems developed due to the technology's limitations when processing speech from students who spoke with distinct regional accents or special speech patterns. This highlights a significant limitation of AI tools in language learning. Modern speech recognition systems face difficulties in understanding diverse linguistic patterns which requires attention from future Artificial intelligence (AI) developers.

Research investigated how AI functions as an additional tool for language education after detecting these results. The positive learning outcomes from the study transformed vocabulary knowledge and listening skills yet students underlined the necessity for personal human instruction during their education. One participant student observed that AI tools help yet requires teacher involvement to help understand complex concepts that students need clarification on. According to Bartolomé et al. (2018), technology performs better than human teachers at enhancing educational outcomes yet it lacks their capability to supply personalized emotional support. Language education transcends vocabulary development and grammatical structure learning because it is fundamentally a social experience where emotional connection and motivation alongside meaning interpretation call for personal interaction between learners (Pintrich, 2003). Human instructors create dynamic learning environments which AI tools fail to duplicate even though they bring worthwhile reinforcement to language skill development. Study results demonstrate the importance of AI integration within a model which merges traditional teaching approaches with automated systems.

Research data showed that the success rate of AI tools in language learning depends on two factors: student drive to learn and their existing skills using technology. The research showed that participants benefiting most from AI tools were students who had trust in technology combined with elevated motivation levels. Research by Naamati-Schneider & Alt (2024), indicated students who demonstrate strong digital literacy receive better outcomes using AI-powered educational tools and this study supports these earlier findings. The AI tools worked better for students who demonstrated skilled use of technology yet students who lacked comfort with technology found navigating these tools quite challenging. The success of AI in language learning depends on combined approaches which account for students' technological expertise and deliver sufficient instructional and support methods. Current conditions challenge educational institutions in Saudi Arabia to establish strategies that will eliminate digital accessibility gaps for AI-powered tool utilization among all student populations.

Participants perceived AI tools to be performance enhancing accessories that especially benefited students who did poorly with conventional language education styles. According to one student's voice "I use AI tools because they offer interactive features and game-like interfaces." Research indicates how AI tools should transform language education through improved interactive approaches that combat student boredom in standard approaches reported in the study by Tai & Chen (2024). Artificial intelligence (AI) tools generate personalized learning experiences that maintain student participation and their willingness to stay motivated in the process. AI-powered tools demonstrate motivational qualities yet their designers need to create frameworks that support sustained student participation and combat technology dependency. Thus, it is needed to use motivational aspects of AI while protecting students from automation-dependence which leads them to lose critical thinking skills.

7. Conclusion and Recommendations

This research demonstrates that AI tools have the potential to substantially improve foreign language (FL) abilities among Saudi Arabian students especially vocabulary and listening skills. The results of the study substantiate research in language education which demonstrates that technology-mediated personalized learning enhances specific language competencies. However, significant restrictions on the use of Artificial Intelligence (AI) in teaching language need to be placed and combined with human educational strategies as these tools still fall short of delivering the interactive and culturally adaptive aspects of traditional classroom instruction.

Technology development requires additional advancements to adapt speech recognition software for diverse accents and regional specifications mainly in Saudi Arabia's bilingual educational setting. The system's boundaries test the boundaries of AI capabilities for language acquisition by basic educational techniques where human contact with culture remains essential.

The study proves that AI technology successfully plays an important function by driving language learning through increased learner engagement and motivation. Students' positive reactions towards AI tools stem from the interactive and personalized features which indicate smart technology implementation offers complementary powerful learning solutions. Educational practitioners need to maintain the right mix of AI systems and human-loaded teaching approaches to build a complete and efficient linguistic education territory.

AI researchers need to concentrate on developing their technologies which effectively support dialogue capabilities and cultural awareness. Additional research combining AI with conventional teaching practices will give us deeper insight into how we can optimize AI systems in language education environments.

8. Limitations of the Study

Despite its contributions to language education research, this study faces several limitations that need acknowledgement. Results from this study only apply to university foreign language (FL) students because the research exclusively focused on this specific population of learners. A study duration of six months proved insufficient to evaluate potential long-term effects of AI tools upon language proficiency development. The complex nature of language acquisition together with its gradual learning trajectory makes it plausible that this research study measures only short-term improvements instead of long-term development. This research only investigated a particular selection of AI tools so the results cannot be generalized to all AI-based language learning technology platforms. The self-reported data from surveys and interviews contained possible biases because participants could have let their personal views about AI tools or study outcome expectations impact their responses. The analysis lacked control methods to eliminate outside variables including students' prior language knowledge or unique learning styles that potentially shaped the research outcomes.

References

Abahussain, M. O. (2016). Implementing communicative language teaching method in Saudi Arabia: Challenges faced by formative year teachers in state schools.

Aldossary, H. S. M. (2023). *Translation Technology and CAT Tools: Addressing Gaps between Pedagogy and the Translation Industry in Saudi Arabia* (Doctoral dissertation, University of Leeds).

Aljabri, N., & Bhutoria, A. (2020). Access to educational technology and its implications on learning outcomes of 15-year-olds in Saudi Arabia empirical evidence from OECD PISA 2018 in the context of COVID-19. *Journal of Educational Science*, *6*(1), 367-400.

Almansour, A. (2019). *An Investigation into The Use of The Smartphone Application' Memrise'in Supporting English Vocabulary Learning among Undergraduate Students in Saudi Arabia*. Liverpool John Moores University (United Kingdom).

Almuhanna, M. A. (2024). Teachers' perspectives of integrating AI-powered technologies in K-12 education for creating customized learning materials and resources. *Education and Information Technologies*, 1-29. https://doi.org/10.1007/s10639-024-13257-y

Al-Nasser, A. S. (2015). Problems of English language acquisition in Saudi Arabia: An exploratory-cum-remedial study. *Theory and Practice in Language Studies*, *5*(8), 1612. http://dx.doi.org/10.17507/tpls.0508.10

Alotaibi, N. S., & Alshehri, A. H. (2023). Prospers and obstacles in using artificial intelligence in Saudi Arabia higher education institutions—The potential of AI-based learning outcomes. *Sustainability*, *15*(13), 10723. https://doi.org/10.3390/su151310723

Bartolomé, A., Castañeda, L., & Adell, J. (2018). Personalisation in educational technology: the absence of underlying pedagogies. *International Journal of Educational Technology in Higher Education*, *15*, 1-17. http://dx.doi.org/10.1186/s41239-018-0095-0

Bernard, R. M., Borokhovski, E., Schmid, R. F., Waddington, D. I., & Pickup, D. I. (2019). Twenty-first century adaptive teaching and individualized learning operationalized as specific blends of student-centered instructional events: A systematic review and meta-analysis. *Campbell Systematic Reviews*, *15*(1-2). https://doi.org/10.1002/cl2.1017

Betaubun, M., Rokhmah, D. E. L., & Budiasto, J. (2023). Personalized Pathways to Proficiency: Exploring the Synergy of Adaptive Learning and Artificial Intelligence in English Language Learning.

Blin, F., & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance to change through the lens of activity theory. *Computers & Education*, *50*(2), 475-490. https://doi.org/10.1016/j.compedu.2007.09.017

Chng, E., Tan, A. L., & Tan, S. C. (2023). Examining the use of emerging technologies in schools: A review of artificial intelligence and immersive technologies in STEM education. *Journal for STEM Education Research*, *6*(3), 385-407.

Cope, B., Kalantzis, M., & Searsmith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational philosophy and theory*, *53*(12), 1229-1245. https://doi.org/10.1080/00131857.2020.1728732

Coșulean, L. (2024). Adaptive learning platforms as a means of individualised learning. *In Probleme de filologie: aspecte teoretice şi practice* (pp. 136-141).

Dennis, N. K. (2024). Using AI-Powered Speech Recognition Technology to Improve English Pronunciation and Speaking Skills. *IAFOR Journal of Education*, *12*(2), 107-126. https://doi.org/10.22492/ije.12.2.05

Dhouib, A., Othman, A., El Ghoul, O., Khribi, M. K., & Al Sinani, A. (2022). Arabic automatic speech recognition: a systematic literature review. *Applied Sciences*, *12*(17), 8898. https://doi.org/10.3390/app12178898

Doghan, M. A. A. (2020). Employees work stress in Saudi Telecom Sector: Examining the role of long working hours, empowerment, work life balance and work environment. *Journal of Economics, Management and Trade*, *26*(2), 81-91. http://dx.doi.org/10.9734/jemt/2020/v26i230230

Dong, W., Pan, D., & Kim, S. (2024). Exploring the integration of IoT and Generative AI in English language education: Smart tools for personalized learning experiences. *Journal of Computational Science*, *82*, 102397. https://doi.org/10.1016/j.jocs.2024.102397

El Zahraa, F. (2025). Leveraging Artificial Intelligence and Digital Technologies to Enhance Sociolinguistic Competence and Arabic Language Skills. In *Proceeding of the International Conference on Religious Education and Cross-Cultural Understanding* (Vol. 1, No. 1, pp. 33-49).

Farr, C. (2024). *Unmasking ChatGPT: The Challenges of Using Artificial Intelligence for Learning Vocabulary in English as an Additional Language* (Doctoral dissertation).

George, A. S. (2023). Preparing students for an AI-driven world: Rethinking curriculum and pedagogy in the age of artificial intelligence. *Partners Universal Innovative Research Publication*, *1*(2), 112-136. http://dx.doi.org/10.5281/z nodo.102 45675

Hassan, M. A., & Fook, F. S. (2014). E-learning modules supported by cooperative learning: Impact on Arabic language achievement among Qatar University students. *Near and Middle Eastern*

Journal of Research in Education, *2012*(1), 1. https://doi.org/10.5339/nmejre.2012.1

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36-53. http://dx.doi.org/10.1016/j.compedu.2015.09.005

Ibrahim, A., & Barnawi, O. Z. (Eds.). (2022). *The past, present, and future of higher education in the Arabian Gulf region: critical comparative perspectives in a Neoliberal Era*. Taylor & Francis.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., ... & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Li, P., & Lan, Y. J. (2022). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition*, *25*(3), 361-378. http://dx.doi.org/10.1017/S1366728921000353

Liu, W., & Wang, Y. (2024). The effects of using AI tools on critical thinking in English literature classes among EFL learners: An intervention study. *European Journal of Education*, *59*(4), e12804. https://doi.org/10.1111/ejed.12804

Naamati-Schneider, L., & Alt, D. (2024). Beyond digital literacy: The era of AI-powered assistants and evolving user skills. *Education and Information Technologies*, 1-31. http://dx.doi.org/10.1007/s10639-024-12694-z

Newton, J. M., & Nation, I. S. (2020). *Teaching ESL/EFL listening and speaking*. Routledge.

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of educational Psychology*, *95*(4), 667. https://psycnet.apa.org/doi/10.1037/0022-0663.95.4.667

Poelling, R. K. (1971). *A Developmental Study of Abraham H. Maslow's Self-actualization Theory*. Northern Illinois University.

Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, *32*(9), 5020-5040. https://doi.org/10.1080/10494820.2023.2208170

Richards, J. C. (2005). *Communicative language teaching today*. Singapore: SEAMEO Regional Language Centre.

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information*, *15*(10), 596. https://doi.org/10.3390/info15100596

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on individual characteristics and foreign language education*, *6*(27), 27-49. https://doi.org/10.1515/9781614510932

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL quarterly*, *41*(2), 255-283. http://dx.doi.org/10.1002/j.1545-7249.2007.tb00059.x

Silverstein, M. (2004). "Cultural" concepts and the language-culture nexus. *Current anthropology*, *45*(5), 621-652. http://dx.doi.org/10.1086/423971

Song, C., Shin, S. Y., & Shin, K. S. (2023). Optimizing foreign language learning in virtual reality: a comprehensive theoretical framework based on constructivism and cognitive load theory (VR-CCL). *Applied Sciences*, *13*(23), 12557. http://dx.doi.org/10.20944/preprints202309.2101.v1

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2024). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*. https://doi.org/10.1002/sd.3221

Taguchi, N. (2008). The role of learning environment in the development of pragmatic comprehension: A comparison of gains between EFL and ESL learners. *Studies in second language acquisition*, *30*(4), 423-452. https://doi.org/10.1017/S0272263108080716

Tai, T. Y., & Chen, H. H. J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, *220*, 105112. https://doi.org/10.1016/j.compedu.2024.105112

Tomlinson, B. (2012). Materials development for language learning and teaching. *Language teaching*, *45*(2), 143-179. https://doi.org/10.1017/S0261444811000528

Xu, Z. (2024). AI in education: Enhancing learning experiences and student outcomes. *Applied and Computational Engineering*, *51*(1), 104-111. http://dx.doi.org/10.54254/2755-2721/51/20241187

Yamijala, S. M. S., Chodisetty, R. M., Chakravorty, C., & Sai, K. P. (2025). AI-Powered Learning Revolutionizing Smart Education With Personalized Learning Styles. In *Internet of Behavior-Based Computational Intelligence for Smart Education Systems* (pp.

191-212). IGI Global. http://dx.doi.org/10.4018/979-8-3693-8151-9.ch007

Yang, J. C., Chen, C. H., & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers & Education*, *55*(3), 1346-1356. https://doi.org/10.1016/j.compedu.2010.06.005

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, *4*, 100134. https://doi.org/10.1016/j.caeai.2023.100134

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 163-170

Efficacy of AI-Powered Adaptive Learning Systems in Enhancing Grammar Acquisition for Arab Learners of English: An Empirical Study

Javed Ahmad*
Mohsin Razza Khan**
Basim Hamdan Ibrahim ***
Kana'an****

Abstract

The challenge of mastering English grammar, particularly the verb tense system, is a well-documented hurdle for Arab learners. This study investigates the efficacy of an AI-powered adaptive learning platform in improving grammar accuracy compared to traditional teacher-led instruction. A quasi-experimental design was employed with 60 intermediate-level Arab ESL learners, divided into an experimental group (using an AI platform) and a control group (receiving standard instruction). Pre-tests and post-tests assessed proficiency in five key grammatical structures: simple present, present continuous, simple past, present perfect, and future tenses. Results, analyzed using a paired samples t-test, indicated a statistically significant improvement ($p < 0.01$) in the experimental group's post-test scores, with a mean increase of 22.5 points compared to 10.3 points in the control group. Furthermore, data from the AI platform's analytics revealed a correlation between the number of adaptive review cycles and final test performance. The study concludes that AI-driven, personalized learning pathways offer a potent method for addressing the persistent grammatical difficulties faced by Arab learners, recommending the integration of such tools as a supplement to traditional classroom teaching.

Keywords: AI in Education, ESL for Arab Learners, Adaptive Learning, Grammar Acquisition, Educational Technology, Empirical Study, Second Language Acquisition

1. Introduction

The global dominance of English has made its mastery a critical skill for academic and professional success in the Arab world. However,

* King Khalid University, KSA; jawed@kku.edu.sa
** King Khalid University, KSA; ncoka@kku.edu.sa
*** King Khalid University, KSA; mokhan@kku.edu.sa
**** King Khalid University, KSA; bkanan@kku.edu.sa

the process of teaching and learning English in Arab countries is fraught with challenges, many stemming from the profound linguistic differences between Arabic and English (Al-Jarf, 2021). These differences are particularly acute in the realm of grammar, where concepts such as verb conjugation, tense consistency, and the use of auxiliaries present significant hurdles. Arabic, a Semitic language, operates on a root-and-pattern system with a fundamentally different approach to expressing time and aspect, leading to frequent and systematic errors in English production (Kharma & Hajjaj, 1997; Scott & Tucker, 2019).

Traditional pedagogical methods, often reliant on rote memorization and teacher-centered instruction, have shown limited success in overcoming these deeply ingrained interferences. The one-size-fits-all approach fails to address the individual learning pace and specific error patterns of each student (Ellis, 2008). In recent years, Artificial Intelligence (AI) has emerged as a transformative force in education, promising personalized and adaptive learning experiences (Luckin et al., 2016). AI-powered platforms can diagnose a learner's weaknesses in real-time, provide immediate corrective feedback, and dynamically adjust the difficulty and focus of subsequent exercises.

This study posits that such AI-driven personalization is uniquely suited to tackle the grammatical challenges faced by Arab learners. While theoretical discussions on AI in education are abundant, there is a scarcity of empirical research measuring its direct impact on specific learner populations with distinct linguistic backgrounds (Zawacki-Richter et al., 2019). This paper aims to fill that gap by addressing the following research questions:

1. Is there a statistically significant difference in grammar acquisition between Arab learners using an AI-powered adaptive learning system and those receiving only traditional instruction?
2. What is the relationship between learner engagement with the AI system (measured by completed exercises and review cycles) and improvement in grammatical accuracy?

2. Literature Review

2.1. Grammatical Challenges for Arab Learners

The grammatical divergence between Arabic and English is a primary source of difficulty. Key areas of struggle identified in the literature include:

- **Verb Tenses:** The Arabic tense system is simpler, often distinguished primarily between past and non-past. This leads to confusion with English compound tenses like the present perfect (e.g., "I have eaten" vs. "I ate") and the misuse of continuous forms (e.g., "I am understanding the lesson") (Mourtaga, 2004; Alzu'bi, 2015).

- **Subject-Verb Agreement:** The concept is less rigid in Arabic, resulting in errors such as "He go to school" (Alsamadani, 2017).
- **Prepositions:** The mapping of prepositions between the two languages is rarely one-to-one, causing errors like "I am good in English" instead of "I am good at English" (Ghani, 2018).

2.2. AI and Adaptive Learning in Language Education

Adaptive learning systems use algorithms to tailor educational content to the needs of individual learners. In language learning, these systems can analyse a student's responses to identify patterns of error and then serve up customized practice exercises targeting those specific weaknesses (Zawacki-Richter et al., 2019). AI enhances this by incorporating Natural Language Processing (NLP) to provide nuanced feedback on writing and speaking, moving beyond simple multiple-choice corrections (Chassignol et al., 2018).

Previous studies have shown the general effectiveness of Computer-Assisted Language Learning (CALL). However, the current generation of AI-driven platforms represents a qualitative leap from earlier, static CALL software (Bax, 2003). They offer a dynamic, responsive learning environment that can, in theory, provide the intensive, personalized practice needed to restructure a learner's interlanguage grammar, a concept central to second language acquisition theory (Ellis, 2008). The principles of spaced repetition and retrieval practice, which are strongly supported by cognitive science (Kang, 2016), can be seamlessly embedded and automated within these AI systems, making them ideal for mastering difficult grammatical structures.

3. Methodology

3.1. Research Design

A quasi-experimental pre-test/post-test control group design was employed. This design was chosen to compare the outcomes of two different instructional interventions under controlled conditions, a standard approach in educational technology research (Creswell & Creswell, 2018).

3.2. Participants

A sample of 60 intermediate-level Arab learners of English (aged 18-25) was recruited from a university foundation program. Participants were randomly assigned to one of two groups:

- **Experimental Group (n=30):** Received instruction supplemented with use of the "LinguaBot AI" adaptive learning platform for 30 minutes daily, three times a week.
- **Control Group (n=30):** Received only the standard, teacher-led grammar instruction for the same duration and frequency.

All participants had a comparable initial proficiency level, as confirmed by a placement test and the pre-study grammar test.

3.3. Instrument and Procedure

The study was conducted over an 8-week period.

- **Pre-test:** All participants completed a 50-item grammar test focusing on the five target structures: simple present, present continuous, simple past, present perfect, and future tenses. The test included multiple-choice, error identification, and sentence completion items, designed based on common error patterns outlined by Al-Jarf (2021) and Mourtaga (2004).
- **Intervention:** The control group attended traditional classes involving explicit rule explanation, group work, and textbook exercises, a methodology still prevalent in many Arab EFL contexts (Al-Mekhlafi, 2011). The experimental group attended the same classes but also used the "LinguaBot AI" platform. This platform provided initial diagnostic tests, generated personalized exercise sets, offered instant feedback with explanations, and forced review of previously missed concepts, embodying the adaptive learning principles described by Chassignol et al. (2018).
- **Post-test:** At the end of the 8 weeks, all participants took a parallel-form post-test identical in structure and difficulty to the pre-test.
- **Data Collection:** The AI platform logged user data, including the number of exercises completed and the number of "adaptive review cycles" triggered for each learner.

3.4. Data Analysis

Pre-test and post-test scores were analysed using a paired samples t-test to determine the statistical significance of the improvement within each group. An independent samples t-test was used to compare the post-test scores between the two groups. Data from the AI platform was analysed using Pearson's correlation to examine the relationship between engagement metrics and learning gains. The alpha level for statistical significance was set at $p < 0.05$.

4. Results and Discussion

4.1. Quantitative Results

The pre-test and post-test scores for both groups are summarized in Table 1.

Group	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Gain
Experimental (AI)	58.4 (5.2)	80.9 (4.8)	**+22.5**
Control (Traditional)	57.9 (5.5)	68.2 (5.1)	**+10.3**

Table 1. Pre-test and Post-test Score Comparison (out of 100)

A paired-samples t-test revealed that the improvement within the experimental group was statistically significant, $t(29) = 15.32$, $p < .001$. The improvement in the control group was also significant, $t(29) = 8.45$, $p < .001$, but the magnitude of gain was substantially smaller. An independent samples t-test confirmed that the post-test scores of the experimental group were significantly higher than those of the control group, $t(58) = 10.11$, $p < .001$.

Participant ID	Exercises Completed	Adaptive Review Cycles	Post-test Score
EXP-01	245	18	85
EXP-02	198	12	78
EXP-03	310	25	92
...	...	...	...
EXP-30	225	15	81
Correlation (r)	**0.75**	**0.82**	-

Table 2. AI Platform Engagement vs. Post-test Score (Experimental Group Only)

Analysis of the AI platform data (Table 2) showed a strong positive correlation between the number of adaptive review cycles a learner completed and their final post-test score ($r = 0.82$, $p < .001$). The correlation between the total number of exercises and the post-test score was also positive and strong ($r = 0.75$, $p < .001$).

4.2. Discussion

The results provide a clear affirmative answer to the first research question. The significantly greater improvement of the experimental group demonstrates that supplementing traditional instruction with an AI-powered adaptive learning system is more effective for grammar acquisition than traditional instruction alone. The AI system's ability to provide unlimited, personalized practice and immediate feedback appears to have accelerated the learning process and led to greater internalization of the target grammatical rules, supporting the theories of personalized learning espoused by Luckin et al. (2016).

The findings related to the second research question are equally compelling. The strong correlation between adaptive review cycles and post-test performance suggests that the system's core adaptive mechanic—revisiting and retesting problematic concepts—is a key driver of its efficacy. This aligns with the pedagogical principles of deliberate practice (Ericsson et al., 1993) and the spacing effect (Kang, 2016), which are difficult to implement consistently in a heterogeneous classroom setting. The AI system automates and personalizes this process, ensuring that learners repeatedly practice their specific weaknesses until mastery is achieved, directly addressing the fossilization of common errors noted by Al-Mekhlafi (2011).

5. Conclusion and Implications

This empirical study provides robust evidence that AI-powered adaptive learning systems are a highly effective tool for teaching English grammar to Arab learners. By offering personalized pathways, instant feedback, and forced review of challenging concepts, these systems directly address the limitations of traditional, one-size-fits-all pedagogy.

The implications for ESL instructors and curriculum designers in the Arab world are significant. AI should not be viewed as a replacement for teachers but as a powerful supplement that can free up classroom time for more communicative, interactive activities. Teachers can leverage these tools to identify common class-wide issues from the aggregated data while the system handles individualized skill-and-drill practice.

5.1. Limitations and Future Research

This study was limited to grammar acquisition and a specific intermediate-level learner profile. Future research should investigate the impact of AI on other skills like writing and speaking, explore its effectiveness with beginner and advanced learners, and conduct longitudinal studies to assess long-term retention. Furthermore, studying the integration models of AI tools within a blended learning framework would be highly valuable for practical applications.

References

Al-Jarf, Reima. "Learning English Grammar Online by Arab ESL Freshmen: Difficulties and Solutions." *Journal of English Language Teaching and Applied Linguistics*, vol. 3, no. 4, 2021, pp. 1-11.

Al-Mekhlafi, A. M. "The Difficulties of Learning English Grammar in the Arab EFL Context: A Case Study of UAE High School Students." *Journal of Education and Practice*, vol. 2, no. 4, 2011, pp. 34-52.

Alsamadani, H. A. "The Effect of Explicit Instruction of English Tenses on Arab ESL Students' Writing Accuracy." *International Journal of Education and Literacy Studies*, vol. 5, no. 3, 2017, pp. 1-7.

Alzu'bi, Mohammad Akram. "The Most Common Grammatical Errors Made by Undergraduate EFL Learners at Al-Balqa Applied University." *International Journal of English Language and Linguistics Research*, vol. 3, no. 6, 2015, pp. 12-33.

Bax, Stephen. "CALL—Past, Present, and Future." *System*, vol. 31, no. 1, 2003, pp. 13-28.

Chassignol, Maëlis, et al. "Artificial Intelligence Trends in Education: A Narrative Overview." *Procedia Computer Science*, vol. 136, 2018, pp. 16-24.

Creswell, John W., and J. David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 5th ed., SAGE Publications, 2018.

Ellis, Rod. *The Study of Second Language Acquisition*. 2nd ed., Oxford UP, 2008.

Ericsson, K. Anders, et al. "The Role of Deliberate Practice in the Acquisition of Expert Performance." *Psychological Review*, vol. 100, no. 3, 1993, pp. 363-406.

Ghani, Miloud A. "A Study of the Errors Made by Arab EFL Learners in the Use of Prepositions." *International Journal of Applied Linguistics and English Literature*, vol. 7, no. 6, 2018, pp. 188-194.

Kang, Sean H. K. "Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction." *Policy Insights from the Behavioral and Brain Sciences*, vol. 3, no. 1, 2016, pp. 12-19.

Kharma, Nayef, and Ali Hajjaj. *Errors in English among Arabic Speakers: Analysis and Remedy*. Longman, 1997.

Luckin, Rose, et al. *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education, 2016.

Mourtaga, Kamal R. "Some Problems Facing Arab Learners of English." *Journal of Al-Aqsa University*, vol. 8, 2004, pp. 75-87.

Scott, Mike, and G. Richard Tucker. *Error Analysis and English-Language Strategies of Arab Students*. ERIC, 2019.

Zawacki-Richter, Olaf, et al. "Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where are the Educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 39, 2019, pp. 1-27.

Appendix A: Sample Grammar Test Items (Pre-test/Post-test)

Section 1: Multiple Choice

Choose the correct verb form to complete the sentence.

1. She _____ to the market every day.

   (a) Go
   (b) Goes
   (c) (c) is going
   (d) (d) went

2. I _____ my homework yet.

   (a) didn't finish

   (b) haven't finished

(c) don't finish

(d) am not finishing

Section 2: Error Identification

Identify the error in the following sentence.

3. He is play football right now.

   A | B | C | D

4. They have lived in Dubai since five years.

   A | B | C | D

Section 3: Sentence Completion

Complete the sentence with the correct form of the verb in parentheses.

5. Last year, we _____ (travel) to Egypt.
6. Look! It _____ (rain) outside.

Appendix B: Sample Data Log from "LinguaBot AI" Platform

Timestamp	User ID	Exercise Type	Target Grammar	User Answer	Correct Answer	Adaptive Review Triggered?
2023-10-26 14:05:12	EXP-15	Fill-in-the-blank	Present Perfect	I have saw	I have seen	Yes
2023-10-26 14:07:30	EXP-15	Multiple Choice	Present Perfect	...	...	No
2023-10-27 09:15:01	EXP-15	Review Exercise	Present Perfect	I have seen	I have seen	No (Concept Mastered)

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 171-184

Negotiating Intimacy and Distance: Pronominal Variation across Kinship Domains in Kashmiri

Javaid Aziz Bhat[*]

Abstract

This study examines how Kashmiri speakers negotiate intimacy and social distance through pronominal variation across consanguineal (blood-based) and affinal (marriage-based) kinship domains. Drawing on data from 240 speakers representing diverse age, gender, education, and settlement backgrounds, the research combines quantitative frequency analysis with qualitative interpretation grounded in sociolinguistic and pragmatic theory. The analysis reveals a systematic contrast between non-deferential pronouns (tsI, yi, su/so) and deferential forms (tohj, yim, tim), demonstrating that pronoun choice serves as a key linguistic strategy for managing social relationships. Consanguineal interactions overwhelmingly favor non-deferential forms (85–90%), indexing solidarity and emotional proximity, whereas affinal interactions show increased deference (40–45%), reflecting social diplomacy and respect. Social variables further condition usage: younger and rural speakers prefer intimate forms, while middle-aged, educated, and urban speakers employ more deferential variants. The findings substantiate Brown and Gilman's (1960) power–solidarity model and Brown and Levinson's (1987) politeness framework, but also extend them by incorporating recent theories of indexicality and relational work (Ochs, 1996; Locher & Watts, 2005; Bucholtz & Hall, 2005). Pronominal choice in Kashmiri thus emerges as an indexical practice through which speakers construct, rather than merely reflect, kinship hierarchies and affective ties. The study contributes to the broader understanding of South Asian address systems and highlights how linguistic politeness functions as an adaptive mechanism in the negotiation of respect and intimacy within kinship networks.

Keywords: Kashmiri, Pronominal, Kinship, Social Variable, Indexicality

1. Introduction

Language is one of the most salient instruments through which social relationships are expressed, maintained, and negotiated. Among the linguistic resources available to speakers, pronominal choice plays a particularly crucial role in indexing interpersonal alignment, authority, and emotional proximity. Across languages, pronouns serve not only as grammatical markers of person and number but also as potent symbols

[*] Higher Education Department, J&K; javaidazizbhat@gmail.com

of social meaning. The alternation between deferential and non-deferential forms is often a subtle yet powerful mechanism for managing the balance between intimacy and respect, especially within kinship systems where social hierarchy and affective ties intersect most visibly.

The sociolinguistic study of pronouns gained early theoretical grounding through Brown and Gilman's (1960) seminal work on the power–solidarity model, which linked second-person pronoun choice to social relations of dominance and intimacy. Later, Brown and Levinson's (1987) theory of politeness expanded this insight by demonstrating how linguistic strategies mitigate threats to face, making pronoun selection a form of pragmatic negotiation. These foundational theories have inspired extensive cross-linguistic research on address forms in European and Asian languages, including studies on Hindi-Urdu, Tamil, Malayalam, and Japanese, each showing how pronouns encode complex social hierarchies.

In the South Asian context, kinship remains a primary locus of linguistic politeness. In Kashmiri, as in many Indo-Aryan languages, pronouns are not interchangeable grammatical options but socially charged expressions that reflect and construct social distance. The language features a rich system of pronominal alternation: *tsI*, *yi*, and *su/so* are generally used in familiar or equal relationships, whereas *tohj*, *yim*, and *tim* serve as deferential or polite forms. However, unlike many Indo-Aryan languages where hierarchical relationships are unidirectionally marked, Kashmiri shows reciprocal deference in certain affinal (marriage-based) relations, suggesting a more symmetrical negotiation of respect.

Person	**Number**	**Form(s)**	**Social Function / Usage**	**Typical Domain of Use**
2nd person	Singular (non-deferential)	*tsI*	Familiar or intimate address; expresses equality and solidarity	Consanguineal relations, peers, children
	Plural (deferential)	*tohj*	Polite or respectful address; signals social distance or deference	Affinal relations, elders, formal contexts
3rd person	Singular (non-deferential)	*yi* (proximate), *su* / *so* (remote)	Informal reference to equals or familiars	Consanguineal and peer contexts
	Plural (deferential)	*yim* (proximate), *tim* (remote)	Honorific or respectful reference; indexes formality	Affinal, hierarchical, or public interactions

Table 1. Kashmiri pronominal structure by person, number, and social meaning

***Note**: In Kashmiri, plural forms (tohj, yim, tim) function as honorific or deferential variants, while singular forms (tsI, yi, su/so) index intimacy and solidarity. This contrast parallels the classic T/V distinction (Brown & Gilman 1960) observed in many other languages.*

Despite these intriguing dynamics, Kashmiri has received limited systematic attention in sociolinguistics. Previous descriptive studies

have documented the structure of Kashmiri pronouns but have not explored their variation across kinship domains or the influence of social variables such as age, gender, education, and settlement type. Similarly, while the pragmatics of politeness in South Asian languages has been widely studied, little is known about how these principles manifest in everyday kin interactions in Kashmir. This gap motivates the present study.

The current research addresses this lacuna by analyzing how Kashmiri speakers use pronominal variation to negotiate intimacy and social distance across consanguineal (blood-based) and affinal (marriage-based) kinship domains. Using data from 240 speakers representing diverse demographic backgrounds, the study combines quantitative and qualitative approaches to uncover systematic patterns of variation. Quantitative frequency counts capture broad social trends, while ethnographic observation reveals the pragmatic nuances underlying pronoun choice.

The study is grounded in the frameworks of Brown and Gilman's (1960) power–solidarity model and Brown and Levinson's (1987) politeness theory but extends them through recent developments in sociolinguistic pragmatics. Following Ochs (1996), Locher and Watts (2005), and Bucholtz and Hall (2005), the analysis views pronominal variation not as a static reflection of social status but as a form of relational work an indexical practice through which speakers continuously construct social relationships. This perspective enables a dynamic interpretation of politeness as an interactional stance, sensitive to local cultural norms and situational context.

By investigating Kashmiri kinship address patterns through this integrated framework, the study contributes both empirically and theoretically: empirically, by providing the first quantitative account of pronominal variation in Kashmiri kin relations; and theoretically, by demonstrating how politeness and solidarity are co-constructed through linguistic choice. Ultimately, the paper shows that in Kashmiri, pronouns are not simply grammatical forms but social acts that sustain the delicate balance between respect and intimacy within family life.

2. Literature Review

The study of pronominal variation has long occupied a central position in sociolinguistics and linguistic anthropology, offering insights into how language both reflects and constructs social organization. Early work in European linguistics, particularly Brown and Gilman's (1960) classic study of the *T/V distinction* in Indo-European languages, established the foundation for understanding the relationship between pronoun choice and social meaning. They proposed that second-person pronouns encode two intersecting dimensions of social relation: *power* (asymmetrical relationships of dominance and subordination) and *solidarity* (symmetrical relationships of intimacy and equality). According to their model, the use of the familiar "T" form signals

solidarity or intimacy, whereas the polite "V" form indexes distance and respect.

Following Brown and Gilman, researchers expanded the investigation of address and politeness systems to a wider range of languages and cultural contexts. Brown and Levinson's (1987) influential theory of *politeness* provided a pragmatic perspective on linguistic behavior, explaining how speakers employ linguistic strategies such as pronoun selection, honorifics, and indirectness to mitigate threats to face. Their framework distinguished between positive politeness (solidarity and involvement) and negative politeness (deference and respect), allowing for a more nuanced interpretation of pronoun choice as a form of *face management*. Later sociolinguistic research further demonstrated that such patterns are socially stratified and contextually variable, aligning with Labov's (1966) principle that variation in speech systematically correlates with social factors such as age, gender, education, and socioeconomic class.

Across many languages, address and pronominal systems have been shown to function as crucial markers of social alignment. Braun's (1988) comprehensive cross-linguistic study of address forms confirmed that pronouns serve not only as referential tools but as indexes of interpersonal stance. Similarly, Fasold (1990) and Wardhaugh (2006) emphasized that pronominal variation provides an observable site for examining the intersection of *linguistic choice* and *social structure*. Subsequent research in socio-pragmatics such as Ervin-Tripp (1972) and Scollon & Scollon (1995) demonstrated that address forms are not merely determined by social hierarchy but are dynamically negotiated within particular interactions.

In the South Asian context, languages such as Hindi-Urdu, Tamil, Malayalam, Punjabi, and Bengali have been extensively studied for their multi-tiered pronominal systems. These languages often feature tripartite distinctions among informal, polite, and honorific forms (e.g., *tu/tum/aap* in Hindi; *nee/neenga* in Tamil), which index both social status and emotional distance (Chandrasekhar, 1970; Shapiro & Schiffman, 1981). Scholars have shown that pronoun choice in these languages is deeply influenced by kinship norms, caste hierarchy, gender roles, and regional variation (Pandharipande, 1992; Nayar, 1997). Despite these parallels, Kashmiri remains underrepresented in comparative discussions, even though it displays equally rich pronominal complexity and social nuance.

Within Kashmiri linguistics, previous studies have focused primarily on phonological, morphological, and syntactic aspects (Kachru, 1969; Koul, 1985), with little systematic attention to variation in pronoun use. Bhat (2017, 2018) provided preliminary analyses of pronominal alternation within consanguineal and affinal kinship contexts, revealing that speakers draw on both deferential and non-deferential forms to navigate familial relationships. These studies point toward the importance of kinship as a sociolinguistic domain, where linguistic politeness and social alignment are most vividly performed.

However, they stop short of offering a comprehensive analysis that integrates both *quantitative variation* and *interactional pragmatics* within a unified framework. The present study builds on this foundation by examining how pronouns function as dynamic social signals across kinship networks in Kashmiri society.

Taken together, the literature establishes two important points: first, that pronouns are sensitive indicators of social distance and intimacy; and second, that kinship offers a crucial arena for studying the interplay of power, solidarity, and politeness. Yet, most existing models treat pronoun choice as a relatively stable reflection of social structure. Recent developments in sociolinguistic and pragmatic theory, however, move beyond this static view to emphasize the *indexical* and *interactional* nature of linguistic politeness. Scholars such as Ochs (1996), Locher and Watts (2005), and Bucholtz and Hall (2005) argue that pronouns and address forms do not merely mirror fixed social hierarchies but *actively index* social stances and relational identities in context. From this perspective, pronominal variation constitutes a form of relational work a continuous process through which speakers co-construct and calibrate social relationships. The reciprocal use of deferential forms among affinal kin, for instance, can be interpreted not as a simple reflection of hierarchical respect but as an *interactional stance of mutual diplomacy*. Conversely, the pervasive use of non-deferential forms among consanguineal kin reflects a stance of solidarity and shared identity.

By situating the analysis of Kashmiri pronouns within this interactional and indexical framework, the present study extends the classic power–solidarity model and contributes to contemporary understandings of linguistic politeness as a socially negotiated practice. This theoretical shift allows for a more nuanced interpretation of how speakers balance *respect and familiarity* through pronominal choice, and how linguistic politeness operates as a form of social agency in everyday kinship communication.

3. Methodology

This study adopts a mixed-methods sociolinguistic approach that combines quantitative frequency analysis with qualitative ethnographic observation to examine pronominal variation across kinship domains in Kashmiri. The goal was to identify how social variables influence the distribution of deferential and non-deferential pronouns and to interpret these patterns within the broader frameworks of politeness and relational work.

3.1 Participants

The data were collected from 240 native Kashmiri speakers, balanced across major sociolinguistic variables: age, gender, education, and settlement type. The sample was stratified into three age groups—AG1 (6–25 years), AG2 (26–45 years), and AG3 (46+ years)—to capture both generational change and age-grading effects. Gender was evenly

represented (120 male and 120 female participants), while education was categorized into educated (secondary level and above) and uneducated (no formal schooling). Settlement type distinguished urban speakers (Srinagar) from rural speakers (from Pulwama, Shopian, and Kupwara districts). All participants were lifelong residents of the Kashmir Valley and reported Kashmiri as their primary home language.

This stratified design ensured coverage of both traditional and modern sociocultural contexts, enabling a robust analysis of linguistic variation. The inclusion of multiple social factors reflects Labov's (1966) multidimensional model of variation, where linguistic forms are systematically conditioned by the speaker's social characteristics.

3.2 Data Collection Procedures

Data were collected through two complementary methods:

1. Structured sociolinguistic questionnaires and

2. Participant observation in natural conversational settings.

The questionnaire elicited pronominal choices in controlled kinship scenarios. Participants were presented with 20 short dialogues representing interactions with specific kin categories—e.g., parents, siblings, uncles/aunts, in-laws, cousins—and asked to indicate the form they would naturally use in each situation. This method allowed for consistent elicitation across respondents and ensured comparability of responses.

Participant observation was conducted during family gatherings, weddings, and informal visits, contexts where kin interaction occurs spontaneously. Naturalistic conversations were audio-recorded with consent, yielding approximately 8,000 tokens of pronoun use across the two kinship domains. Field notes supplemented recordings, capturing contextual information such as setting, tone, and emotional tenor of interaction.

3.3 Data Coding and Classification

The collected data were coded according to two primary dimensions:

1. Kinship Domain
    - Consanguineal (blood-based): relations by birth, such as parents, siblings, uncles, aunts, and cousins.
    - Affinal (marriage-based): relations by marriage, such as spouse, in-laws, and extended marital kin.
2. Pronominal Form
    - Non-deferential forms (*tsI*, *yi*, *su/so*) – marking familiarity, intimacy, and equality.
    - Deferential forms (*tohj*, *yim*, *tim*) – marking politeness, respect, or social distance.

Each occurrence of a pronoun was categorized based on both context and social variable. Ambiguous or non-pronominal references were excluded. Tokens were tabulated to generate frequency distributions by kinship type and social variable (age, gender, education, and settlement).

4. Analytical Framework

The quantitative component involved descriptive statistical analysis of pronoun frequencies and proportions within and across kinship domains. Data were summarized in five key tables (Tables 2–6) showing distributions by domain and social variable. While inferential statistics were not the focus, relative frequencies and percentage contrasts were used to identify salient trends.

The qualitative component employed discourse analysis informed by Brown and Levinson's (1987) politeness theory and Locher and Watts' (2005) notion of *relational work*. Transcribed interactions were analyzed for patterns of reciprocity, mitigation, and stance-taking to understand how pronoun choice indexes social alignment. The integration of both components allowed for what Tannen (1984) terms a "complementary analytic lens"—where numerical trends illuminate social tendencies, and ethnographic detail explains pragmatic motivation.

5. Reliability and Validity

Triangulation between elicited and observed data enhanced the validity of findings. Questionnaire results provided controlled comparability, while natural data ensured ecological authenticity. Coding reliability was tested by double-coding a random 10% subset of data, yielding an inter-coder agreement rate of 94%, which meets accepted standards for sociolinguistic analysis (Milroy & Gordon, 2003). The balance of demographic variables and inclusion of both rural and urban speakers further strengthened the representativeness of the dataset. The methodological design thus integrates variationist and pragmatic approaches to capture both the structural regularities and interactional subtleties of pronominal choice. Quantitative evidence highlights how pronoun usage correlates with social variables, while qualitative insights reveal how speakers deploy these forms as strategic resources for negotiating kinship identities. This comprehensive design provides the empirical foundation for the subsequent analysis of how intimacy and deference are linguistically constructed in Kashmiri.

6. Results and Discussion

This section presents the findings of the study and interprets them within the frameworks of power, solidarity, politeness, and relational work. The analysis proceeds in two parts. The first compares overall pronominal usage across consanguineal and affinal kinship domains. The second examines how four key social variables age, gender, education, and settlement condition variation in pronoun choice. Together, these findings reveal how Kashmiri speakers use linguistic

forms to negotiate the fine balance between intimacy and deference, thereby constructing kinship as both a social and pragmatic category.

6.1 Overall Pronominal Usage across Kinship Domains

At the broadest level, Kashmiri speakers distinguish between two social orientations in pronoun use: solidarity in consanguineal relations and deference in affinal relations. The distribution of non-deferential (*tsI, yi, su/so*) and deferential (*tohj, yim, tim*) forms is summarized in Table 2.

Domain	Non-Deferential Form (% tsI / yi / su/so)	Deferential Form (% tohj / yim / tim)	Interpretation
Consanguineal Kinship	88%	12%	High solidarity and physical proximity among blood relatives lead to predominant non-deferential forms.
Affinal Kinship	58%	42%	Greater social distance and respect-based hierarchy create higher use of deferential forms.

Table 2. Overall Pronominal Usage across Kinship Domains

Non-deferential forms dominate in consanguineal contexts, reflecting intimacy, equality, and shared identity, while affinal relations exhibit a greater proportion of deferential forms that signal politeness and respect. Importantly, deferential usage in affinal contexts is reciprocal rather than strictly hierarchical a marked difference from the unidirectional politeness observed in Hindi or Tamil. This reciprocity demonstrates what Locher and Watts (2005) term *relational work* a cooperative management of social distance through linguistic choices.

6.2 Influence of Social Variables

To understand how pronominal variation is shaped by social structure, the following subsections analyze each variable in turn.

Age strongly correlates with patterns of pronominal choice, producing a distinctive U-shaped curve across the lifespan (Table 3).

Age Group	Consanguineal (tsI/tohj %)	Affinal (tsI/tohj %)	Interpretation
AG1 (6–25 years)	87 / 13	55 / 45	Younger speakers use more non-deferential forms overall but show heightened politeness toward affinal kin.
AG2 (26–45 years)	81 / 19	43 / 57	Middle-aged speakers are most deferential, reflecting peak conformity to social norms.
AG3 (46+ years)	85 / 15	65 / 35	Older speakers revert to intimate forms, signaling reduced social distance and familial authority.

Table 3. Age-Based Variation across Kinship Domains

This age pattern reflects age grading rather than generational change. Young speakers display linguistic solidarity within their peer and family networks but express deference to elders and in-laws. Middle-aged speakers especially those engaged in formal work or social institutions adhere most strongly to polite norms. Older speakers, occupying positions of authority, can afford to use more familiar forms without breaching decorum. This cyclical pattern supports Labov's (1966) finding that social evaluation of linguistic variants fluctuates across life stages.

Gender plays a decisive role in shaping pronominal behavior. As shown in Table 4, women use deferential forms more frequently than men in both kinship domains.

Gender	Consanguineal (tsI/tohj %)	Affinal (tsI/tohj %)	Interpretation
Male	86 / 14	59 / 41	Males favor non-deferential forms, aligning with informal authority.
Female	82 / 18	49 / 51	Females use more deferential forms, reflecting prestige consciousness and relational sensitivity.

Table 4. Gender-Based Variation

This gendered asymmetry mirrors global sociolinguistic trends (Trudgill, 1983; Holmes, 1995), wherein women tend to use prestige or polite variants more often than men. In Kashmiri culture, this pattern aligns with expectations of verbal decorum and relational modesty in female speech, particularly in affinal relations where women navigate respect-based hierarchies through linguistic politeness. Women's higher use of *tohj* and *yim* forms thus indexes social awareness and empathy, confirming the role of gender as both a sociolinguistic and cultural variable.

Education exerts a subtle but consistent influence on pronoun choice (Table 5). Educated speakers show greater preference for deferential forms across both kinship types.

Education Level	Consanguineal (tsI/tohj %)	Affinal (tsI/tohj %)	Interpretation
Educated	77 / 23	47 / 53	Educated speakers show higher politeness and awareness of social hierarchy.
Uneducated	92 / 8	60 / 40	Uneducated speakers favor solidarity-driven forms, minimizing social distance.

Table 5. Education-Based Variation

Education introduces metapragmatic awareness and understanding of how language can strategically convey refinement or respect. Educated speakers, especially those exposed to multilingual or urban environments, exhibit sensitivity to formality norms. This aligns with Brown and Levinson's (1987) concept of *negative politeness*, wherein speakers consciously mitigate potential impositions through linguistic respect.

Place of residence also influences variation, with urban speakers displaying a higher tendency toward polite forms than rural speakers (Table 6).

Residence	Consanguineal (tsI/tohj %)	Affinal (tsI/tohj %)	Interpretation
Rural	87 / 13	56 / 44	Tight-knit rural networks maintain solidarity-based speech norms.
Urban	81 / 19	52 / 48	Urban speakers exhibit slightly greater deference, reflecting social sophistication and weaker kin proximity.

Table 6. Settlement-Based Variation

Urban environments, characterized by social mobility and weaker kinship density, tend to promote linguistic politeness as a generalized social strategy. Rural contexts, by contrast, favor non-deferential forms grounded in familiarity and communal belonging. This urban–rural gradient echoes Labov's (1972) findings in urban dialectology, where social anonymity correlates with increased formality.

7. Comparative Analysis: Consanguineal vs. Affinal Kinship

When viewed together, these quantitative patterns reveal two distinct but overlapping pragmatic orientations. Consanguineal kinship is dominated by solidarity and emotional proximity, where linguistic intimacy sustains affective bonds. Affinal kinship, on the other hand, foregrounds negotiated respect a space where politeness functions as symbolic diplomacy. Yet, rather than expressing rigid hierarchy, affinal relationships in Kashmiri exhibit reciprocal politeness: both interlocutors often employ deferential forms, signaling balanced social distance. This reciprocity contrasts with the unidirectional politeness systems of many Indo-Aryan languages and underscores Kashmiri's relational distinctiveness.

These results support Brown and Gilman's (1960) power solidarity model while also expanding it through the lens of *indexicality* and *relational work*. Pronouns in Kashmiri not only reflect static social relations but index evolving stances of respect and familiarity that shift with context, emotion, and interactional intent. For instance, an affinal

dyad might alternate between *tohj* and *tsI* within a single conversation, depending on mood or topic—illustrating the dynamic and negotiated character of pronominal politeness.

To visualize this relationship, Figure 4 summarizes the contrasting proportions of deferential and non-deferential forms across the two kinship domains.

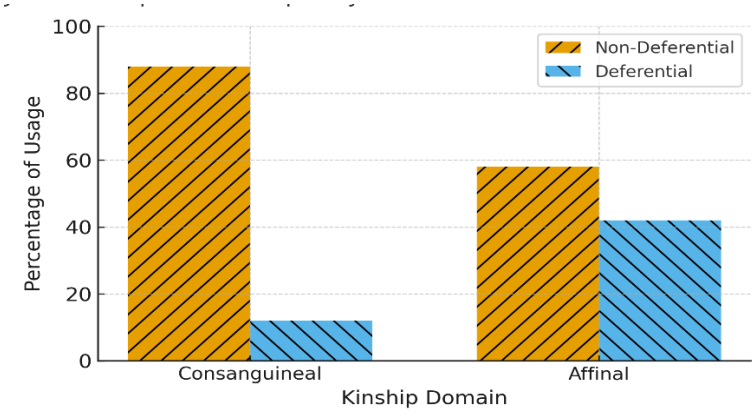

Figure 4. Comparative frequency of deferential vs. non-deferential pronouns across kinship domains. Non-deferential forms dominate in consanguineal kinship (≈88%), while deferential forms prevail in affinal kinship (42%), illustrating the linguistic negotiation between intimacy and respect.

8. Cultural Interpretation

The sociolinguistic patterns described above are deeply embedded in Kashmiri cultural norms. In traditional Kashmiri society, kinship interactions are structured by ideals of *izzat* (honor) and *adab* (decorum), values that are linguistically encoded through pronominal choice. Non-deferential forms among close kin express emotional transparency and trust, while deferential forms in affinal contexts protect face and uphold mutual respect. Importantly, the data demonstrate that politeness in Kashmiri is cooperative rather than submissive—a shared performance of social grace rather than an index of inferiority.

This aligns with Locher and Watts' (2005) argument that politeness is best understood as *relational work*—an active, context-sensitive process through which speakers continuously construct social meaning. Thus, Kashmiri pronominal usage exemplifies the way linguistic forms both sustain and recalibrate social hierarchies, blending intimacy with restraint in a culturally specific equilibrium.

Across all variables, the data demonstrate that pronominal variation in Kashmiri is systematic, socially ordered, and pragmatically motivated. The findings can be summarized as follows:

8.1 Kinship Domain Effect:

Consanguineal relations privilege solidarity (88% non-deferential), while affinal relations privilege deference (42% deferential).

8.2 Social Variable Effects:

Age: Middle-aged speakers most deferential.

Gender: Females more polite overall.

Education: Educated speakers more formal.

Settlement: Urban speakers show higher deference.

8.3 Theoretical Implication:

Pronominal choice in Kashmiri operates as a form of *indexical stance-taking*, where speakers dynamically express respect or intimacy according to social and situational demands.

These findings reinforce the view that linguistic politeness in Kashmiri is not a fixed code but an adaptive social strategy—a nuanced, context-dependent means of negotiating interpersonal distance within kinship networks.

9. Conclusion

This study has examined how Kashmiri speakers negotiate intimacy and social distance through pronominal variation across consanguineal and affinal kinship domains. By integrating quantitative frequency data with qualitative sociopragmatic interpretation, it provides the first comprehensive account of how pronouns in Kashmiri function not merely as grammatical markers but as indexical resources through which speakers construct kinship relations, identities, and affective alignments.

The findings confirm that pronominal variation in Kashmiri is both systematic and socially conditioned. In consanguineal relations, non-deferential pronouns (*tsI*, *yi*, *su/so*) predominate, indexing solidarity, emotional closeness, and shared identity. In contrast, affinal relations exhibit a markedly higher use of deferential forms (*tohj*, *yim*, *tim*), which convey respect and social diplomacy. Crucially, this deferential usage is reciprocal rather than hierarchical, signaling mutual awareness of social distance rather than one-sided subordination. Such reciprocity distinguishes Kashmiri from many Indo-Aryan languages, where politeness typically flows unidirectionally from the socially subordinate to the superior.

The analysis of social variables age, gender, education, and settlement further demonstrates that linguistic politeness is embedded in broader social structure. Age produces a U-shaped pattern of deference: young speakers favor informal address, middle-aged speakers exhibit the greatest deference, and older speakers revert to familiarity as a function of authority. Women use deferential forms more frequently than men, reflecting both cultural expectations of verbal modesty and the broader sociolinguistic pattern of female

prestige orientation. Education and urban residence correlate positively with polite usage, suggesting that exposure to formal or cosmopolitan environments heightens meta-pragmatic awareness of linguistic decorum. Together, these patterns confirm that pronominal choice in Kashmiri operates as a socially ordered variable, governed by both structural and pragmatic constraints.

Beyond confirming Brown and Gilman's (1960) power–solidarity model, the present study contributes theoretically by reframing pronominal variation through the lens of indexicality and relational work (Ochs, 1996; Locher & Watts, 2005; Bucholtz & Hall, 2005). Pronoun use in Kashmiri does not simply reflect static hierarchies; it dynamically constructs and reconfigures social relationships in interaction. Each utterance of *tsI* or *tohj* is an act of stance-taking, through which speakers negotiate the delicate equilibrium between *affection and respect*. In this sense, linguistic politeness in Kashmiri is best understood as a cooperative achievement—a shared performance of harmony and social grace that continually reaffirms kinship ties.

The study's methodological approach also demonstrates the value of combining variationist and pragmatic analysis. Quantitative data illuminate broad sociolinguistic patterns, while ethnographic observation captures the fluidity and contextual sensitivity of politeness in practice. This integrated framework may serve as a model for future research on other under-described languages of South Asia, where pronominal systems remain rich but analytically underexplored.

Future investigations could extend this work to non-kinship domains such as workplace, educational, or public interactions, where the boundaries of deference and intimacy are negotiated differently. Longitudinal or experimental studies might also explore how modernization, migration, and digital communication reshape the norms of politeness and address among younger Kashmiri speakers.

In conclusion, the study reveals that the Kashmiri pronominal system is not merely a reflection of social hierarchy but a living instrument of relational expression. Through pronominal choice, speakers continuously balance solidarity and respect, intimacy and distance—sustaining both linguistic and social harmony within the intricate web of Kashmiri kinship.

References

Bhat, J. A. (2017). *Pronominal variation in consanguineal kinship relations in Kashmiri*. *Interdisciplinary Journal of Linguistics*, 10(1), 45–67.

Braun, F. (1988). *Terms of address: Problems of patterns and usage in various languages and cultures*. Mouton de Gruyter.

Brown, P., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style in language* . pp. 253–276. MIT Press.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614. https://doi.org/10.1177/1461445605054407

Chandrasekhar, A. (1970). *Social meaning of address forms in South Indian languages*. Central Institute of Indian Languages.

Ervin-Tripp, S. (1972). Sociolinguistic rules of address. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 225–240). Penguin.

Fasold, R. (1990). *The sociolinguistics of language*. Basil Blackwell.

Holmes, J. (1995). *Women, men and politeness*. Longman.

Kachru, B. B. (1969). *Kashmiri: A grammatical sketch*. Mouton.

Koul, O. N. (1985). *Modern Kashmiri grammar*. Indian Institute of Language Studies.

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Locher, M. A., & Watts, R. J. (2005). Politeness theory and relational work. *Journal of Politeness Research*, 1(1), 9–33. https://doi.org/10.1515/jplr.2005.1.1.9

Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation*. Blackwell.

Nayar, P. (1997). Politeness in Indian English: A sociolinguistic perspective. *Language Sciences*, 19(3), 313–333. https://doi.org/10.1016/S0388-0001(96)00026-5

Ochs, E. (1996). Linguistic resources for socializing humanity. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 407–437). Cambridge University Press.

Pandharipande, R. (1992). Language choice, agency and social identity in Indian multilingualism. *Language in Society*, 21(4), 521–553. https://doi.org/10.1017/S0047404500015525

Scollon, R., & Scollon, S. W. (1995). *Intercultural communication: A discourse approach*. Blackwell.

Shapiro, M., & Schiffman, H. (1981). *Language and society in South Asia*. Motilal Banarsidass.

Tannen, D. (1984). *Conversational style: Analyzing talk among friends*. Ablex.

Trudgill, P. (1983). *Sociolinguistics: An introduction to language and society*. Penguin.

Wardhaugh, R. (2006). *An introduction to sociolinguistics* (5th ed.). Blackwell.

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 185-202

Agreement in Khortha

Siddharth Sanket *
Vadthya Kishore[**]

Abstract

Khortha is an Indo-Aryan language primarily spoken in the state of Jharkhand. According to the 2011 census, Khortha is considered to be a dialect of Hindi, similar to many other regional languages. Despite having over 8.04 million speakers, Khortha lacks comprehensive linguistic documentation and is considered endangered by some scholars (Priya & Singh, 2018). Previous research on Khortha has been limited and insufficient.

There are six distinct varieties of Khortha spoken in Jharkhand. This paper focuses on the agreement patterns of the Ramgarhiya 'ramgəɽʰɪya' variety spoken in the Ramgarh district of Jharkhand (Ohdar, 2012). Unlike Hindi, which employs a 'dual system of agreement', Khortha uses a 'single system of agreement' (Das, 2006). This study examines various aspects of agreement in terms of number and person, highlighting that Khortha, unlike Hindi, does not use grammatical gender. In Khortha, agreement markers can appear as verbal elements or auxiliaries.

By analyzing these patterns, we gain a deeper understanding of Khortha's unique features and its distinctions from Hindi. This paper aims to contribute to linguistic typology and language documentation by thoroughly examining Khortha's agreement pattern. Enhanced understanding and documentation of Khortha are crucial for its preservation and for enriching our knowledge of these under-resourced languages.

Keywords: Khortha, Ramgarhiya, Indo-Aryan, Austro-Asiatic, Dravidian, Grammatical Agreement, Single-System of Agreement, Dual-System of Agreement

1. Introduction

Khortha is a language variety spoken primarily in the state of Jharkhand and in some parts of Bihar and West Bengal. It is the second largest language in terms of the number of speakers spoken in the state of Jharkhand, boasting about 8.04 million speakers (Census, 2011). The census of India also includes Khortha as a dialect of Hindi. Jharkhand, being a linguistically rich region which has speakers of languages from

* Jawaharlal Nehru University New Delhi; siddarthraj1302@gmail.com
** Jawaharlal Nehru University New Delhi; vadthyakishore0@gmail.com

three different language families, namely Indo-Aryan, Austro-Asiatic and Dravidian, the languages present in the state need to be studied extensively for their unique linguistic characteristics.

Grierson (1927), while formulating the Linguistic Survey of India volumes, describes Khortha to be a dialect of Magahi and calls it *Eastern Magahi*. Prasad and Shastri (1958) also identify it as the same (a dialect of Magahi). Khortha or Khotta is considered to be a dialect of *Magahi* (Language Code: *mag*)[1] as well as *Maithili* (Language Code: *mai*)[2], according to the Ethnologue. Das (2013) also considers that it is a dialect of Angika, which in turn is a dialect of Maithili itself. So, which one is correct? and which one should we consider? These were aptly answered by Prasad & Shastri (1958), Dangi (2012), Ohdar (2012) (as cited by Aman et al. 2020) and Aman et al. (2020) itself that there exist 6 different varieties of Khortha. (Aman et al. 2020)[3] tried to give a linguistic profile of each of these varieties and described how they can be distinguished with respect to the use of different phonemes, verb endings and vocabulary. Given below is a table (*Table 1*) which lists all the different varieties, where they are spoken and which are the languages they are associated with.

Sr. No.	Varieties	Region	Neighbouring Languages
1.	Deshwali [desʰvɑli]	Places within the Patratu border to the western side of Palamu as in Daltonganj, Lohardaga, etc.	Magahi and Bhojpuri
2.	Ramgarhiya [rɑmgəṛʰiyɑ]	Ramgarh, Hazaribagh, Bhurkunda, Eastern Patratu, etc.	Kurmali and Nagpuri
3.	Parnadiya [pərnədiyɑ]	The northern side of Damodar river in and around Nai Sarai situated opposite to Ramgarh.	
4.	Sikhariya [sikʰəriyɑ]	Jainamore, Parasnath Pahad, Topchanchi, Giridih upto Eastern Bengal borders.	Bengali
5.	Santhal	Deoghar, Madhupur, Dumka,	Maithili or

[1] Ethnologue code for *Magahi*

[2] Ethnologue code For *Maithili*

[3] Aman et al. (2020) gives a very good detailed overview of all the varieties of *Khortha* spoken in the state of Jharkhand. It is very interesting to look at the distributional properties of these varieties. Two of them are identified as distinct because of the use of different measuring device called 'pəilɑ'. The *Sikhariya* variety can be traced to some parts of Bengal and it was found that the variety has a lot of Bangla influence, not just in lexical borrowing but also with the rounding of vowels and others. The paper also discusses the differences of other varieties of *Khortha* in linguistic terms. *Golvari*, for example uses *'a:he'* instead of *'hɛ'* for the 'be' verb. Ohdar (2012) also claimed that *Parnadiya* variety adds '-o' suffix to the verb phrases and *'-a'* to the noun phrase. *Santhal Parganiya* variety is also known to have *'čʰika-čʰiki'* property which is only noticed in this variety. It is called so due to the use of *'čʰi'* as auxiliary verb 'be'.

	Parganiya [sənt^{h}ɑl pərgəniyɑ]	and other North-Eastern areas of Jharkhand Bordering to Bihar.	Angika
6.	Golwari [golvɑri]	South of Ramgarh area from Kasmar, Kherachatta, Gola, Rajrappa, etc.	Kurmali

Table 1: Varieties of Khortha along with the languages which are neighbouring those varieties, hence affecting the varieties. Adopted from Aman et al. (2020). Although the neighbouring languages affect the respective varieties of Khortha, it is not strictly confined to that variety or that area only.

On studying the table above, it is easier to see that both the claims of Khortha being a dialect of Magahi and Maithili are true, and it should be so that we should look more deeply into these varieties that exist and analyse the structural differences. Defining Khortha to be in strict boundaries cannot do justice to all these varieties. This paper is an attempt to do that exactly.

Since, the author is a native speaker of this language, the data collection process for this study went smoothly. The data was collected in the district of Ramgarh, in the state of Jharkhand. In Ramgarh district, the data was collected from the regions surrounding the town called Kuju such as Kuju Basti, Murpa, Ara, Sarubera, etc. The informants include people from age group 18-60 from both the genders.

The method used to collect data was primarily an interview with a pre-approved questionnaire, which included all the types of sentences that were required to get all the types of sentences to understand the agreement pattern in this language. The sentences were in Hindi, and the informants were asked to translate these into Khortha. Since, the questionnaire used for this data collection was build and adapted from various sources but majorly from Das (2006). The questionnaire included Intransitive, transitive, ditransitive sentences with the arguments varying in Person, Number and Gender.

The collected data then was transcribed and properly glossed. The data was transcribed using Bloch & Trager (1942). The glossing was done using Leipzig glossing rules.

2. Grammatical Agreement

Grammatical agreement depicts the relationship of a constituent (generally NPs) to another constituent (generally the verbal elements) in the clause. The term ‘agreement’ in general has been defined by many scholars over the years. Lehmann (1985) discusses how grammatical agreement is ‘referential’ in the sense that it helps to identify the referents and keep track of the referents in a discourse. Agreement does so by mapping certain features (PNG or phi features) of the NPs (Subject/Object) to the Verbal element, making the verbal element change with respect to the change in the phi features of the NP(S) it agrees with.

Das (2006), in his seminal work, has proposed a theoretical framework for the typological study of the Indian languages with

respect to their agreement patterns. He suggests that there are languages which follow a 'Dual System of Agreement', such as Hindi, Punjabi, Marathi, Nepali etc., where the verb agrees with the covertly case-marked argument, because the case markers are the blockers of agreement. So, these languages show 'Subject-Verb Agreement' as well as 'Object-Verb Agreement'. It is also worth noting that these two are in complementary distribution, meaning that when there is Subject-Verb Agreement, there won't be Object-Verb Agreement and vice versa. This change in the agreement system will occur in different syntactic domains.

Khortha[4] from hereby meaning to Ramgarhiya [rɑmgəɽʰiyɑ] in this paper, shows a pattern similar to the 'Single System of Agreement' as defined by Das (2006) in which there is a clear marking of the subject onto the verb despite it being marked by case (either overtly or covertly). The subject is the one that will always map its phi features onto the verb, making it finite along with Tense and Aspect.

This paper aims not to investigate the different varieties of Khortha and analyse the agreement patterns in all these varieties, rather this paper is an attempt to look into one of these varieties, namely Ramgarhiya [rɑmgəɽʰiyɑ] and analyse the agreement pattern followed by this variety of Khortha. Similar works on agreement patterns on the other varieties of Khortha have been done by many scholars such as Das (2017), Ray & Kumar (2022) and others. But there is no work to be found on agreement in this variety of Khortha.

We will now look into the basic agreement pattern followed by Khortha in different tenses and aspects and how the markings of number and person are mapped onto the verb. As stated earlier, Khortha doesn't have grammatical gender and hence does not differentiate between Masculine and Feminine. So, the markings are also not present for the same, and we do not have to investigate gender as this is not the topic of this paper.

When I say that Khortha doesn't have 'grammatical gender', what I mean is that apart from the biological gender assigned to the animate objects, the language does not differentiate between Masculine and Feminine non-animate nouns. This can only be seen in agreement when we talk about NP internal agreement, meaning the agreement within the Noun Phrase (NP) between the head noun, where the noun is animate, and the dependents, i.e., adjectives. Agreement of gender between noun(s) and verbs is not seen in Khortha. Consider the following examples (a-c).

a. moṭ-kɑ lərk-wɑ
   fat-MS boy-MS

   'The fat boy'

[4] All references to Khortha henceforth mean Ramgarhiya [rɑmgəɽʰiyɑ] Khortha

b. moṭ-ki lərkı-yɑ
fat-FS girl-FS
'The fat girl'

c. ʊ lərk-wɑ / lərkı-yɑ səb-se moṭ h-o
that.DEM boy-DEF.3MS / girl-DEF.3FS all-than fat be-PRS.3S
'That boy/girl is fattest among all'

In examples (a&b), we see that there is gender agreement between the head noun and the adjective, meaning the adjective agrees with the head in terms of gender. This is because the head noun is animate. Also, this gender agreement is nuanced because it's only seen NP internally and with animate nouns. As one can see in example (c), we do not find any agreement between the adjective and the noun, even though the noun is animate. This is due to the fact that the adjective is predicative.

So, with this out of the way, we will now move on to the clausal agreement pattern seen in Ramgarhiya Khortha.

3. Basic Clausal Agreement Pattern in Khortha (*Ramgarhiya*)

The following section deals with the basic agreement patterns followed by Khortha. The section gives us an in-depth analysis of all the markings used and how these markings and the form of the verb change with respect to person, number, and tense.

3.1. Simple-Present Clauses

1. həm gʰər jɑ-vo h-ıyo
i.1S.NOM house.3S.LOC go-PRES be-PRES.1S
'I go home'

2. tʊ gʰər jɑ-vo h-ĩ
you.2S.NOM house.3S.LOC go-PRES be-PRES.2S
'You go home'

3. ʊ gʰər jɑ-vo h-o
(s)he.3S.NOM house.3S.LOC go-PRES be-PRES.3S
'(S)he goes home'

Looking into the simple clauses (1-3), we see that the subject is singular and we find that there are two constituents in the verbal component, the first constituent is the main verb (jɑ meaning 'go') and the second is the 'be' verb (ho). In the simple clauses, the main verbs have only tense markings, and the agreement markings are all inflected

on the 'be' verb. We see that with respect to person, the inflection in the 'be' verb changes. Eg, for 1st person, the inflection in the 'be' verb is '-ɪyo'; for 2nd person, it is '-ĩ'; and for the 3rd person, it is '-o'.

So, from the above examples we can see that one of the phi features, i.e., person has an impact on the markings on the verbal constituent. We can also now check the plural counterparts for further analysis:

4. həm-ɪn gʰər jɑ-vo h-ɪyo
   we-1PL.NOM house.3S.LOC go-PRES be-PRES.1PL
   'We go home'

5. toh-ɪn gʰər jɑ-vo h-ĩ
   you-2PL.NOM house.3S.LOC go-PRES be-PRES.2PL
   'You (all) go home'

6. okʰ-ɪn gʰər jɑ-vo h-ətʰʊn
   they-3PL.NOM house.3S.LOC go-PRES be-PRES.3PL
   'They go home'

Examples (4-6) are the same examples as in (1-3), the only difference being in the number feature of the Subject. Here we can see that, in respect to their singular counterparts, again, the marking of plurality can be seen on the 'be' verb. There are a few things to note here, the markings for number in 1st and 2nd person are the same, meaning that there is no change in the form of the 'be' verb, which is otherwise proof for person agreement in Khortha. Nonetheless, we find that there is a change in the form of the 'be' verb in Number when it comes to the 3rd person. The inflection for 3rd plural is '-ətʰʊn', which is not the same as the inflection for 3rd singular, i.e., '-o'.

Examples (1-6) give us a general overview of the agreement pattern followed by Khortha. The observation till now can be summarized as follows:

In Khortha, there is agreement between the subject and the verbal element of the sentence; this agreement is realised in terms of Person and Number features of the NP(s) onto the verbal elements. When there are two elements in the Verb Phrase, the main verb bears the Tense markings and the 'be' verb takes the Person and Number inflection. The form of the 'be' verb changes with respect to the Person feature. It is also seen that the Number markings for 1st and 2nd person are the same, i.e., there is no change in the form of the 'be' verb for 1st singular and plural, as well as 2nd singular and plural. But, this is not the case for 3rd person. The number markings are distinct for 3rd person in Khortha.

In the above explanation, I've said that the main verb is marked for Tense, but from the examples (1-6), one can argue it's not the case. This can be solved when we look into other examples, where there is only one element in the VP.

3.2. Simple-Past Clauses

7a. həm gʰər g-əl-ɪo
i.1S.NOM house.3S.LOC go-PST-1
'I went home'

7b. həm-ɪn gʰər g-əl-ɪo
we-1PL.NOM house.3S.LOC go-PST-1
'We went home'

8a. tʊ gʰər g-əl-ĩ
you.2S.NOM house.3S.LOC go-PST-2
'You went home'

8b. toh-ɪn gʰər g-əl-ĩ
you-2PL.NOM house.3S.LOC go-PST-2
'You(all) went home'

9. ʊ gʰər g-əl-o
(s)he.3S.NOM house.3S.LOC go-PST-3S
'(s)he went home'

10. okʰɪn gʰər g-əl-tʰʊn
they.3PL.NOM house.3S.LOC go-PST-3PL
'They went home'

The markings in these examples (7a-10) again conform to the findings and observations about the number and person agreement in Khortha in the simple past tense. From the above examples, we see that that verb is inflected for person and number. We also see that there is no change in markings for number in 1st and 2nd person (7a&7b; 8a&8b).

However, there is a slight difference between the examples of the simple past tense and the examples in the simple present tense. In the latter, we see that the verb was accompanied by a 'be' verb, which was carrying the markers for person and number, but in the case of simple past clauses, we see that all the markers are carried by the main verb itself. The tense, number and person markers are analysed onto the main verb through suffixes. The suffix for the past tense is '-əl', and the person and number markings are the same as discussed above. It is also important to note the order of the inflection. The tense marker is the first morpheme to inflect with the verb, and then the number and person marker can come.

3.3. Simple-future Clauses

11a. həm gʰər j-ɪb-o
i.1S.NOM house.3S.LOC go-FUT-1
'I will go home'

11b. həm-ɪn gʰər j-ɪb-o
we-1PL.NOM house.3S.LOC go-FUT-1
'We will go home'

12a. tʊ gʰər j-ɪb-ĩ
you.2S.NOM house.3S.LOC go-FUT-2
'You will go home'

12b. toh-ɪn gʰər j-ɪb-ĩ
you-2PL.NOM house.3S.LOC go-FUT-2
'You(all) will go home'

13. ʊ gʰər j-ɪt-o
(s)he.3S.NOM house.3S.LOC go-FUT-3S
'(s)he will go home'

14. okʰɪn gʰər j-ɪb-tʰʊn
they.3PL.NOM house.3S.LOC go-FUT-3PL
'They will go home'

The examples (11a-14) are given here to show the agreement patterns followed by Khortha in simple future tense. After analysing the data, we see that the agreement pattern is consistent throughout the tenses, apart from the 1st person, as there is a slight change in the morpheme marking person. Earlier, it was seen that for 1st singular, the morpheme used was '-ɪyo'. But in future tense it is '-o'. Apart from this, we can say that agreement is an indication of structural uniformity in Khortha language. Although there needs to be a discussion about the 3rd person marking of agreement in the language. Looking into the examples (11a-14), we see that the future tense is marked upon the verb by a suffix, i.e., '-ɪb' in all the number and person in the simple future tense, apart from the 3rd person singular, where we find that the tense marker for future tense in 3rd person is '-ɪt'.

'Be' verb (ho)

	Present		Past		Future	
	S	**PL**	**S**	**PL**	**S**	**PL**
1st Person	h-ɪyo	h-ɪyo	h-əl-ɪo	h-əl-ɪo	h-əb-o	h-əb-o
2nd Person	h-ĩ	h-ĩ	h-əl-ĩ	h-əl-ĩ	h-əb-ĩ	h-əb-ĩ
3rd Person	h-o	h-ətʰʊn	h-əl-o	h-əl-tʰʊn	h-ət-o	h-əb-tʰʊn

Table 2: The verb declensions for different tenses for the 'be' verb in Khortha. We can also see the changes in the form of the verb with respect to person and number.

'Go' verb (jɑnɑ)

	Past		Future	
	S	**PL**	**S**	**PL**
1st Person	g-əl-ɪo	g-əl-ɪo	j-ɪb-o	j-ɪb-o
2nd Person	g-əl-ĩ	g-əl-ĩ	j-ɪb-ĩ	j-ɪb-ĩ
3rd Person	g-əl-o	g-əl-$t^hʊ$n	j-ɪt-o	j-ɪb-$t^hʊ$n

Table 3: The verb declensions in different tenses for the 'go' verb in Khortha.

The following table (*Table 1&2*) shows the verb declensions of *be* and *go* verbs, respectively. We can see from the table the paradigm that is being followed in Ramgarhiya Khortha.

4. Aspectual Agreement in Khortha

After having looked at the agreement patterns in different tenses in Khortha, it is necessary that we investigate these patterns in different aspects as well. We will see how aspectual markings in Khortha work and how they are analysed in the language.

4.1. Present Simple/Indefinite Aspect

15a. həm roṭi $k^hɑ$-vo h-ɪyo
i.1S.NOM chapatti.3S.ACC eat-PRES be-PRES.1
'I eat chapatti'

15b. həm-ɪn roṭi $k^hɑ$-vo h-ɪyo
we-1PL.NOM chapatti.3S.ACC eat-PRES be-PRES.1
'We eat chapatti'

16a. tʊ roṭi $k^hɑ$-vo h-ĩ
you.2S.NOM chapatti.3S.ACC eat-PRES be-PRES.2
'You eat chapatti'

16b. toh-ɪn roṭi $k^hɑ$-vo h-ĩ
you-2PL.NOM chapatti.3S.ACC eat-PRES be-PRES.2
'You(all) eat chapatti'

17. ʊ roṭi $k^hɑ$-vo h-o
(s)he.3S.NOM chapatti.3S.ACC eat-PRES be-PRES.3S
'(s)he eats chapatti'

18. ok^{h}ɪn roṭi $k^hɑ$-vo h-ə$t^hʊ$n
they.3PL.NOM chapatti.3S.ACC eat-PRES be-PRES.3PL
'They eat chapatti'

The examples from (15a-18) are from present simple or indefinite aspect in Khortha. We can see that these examples, all conform to the previously discussed agreement patterns in the language. The verbs chosen are transitive in nature to show that transitivity doesn't have an effect on the agreement patterns in Khortha.

4.2. Present Progressive/Continuous Aspect

19a. həm roṭi kʰɑv-et h-ɪyo
i.1S.NOM chapatti.3S.ACC eat-PROG be-PRES.1
'I am eating chapatti'

19b. həm-ɪn roṭi kʰɑv-et h-ɪyo
we-1PL.NOM chapatti.3S.ACC eat-PROG be-PRES.1
'We are eating chapatti'

20a. tʊ roṭi kʰɑv-et h-ĩ
you.2S.NOM chapatti.3S.ACC eat-PROG be-PRES.2
'You are eating chapatti'

20b. toh-ɪn roṭi kʰɑv-et h-ĩ
you-2PL.NOM chapatti.3S.ACC eat-PROG be-PRES.2
'You(all) are eating chapatti'

21. ʊ roṭi kʰɑv-et h-o
(s)he.3S.NOM chapatti.3S.ACC eat-PROG be-PRES.3S
'(S)he is eating chapatti'

22. okʰɪn roṭi kʰɑv-et h-ətʰʊn
they.3PL.NOM chapatti.3S.ACC eat-PROG be-PRES.3PL
'They are eating chapatti'

The examples above (19a-22) are given to showcase the progressive aspect in present tense in Khortha. The progressive marker for marking continuity in Khortha is '-et', which attaches to the main verb to show that the action is not complete. The markers for person and number remain the same throughout the continuous aspect, as it is in simple tense. The markings for number in 1st and 2nd person remain the same, i.e., '-ɪo' and '-ĩ' respectively. The markings for number in the 3rd person change, i.e., '-o' for singular and '-ətʰʊn' for plural.

4.3. Past Progressive Aspect

23a. həm roṭi kʰɑv-et h-əl-ɪo
i.1S.NOM chapatti.3S.ACC eat-PROG be-PST-1
'I was eating chapatti'

23b. həm-ɪn roṭi kʰɑv-et h-əl-ɪo
we-1PL.NOM chapatti.3S.ACC eat-PROG be-PST-1
'We were eating chapatti'

24a. tʊ roṭi kʰɑv-et h-əl-ĩ
you.2S.NOM chapatti.3S.ACC eat-PROG be-PST-2
'You were eating chapatti'

24b. toh-ɪn roṭi kʰɑv-et h-əl-ĩ
you-2PL.NOM chapatti.3S.ACC eat-PROG be-PST-2
'You(all) were eating chapatti'

25. ʊ roṭi kʰɑv-et h-əl-o
(s)he.3S.NOM chapatti.3S.ACC eat-PROG be-PST-3S
'(S)he was eating chapatti'

26. okʰɪn roṭi kʰɑv-et h-əl-tʰʊn
they.3PL.NOM chapatti.3S.ACC eat-PROG be-PST-3PL
'They were eating chapatti'

The examples (23a-26) are examples to show the agreement patterns in past progressive aspect. As in the present progressive, the main verb is inflected with the progressive marker '-et' to show the progressive nature of the action. Also, compared to the present progressive, the form of the 'be' verb is also changed. It has taken the past tense marker '-əl' to show that the action is done in the past. In these examples (23a-26) as well, we can see that the transitivity of the verb doesn't have any impact on the markings for past progressive aspect in Khortha.

4.4. Future Indefinite Aspect

27a. həm roṭi kʰə-ɪb-o
i.1S.NOM chapatti.3S.ACC eat-FUT-1
'I will eat chapatti'

27b. həm-ɪn roṭi kʰə-ɪb-o
we-1PL.NOM chapatti.3S.ACC eat-FUT-1
'We will eat chapatti'

28a. tʊ roṭi kʰə-ɪb-ĩ
you.2S.NOM chapatti.3S.ACC eat-FUT-2
'You will eat chapatti'

28b. toh-ɪn roṭi kʰə-ɪb-ĩ
you-2PL.NOM chapatti.3S.ACC eat-FUT-2
'You(all) will eat chapatti'

29. ʊ roṭi kʰə-ɪt-o
(s)he.3S.NOM chapatti.3S.ACC eat-FUT-3S
'(S)he will eat chapatti'

30. okʰɪn roṭi kʰə-ɪb-tʰʊn
they.3PL.NOM chapatti.3S.ACC eat-FUT-3PL
'They will eat chapatti'

Examples (27a-30) are in future-indefinite, but we have already seen examples from future indefinite in (11a-14). This is because in examples (27a-30), the verb is transitive. However, we see that the previous discussion regarding the aspect marking of verbs in Khortha holds true in these examples, regardless of the transitivity of the verb.

4.5. Future Progressive Aspect

31a. həm roṭi kʰɑv-et h-əb-o
i.1S.NOM chapatti.3S.ACC eat-PROG be-FUT-1
'I will be eating chapatti'

31b. həm-ɪn roṭi kʰɑv-et h-əb-o
we-1PL.NOM chapatti.3S.ACC eat-PROG be-FUT-1
'We will be eating chapatti'

32a. tʊ roṭi kʰɑv-et h-əb-ĩ
you.2S.NOM chapatti.3S.ACC eat-PROG be-FUT-2
'You will be eating chapatti'

32b. toh-ɪn roṭi kʰɑv-et h-əb-ĩ
you-2PL.NOM chapatti.3S.ACC eat-PROG be-FUT-2
'You(all) will be eating chapatti'

33. ʊ roṭi kʰɑv-et h-ət-o
(s)he.3S.NOM chapatti.3S.ACC eat-PROG be-FUT-3S
'(S)he will be eating chapatti'

34. okʰɪn roṭi kʰɑv-et h-əb-tʰʊn
they.3PL.NOM chapatti.3S.ACC eat-PROG be-FUT-3PL
'They will be eating chapatti'

Examples (31a-34) are examples to show the future progressive aspect. We can see from the examples that the progressive aspect in the future tense is also marked onto the main verb. The marker is the same, which is used in all other tenses to depict continuity, i.e., '-et'. We can also see that the markers for person and number are consistent throughout different aspects. The tense marker is also uniform. In these examples (31a-34), we see that the tense marker for future tense is '-əb' in 1st (Singular and Plural); 2nd (Singular and Plural); and 3rd Plural. The marker for tense in 3rd Singular is '-ət'.

In the following section, I'll discuss the perfective aspect markings in Khortha. Giving a separate space for discussion in Khortha for the perfect aspect is important as we must understand how Khortha behaves differently from many Indo-Aryan languages and especially Hindi, to which Khortha is considered a dialect (census 2011). Many Indian languages such as Hindi, Punjabi, Gujarati, Marathi and others, show a typological similarity in that they all exhibit something called 'split-ergativity'. The languages mentioned are called split ergative languages as they partially display features of 'Nominative-Accusative' type of case markings and partially show 'Ergative-Absolutive type' case markings. These changes are conditioned due to changes in tense, aspect, number and person markings. Generally, the marking of the ergative case with the subject is seen in 'Simple Past' and all 'Perfective' aspects.

I'll not go into details about ergativity and how ergativity in Indian languages is conditioned, as it is not the topic of this paper. But from the examples, one would be able to see how Khortha is different from the rest of the Indo-Aryan languages that are mentioned above. We will look into the examples of Simple past and Perfective Aspect in Khortha, which will give us a clear indication that Khortha doesn't belong to the typology of 'Split-Ergative languages.' We see that there is no difference in the marking of the subject (either overt or covert) with any other case other than Nominative, when there is a change in the transitivity of the verb, which would otherwise be true for other Indian languages such as Hindi, Marathi, Punjabi, Gujarati, Nepali, etc. So, Khortha is, strictly speaking, a 'Nominative-Accusative' type language.

4.6. Present Perfect Aspect

35a. həm roṭi kʰə-ɪle h-ɪyo
i.1S.NOM chapatti.3S.ACC eat-PERF be-PRES.1
'I have eaten chapatti'

35b. həm-ɪn roṭi kʰə-ɪle h-ɪyo
we-1PL.NOM chapatti.3S.ACC eat-PERF be-PRES.1
'We have eaten chapatti'

36a. tʊ roṭi kʰə-ɪle h-ĩ
you.2S.NOM chapatti.3S.ACC eat-PERF be-PRES.2
'You have eaten chapatti'

36b. toh-ɪn roṭi kʰə-ɪle h-ĩ
you-2PL.NOM chapatti.3S.ACC eat-PERF be-PRES.2
'You(all) have eaten chapatti'

37. ʊ roṭi kʰə-ɪle h-o
(s)he.3S.NOM chapatti.3S.ACC eat-PERF be-PRES.3S
'(S)he has eaten chapatti'

38. okʰɪn roṭi kʰə-ɪle h-ətʰʊn
they.3PL.NOM chapatti.3S.ACC eat-PERF be-PRES.3PL
'They have eaten chapatti'

The examples (35a-38) are examples in the perfect aspect in the present tense. The perfective aspect indicates that an action is completed or the state of an action is finished. We see that there is no change in the markings on the subject, indicating that there is no ergativity in Khortha. Secondly, we see that the markings for the perfective aspect are in the main verb. The suffix '-ɪle' is used in the present tense to denote the perfective nature of the action. It is interesting to see that the marker for perfective aspect in the present tense and the marker of past, i.e., '-əl/ɪl' in Khortha, are very much similar to each other. Apart from this, the rest of the markings of

person, number and tense conform to the structure of agreement in the language. Also, since the subject is still in the Nominative case, the agreement is done with the subject itself.

4.7. Past Simple Aspect (Transitive)

39a. həm roṭi kʰə-ɪl-ɪo
i.1S.NOM chapatti.3S.ACC eat-PST-1
'I ate chapatti'

39b. həm-ɪn roṭi kʰə-ɪl-ɪo
we-1PL.NOM chapatti.3S.ACC eat-PST-1
'We ate chapatti'

40a. tʊ roṭi kʰə-ɪl-ĩ
you.2S.NOM chapatti.3S.ACC eat-PST-2
'You ate chapatti'

40b. toh-ɪn roṭi kʰə-ɪl-ĩ
you-2PL.NOM chapatti.3S.ACC eat-PST-2
'You(all) ate chapatti'

41. ʊ roṭi kʰə-ɪl-o
(s)he.3S.NOM chapatti.3S.ACC eat-PST-3S
'(S)he ate chapatti'

42. okʰɪn roṭi kʰə-ɪl-tʰʊn
they.3PL.NOM chapatti.3S.ACC eat-PST-3PL
'They ate chapatti'

The examples (7a-10) discussed earlier already describe the agreement patterns in simple past in Khortha. Examples (39a-42) are also in simple past, but they are transitive constructions as compared to examples (7a-10). The agreement patterns are identical as discussed earlier, but these constructions are required to show that Khortha doesn't have ergativity at all.

4.8. Past Perfect Aspect

43a. həm roṭi kʰɑy lələ h-əl-ɪo
i.1S.NOM chapatti.3S.ACC eat take.PERF be-PST-1
'I had eaten chapatti'

43b. həm-ɪn roṭi kʰɑy lələ h-əl-ɪo
we-1PL.NOM chapatti.3S.ACC eat take.PERF be-PST-1
'We had eaten chapatti'

44a. tʊ roṭi kʰɑy lələ h-əl-ĩ
you.2S.NOM chapatti.3S.ACC eat take.PERF be-PST-2
'You had eaten chapatti'

44b.	toh-ın	roṭi	kʰɑy	ləle	h-əl-ĩ
	you-2PL.NOM	chapatti.3S.ACC	eat	take.PERF	be-PST-2
	‘You(all) had eaten chapatti’				

45.	ʊ	roṭi	kʰɑy	ləle	h-əl-o
	(s)he.3S.NOM	chapatti.3S.ACC	eat	take.PERF	be-PST-3S
	‘(S)he had eaten chapatti’				

46.	okʰın	roṭi	kʰɑy	ləle	h-əl-tʰʊn
	they.3PL.NOM	chapatti.3S.ACC	eat	take.PERF	be-PST-3PL
	‘They had eaten chapatti’				

If we look at the past perfect examples (43a-46) in Khortha, we see something interesting. The verbs in all these examples do not carry any aspectual or tense marker at all. Rather, there seems to be an addition of another verbal element which specifically carries the aspectual marker (perfective here). The verbal element here is ‘ləle’ (meaning ‘take’), which gives the sentence a perfective reading. The ‘be’ verb, which is also present, carries the tense as well as person and number agreement. The tense, person and number agreement is consistent with the structural uniformity in the language.

4.9. Future Perfect Aspect

47a.	həm	roṭi	kʰɑy	ləle	h-əb-o
	i.1S.NOM	chapatti.3S.ACC	eat	take.PERF	be-FUT-1
	‘I will have eaten chapatti’				

47b.	həm-ın	roṭi	kʰɑy	ləle	h-əb-o
	we-1PL.NOM	chapatti.3S.ACC	eat	take.PERF	be-FUT-1
	‘We will have eaten chapatti’				

48a.	tʊ	roṭi	kʰɑy	ləle	h-əb-ĩ
	you.2S.NOM	chapatti.3S.ACC	eat	take.PERF	be-FUT-2
	‘You will have eaten chapatti’				

48b.	toh-ın	roṭi	kʰɑy	ləle	h-əb-ĩ
	you-2PL.NOM	chapatti.3S.ACC	eat	take.PERF	be-FUT-2
	‘You(all) will have eaten chapatti’				

49.	ʊ	roṭi	kʰɑy	ləle	h-ət-o
	(s)he.3S.NOM	chapatti.3S.ACC	eat	take.PERF	be-FUT-3S
	‘(S)he will have eaten chapatti’				

50.	okʰın	roṭi	kʰɑy	ləle	h-əb-tʰʊn
	they.3PL.NOM	chapatti.3S.ACC	eat	take.PERF	be-FUT-3PL
	‘They will have eaten chapatti’				

These examples (47a-50) are the last set of examples to show the agreement pattern Khortha. The above examples are of perfect aspect in future. Here, again, we see that the main verb is bare and doesn't carry any of the aspectual or tense markings. Similar to past perfect, we see that there is an addition of a verbal element 'ləle' which gives the sentence a perfective reading. The rest of the agreement markings such as '-əb' and '-ət' for future; '-o' for 1st (S&PL) and 3rd singular future; '-ĩ' for 2nd; and '-tʰʊn' for 3rd plural, are being followed consistently. The tables below (*Table 4 – 6*) provide us the inflected forms of the verb *kʰa* 'to eat' in different aspects in each of the tenses.

Person	**Present Indefinite**	**Present Continuous**	**Present Perfect**
1st (S/PL)	kʰɑvo h-ɪyo	kʰɑv-et h-ɪyo	kʰə-ɪle h-ɪyo
2nd (S/PL)	kʰɑvo h-ĩ	kʰɑv-et h-ĩ	kʰə-ɪle h-ĩ
3rd Singular	kʰɑvo h-o	kʰɑv-et h-o	kʰə-ɪle h-o
3rd Plural	kʰɑvo h-ətʰʊn	kʰɑv-et h-ətʰʊn	kʰə-ɪle h-ətʰʊn

Table 4: Inflections on the verb in different aspects in present tense.

Person	**Past Continuous**	**Past Perfect**
1st (S/PL)	kʰɑv-et h-əl-ɪo	kʰɑy ləle h-əl-ɪo
2nd (S/PL)	kʰɑv-et h-əl-ĩ	kʰɑy ləle h-əl-ĩ
3rd Singular	kʰɑv-et h-əl-o	kʰɑy ləle h-əl-o
3rd Plural	kʰɑv-et h-əl-tʰʊn	kʰɑy ləle h-əl-tʰʊn

Table 5: Inflections on the verb in different aspects in past tense.

Person	**Future Indefinite**	**Future Continuous**	**Future Perfect**
1st (S/PL)	kʰə-ɪb-o	kʰɑv-et h-əb-o	kʰɑy ləle h-əb-o
2nd (S/PL)	kʰə-ɪb-ĩ	kʰɑv-et h-əb-ĩ	kʰɑy ləle h-əb-ĩ
3rd Singular	kʰə-ɪt-o	kʰɑv-et h-ət-o	kʰɑy ləle h-ət-o
3rd Plural	kʰə-ɪb-tʰʊn	kʰɑv-et h-əb-tʰʊn	kʰɑy ləle h-əb-tʰʊn

Table 6: Inflections on the verb in different tenses in future tense.

5. Conclusion

In this paper, we discussed the agreement pattern shown by the Ramgarhiya variety of Khortha. We saw the agreement patterns in this language with respect to Tense and Aspect and discussed how person and number features of the Subject influence the form of the verb and how these features are mapped onto the verbal element. We also discussed how the agreement pattern in this language is different from the major Indian languages, such as Hindi, Marathi, Bengali, Punjabi, etc., by explaining how this language exhibits a 'single system of agreement' as per the definitions given by Das (2006). We also discussed how ergativity is not present in this language and it is strictly a 'Nominative-Accusative' type language.

The fact that both *Magahi* and *Maithili*, and for that matter, the rest of the Eastern Indo-Aryan languages in the region, do not show what is called 'ergativity', but one of the varieties of Khortha, i.e., *Santhal*

Parganiya[5], shows full ergativity and is strictly an "ergative-absolutive" type language. This variety exhibits a 'dual system of agreement.' This not only sets precedence that that languages are so nuanced, and hence we need to be more investigative, but also to bunch them into major languages doesn't do justice to their unique structural features.

References

Aman, A., et al. (2020). A preliminary account on the Khortha speech prosody. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 6440–6458

Aman, A., et al. (2020). Consonants in Khortha. In *Alternative horizons in linguistics: A festschrift in honour of Prof. Panchanan Mohanty* (pp. 336–352).

Aman, A., & Dash, N. S. (2020). Designing a linguistic profile of Khortha: A less resourced language spoken in the state of Jharkhand, India. *Dialectologia, 25*, 25–43.

Baker, M. (2011).When agreement is for number and gender but not person. In *NLLT Springer Science, Business Media*, 875–915

Bhatt, R. M. (1997). On agreement in Hindi. *South Asian Language Review: Special Issue on Studies in Hindi-Urdu*, 73–82.

Comrie, B. (1984). Reflections on verb agreement in Hindi and related languages. *Linguistics, 22*, 857–864.

Dangi, A. K. (2012). *Khortha bhasha: Ek parichay (Khortha language: An identity).* Spardha Publications.

Das, P. K. (1997). *Agreement in Hindi-Urdu*. Ph.D Dissertation, centre for Linguistics and English (SLL & CS) Jawaharlal Nehru University New Delhi, Delhi

Das, P. K. (2006). *Grammatical agreement in Hindi-Urdu and its major varieties*. Vol.33, Lincom Europa.Munich Germany

Das, P. K. (2013). Ergativity in Khortha: An enigmatic phenomenon. *The Journal of South Asian Studies, 18*, 321–350.

Das, P. K. (2018). Ergativity and agreement in Kinnauri. *Interdisciplinary Journal of Linguistics, 11*, 36–53.

Dash, N., & Aman, A. (2015).Generation of a dialect corpus in Khortha used in Jharkhand, India: Some empirical observations and theoretical postulations. *The Journal of Advanced Linguistic Studies*, 151–171.

Davison, A. (2002). Agreement features and projections of tense and aspect. *Yearbook of South Asian Languages and Linguistics*.

Dixon, R. M. (1979). Ergativity.*Language, 55*(1), 59–138.

[5] To know more about this language, please refer to (Das, 2013).

Dixon, R. M. (1994).*Ergativity*.Cambridge, University Press

Grierson, G. A. (1927). *Linguistic survey of India, 1927: Introduction*. Central Publication Branch, Government of India.

Kaul, O. N. (2008). *Modern Hindi grammar*. Dunwoody Press.

Kumari, A. (2021). Folklore and language: Ergativity and agreement of Khortha language in Jharkhand. *International Journal of Humanities, Engineering, Science and Management, 2*, 35–44.

Lapointe, S. (1985). *A theory of grammatical agreement*. Garland Publishing.

Lehmann, W. P. (1978). *Syntactic typology: Studies in the phenomenon of language*. University of Texas Press.

Ohdar, B. (2012). *Khortha bhasha evam sahitya: Udbhav evam vikash (Khortha language and literature: Origin and growth).* Khortha Bhasha Sahitya Academy.

Prasad, B. N., & Shastri. (1958). *Linguistic survey of the Sadar Subdivision of Manbhum and Dalbhum (Singhbhum).* Bihar Rastrabhasha Parishad.

Priya, S., & Singh, R. (2018). Khortha, a dying language and the urgency to retain its pure variety. *Rupkatha Journal on Interdisciplinary Studies in Humanities, 5*(1), 171–190.

Ray, B., & Kumar, A. (2022). Person agreement in Khortha: A language of Jharkhand. *Interdisciplinary Journal of Linguistics, 15*, 85–98.

Saksena, S. (1981). Verb agreement in Hindi.*Linguistic,*14, 467–474.

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 203-218

A Sketch of Tense, Aspect and Modality in Bongcher

Niloy Chakraborty*

Abstract

The paper has examined tense, aspect and modality in Bongcher, a Tibeto-Burman language of Kuki-Chin branch. The aspect and modality has been discussed with reference to past, present and future tenses, and simple, completive, incompletive, irrealis, and progressive tenses in Bongcher. In addition, few distinctive modality markers, such as /-tʰəi/, /raŋnaŋ/, /-inlaŋ/, /val/ etc. have also been discussed. Both primary and secondary data were analysed for this study, with the majority of the data coming from the speakers of Ompi and Nelsipara village in the district of Gomati. The study focused on Bongpui clan for data which was collected during a field visit in September and October, 2025 in Nelsipara and Tuicholong (Near Ompi) villages, in Amarpur sub-division.

Keywords: Bongcher, Tense, Aspect, Mood, Modality, Irrealis Aspect

1. Introduction

In recent years, modality has been acknowledged as a valid cross-linguistic grammatical category for linguistic typological studies[1]. It is a category that is closely associated with tense and aspect in that all three categories are categories of the clause and are generally, but not always, marked within the verbal complex (Palmer, 2001, p.3). In a theoretical sense, all three are involved in the circumstance or occurrence that is reported by the utterance. Aspect is concerned with the nature of the event, specifically with its 'internal temporal constituency', whereas tense is plainly concerned with the time of the event (Comrie, 1976, p.3). Modality is concerned with the status of the proposition that described the event (Palmer, 2001, p.3). In contrast to tense and aspect, modality just refers to the condition of the proposition rather than any specific aspect of the event. Making a binary[2] distinction between 'non-modal' and 'modal' or 'declarative' and 'non-declarative' and linking this distinction to the notional contrast of 'factual' and 'non-factual' or 'real' and 'unreal' is one way to analyse it (Palmer, 2001, p.4).

The word 'tense' comes from the Latin translation of the Greek word 'khronos', which means 'time' (Lyons, 1968). The grammatical system of a language that expresses tense, aspect, and mood or modality is known as tense-aspect-mood or TAM in short. It is also

* Tripura University; nilaychk@gmail.com

frequently referred to as tense-modality-aspect, or TMA. According to Bybee, Perkins, and Pagliuca (1994), tense conveys 'location in time', aspect deals with the 'fabric of time a single block of time, continuous flow of time, or repetitive occurrence', and mood or modality illustrates the 'degree of necessity, obligation, probability, ability, etc.'

Grammatical categories of tense[3], aspect[4] and mood (TAM) are incorporated into verbs to represent time, temporal structure, and modality. TAM typically takes the shape of morphemes in verbs. Aspect describes the dynamism of an action, whether it is finished or on-going, while tense describes the manifestation of time, whether it is past or not. Similarly, speakers' attitudes regarding the action such as command, certainty, prohibition, wish, or will are reflected in their mood. Along with aspect and mood, Crystal (2008) challenges the use of tense as a category in the grammatical description of verbs, focusing mainly on how the grammatical category indicates time. However, Comrie (1985) notes that aspect is the grammaticalization of the expression of internal temporal constituency, which is concerned with the nature of the event, whether it is on-going or completed, and tense is concerned with time, or the grammaticalization of its placement in time. According to Palmer (2001), the status of the notion describing the event is what modality is all about. Similarly, Bybee et al. (1994) point out that mood and modality can be refused as the grammaticalization of speakers' (subject) emotions and beliefs because they are not as easily defined as tense and aspect. Aspects are several perspectives on a situation's internal temporal constituency (Holt, 1943 as cited by Comrie, 1976). In conclusion aspect relates the internal temporal structure of an action, mood connects the purpose toward the proposition, and tense relates the time of an action.

2. The Language

Sailo (1992) states that Tripura is home to nineteen recognized tribal communities, among which Bongchers identify themselves as the sixth-largest sub-tribe of the Halam community. However, the exact number of Halam sub-tribes remains uncertain and continues to be a matter or scholarly discussion. Historical records provide varying classifications of the Halam sub-groups. The Tripura State Gazetteer (Vol. 16, Part 29, Census 1931) lists nine Halam tribes: Morsum, Rangkhol, Koloi, Kaibeng, Bongshel, Rupini, Korbong, Thangchep, and Chorai. Furthermore, the 1931 Census identified as many as nineteen Halam sub-tribes, indicating the complexity and fluidity of Halam ethnographic classification. Despite these variations, there is broad agreement among scholars and community members that Bongchers constitute one of the sub-tribes of the Halam tribe. The community is also referred to as *Bongshel* or *Bongshi* (Sailo, 1991, p.3). Etymologically, the term Bongcher derives from two native lexical items: *bong* means 'to cut' and *cher* means 'sprout'. The Bongchers represent one of the smallest indigenous communities in Tripura, with an estimated population ranging between 680 to 1000 individuals (Bongcher, 2011, p.vi). Most Bongchers are concentrated in settlements

located at the foothills of the Baramura range, particularly in Kulai, Ambassa, Nelsibongcher Para of Ompi. Smaller populations are also found in Agartala, Dhanalekha and Teliamura. Linguistically the language comes under Old-Kuki-Chin group. Schematically it can be represented as Sino-Tibetan > Tibeto-Burman > Kuki-Chin > Old Kuki-Chin > Bongcher.

3. Research Objectives and Methodology

The paper has discussed the TAM in the Bongcher language. The paper is divided into four sections such as tense markers in the language, aspectual discussion on Bongcher, moods in Bongcher and lastly different modal in Bongcher, which has been discussed under a single section called discussion. For this study, both the primary and secondary sources of data were employed. The secondary data has been collected from books, journals, articles, records, and the digital archive. Community speakers from Gomati district (Tripura) were interviewed during fieldwork, conducted in September and October, 2025 which served as the main source of data. Nikon D5300 digital camera and a Zoom H4N Pro audio recorder were used to record and digitize the data. We primarily distributed questionnaire related to tense, aspect and mood. Apart from this, we have also recorded folk narratives to get comprehensive data. In this data collection, total 8 community members have involved from different age groups, gender and literacy backgrounds.



Figure 1: Map of Ampi (Ompi), place of data collection

This paper aims to provide a comprehensive analysis of the distribution of mood and modality with special reference to tense and aspect in the Bongcher language. The primary goals of this paper are as follows—

a. To describe how tense and aspectual qualities are expressed in the language.

b. To analyse moods and modality in Bongcher, especially the markers.

c. To identify the individual lexical elements (if any) and morphemes (free and bound) that conveys the aspect and tense of the language.

4. Discussion

4.1. Tense

According to Payne (1997), tense is "the grammatical expression of the relation of the time of an event at some reference point of time, usually the moment the clause is uttered." Languages are able to categorize tenses according to two-way distinctions, such as past vs. non-past, future vs. non-future, and present vs. non-present) and three-way distinctions (past, present, and future). Among the Tibeto-Burman languages that regularly differentiate between future and non-future tenses are Angami (Giridhar, 1980), Dimasa (Longmailai, 2014) from the Bodo-Garo group, and Mongen Ao (Coupe, 2007) from the Naga group. Moreover, tense is a category used in the grammatical description of verbs (along with aspect and mood), referring primarily to the way the grammar marks the time at which the action denoted by the verb took place (Crystal, 2000, p.384).

The characteristics of Bongcher TAM are covered in the further details below. We can characterize the present tense in Bongcher as a situation situated temporally with or during the moment of utterance, the past tense as located prior to the moment of utterance, and the future tense as located subsequent to the moment of utterance because tense locates the time of a situation in relation to the moment of utterance (Comrie, 1976). Bongcher makes a distinction between the past, present, and future tenses. Here are some sentences from Bongcher that demonstrate the use of tense markers and basic tenses—

1. *sita* *a-se*
   [Present]
   3SG PRO PRE-go
   'Sita goes'

2. *ram* *a-se-laijiŋ*
   3SG PRO PRE-go-PROG
   'Ram is going'

3. *sita* *a-se-ta*
   [Past]
   3SG PRO PRE-go-PST
   'Sita went'

4. *ram* *a-se-jei-ta*
   3SG PRO PRE-go-COMPL-PST
   'Ram had gone'

5. *sita* *ələi-se-laijiŋ*
   3SG PST-go-PROG
   'Sita was going'

6. *sita* *se-a-ti*
   [Future]
   3SG go-PRO PRE-FUT

'Sita will go'

7. *atak* *se-a-ti-laijiŋ*
   3SG go-PRO PRE-FUT-PROG
   'S/he will be going'

In the aforementioned examples, the present tense is unmarked in Bongcher, which is really an interesting feature of the language. In examples 3—5, the past tense is marker by two separate morphemes, such as suffix */-ta/* and prefix */alәi-/*, and lastly the future tense is marked by the suffix */-ti/* in examples 6—7. Tense in Bongcher is marked by different suffixes which is again elaborated in the following examples 8—10.

8. *rubi* *minjani* *ompipara* *a-se-ta*
   3SG yesterday ompipara PRO PRE-go-PST
   'Rubi went to Ompipara, yesterday'

9. *laksmi* *avani* *udaipur* *a-se*
   3SG today udaipur PRO PRE-go
   'Laxmi goes to Udaipur, today'

10. *ram* *nautuka* *sʌbrum* *se-a-ti*
    3SG tomorrow sabroom go-PRO PRE-FUT
    'Ram will go to Sabroom, tomorrow'

The above sentences (8—10) demonstrate that the underlying verb */se/*, means 'go' stays the same even though it appears with distinct adverbial times, */minjani/* 'yesterday', */avani/* 'today', and */nautuka/* 'tomorrow'. Nevertheless, the morphemes */-ta/*, */-ti/* are added to indicate different locations in time. Separate suffixes indicate each tense; in other words, different markers are used to inflect the verb. The tense system in Bongcher is displayed in the table—1 below:

Tense	**Tense-Marker**
Present	*Ø*
Past	*/alәi-/, /-ta/*
Future	*/-ti/*

Table 1: Tense Markers

However, Bongcher's use of tense is frequently ambiguous, like other Tibeto-Burman languages. One of Bongcher's structural characteristics is the indistinctiveness of tense. Here are few examples—

11. *avani* *ku-atak* *in-a* *om-mak*
    today PRO PRE-3SG house-LOC stay-NEG
    'S/he is not at home today'

12. *minjani* *ku-atak* *in-a* *alәi om-mak*
    yesterday PRO PRE-3SG house-LOC PST stay-NEG
    'S/he was not at home yesterday'

13. *nautuka* *ku-atak* *in-a* *om-nani*
tomorrow PRO PRE-3SG house-LOC EXIS-FUT-NEG
'S/he will not be at home tomorrow'

The aforementioned examples demonstrate how the verb ends in the first two sentences */avani/*, which means 'today', and */minjani/*, which means 'yesterday'—are identical despite occurring at separate adverbial times. In other words, time reference is not grammaticalized since there is no clear agreement between the verb ends and the adverbial time. Additionally, the above examples demonstrate that the verb primarily distinguishes between the realis and irrealis forms. Therefore, it is more convenient to talk about aspect rather than tense in Bongcher.

4.2. Aspect

Aspect in Bongcher denotes an action's time distribution. Aspects are several perspectives on the internal temporal components of a situation or an event. In other words, aspect shows the internal organization of a situation or occurrence. Bongcher contains the following kinds of aspects.

4.2.1. Simple Aspect

In Bongcher, simple aspect shows simple statement, habitual, universal truth. Following examples 14—19 shows the simple aspect in Bongcher—

14. *sərat* *əini-raŋ* *dut* *a-pek*
[General statement]
cow we-DATmilk PRO PRE-give
'Cow gives us milk'

15. *kʰomol* *ni-raŋ* *a-rebel*
[Universal truth]
earth sun-ACC PRO PRE-revolves
'The earth goes round the sun'

16. *atak* *bu* *ke-nek*
[Habitual aspect]
3SG rice PRO PRE-eat-PRS
'S/he eats rice'

17. *atak* *naŋ-raŋ* *ke-di*
[Stative verbs]
3SG 2SG-DAT PRO PRE-love
'S/he loves you'

18. *atak* *bu* *ke-nek-ta*
[Simple past]
3SG rice PRO PRE-eat-PST
'S/he ate rice'

4.2.2. Completive Aspect

The completion of an action is shown by the completive aspect in Bongcher. Therefore, a completed action is referred to as the

completive aspect. As in the following examples, the completive form of the verb is created by appending the past suffix */-ta/* to the active verb together with the completive aspect marker */-jəi/*, which means 'complete or finish'.

19. *atak* *bu* *ke-nek-jəi-ta*
    3SG rice PRO PRE-eat-COMPL-PST
    'S/he has eaten rice'

20. keima a-se-jəi-ta
    1SG PRO PRE-come-COMPL-PST
    'I have gone'

21. *ani* *a-se-jəi-ta*
    3PL PRO PRE-come-COMPL-PST
    'They have gone'

4.2.3. Incompletive Aspect

In Bongcher, an incompletive aspect denotes an incomplete action. A faulty action at a specific point in the past is referred to as the incompletive aspect. The suffix */-niŋ/* is added to the verbal root to produce the verb's incomplete form, as in the following examples (22) & (23)

22. *atak* *tʰeihai* *ročok-niŋ* *ka-sak-ta*
    3SG mango buy-ICPL PRO PRE-eat-PST
    'S/he bought mango and ate'

23. *kəima* *kančili* *pak-ka* *ročok-niŋ* *ki-ni-pek*
    1SG shirt CLF-one buy-IPLC me-to-give
    'I bought a shirt for myself'

4.2.4. Progressive Aspect

In Bongcher, the progressive aspect denotes the continuation of an action. Regarding the progressive aspect, Bongcher presents a two-way contrast in tense, such as past and present but no future. The marker */-laijiŋ/* expresses the progressive aspect. A few examples that illustrate Bongcher progressive aspect are provided below.

24. *kəima* *twi* *a-in-laijiŋ*
    1SG water PRO PRE-drink-PROG
    'I am drinking water'

25. *kəima* *twi* *aləi-in-laijiŋ*
    1SG water PST-drink-PROG
    'I was drinking water'

4.2.5. Irrealis Aspect

In Bongcher, the term 'irrealis' or 'unrealized aspect' refers to an action that will occur soon; as a result, it is exclusively used in the future tense with the marker */-ti/*. In the irrealis aspect, the following examples discuss in details.

26. *naŋ* *kančili* *pak-ka* *ročok-ki-ti*
    2SG shirt CLF-one buy-PRO PRE-FUT
    'You will buy a shirt'

27. *kəima* *čə* *putla* *le* *rədai-a-ti*
    1SG DEM doll with play-PRO PRE-FUT
    'I will play with the doll'

28. *ram* *bazar-a* *se-a-ti*
    ram market-LOC go-PRO PRE-FUT
    'Ram will go to bazaar'

4.3. Mood

Mood indicates syntactic and morphological characteristics of verbs, instead of verbal inflections that enable speakers to convey their attitude toward what they are saying (Palmer, 2001). Following section gives a detailed discussion of the various moods found in Bongcher language.

4.3.1. Indicative Mood

In Bongcher, factual and habitual assertions are made using the indicative mood. It also describes the status of the action in connection to the verbal act or the moment the action occurs. As will be covered below, the indicative mood in Bongcher is expressed for three tenses—present, past and future.

In Bongcher, the present indicative describes the action occurring or the situation at the moment of the speech act. It also refers to statements that are factual and actions that are routinely performed. Here are the few examples—

29. *sita* *a-er*
    [Active]
    3SG PRO PRE-cook
    'Sita cooks'

30. *hima* *in* *pak-ka* *a-om*
    [Stative]
    DEM house CLF-one PRO PRE-EXIS
    'This is a house'

31. *kʰomol* *a-čeŋ*
    [Factual]
    earth PRO PRE-round
    'Earth is round'

32. *ram* *ani* *librari* *a-se*
    [Habitual]
    3SG everyday library PRO PRE-go
    'Ram goes to library everyday'

The activity that occurred prior to the speech act is referred to as past indicative. In Bongcher, the non-future suffix */-ta/* is added to the past indicative suffix for both stative and active verbs. For instance—

33. *ram* *a-mu-ta*
    [Active]
    3SG PRO PRE-see-PST
    'Ram saw'

34. *naŋ* *kə-rətəi-ta*
    2SG PRO PRE-run-PST
    'You ran'

35. *atak* *kolom* *kə-nəi-ta*
    [Stative]
    3SG pen PRO PRE-has-PST
    'S/he had a pen'

36. *niŋ* *aləi* *daktor*
    2SG PST doctor
    'You were a doctor'

The future indicative describes a certain activity that will undoubtedly be carried out after the speech act. In Bongcher, the future tense marker */-ti/* indicates both the definite and the dubitative future suffix. Both active and stative verbs utilize the future suffix */-ti/*. Here are some instances—

37. *ani* *rəsəi* *a-ti*
    [Active]
    3PL learn PRO PRE-FUT
    'They will learn'

38. *kəima* *bazar-a* *se-ki-ti*
    1SG bazaar-LOC go-PRO PRE-FUT
    'I will go to the market'

39. *atak* *kolom* *kʰat-ka* *nəi-ki-ti*
    [Stative]
    3SG pen CLF-one have-PRO PRE-FUT
    'S/he will have a pen'

40. *ram* *daktor* *om-a-ti*
    3SG doctor EXIS-PRO PRE-FUT
    'Ram will be a doctor'

It should be noted that, as demonstrated in the examples below 41—43, Bongcher expresses the progressive tense in indicative mood by appending the progressive aspect suffix marker */-laijiŋ/* to the verb.

41. *kəima* *a-rətəi-laijiŋ*
    1SG PRO PRE-run-PROG

'I am running'

42. *kəima* *ələi* *a-rətei-laijiŋ*
    1SG PST PRO PRE-run-PROG
    'I was running'

43. *kəima* *rətəi-laijiŋ* *a-ti*
    1SG run-PROG PRO PRE-FUT
    'I will be running'

4.3.2. Imperative Mood

The imperative mood in Bongcher denotes command or request to perform certain action by the listener. The subject of the imperative mood is always in the second person and often may not be overtly expressed. The imperative form of the verb is obtained by adding the imperative suffix */-ro/* to the active verb. On the other hand, the polite form of the imperative is created by prefixing the active verb with */rəŋmuŋin/* stands for 'please' as in examples (46) & (47). The following examples illustrate this—

44. *rətəi-ro*
    run-IMP
    '(You) run

45. *rəsəi-ro*
    learn-IMP
    '(You) learn

46. *rəŋmuŋin* *ram-raŋ het-ro*
    please 3SG-ACC call-IMP
    'Please call Ram'

47. *rəŋmuŋin* *ink^{h}a* *rohoŋ-ro*
    please door open-IMP
    'Please open the door'

4.3.3. Hortative Mood

In Bongcher, the hortative mood is employed to express a certain desire or to imply that a certain action will be taken. The hortative suffix */-ri/*, or */-rese/* is added to the active verb in Bongcher to produce the hortative form of the verb. Here are such instances—

48. *mastor* *ačaŋ* *ri-ki-ti*
    teacher become HORT-1SG-FUT
    'Let me become a teacher'

49. *atak-raŋ se-ri-an-ti*
    3SG-ACC go-HORT-PRO PRE-FUT
    'Let her go'

50. *ani* *bu* *a-sak-rese*
    3PL rice PRO PRE-eat-HORT
    'Let them eat'

4.4. Modal Construction in Bongcher

Bongcher uses specific 'modal structures' to convey the subjunctive mood. The simple finite verb is used with specific affixes to create the modal forms in Bongcher. Below is a detailed discussion of various modal-formations in Bongcher—

51. *kəima la a-sak-tʰəi-ta*
    [Modal of Ability][5]
    1SG song PRO PRE-sing-MOD ABL-PST
    'I could sing'

52. *sita la a-sak-tʰəi*
    3SG song PRO PRE-sing-MOD ABL
    'Sita can sing'

53. *niŋ raŋnaŋ a-nek*
    [Modal of Certainty][6]
    2SG MOD CER PRO PRE-eat
    'You must eat'

54. *atak la raŋnaŋ a-sak-ta*
    3SG song MOD CER PRO PRE-sing-PST
    'S/he must have sang the song'

55. *niŋ kʰɔm a-nek-jəi-ta val*
56. [Modal of Probability][7]
    2SG also PRO PRE-eat-COMPL-PST MOD PRO
    'You might have eaten'

57. *laksmi kʰɔm a-hoŋ val*
    3SG also PRO PRE-come MOD PRO
    'Laxmi may also come'

58. *atakma hoŋ-inlaŋ kəima se-ki-ti*
    [Modal of Condition][8]
    3SG come-MOD CON1SG go-PRO PRE-FUT
    'If she comes, I will go'

59. *hana naŋma om-inlaŋ*
    DEM 2SG EXIS-MOD CON
    'If you were there'

60. *atak kʰamkʰa tʰo-ki-ti*
    [Modal of Determination][9]
    3SG MOD DET do-PRO PRE-FUT
    'S/he will do it'
61. *naŋ kʰamka se-ki-ti*
    2SG MOD DET go-PRO PRE-FUT
    'You will go

62. *atak* *bu* *ka-nek-nəŋ*
[Modal of Necessity][10]
3SG food PRO PRE-eat-MOD NEC
'S/he needs to eat food'

63. *naŋ* *boi* *kʰat-ka* *ke-nəŋ*
2SG book CL-one PRO PRE-MOD NEC
'You need a book'

5. Conclusion

The paper presents a descriptive and analytical study of the grammatical categories of tense, aspect, mood, and modality in Bongcher, a lesser-documented Tibeto-Burman language belonging to the Kuki-Chin branch and spoken by Bongcher community of Tripura, India. The study is significant because linguistic descriptions of Bongcher remain scarce, and the paper contributes to the documentation and preservation of an endangered indigenous language. The analysis demonstrates that Bongcher formally distinguishes three tenses present, past, and future. The present tense is generally unmarked, while the past is expressed through the prefix /-*alə*i/ and the suffix /-*ta*/, and the future is marked by /-*ti*/. Nevertheless, the study observes that tense distinctions are frequently ambiguous and that temporal reference often depends on contextual adverbials rather than verbal inflection alone. This feature aligns Bongcher with many other Tibeto-Burman languages where the grammatical opposition between realis and irrealis is often more prominent than a rigid tense system. Consequently, the author suggests that aspectual distinctions provide a more reliable framework for understanding temporal reference in the language. Five aspectual categories are identified: simple, completive, incompletive, progressive, and irrealis. The simple aspect is used for habitual actions, universal truths, and general statements. The completive aspect, formed through the marker /-*jə*i/, denotes actions that have been fully completed. The incompletive aspect, marked by /-*niŋ*/, signals unfinished or partially completed events. Progressive aspect is expressed through /-*laijiŋ*/ and highlights actions in progress, while irrealis aspect is associated with future or unrealized events and is marked by /-*ti*/. The paper further identifies three major mood categories indicative, imperative, hortative. Subjunctive moods are realized through modal constructions. Particularly noteworthy is the rich inventory of modal markers. The suffix /-*tʰə*i/ express ability, /-*raŋnaŋ*/ indicates determination, and /-*nəŋ*/ conveys necessity. These markers illustrate how Bongcher grammaticalizes different shades of epistemic and deontic modality. The study demonstrates that Bongcher possesses a highly structured TAM system characterized by the

interaction of tense, aspect, mood, and modality. The findings not only enrich our understanding of Kuki-Chin linguistics but also provide valuable documentation of an underrepresented language, thereby contributing to broader typological and comparative studies of Tibeto-Burman languages.

Abbreviations

1SG—first person singular
2SG—second person singular
3SG—third person singular
3PL—third person plural
ACC—accusative case
CLF—classifier
COMPL—completive
EXIS—existential copula
FUT—future
HORT—hortative
ICPL—incompletive
MOD ABL—modal of ability
MOD CER—modal of certainty
MOD CON—modal of condition
MOD NEC—modal of necessity
MOD PRO—modal of probability
MOD DET—modal of determination
PRS—present
PROG—progressive
PST—past

References

Abbi, A. (2001). *A manual of linguistic fieldwork and structures of Indian languages*. Lincom Europa.

Bauman, J. J. (1975). *Pronouns and pronominal morphology in Tibeto-Burman* [Unpublished doctoral dissertation]. University of California, Berkeley.

Bhat, D. N. S. (1999). *The prominence of tense, aspect and mood*. John Benjamins.

Bhat, D. N. S., & Ningomba, M. S. (1997). *Manipuri grammar*. Lincom Europa.

Binnick, R. I. (1991). *Time and the verb*. Oxford University Press.

Bohnemeyer, J. (2019). Elicitation and documentation of tense and aspect. *Language Documentation and Conservation*.

Brahma, D. (2014). Inflectional processes of tense and aspect in Bodo. *Language in India, 14*(1).

Chung, S., & Timberlake, A. (1985). Tense, aspect, and mood. In T. Shopen (Ed.), *Language typology and Syntactic Description: Vol. 3. Grammatical categories and the lexicon* (pp. 202–258). Cambridge University Press.

Comrie, B. (1985). *Tense*. Cambridge University Press.

Coupe, A. R. (2013). Tense, but in the mood: Diachronic perspectives on the representation of time in Ao. *Language and Linguistics, 14*(6), 1105–1138.

Crystal, D. (2000). *A dictionary of linguistics and phonetics*. Blackwell.

Dahl, Ö. (1985). *Tense and aspect systems*. Basil Blackwell.

Givón, T. (1984). *Syntax: A functional-typological introduction*. Vol.1 John Benjamins.

Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press. https://doi.org/10.1017/CBO9781139167178

Singh, C. Y. (1999). Tense and aspect in Kuki-Chin. *Linguistics of the Tibeto-Burman Area, 22*(2)

Sailo, S. 1992. *The Bongchers.* Directorate of research, government of Tripura.

[1] People in contemporary Western civilization are used to precise descriptions of time, locations and other time-related phenomena. At home, a lot of people talk about minuscule intervals of time, like minutes, seconds etc. It is theoretically possible to achieve an unlimited degree of precision by combining the linguistic possibilities with those of normal mathematical motion. However, such accuracy is not achievable in many other cultures, possibly only through direct appropriation of phrases from more advanced societies. In fact, some cultures place relatively little importance on temporal location accuracy; in Yidiny, for example, it is impossible to lexically distinguish between the concepts of 'today' and 'now' (Dixon, 1977, p.498-499).

[2] Apart from the aforementioned binary distinctions, there is one more distinction i.e. 'realis' and 'irrealis'. According to Mithun (1999: 173), 'the realis portrays situations as actualized, as having occurred or actually occurring, knowable through direct perception' On the other hand, 'irrealis portrays situations as purely within the realm of thought, knowable only through imagination' (Palmer 2001: 1). The concept of realis and irrealis can be better realised in Papuan language Amele, which can distinguish it 'remote past', 'habitual past' and 'present' from 'future', 'imperative', 'hortative', 'counterfactual' and 'negative'. For instance—Palmer (1990, p.371-405)

a. *ho* *bu-busal-en* *age qo-in*
pig SIM run out-3SG+DS+REAL 3PL hit-PL+REM.PAST
'They killed the pig as it ran out'

b. *ho* *bu-busal-eb* *age qo-qag-an*
pig SIM run out-3SG+DS+IRR 3PL hit-3PL-FUT
'They will kill the pig as it runs out'

The above examples show that the realis marker *–en* is used to mark the 'remote past', whereas the irrealis marker *–eb* is used to denote the 'future'.

[3] From a linguistic point of view, tense refers to how situations are described linguistically in terms of their temporal position; in many languages, this is done through grammatical phenomena. However, the grammatical categories that express time as verbal particles make comprise a group that makes it possible to express information about the temporality of the situation in a less accurate way than other ways. Only the anteriority, simultaneity, and posterity of a circumstance in relation to another on a timeline are typically coded by the grammatical categories that represent time in languages (Comrie, 1985).

[4] According to Smith (1991, p.22), 'The concepts of aspect play a role in all languages, so far as we know. And the aspectual systems of different languages are strikingly similar […]. We define in Universal Grammar a maximally general aspectual system. It has the basic aspectual categories that generally occur in languages of the world, and on which particular variation depends.

[5] In Bongcher, the use of */-tʰəi/* typically indicates the modal of ability. In Bongcher, the modal of ability is formed by the tense-marking the base verb with suffix */-tʰəi/*.

[6] Modal of Certainty is structured out of prefixing */raŋnaŋ/* to the following finite verbs of present, past and future indicative mood.

[7] In Bongcher language, a different modal of probability marker */val/* is employed in relation to food and beverages, whereas the modal of probability can also be expressed by the suffix marker */-tʰəi/*, which is again polysemous to modal of ability. For instance—

c. *ram* *kʰɔm* *a-hoŋ-tʰəi*
3SG also PRO PRE-come-MOD PRO
'Ram will also come'

On the other hand, the optional addition of the word */ kʰɔm/* fulfils the grammatical requirements.

[8] In Bongcher, the conditional mood is conveyed by appending /*–inlaŋ/*, which means 'if', to either active verb */čaŋ/* or */om/* which states 'remain or stay'.

[9] Bongcher expresses modal of determination by the suffix */-kʰamkʰa/*, which has been shown in the examples 59—60.

[10] In Bongcher, present, past and future tense markers are combined with the suffix */-nəŋ/*, which means 'need' to obtain modal of necessity.

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 219-228

Nominal Inflection and Case Alignment in Bhadarwahi: A Descriptive Study

Aqib Majeed[*]
Sabiya Rafiq[**]

Abstract

Bhadarwahi is a Northern Indo-Aryan language belonging to the Western Pahari subgroup, spoken primarily in the mountainous Bhadarwah and Chinta valleys within the Doda district of Jammu and Kashmir, India. Despite its substantial speaker population and complex regional contact ecology, descriptive documentation of its inflectional morphology remains limited. This study examines the nominal inflection and syntactic case alignment of Bhadarwahi using primary field data gathered from 200 native speakers through structured wordlists and sentence-elicitation schedules. The findings demonstrate that Bhadarwahi organizes nominals along a binary gender opposition (masculine and feminine) and a two-tier number distinction (singular and plural). These categories are realized through terminal vocalic alternations, overt suffixation, and suppletive stems. Syntactically, Bhadarwahi exhibits a split-ergative case alignment conditioned by transitivity and aspect. Core arguments and peripheral roles are indexed via a combination of bound case markers most notably the homophonous ergative and dative suffix /-e:/ and postpositional case assigners governing oblique bases. This paper provides a descriptive and typological analysis of nominal declension, pronominal paradigms, and case marking in the language.

Keywords: Bhadarwahi, Western Pahari, Nominal Inflection, Split Ergativity, Case Alignment, Indo-Aryan

1. Introduction

Bhadarwahi is spoken predominantly in the Bhadarwah subdivision and surrounding areas in the Doda district of Jammu and Kashmir, situated along the northern slopes of the middle Himalayas. In classical classifications, notably George A. Grierson's *Linguistic Survey of India* (1916, 1919), Bhadarwahi was placed within the Western Pahari cluster of the Northern Indo-Aryan group, grouped alongside neighboring varieties such as Bhalesi, Paddari, and Pangwali. While early comparative surveys occasionally classified it with mixed border

[*] University of Kashmir; akibmajeedakhoon@gmail.com
[**]University of Kashmir; sabiarafiq11@gmail.com

dialects of Kashmiri south of the Pir Panjal, modern typological accounts consistently group it within Western Pahari (Masica 1991; Dwivedi 2013, 2015).

Bhadarwahi functions as the primary regional lingua franca for more than 120,000 speakers across the Bhadarwah and Chinta valleys. It follows an unmarked Subject-Object-Verb (SOV) clause structure and shares considerable lexical similarity (approximately 50% to 55%) with contiguous varieties such as Bhalesi, Padderi, and Siraji, while remaining mutually unintelligible with official regional standards like Dogri, Kashmiri, and Standard Hindi.

Prolonged sociolinguistic contact with Hindi/Urdu, Dogri, and Kashmiri has introduced significant lexical borrowings and writing practices; however, systematic empirical accounts of its core morphology remain sparse. Extant studies have addressed phonology (Sheikh & Zarger, 2017), broad typological features (Dwivedi, 2013, 2015), and derivational processes (Khan & Nazir, 2018). This study provides an empirical descriptive account of nominal inflection and case alignment in Bhadarwahi, detailing its gender system, pluralization processes, pronominal paradigms, and syntactic case mechanisms.

2. Methodology and Data Collection

The empirical findings presented in this paper are based on fieldwork conducted across educational institutions in Bhadarwah town and the Chinta valley, specifically Bhaderwah Campus of the University of Jammu, Government Degree College Bhaderwah, and Government High School Chinta Valley.

The sample comprised 200 native Bhadarwahi speakers (males and females aged 13 to 22) who use Bhadarwahi as their first language across primary social and domestic domains. Data elicitation employed both qualitative and quantitative techniques using two formal instruments:

- A 500-word nominal questionnaire covering semantic fields including kinship, tools, material objects, natural topography, animals, and abstract qualities.
- A 250-sentence schedule designed to evaluate syntactic case marking, argument alignment, postpositional governance, and verbal agreement.

The elicitation procedure proceeded through three distinct stages:

- **Phase 1:** Preliminary administration and clarification of items in the contact language (Urdu/Hindi).
- **Phase 2:** High-fidelity audio recordings where each lexical item and sentence structure was repeated three times by native informants.

- **Phase 3:** Narrow phonetic transcription, morphemic segmentation, and cross-informant validation to verify morphological consistency.

Analytical conventions adhere to standard descriptive linguistic methodology and the Leipzig Glossing Rules.

3. Nominal Classes and Derivational Morphology

Nouns in Bhadarwahi fall into simple (uninflected monomorphemic roots, e.g., */da:r/* 'door') and complex classes (derived through affixation or compounding). Semantically and structurally, they are organized as follows:

- **Countable vs. Uncountable Nouns:** Countable nouns freely pluralize and co-occur with numeral qualifiers (e.g., */tsajjõ:/* 'apple', */tshali://* 'goat', */khide:no:/* 'toy'). Uncountable nouns denote mass, liquid, or ambient substances and reject direct plural marking without measure words (e.g., */pa:ni:/* 'water', */ba:t/* 'air', */dhu:d/* 'dust', */dud/* 'milk').

- **Concrete vs. Abstract Nouns:** Concrete nouns denote tangible, perceptual entities (e.g., */tʃʰamtʃo:/* 'spoon', */gho:ɖ/* 'stone', */bi:do:/* 'button'). Abstract nouns designate emotional states, periods of life, or qualities (e.g., */ɖar/* 'fear', */nikda:re:/* 'childhood', */khuʃi://* 'happiness', */a:za:di://* 'freedom') and are typically non-inflecting for number.

- **Verbal Nouns (Gerunds):** Gerundial forms functioning as nominal heads are productively derived by suffixing the infinitival morpheme */-no/* to the verb root: */natsno/* 'dancing', */kha:no/* 'eating', */likhno/* 'writing', */tʃalno/* 'walking'.

- **Compound Nouns:** Compounds are formed through stem juxtaposition, particularly in complex kinship nomenclature where the prefix */par-/* denotes an ascending or descending genealogical remove:

  - */par-daddʒi:/* 'paternal great-grandfather'
  - */par-da:di:/* 'paternal great-grandmother'
  - */par-nandʒi:/* 'maternal great-grandfather'
  - */par-na:ni:/* 'maternal great-grandmother'

4. Nominal Inflectional Systems

4.1 Grammatical Gender

Bhadarwahi maintains a two-gender system: masculine (MAS) and feminine (FEM). Grammatical gender is assigned to all nouns; for inanimate objects, gender assignment is largely arbitrary (e.g., */pəkkho/* 'fan' is masculine, whereas */andəç/* 'eye' is feminine). For animate nouns, grammatical gender typically corresponds to biological sex. Feminine counterparts are derived from masculine stems through three

primary morphological processes:

4.1.1. Stem-Vowel Modification (/-o:/ → /-i:/) Masculine nouns ending in the back vowel */-o:/* systematically shift to the high front vowel */-i:/* to form the feminine:

- */benezo:/* 'sister's son' → */benezi:/* 'sister's daughter'
- */sa:lo:/* 'wife's brother' → */sa:li:/* 'wife's sister'
- */ʧa:ʧo:/* 'paternal uncle' → */ʧa:ʧi:/* 'paternal uncle's wife'
- */go:ɖo:/* 'horse' → */go:ɖi:/* 'mare'

4.1.2. Suffixation

- **Honorific Root Substitution:** In kinship roots marked by the masculine honorific morpheme */-dzi:/*, this marker is elided and replaced with the feminine suffix */-i:/*:
  - */da:ddzi:/* 'paternal grandfather' → */da:di:/* 'paternal grandmother'
  - */na:ndzi:/* 'maternal grandfather' → */na:ni:/* 'maternal grandmother'
  - */ma:mdzi:/* 'maternal uncle' → */ma:mi:/* 'maternal uncle's wife'
  - */ma:sedzi:/* 'maternal aunt's husband' → */ma:si:/* 'maternal aunt'

- **Occupational/Social Suffix /-ni:/:** Attached to consonant-final stems denoting trades and animals:
  - */suna:r/* 'goldsmith' → */suna:rni:/* 'female goldsmith'
  - */loha:r/* 'blacksmith' → */loha:rni:/* 'female blacksmith'
  - */kuma:r/* 'potter' → */kuma:ni:/* 'female potter'
  - */mu:ʃo:/* 'male rat' → */mu:ʃni:/* 'female rat'

4.1.3. Suppletion Several core animate kinship terms and domestic animal names rely on distinct lexical roots without morphophonological relatedness:

- */ʃe:ro/* 'father-in-law' */ʃikʰ//* 'mother-in-law'
- */ɖila/* 'brother' */beʲn/* 'sister'
- */mũ:ʃ/* 'husband' */laɖi/* 'wife'
- */garza:l/* 'male spider' */gra:ʃɖi:/* 'female spider'

4.2 Number Marking

Grammatical number is categorized into Singular (SG) and Plural (PL). Plurality is realized via terminal vocalic shifts, suffixation, or zero marking:

4.2.1. Vocalic Shift (/-o:/ \→ /-e:/) Masculine nouns ending in */-o:/* alternate to */-e:/* in the plural nominative:

- */ʃi:ʃo:/* 'mirror' → */ʃi:ʃe:/* 'mirrors'

- */gide:lo:/* ‘pillow’ → */gide:le:/* ‘pillows’
- */go:ɖo:/* ‘horse’ → */go:ɖe:/* ‘horses’
- */dando:/* ‘stick’ → */dande:/* ‘sticks’
- */kango:/* ‘comb’ → */kange:/* ‘combs’
- */basto:/* ‘bag’ → */baste:/* ‘bags’

4.2.2. Suffixation

- **Suffix /-a:/:** Productively added to consonant-final bases:
  - */kutar/* ‘dog’ → */kutra:/* ‘dogs’
  - */ligəd/* ‘cloth’ → */ligəda:/* ‘clothes’
  - */kantʰ/* ‘wall’ → */kantʰa:/* ‘walls’
  - */zib/* ‘tongue’ → */ziba:/* ‘tongues’
- **Nasalized Suffix /-ã:/:** Attached to specific nominal stems:
  - */ka:kil/* ‘paper’ → */ka:kilã:/* ‘papers’
  - */ghar/* ‘house’ → */ghara:/* | */gharã:/* ‘houses’
  - /mətthu/ ‘boy’ → */mətthã:/* ‘boys’
  - */da:r/* ‘door’ → */da:rã:/* ‘doors’

4.2.3. Zero Modification

Certain nominal classes maintain invariant phonological forms across both numbers:

- */tama:tar/* ‘tomato / tomatoes’
- */kentʃi:/* ‘scissors (sg. / pl.)’
- */tʃe:ko/* ‘knife / knives’

4.2.4. Pronominal Paradigms

Bhadarwahi personal pronouns display inflectional sensitivity to person (1st, 2nd, 3rd), number (singular, plural), case (nominative, ergative, possessive/genitive), and social honorificity.

Person / Honorificity	Subject (NOM)	Ergative (ERG)	Possessive (GEN)
1SG	*âũ* (‘I’)	*mĩ*	*mero / meri / mere*
1PL	*əsa* (‘we’)	*əseĩ*	*iʃo / iʃi / iʃe*
2SG (-HON)	*tu* (‘you’)	*tusəĩ*	*təro / təri / təre*
2SG (+HON)	*tus* (‘you’)	*tusəĩ*	*tuʃo / tuʃi / tuʃe*
2PL (-HON)	*tu* (‘you’)	*tuseĩ*	*tero / teri / tere*
2PL (+HON)	*tusa* (‘you’)	*tuseĩ*	*tuʃõ / tuʃĩ / tuʃẽ*
3SG (Prox./Dist.)	*tu / u* (‘he/she’)	*teni* (M), *tesa* (F)	*usero / tesero*
3PL (-HON)	*ten* (‘they’)	*unei*	*unkero*
3PL (+HON)	*tena* (‘they’)	*unei*	*tenkero*

(Note: Possessive clitics alternate in /-o, -i, -e/ in gender and number concord with the possessed head noun).

5. Case Marking and Syntactic Alignment

Nouns inflect across three primary structural case forms: Nominative, Oblique, and Vocative. The oblique form serves as the base for bound case inflections and free postpositions.

Declension Paradigm of Masculine Noun *mətthu* ('boy'):

- **Nominative:** Singular *mətthu*, Plural *mətthe*
- **Oblique:** Singular *mətthe*, Plural *mətthɑ*
- **Vocative:** Singular *mətthɑ*, Plural *mətthɔ*

In this declension matrix, syncretism is evident: the nominative plural is phonologically identical to the oblique singular (*mətthe*), while the oblique plural merges with the vocative singular (*mətthɑ*).

Bhadarwahi displays a split-ergative alignment typical of Western Pahari languages. In past/perfective aspectual domains, subjects of transitive verbs receive overt ergative marking, while intransitive subjects and direct objects occur in the unmarked nominative/absolutive. In imperfective/non-past domains, nominative-accusative alignment is observed.

5.1 Ergative Case (/-e:/)

The ergative case identifies the noun as a subject of a transitive verb in ergative-absolutive languages. This is a marked case. In Bhaderwahi / e: /, ergative marker is used in a subject with a transitive verb.

Examples:

rəmeʃ - e: kəm kero
Ramesh -ERG work do.PST
'Ramesh did the work.'

rəmeʃ - e: hari: pe s ditto
Ramesh - ERG Hari money give.PST
'Ramesh gave money to Hari.'

ten – i məʈʈʰero kʰija:l rəkʰo tʰu:
He-ERG child.POSS care put.PERF AUX.PST
'He took care of the child.'

məʈʈʰ e: pa:ɳi pijo:
Boy -ERG water drink.PST
'The boy drank water.'

5.2 Dative Case

Dative case is used to indicate the noun to which something is given i.e. it is assigned to indirect object of a verb. / e:/ marker is used for (Dat) in Bhaderwahi language. Some examples are given below:

ten - i məʈʈʰ - e: sabak paɖʰ o
He - ERG boy - DAT lesson teach.PST

‘He taught a lesson to the boy.’
kriʃ - e: oːmkaːr - e: pe s ditto
Krish - ERG omkaar- DAT money give.PST
‘Krish gave money to Omkaar.’

raːm - e: məʈʈʰ - e: kitaːb ditto
Ram - ERG boy - DAT book give .PST
‘Ram gave book to the boy.’

siːtaː oːmkaːr- e: pen ditto
Sita Omkaar- DAT pen give.PST
‘Sita gave a pen to Omkaar.’

5.3 Nominative Case

Nominative case generally marks the subject of a verb or the predicate noun or predicate adjective. There is no overt marker for nominative case in Bhadarwahi language. The nominative case marker is usually denoted by the symbol ø.

Examples are given below:

əsa baːzaːre ʤoːjeː
1P-PL-NOMø market go.PRST
‘We go to the market.’

ə iɖiː eʤtaː
1P-SG-NOMø here come.PRST
‘I come here.’

baʧʧoː iɖiː orotʰu eː
Child- NOMø here come.PST
‘The child came here.

5.4 Accusative Case

Accusative case is used to mark the direct object of a transitive verb. Some of the examples in Bhaderwahi language are as under:

teni goːɖoː maːroː
3P-M-S horse-ACC kill.PST
‘He killed a horse.’

unei tsoːr tsillə
3P-PL thief- ACC catch.PST
‘They caught the thief.’
bəʧʧeːni matsʰliː tsiloji:

3P-PL fish-ACC catch.PST
‘The children caught the fish.’

5.5 Instrumental Case

Instrumental case is used to indicate the noun as an instrument or means by which the action is accomplished by the subject. In Bhaderwahi language a postposition /si:/ usually assigns an

instrumental case to the noun. Some examples are given below:

raːm - eː kʰat likʰoː peneː siː
Ram - DAT letter write.PST pen –INS with (P)
'Ram wrote the letter with a pen.'

oːmkaːr- eː gejəɖ kaːʈo niku siː
Omkaar tree cut.PST axe –INS with(P)
'Omkaar cut the tree with an axe.'

teni ganɖoː kaːʈo ʧeːko siː
He onion cut.PST knife- INS with(P)
'He cut an onion with a knife.'

5.6 Locative Case

Locative case indicates a location of a subject. In Bhadarwahi the proposition / :/, /puɖ/, etc are the case assigners. Some of the examples are given below:

ə apneː gʰareː :
1P-Sg my house-LOC in(P)
'I am in my house.'

toːtoː geːɖeː puɖ eː
Parrot tree-LOC on(P) be.PRST
'The parrot is on the tree.'

5.7 Ablative Case

Ablative case is used to express the motion of the subject away from something. Bhadarwahi language has proposition /puɖɑː/ as case assigner. Examples are given below:

teni geːɖeː puɖaː ʧʰaːl maːriː
3P-SG tree- ABL from jump.PST
'He jumped from the tree.'

paʈʰla lan zaɖne geːɖeː puɖaː
Leaves fall.PRST tree-ABL from(P)
'The leaves fall from the tree

6. Conclusion

Bhadarwahi nouns are inflected for gender, number, and case by morphosyntactic categories. Nouns are masculine or feminine, and count or non-count. All nouns in Bhadarwahi are assigned a grammatical gender, which may or may not coincide with natural gender e.g. pəkkʰo 'fan' is masculine; andəç 'eye' is feminine. The distribution of gender to inanimate objects is arbitrary, i.e. any inanimate object may get either masculine or feminine gender and we can't reason why we are assigning either masculine or feminine to them. Human and animal nouns sometimes have a single lexical form which represents both the genders as in kaw 'crow', dlɑg 'cheetah', bəʈʃʈʃo 'kid', etc. A borrowed noun from a foreign language also takes only one form to representeither of the genders e.g. grɑunɖ 'ground'.

Common nouns inflect for gender, number and case. Proper nouns are assigned to specific gender categories and remain the same in all numbers and cases. Bhadarwahi has two genders: masculine and feminine. Both animate nouns and inanimate nouns get a gender.

According to Masica (1991: 219): "In the New IndoAryan…gender is an inherent and classificatory property of one class of words (nouns) and a variable or inflectional property of others (adjectives, certain verbal forms, sometimes pronouns and adverbs), and one extremely important postpositions."Assigning gender to nouns is arbitrary. Abstract nouns generally do not inflect for numbers with some exception, but concrete nouns inflect. Non-count concrete nouns are used with partatives e.g. the gender formation involves two processes: (a) suffixation, and (b) phonological changes (suppletion, etc.). Most males are masculine nouns. Most of the masculine nouns get suffixed with /o/ sound, and the rest is consonant ending. For example, gʰoɽo 'horse', ʧaʧo 'uncle', bəʧʧo 'kid', kɑw 'crow', səpp 'snake', luhɑr 'black-smith', kumhɑr 'potter', pəkkʰo 'fan' alike. Most females are feminine nouns. Most of the feminine nouns get suffixed with /i/ sound, and rest are consonant ending, such as gʰoɽi 'mare', ʧaʧi 'aunt', luhɑrni 'blacksmith's wife', kumhɑrni 'potter's wife', ʧəpli 'slippers' kutəri 'bitch', bəççʰi 'cow' etc. There are nouns which are inherently masculine or feminine as miʈl 'friend' a masculine noun (used for both masculine and feminine), bidɑli 'ant' a feminine noun (used for both masculine and feminine), etc. In Bhaderwahi language, case morphology plays an important role in indicating the grammatical relationships between nouns and other parts of the sentence. The language has a rich case system with various cases that are marked through inflectional endings. Here are some common cases found in Bhaderwahi:

1. Nominative case: This case is used for the subject of a sentence.
2. Accusative case: This case is used for the direct object of a transitive verb.
3. Ergative case: The ergative case identifies the noun as a subject of a transitive verb in ergative-absolutive languages. This is a marked case. In Bhaderwahi / eː /, ergative marker is used in a subject with a transitive verb.
4. Dative case: This case is used for the indirect object or recipient of an action. / eː/ marker is used for (Dat) in Bhaderwahi language.
5. Genitive case: This case is used to indicate possession or association.
6. Locative case: This case is used to indicate location or place. In Bhadarwahi the proposition / eː/, /puɖ/, etc are the case assigners.
7. Ablative case: This case is used to indicate movement away from a location or source. Bhadarwahi language has proposition /puɖɑː/ as case assigner.

8. Instrumental case: This case is used to indicate the means or tool of an action. In Bhaderwahi language a postposition /si:/ usually assigns an instrumental case to the noun.

Each case has its own distinct forms, which are added as suffixes to the nouns to indicate their grammatical function in a sentence. These inflectional endings vary depending on factors like noun gender, number, and animacy.

Pronoun morphoplogy is connected with the different forms of pronons, gender, number, animacy, and deictic categories.

This study summarizes some grammatical features of the Bhadarwahi language. Due to prolonged contact with Hindi, Dogri and Kashmiri speakers, the natives of Doda and Bhadarwah have borrowed script and few lexical items from these languages. Despite this, Bhadawahi is a different language and the language does not show any mutual intelligibility with Dogri, Hindi and Kashmiri. However, other languages spoken it the region such as Pangwali, Siraji, Padri and Bhalesi show lexical similarity up to fifty to fifty five percent.

References

Bailey, T. G. (1908). *The languages of the Northern Himalayas: Being studies in the grammar of twenty-six Himalayan dialects* (Asiatic Society Monographs, Vol. 12). Royal Asiatic Society.

Dwivedi, A. V. (2013). *A grammar of Bhadarwahi*. Lincom Europa.

Dwivedi, A. V. (2015). *Bhadarwahi: A typological sketch. Acta Linguistica Asiatica, 5(1),* 125–148.

Grierson, G. A. (1916). Linguistic *survey of India: Indo-Aryan family, North-Western group* (Vol. 8, Part 1). Office of the Superintendent of Government Printing.

Grierson, G. A. (1919). *Linguistic survey of India: Indo-Aryan family, Central group* (Vol. 9, Part 4). Office of the Superintendent of Government Printing.

Khan, H., & Nazir, A. (2018). Word formation processes in Bhaderwahi. *The Literary Herald*, 3(5), 416–428.

Koul, P. K. (2014). Bhaderwahi. In O. N. Koul (Ed.), *The languages of Jammu & Kashmir* (People's Linguistic Survey of India, Vol. 12). Orient BlackSwan.

Masica, C. P. (1991). *The Indo-Aryan languages*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486883

Sheikh, A. M., & Zarger, A. A. (2017). Phonological description of Bhaderwahi. *Interdisciplinary Journal of Linguistics, 10*, 85–92.

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 229-240

Between Languages, Beyond Voices: Examining the Role of Code-Switching as Tool for Silencing Women through Appraisal Theory in "*A Thousand Splendid Suns"*

Vijay Bernard[*]
Arpana[**]

Abstract

Linguistic choices produce evaluative meanings within a discourse by expressing emotions and judgements to shape social relationships. In this light, this paper examines the intentional use of code switching in Khalid Hosseini's novel A Thousand Splendid Suns. It further explores the alternation of codes from English to local Afghan languages applying the Appraisal theory, a part of systemic functional linguistics developed by Jim Martin and Peter Rose. The results indicate that conscious employment of code switching emerges as a powerful tool to subjugate and marginalise women. The text reveals how patriarchal authority manipulates language to control women characters especially Mariam and Laila. Thus, the paper adds to the existing knowledge of Appraisal Theory, feminist linguistics and literary analysis by revealing how language use is not neutral but rather a source through which power is exercised, negotiated and contested.

Keywords: Affect Appraisal Code Switching, Judgement, Marginalisation, feminist, Linguistics

1. Introduction

Language creates and maintains social reality. It conveys not only meanings, but also enforces and reinforces cultural norms, moral judgments, and power structures. In doing so, it serves as a means of performing a wide range of social actions. M. A. K. Halliday conceives language as a social semiotic system that "makes meaning within specific social contexts" (Halliday, 1978, p. 1–4). He notes three functions of language, namely ideational, interpersonal, and textual. These functions enable human beings to express emotions, make requests, give commands, and perform various other acts. However, language is never used without a context or in isolation. It always operates within certain specific discourse situations. As Paul Simpson observes, linguistic form "is never purely neutral but is implicated in ideology and power relations" (Simpson, 1993, p. 5). Since evaluative meaning in a discourse is realized through linguistic choices made by

[*] Khalsa College, Amritsar; shahbernard46@gmail.com
[**] G.N.D.U College Pathankot; arpana65@gmail.com

the speaker, code-switching can be considered as an appraisal resource that expresses emotions, judgments, and ideology. Drawing upon Appraisal Theory, an extension of Systemic Functional Linguistics developed by J. R. Martin and Peter White, the present paper examines the use of code-switching in *A Thousand Splendid Suns* by Khaled Hosseini. The study argues that code-switching in the novel functions not merely as a marker of cultural identity or indigenous expression, but also as an evaluative resource through which patriarchal authority is enacted and women are silenced and marginalised. Through culturally embedded expressions carrying judgemental and affective meanings, the novel exposes how language becomes a site where gendered power relations are exercised, negotiated, and contested within Afghan society.

1.2 Code Switching

Scholars have explained code-switching differently across different years. Auer (1998) defines it as the alternation between two or more languages within a discourse, while Esch and Riley (2003) are concerned with its occurrence in bilingual discourse where speakers switch between the two. Bock defines code-switching as a thoughtful and intentional shift between language that speakers use to express more effectively or fulfil social communicative functions in a conversation. Winkler (2007) refers to this term as the practice of incorporation of the elements of one language by the speakers of the bi or multilingual speech communities. C. V. Bhuvaneswari and Anvita Abbi (1986) explain it as a "juxtaposition of larger stretches of mother tongue and other tongue elements in one exchange as interaction device common to all multilingual speech communities" (Abbi, 1986, p. 129). But according to Hudson (1980), Wardhaugh (1986), and Trask (2004), code-switching is not another use of two languages but it may be a use of two or more varieties of one language. Hudson's definition is narrow; in his view it is a case in which a speaker uses various varieties of a language at different periods (Hudson, 1980, p. 56). Wardhaugh (1986) and Trask (2004) quote it not only as the switching of different varieties at varying times but as "altering back and forth between two language varieties particularly in a single conversation" (Trask, 2004, p. 56). Trask also notes that most people are proficient in a minimum of two varieties of their language. Dr. Sarangi (2012) describes this practice as "the simultaneous use of more than one language variety in communication" (Sarangi, 2012, p. 28).

All these diversely formulated definitions point to the fact that code-switching is not only a linguistic phenomenon but also a socially and ideologically loaded communicative device as a way of negotiating identity, authority, inclusion/exclusion. In literary texts, then, alternations between languages or language varieties can be used to indicate power relations and social positioning. In this theoretical background the present study examines the role of code switching as a linguistic weapon for silencing and marginalisation of female voices in *A Thousand Splendid Suns* by the lens of Appraisal Theory.

2. Literature Review

The connection between language and ideology has been a prominent issue in literary and linguistic analysis since long. Literary scholars in the area of Sociolinguistics, Stylistics, Postcolonial Studies and Discourse Analysis have studied the ways in which multilingualism is used in literary texts, not only as a means of communication, but also as an indication of identity, resistance, cultural memory and social positioning. A particular focus of study has been code switching, which is seen to be related to cultural representation and ideological negotiation among these multilingual practices. But despite the huge body of research, comparatively little has been written on code-switching from the systemic functional linguistics (SFL) point of view, especially within J. R. Martin and P. R. R. White's (2005) framework.

The early sociolinguistic research carried out by John J. Gumperz viewed code-switching as a cue for the contextualisation of meaning, that is, a means through which the speaker indicates context to the listener (Gumperz, 1982). Similarly, Carol Myers-Scotton, in her Markedness Theory and Matrix Language Frame Model, argued that language alternation results from social motives and negotiations of identity (Myers-Scotton, 1993). Building on these sociolinguistic and literary approaches, other scholars turned to code-switching as a means by which to convey evaluative stance or interpersonal positioning. Bambi B. Schieffelin discussed the sociocultural aspect of code-switching, which involves communicating attitudes and cultural identity (Schieffelin, 1989) Susan Hunston and Geoff Thompson (1999) view evaluation as the way in which speakers convey attitudes, stances, and opinions towards people, actions, and events. On this theme, Zannie Bock (2008) explored code switching as an appraisal resource in the statements of the TRC. Her research has shown that the alternation between English and Africans allows the speakers to enhance their affect, express their moral evaluations, and position themselves in interpersonal space. In this context, code switching served as a tool to the expression of affect, judgment and appreciation. . This view is also supported through further empirical evidence from Nicole Baumgarten and Inke Du Bois (2012), who study dialogues in interviews focusing on third-party references. Their results suggest that code switching helps speakers to depict other speakers in conversation and to build solidarity/distance from interlocutors. Code-switches cannot be considered evaluative, but they become evaluative when they are used in combination with appraisal markers and other contextually appropriate features, which further support the function as part of the negotiation of interpersonal meanings.

More research has also shown the effectiveness of M. A. K Halliday's Systemic Functional Linguistic approach in elucidating the role of code switching in discourse. In analysis of Chinese magazines, Zhao and Liu (2015) have claimed that the functions of code-switching are all the Meta functions of language. It is lexically complete at the experiential level, evaluative at the interpersonal level and cohesive at

the textual level. Wang Huabin (2018) also uses a similar method in his work on Chinese television series, which shows that code-switching in TV language mainly, serves to create interpersonal meanings, such as conveying emotion, and to negotiate interpersonal relations between the participants. Similarly, Hanan Ben Nafa (2018) and Romansyah and Adam (2023) accentuate the position taking function of code switching. Ben Nafa shows the ways in which bilingual speakers switch languages to convey irony, affection, and identity, and Romansyah and Adam provide an analysis of *How the García Girls Lost Their Accents* by Julia Alvarez, an exploration of the complex interplay between language choice, gender, and power for the female characters in the novel. These foundational studies established that code-switching is not a random interference in language but a meaningful social and ideological strategy.

In the diaspora literature, the use of code-switching to portray multilingual realities and indigenous cultural experiences has received increasing scholarly attention. Ngũgĩ wa Thiong'o emphasises the close relationship between language, cultural memory, and power, arguing that postcolonial writers often appropriate colonial languages to articulate native experiences (Ngũgĩ wa Thiong'o, 1986). Such arguments have opened new directions in the study of multilingualism in literature, particularly in understanding code-switching as a mode of resistance and identity construction.

Studies on code switching in literature done by South Asian and diasporic people have shown that code switching can have a number of literary and ideological uses. Words from native languages, kinship terms, terms used for religious purposes, and culturally specific words have been noted in the writing of writers to evoke local colour, authenticity, and cultural hybridity. In Indian, Pakistani, and Afghan English fiction, code-switching has been further explored as a linguistic tool that enables the author to deal with the opposite themes of tradition and modernity, East and West and local and global identities.

The research on the works of Khaled Hosseini has been mainly dedicated to multilingual representation and cultural translation work as he utilizes Persian, Dari, Pashto and Islamic expressions in English to mirror the Afghan society, culture and emotional memory. The article entitled 'Strategies of Language Appropriation in *A Thousand Splendid Suns'* published in 'Language in India'(Awan, M. S., & Ali, M. 2012) is one of the important contribution in this field which explores Afghan lexical items, kinship terms, religious expressions, and culturally embedded discourse patterns as strategies of indigenizing English and expressing Afghan identity. The study claims that code switching and lexical borrowing result in authenticity in character and setting, but fails to focus on the evaluative meaning that is produced by code switching. It therefore does not discuss code-switching as part of the Appraisal framework in connection with Judgement or Attitude or interpersonal positioning, a topic also not addressed in the broader

literature on the code-switching of fiction with regard to exile, trauma, gender oppression, war, and cultural identity.

Although the studies on multilingualism and code switching from language appropriation approach have been carried out, the study based on the Systemic Functional Linguistics and Appraisal Theory on code switching has been limited. Much previous work on South Asian and Afghan English literature has focused on the phenomenon of code-switching in terms of aspects of authenticity, cultural identity, postcolonial resistance and linguistic appropriation, whereas several studies have addressed the phenomenon in terms of an appraisal resource in spoken discourse. Because of this, there is little research on how code switching is used in *A Thousand Splendid Suns* in relation to the Appraisal framework to build Judgement, Attitude, moral evaluation and interpersonal positioning between characters. The present study aims to fill this void by analysing code-switching in the novel from the Judgement in the Appraisal Theory.

3. Methodology

The methodology used in the present study is qualitative and interpretative method based on the theory of Systemic Functional Linguistics (SFL) Appraisal theory. *A Thousand Splendid Suns* has numerous code switched expressions that belong to Persian, Dari, Pashto and Islamic discourse, but due to the limited scope of the specific issues the study focused on, and constraints of length, only twelve expressions have been discussed more in-depth. The examples selected are discussed to understand the role of code switching as a tool to express emotion, moral values, interpersonal relationship, and ideological positioning in the narration. The analytical procedure of the text consists of three major steps: identification of code-switched expressions in the text, contextual interpretation of the identified expressions and their cultural and narrative interpretation and classification in the categories of Appraisal Theory, specifically Judgement subsystem. This approach is used to explore the role of multilingual expressions in the representation of culture as well as the creation of evaluative meanings in the novel.

3.1 Appraisal Framework

Appraisal theory is an extension of Systemic Functional Linguistics that accounts for evaluation in language. It is an outcome of the extensive work on the use of evaluation in language carried out by J. R. Martin, Peter White, and David Rose. The origins of this theory can be traced back to M.A.K. Halliday's Meta functions of language, specifically the Interpersonal Function. According to Martin and White, appraisal functions as an interpersonal system within discourse semantics (33). Therefore, within the framework of Systemic Functional Linguistics (SFL), appraisal is a component of the broader system of discourse semantics. Bock (183) observes that it relates to how speakers or writers embed attitudes, feelings, and personal perspectives into a text. The concept of appraisal in discourse analysis

goes beyond merely describing attitudes and feelings. It also examines how a text interacts with its audience regarding issues of solidarity and power. This means it looks at how a text positions its readers in relation to the opinions or experiences presented, influencing whether they feel sympathetic and supportive or dismissive and critical.

Martin and Rose explain this by stating that Appraisal is focused on evaluation. This involves analysing the types of attitudes expressed in a text, the intensity of the emotions conveyed, and how values are presented and connected with the readers. Essentially, it's about understanding how texts shape and reflect the attitudes and alignments between the writer and the audience (22).

Appraisal theory is divided into three primary sub-systems: the Attitudinal System, the Graduation System, and the Engagement System. These sub-systems are distinguished by their semantic characteristics rather than grammatical ones. The Attitudinal System, in particular, focuses on expressing values or emotions. Graduation system refers to the intensity of attitudes being felt by the reader or listener. Engagement refers to the source to which attitudes are attributed.

Due to the scope of the present study, which is limited to Attitude system of Appraisal Theory, the following three subcategories are highlighted here: Affect, Judgement, Appreciation, and the domains of Engagement and Graduation are not considered. Affect is the outward display of feelings or emotional responses like fear, anger, grief, joy or annoyance. In literary speech, code switching may enhance emotional communication by bringing into consciousness culturally significant words and expressions with more emotional resonance than the translated equivalents, thereby creating a stronger emotional reaction. Judgement is a judgment of human actions in relation to morality, ethics, social respect or social approval. The use of code switching can allow characters to commend, denounce, mock or morally critique people or communities and can expose their ideological and social stance. Appreciation, on the other hand, is concerned with the evaluation of things, events, phenomena, cultural practices, situations and objects according to their aesthetic, social or cultural value. Indigenous lexical elements in multilingual literary texts can play a role as a marker of cultural appreciation, which brings to light traditions, beliefs, customs, and sociocultural realities. The present paper is based on the insights of Zannie Bock and investigates the function of code switching as an evaluative resource in these three attitudinal categories, but not in the wide-ranging dialogic and scaling function of Engagement and Graduation.

The brief introduction to Appraisal Theory will provide the researcher with a theoretical framework to analyse the text under investigation. This framework will help to explore lexico-grammatical resources as code switches used as sources of inserting feelings and subjectivities.

3.2 Data

A Thousand Splendid Suns, is an acclaimed novel written by the Afghani American writer which deals with the entwined stories of Marriam and Laila entangled with patriarchal and Afghan struggle for political supremacy. Marriam, an illegally begotten daughter of a rich man and his maid, is out-casted along with her mother. After her mother's death, she is married to a middle-aged shoemaker, a widower, against her will at the teen age. He deals with her very harshly when she fails to give birth to a son for him. She kills her husband Rasheed for his co-wife Laila when he tries to strangle her. It leads to her tragic death.

The second story of the novel deals with the suppression of women together with the suppression of Afghanistan through the hands of the fundamentalist regime. Laila is separated from her lover as he migrates under fearful circumstances. Her family gets killed with a rocket blast. She marries Rasheed against her wishes to hide her pregnancy, taking her lover to be dead. On her lover's return, she meets him; this annoys Rasheed. In a fit of rage, he attempts to kill her; however, her husband's first wife saves her by killing him. She is reunited with her lover, marries him, and works for the rehabilitation of Afghani people.

S. No.	Code-switch	Type	Appraisar	Type of judgement	+/-
1	You are a clumsy little harami (4)	Judgement	Nanna to Marriam	Social Esteem normality	Normality
2	For what ! To watch him drive his Kinchini wives around (8)	Judgement	Nanna	Social Sanction Propriety	–
3	How dare you abandon me like this, you treacherous little Harami (27)	Judgement	Nanna	Social Esteem Normality	–
4	This is enough , You ‘ve made a scene, Bas, It’s time to go.	Affect	Jalil’s Driver	Displeasure	–
5	They don’t see that they ‘re spoiling their own nang and namoos	Judgement	Rasheed	Social Sanction Propriety	–
6	Bebakhsh. Rasheed. Iam sorry. (93)	Affect	Marriam	Unhappiness	–
7	You don’t….. Bah(92)	Affect	Rasheed	Displeasure	–
8	Chup ko Shut up (97).	Affect	Rasheed	Displeasure/ anger	–
9	She is a dehati girl.(216)	Judgement	Rasheed	Social Sanction Propriety	–
10	Have you told her that you are a harami (216)	Judgement	Rasheed	Social Esteem normality	–

11	You on the other hand would be a Benz a brand – new first class shiny Benz. Wah wah. (216)	Appreciation	Rasheed		_
12	Wallah O billah! I will go down and teach her a lesson(223)	Affect	Rasheed	Displeasure	_

Table 1. An overview of distribution of Code switches as source of Appraisal

4. Analysis and Discussion

The table above indicates half of the code-switched examples express negative Judgement and all of these are strongly negative. Four of the code-switches reflect Affect to express emotional responses. These Affect instances show negative emotions revealing emotions such as frustration, irritation, and anger. The data shows only two instances of Code switching occurring as a source of Appreciation. It is Significant to observe that no instances of positive Judgement and Affect are found in the data analysed. This highlights the power imbalance between male and female characters of the novel.

When the characters in novel use culturally loaded word *harami*, like "You are a clumsy little harami" (4), "How dare you leave me like this, you treacherous little harami" (27), and "Have you informed her that you are a harami?" These dialogues containing Code -switching denote the illegitimacy of an individual in South Asian context. It serves as a moral tag which marks Mariam as socially deviant. Such language becomes a label that justifies her exclusion and silencing. It is as in Appraisal Theory, such statements realize negative Judgement of Normality, in which deviance from assumed social norms constitutes a ground for exclusion or silencing. This can be seen in the novel through Mariam's isolated Kobla (mud House) outside the city and away from the locality that reinforces her social marginality. Rashid's remark "She is a dehati (rural or unsophisticated) girl" (216) similarly confers a dismissive and derogatory Judgement that signifies how code-switching makes hierarchical positioning and marginalizes her from the mainstream. The use of regional or vernacular terms, such as *dehati*, reflects disdain and class rooted discrimination. It further contributes to Mariam's marginalization. All these examples illustrate how code switching turns language into a formidable tool to social control to exert patriarchal authority and marginalize people, specifically women, in the Afghani context.

The expression, "For what! To see him take his Kinchini (Whores) wives out for a spin" (8) is the code switch by which women have been attributed with being mean and lower than men. The word employed in the novel for Jalil's wives, *Kinchini*, means "prostitute". In Appraisal theory, this is a source of Judgement. The change in the code again targets the morality of women for marrying a man who is already married, but this is not the case with males. No insulting language is

directed towards men in the text who practice polygamy. Therefore, the code switch does not function merely as a marker of Judgement, but it reveals how language becomes a device to silence women conveying their moral corruption and shame.

The Code switching in the sentence, “They don't see that they‘re ruining their own nang and namoos” (Pride and Honour) from English to Pashto draws attention towards two culturally loaded terms. Both the words encapsulate deeply held Afghan ideas about honour and moral conduct of females. From an Appraisal Theory perspective, the example reflects Judgement in the domain of Social Sanction, particularly Propriety. Rasheed evaluates others’ behaviour according to cultural expectations of moral decency, criticising anyone unable to live up to accepted norms of respectability. The statement given by the character is socially very forceful. It embodies the cultural belief that a family's honour depends on the modesty and chastity of women. Thus, the code switch functions to uphold the power structure working in the Afghan culture.

Along with the Judgement instances, Affect indicating statements can also be observed. These shifts to local languages amplify emotions of displeasure, anger, and unhappiness. The code-switched phrases such as “Bas (Stop), it's time to go,” “Bah!,” “Chup ko,(shut up) !,” and “Wallah o billah (I swear by God), I will go down and teach her a lesson” (223) heighten the emotional force of the interactions. All these dialogues with instances of code switching explicitly demonstrate how language use could be employed to promote patriarchal dominance over the female characters. In this case, the first dialogue presents Mariam’s denial of her rights to demand to see her father. The use of code switching, in this dialogue, acts as a device to weaken her resistance. The resistance in this case, becomes an element to be pushed to the margins in suggesting that Mariam does not have any rights to resist in patriarchal dominant culture. Code switching is also used in silencing the voice of women in these dialogues. The dominance of men in this case, over Mariam, is brought about by another dominant name, “Rasheed,” in his dialogue. The dialogue employs code switching in silencing Mariam’s voice. The dialogue employs “Bah!” and “Chup ko, shut up!” to make Mariam mute, which reflects her limited knowledge in suggesting that her voice does not matter.

The code-switch “Wallah o billah (I swear by God) I am going down and will give her a lesson” (223) also arises from male anger. It reflects patriarchal control and the threat of violence used to discipline women. Thus, code-switching becomes a means of subjugating women. The codes switching the words of the female character Laila, “Bebakhsh (forgive me) Rasheed. I am sorry” (93), is a display of Affect. It conveys Mariam's fear, guilt, and longing for acceptance. It communicates not only the apology but also the emotional restraint expected of women in a male-dominated society. Thus, language reflects the limited emotional space allowed to women in this context.

"You would be a Benz a brand new, first-class, shiny Benz. Wah wah" can be recognised as an example of the code switching serving as a token of Appreciation because it appears as though Rasheed admires Laila's youthfulness as well as her physicality. The code-switched word "Wah wah" in particular highlights the appreciation because it also happens to be an expression that might be particular to Afghan culture as a representation of extreme appreciation or admiration. Nevertheless, reducing a human being to a status object by regarding her as a luxury item only serves to render her an object meant for display. It does not single her out as an individual but rather values her based on her appearance. This particular code-switching highlights the function of words as an embodiment of patriarchy because it upholds patriarchy by associating value in women with their appeal.

5. Conclusion

The analysis of the findings shows that the role of code-switching is primarily as Judgement and Affect in "A Thousand Splendid Suns," which confirms societal stratification with the exclusion of the female protagonists. Male characters employ code-switching intentionally with the intent of dominating, regulating social behaviour, and punishing morally. Through the identification of the patterns of code-switching linking Judgement/Affect sources, the study makes clear the manner in which multilingual code-switching in literary discourse marks the moral dimension, affect, or authority. These results are significant in the development of Appraisal Theory, feminist linguistics, as well as literary theory, further explaining the manner in which language is utilized as the instrument of oppression as well as the nuanced arena of subversiveness within the patriarchal society of Afghanistan.

References

Abbi, A., & Bhuvaneswari, C. V. (1986). *Language contact and multilingualism in India*. Sage Publications.

Auer, P. (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge.

Awan, M. S., & Ali, M. (2012). Strategies of language appropriation in Khaled Hosseini's A Thousand Splendid Suns. *Language in India*, *12*(8), 478–495. http://www.languageinindia.com

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin* (C. Emerson & M. Holquist, Eds.; C. Emerson, Trans.). University of Texas Press. (Original work published 1975)

Baumgarten, N., & Du Bois, I. (2012).Code-switching as appraisal resource in talking about third parties. *Linguistik Online*, *51*(1), 39–61. https://doi.org/10.13092/lo.51.304

Ben Nafa, H. (2018). *Code-switching as an evaluative strategy: Identity construction among Arabic-English bilinguals in Manchester* [Doctoral dissertation, Manchester Metropolitan

University]. Manchester Metropolitan University e-space. https://e-space.mmu.ac.uk/620931/

Bock, Z. (2008).*Code-switching as a resource for evaluation: An appraisal perspective*. UNISA Press.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2009). *Educational research: Competencies for analysis and Application* (9th ed.). Pearson.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Harding-Esch, E., & Riley, P. (2003). *The bilingual family: A handbook for parents* (2nd ed.). Cambridge University Press.

Hosseini, K. (2013). *A thousand splendid suns*. Bloomsbury.

Hudson, R. A. (1980). *Sociolinguistics*. Cambridge University Press.

Hunston, E., & Thompson, G. (Eds.). (1999). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford University Press.

Martin, J. R., & White, P. R. R. (2005).*The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Ochs, E., & Schieffelin, B. B. (1989).Language has a heart. *Text - Interdisciplinary Journal for the Study of Discourse*, *9*(1), 7–25. https://doi.org/10.1515/text.1.1989.9.1.7

Romansyah, N. D., & Adam, M. (2023). The use of code-switching and code-mixing to express social identity in the novel How the Garcia Girls Lost Their Accents by Julia Alvarez. *Prologue: Journal on Language and Literature*, *9*(1), 115–127. https://doi.org/10.36277/jurnalprologue.v9i1.86

Sarangi, J. (2012). Towards theorizing Indianization of English: A sociolinguistic study. In R. S. Sharda (Ed.), *Proceedings of the International Conference on English Language and Literary Studies* (pp. 1–8). Jalandhar.

Simpson, P. (1993). *Language, ideology and point of view*. Routledge.

Trask, R. L. (2004). *A dictionary of phonetics and phonology*. Routledge.

Wa Thiong'o, N. (1986). *Decolonising the mind: The politics of language in African literature*. James Currey.

Wang, H. (2018). Interpersonal meaning of code-switching: An analysis of three TV series. *Australian Journal of Applied Linguistics*, *1*(1), 3–19. https://doi.org/10.29140/ajal.v1n1.4

Wardhaugh, R. (1986). *An introduction to sociolinguistics*. Blackwell.

Winkler, E. G. (2007). *Understanding language: A basic course in linguistics*. Continuum.

Zhao, X. A., & Liu, R. X. (2015). A systemic functional approach to code-switching in Chinese magazines. In *Proceedings of the 2015 International Conference on Social Science, Education Management and Sports Education* (pp. 109–112). Atlantis Press. https://doi.org/10.2991/ssemse-15.2015.27

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 241-250

Romantic Tales in Pashtoon Tradition: A Comparative Study of *Adam Khan-Durkhanai* and *Momin Khan-Sherino*

Saima Jan*
Bushra**
Aabida Zahoor***

Abstract

Romantic tales constitute an important genre of oral Pashtoon literature through which they have preserved narratives of love, honour, sacrifice, social obligation and human suffering. Romantic tales of Adam Khan and Durkhanai, and Momin Khan and Sherino are significant. This paper presents a comparative analysis of the two tales with regard to their plot development and narrative structure; treatment of love, social restrictions, family honour, sacrifice, heroic action and tragic resolution. Adam Khan and Durkhanai is primarily structured around love at first sight, forbidden attachment, music and separation, while Momin Khan and Sherino combines romantic love with a journey, heroic adventure, economic hardships and mistaken identity.

Keywords: Folklore, Pashto, Romantic Tales, *Adam Khan* and *Durkhanai*, *Momin Khan* and *Sherino*, Oral Tradition, Narrative

1. Introduction

Storytelling occupies an important position in Pashtoon cultural life. Traditional narratives preserve collective memories, social values and cultural concepts while simultaneously providing entertainment and emotional expression. Romantic tales are particularly significant because they bring together personal emotions, family, honour, social obligation and individual choice. Pashtoon folklore contains several celebrated romantic narratives, including *Adam Khan* and *Durkhanai*, *Momin Khan* and *Sherino*, *Yusuf Khan* and *Sherbano*, and others.

Among these, *Adam Khan* and *Durkhanai* is regarded as a classic Pashto romance. The tale is associated with Bazdara in the Malakand region and has been preserved through both oral and written traditions. One scholarly account identifies its early versification with Sadar Khan Khattak and reports that the narrative was composed in thousands of couplets before being published by Pashto Academy, University of

* University of Kashmir; saima@uok.edu.in
** University of Kashmir; bushra.lnscholar@gmail.com
*** University of Kashmir; abidazahoor108@gmail.com

Peshawar. *Momin Khan* and *Sherino*, meanwhile, is another widely circulated Pashtoon tragic romance. Research on the tale emphasizes that it has travelled through different Pashtoon regions and exists in both written and oral variants.

The two tales are therefore valuable from a linguistic perspective because they demonstrate how oral narrative is constructed through language and performance. Their significance lies not merely in their plots but in the ways storytellers employ dialogue, repetition, poetic forms, emotional vocabulary, figurative language, etc to maintain audience involvement.

2. Literature Review

The romantic narratives of Pashtun oral tradition have attracted attention from folklorists, literary scholars, and researchers of Pashto language and culture for more than a century. Early documentation of Pashtun folklore is associated with James Darmesteter (1888), whose Chants Populaires des Afghans contains an early published version of *Adam Khan* and *Durkhanai*. The inclusion of the tale in a scholarly collection is significant because it demonstrates the transition of a traditionally oral narrative into a documented literary form. Darmesteter's work is therefore an important early source for examining the relationship between Pashtun oral literature, language, romance, and cultural memory. The tale was subsequently presented in English by Edward Emmerson Oliver (1890) in Across the Border: Or Pathan and Biloch, where it was described as a "Border Romance" and compared with the Romeo and Juliet tradition.

The early twentieth-century literary history of *Adam Khan* and *Durkhanai* is particularly important because the narrative existed simultaneously in oral and poetic forms. Sadar Khan Khattak, son of Khushal Khan Khattak, is traditionally associated with the poetic rendering of the tale. According to Suliman Sayad (2020), the story was initially versified in couplets, with each line consisting of eight syllables, and the version attributed to Sadar Khan Khattak contains 2,587 couplets. Sayad further records that the tale was published by the Pashto Academy, University of Peshawar, in 1959, making the Academy's edition an important stage in the movement of the story from oral tradition into published Pashto literature.

Several published collections subsequently preserved and popularized *Adam Khan* and *Durkhanai*. Zainab Ghulam Abbas (1957) included an English rendering of the tale in Folk Tales of Pakistan, while Masudul Hasan (1976) included it in Famous Folk Tales of Pakistan. Richard M. Dorson (1978) also included the tale in Folktales Told around the World. Later, Wilma Louise Heston and Mumtaz Nasir (1989) included "*Adam Khan* and *Durkhanai*" in The Bazaar of the Storytellers. These publications are important because they demonstrate the continuing circulation of the narrative across languages and literary forms. In particular, The Bazaar of the Storytellers places the tale within a broader collection of Pashtun storytelling traditions,

emphasizing the importance of romantic narratives in the cultural repertoire of the region.

Another important contribution is the comparative linguistic-literary study by Mohammad Muazzam Sharif et al. (2020), "Linguistics Largeness: From Romeo and Juliet to *Adam Khan aw Durkhanai.*" The study approaches the Pashtun tale through qualitative textual analysis and comparison with Shakespeare's Romeo and Juliet. Such comparative scholarship demonstrates that *Adam Khan* and *Durkhanai* can be studied not simply as a love story but as a narrative in which language, cultural values, social restrictions, characterization, and tragedy interact. This provides a useful basis for the present study's focus on narrative structure and linguistic representation.

Compared with *Adam Khan* and *Durkhanai*, *Momin Khan* and *Sherino/Sherinai* has received relatively less sustained scholarly attention, although it occupies an important position in Pashtun romantic folklore. A particularly relevant publication is the 2019 study "An Analysis of Oral Tradition and Written Text of Tragic-Romance Story of *Momin Khan-Sherino*," published in Takatoo, University of Balochistan. The study explicitly examines the relationship between oral and written versions of the narrative. It compares a written version compiled by Muhammad Gul Noori of Kandahar with oral accounts preserved among people in villages of Pishin. The study argues that, as the story travelled through oral tradition, changes occurred in its textual form, poetic expression, and sometimes its characters. This is particularly important for the present research because it confirms that *Momin Khan* and *Sherino* should be treated as a variant-based oral narrative rather than a single fixed text.

The publication history of *Momin Khan* and *Sherino* is also associated with Muhammad Gul Noori's Milli Hindara, a collection of indigenous Pashto dastans. A discussion published in Pashto by the Pashto Academy notes that Milli Hindara brought a number of indigenous Pashtun narratives into prose, including *Momin Khan-Sherino* and *Adam Khan-Durkhanai.* This is important from a narrative perspective because it demonstrates the transformation of stories traditionally circulated through oral and poetic forms into prose narratives. The transition also provides an opportunity to study how plot organization, dialogue, characterization, description, and linguistic expression change when an oral/poetic narrative is converted into written prose.

The available scholarship consequently indicates that both romances have a substantial oral foundation but have undergone multiple processes of textualization, reinterpretation, and variation. *Adam Khan* and *Durkhanai* has a comparatively longer history of documentation and publication, beginning with nineteenth-century European documentation and continuing through Pakistani and Pashto literary collections and recent academic studies. *Momin Khan* and *Sherino*, on the other hand, has particularly strong evidence of variation between oral accounts and written versions. The existing studies,

however, generally examine the two narratives separately. There remains scope for a systematic comparative study of their plot development, narrative methodology, and storytelling techniques. The present study therefore seeks to fill this gap by placing the two major Pashtun romantic tales within a common analytical framework and examining how similar cultural values—love, loyalty, honour, family, social obligation, and sacrifice—are constructed in two different tragic-romantic traditions.

3. Methodology

This study employs a comparative textual analysis to examine two prominent Pashtoon romantic folktales, *Adam Khan and Durkhanai* and *Momin Khan and Sherino*. Since both narratives originate within an oral storytelling tradition while also circulating in documented written forms, the study relies on published textual versions of the tales as its primary source material, drawing on established scholarly collections, translations, and academic treatments that have preserved and transmitted these stories over time.

The analysis focuses on three interrelated dimensions: plot development and structure, narrative and storytelling techniques, and the treatment of core cultural themes of love, family honour, social obligation, sacrifice, and tragic resolution. Each tale is first examined individually to trace its plot trajectory and identify the specific narrative devices employed (such as dialogue, repetition, suspense, poetic form, and characterization).

Following this individual analysis, a comparative method is applied to place the two narratives within a shared analytical framework. This comparative step identifies points of structural and thematic convergence like their shared use of tragedy and idealized love as well as elements of divergence, including differences in plot triggers, central obstacles, and narrative emphasis (romance-centred versus romance-adventure structures). This approach allows the study to illuminate both the shared cultural values and the distinct storytelling strategies characteristic of Pashtoon oral romantic tradition.

4. *Adam Khan* and *Durkhanai*: The Tale and Plot Development

Adam Khan and *Durkhanai* is one of the best-known romantic narratives of Pashtoon tradition. *Adam Khan* belongs to Lower Bazdara, while *Durkhanai* belongs to Upper Bazdara. Traditional accounts portray *Adam Khan* as a handsome young man and talented rabab player, while *Durkhanai* is represented as beautiful, intelligent and educated lady.

The narrative begins with the first encounter between the two protagonists. *Durkhanai* hears of *Adam Khan* and catches a glimpse of him while he is performing. *Adam Khan* also sees *Durkhanai*, and the encounter immediately establishes the romantic bond between them. Music is therefore not merely decorative in the narrative; it becomes one of the mechanisms through which attraction is created.

The major complication is that *Durkhanai* is already betrothed to Payu Khan. Her attachment to *Adam Khan* therefore conflicts with the established expectations surrounding betrothal, family honour and social obligation. The story develops through meetings, communication and increasing opposition to the lovers' relationship.

As the narrative progresses, separation intensifies the lovers' emotional suffering. In one of the circulated versions, *Adam Khan* becomes deeply distressed and eventually dies before the lovers can achieve their desired union. *Durkhanai*, devastated by his death, also dies. Their burial together becomes a symbolic representation of the union that they were denied during life.

The plot can therefore be represented as:

First encounter → attraction → love → existing betrothal → social conflict → separation → emotional suffering → death → symbolic union.

5. Narrative Techniques in *Adam Khan* and *Durkhanai*

5.1 Oral and Performative narration

The story has been transmitted through oral as well as written traditions. Its early poetic form is particularly significant because the narrative was organized into couplets, making it suitable for oral recitation and memorization. One study reports that Sadar Khan Khattak versified the tale in couplets with an eight-syllable line structure.

5.2 Linear Chronological Narration

Events are presented sequentially: the lovers' encounter → development of love → obstacles/separation → attempts to overcome the obstacles → tragic resolution.

The sequence makes the tale relatively easy to follow.

5.3 Music as a narrative device

The rabab is more than an object associated with *Adam Khan*; it functions as a narrative instrument through which attraction and identity are established. *Adam Khan's* musical ability contributes to his characterization and becomes part of the romantic atmosphere.

5.4 Dialogue and direct speech

Dialogue allows the lovers and other characters to communicate directly. It provides immediacy and allows emotional states such as affection, anxiety, persuasion and grief to be represented through speech. Direct conversations between characters dramatize important episodes. Instead of simply saying that a character id distressed, the narratives made in the text allows the character to express that distress directly.

5.5 Repetition

Repetition as one of the devices is very frequently seen in the documented written version. Repetition of emotional expressions and important events serves two functions: it reinforces the central theme and facilitates oral memorization. It also creates rhythm and gives emotional passages greater intensity.

5.6 Suspense

Information and outcomes in the text are sometimes delayed, in order to keep the readers' curiosity at its peak as what is going to happen next. The audience is gradually made aware of the difficulties surrounding the lovers. The tension between their love, *Durkhanai's* existing betrothal and obstacles to the lovers' union provides a continuing source of suspense as they progress gradually.

5.7 Hyperbole and Idealization

Descriptions of beauty, love and suffering are intensified in the story. *Durkhanai* is represented as exceptionally beautiful and intelligent and her beauty is compared with that of moon, sunlight and other radiant objects, while *Adam Khan's* love and suffering are elevated beyond ordinary experience, as is mentioned in the text that *Adam Khan* becomes incapable of functioning normally without *Durkhanai*. Moreover, following separation, the lovers' grief is described as overwhelming, with expressions suggesting that his tears could fill a river or that his sorrow is capable of destroying him.

6. *Momin Khan* and *Sherino*: The Tale and Plot Development

Momin Khan and *Sherino* is another important Pashtoon tragic romance. The story presents *Momin Khan* as the only son of his father and *Sherino* as his cousin. Their families arrange their betrothal in childhood. After the death of their fathers, a conflict concerning leadership develops between *Momin Khan* and *Sherino's* brother, Zabardast Khan.

Momin Khan's family sends a delegation to request *Sherino's* hand. Zabardast Khan eventually agrees but demands a very large walwar, creating an economic obstacle to the marriage. Unable to meet the demand, *Momin Khan* decides to travel to Hindustan to acquire the wealth required for marriage. This introduces an important journey motif into the narrative.

During his journey, *Momin Khan* becomes involved in a series of heroic adventures. The tale describes his encounter with a royal court, his victory over wrestlers and his intervention in a political conflict. He eventually encounters a kingdom threatened by a dragon or serpent that is consuming young women. *Momin Khan* confronts the creature and saves the princess. The episode transforms the romantic protagonist into a heroic figure, combining romance with adventure and epic narrative.

Although the foreign ruler offers *Momin Khan* wealth and marriage, he remains committed to *Sherino*. He returns to his homeland after acquiring the wealth necessary for his marriage. However, the final reunion produces tragedy rather than fulfilment. In a documented version, *Momin* arrives at night and approaches *Sherino's* house. She cries out when she encounters the unexpected intruder, and her brother Zabardast Khan attacks him without recognizing him. *Momin Khan* is fatally wounded. *Sherino* subsequently dies from grief, producing the tragic conclusion characteristic of the tale.

The plot can therefore be represented as:

Childhood betrothal → family conflict → demand for walwar → departure for Hindustan → heroic adventures → acquisition of wealth → return home → mistaken identity → death → tragic ending.

This structure makes *Momin Khan* and *Sherino* more complex in terms of plot movement than a purely romantic narrative. It combines romance, family conflict, economic difficulty, migration, heroic adventure and tragedy.

7. Narrative Techniques in *Momin Khan* and *Sherino*

7.1 Journey narrative

The journey to Hindustan provides the story with an extended narrative movement. It allows the protagonist to encounter new characters, locations and challenges while simultaneously demonstrating his loyalty to *Sherino*.

7.2 Heroic episodes

The dragon episode and other adventures in the tale of *Momin Khan* and *Sherino* introduce elements of heroic storytelling. *Momin Khan's* courage becomes a means of establishing his worthiness as a lover and hero.

7.3 Dialogue and direct speech

Dialogue is important in establishing relationships between *Momin Khan*, *Sherino*, family members, elders and rulers. It also reveals social expectations and disagreements.

7.4 Suspense and dramatic irony

The final episode depends heavily on suspense and dramatic irony. The audience knows *Momin's* identity and purpose, while the characters involved in the fatal encounter do not fully recognize him. This makes the eventual tragedy particularly powerful. Moreover, as the story progresses the level of suspense keeps on rising, for instance, will *Momin Khan* and *Sherino* be united in the end?, will social restrictions overcome their love?, Will he reach *Sherino* or will another obstacle intervene?, etc.

7.5 Repetition and formulaic expressions

Repetition is seen in the story as it supports memory and performance. Recurrent references to love, separation, honour, journey and sacrifice reinforce the central themes.

7.6 Tappa and lyrical expression

The use of tappa is the characteristic Pashto poetic form and in a documented version, *Sherino's* distress is expressed through a tappa that seems very appealing.

8. Comparative Analysis

8.1 Similarities

The two tales share several fundamental characteristics:

1. Both are tragic romances. The lovers fail to achieve a normal marital union in life.
2. Both place love within a social framework. Love is influenced by family, honour, marriage arrangements and social expectations.
3. Both employ separation as a major narrative device.
4. Both idealize the lovers and present their devotion as extraordinary.
5. Both combine personal emotion with cultural values, particularly loyalty, honour and sacrifice.
6. Both use death as the final transformation of romantic love, allowing the lovers to achieve a symbolic union beyond the restrictions that prevented them from being together during life.
7. Both have undergone textual and oral transmission, resulting in different versions.

Differences

The two narratives differ substantially in their plot structures and narrative emphasis.

1. In *Adam Khan* and *Dukhanai*, love develops after first encounter, while as in *Momin Khan* and *Sherino* lovers are associated through childhood betrothal.
2. The central obstacle in *Adam Khan* and *Durkhanai* is the existing betrothal of *Durkhanai*, thereby leading to social restrictions; while as in *Momin Khan* and *Sherino* walwar, clan conflict, distance and circumstances served as the central obstacles.
3. *Adam Khan's* rabab is central to his characterization and in *Momin Khan-Sherino* music is less central to the main plot.
4. Journey is major structural element in *Momin Khan* and *Sherino*.

5. *Adam Khan* and *Durkhanai's* plot ended with the separation and death of the lovers, while as in *Momin Khan* and *Sherino* mistaken identity and fatal encounter added to the final tragedy.

9. Narrative Structure: A Comparative Interpretation

The two tales can be understood as representing two different models of romantic tragedy.

Adam Khan and *Durkhanai* follows a romance-centred structure. The narrative begins with attraction, develops through emotional attachment and focuses increasingly on the social forces preventing union. Music and beauty provide the initial mechanism of attraction, while separation and death provide the tragic resolution.

Momin Khan and *Sherino*, in contrast, follows a romance-adventure structure. Romantic attachment provides the motivation for the protagonist, but the plot expands into a journey involving heroic trials, foreign settings, wealth, political encounters and supernatural danger. The protagonist's return does not produce reunion; instead, the narrative turns suddenly from anticipated happiness to tragedy.

This difference is particularly important because it demonstrates the generic flexibility of Pashtoon romantic storytelling. A romantic tale can incorporate features of heroic narrative, social drama, adventure and tragedy without losing its central romantic identity.

10. Conclusion

Adam Khan and *Durkhanai* and *Momin Khan* and *Sherino* represent two significant manifestations of Pashtoon romantic storytelling. Both narratives centre on love that encounters social, familial or circumstantial barriers and ultimately assumes a tragic form. They use narrative techniques such as dialogue, repetition, suspense, emotional characterization, poetic expression, etc., to maintain audience interest and transmit cultural meanings.

The principal difference lies in their narrative organization. *Adam Khan* and *Durkhanai* is predominantly a story of romantic attraction, forbidden attachment, music, separation and tragic love. *Momin Khan* and *Sherino* expands the romantic plot into a journey of economic struggle, heroic adventure, supernatural conflict and mistaken identity. Yet both ultimately use tragedy to emphasize the intensity and endurance of love.

From a linguistic perspective, the two narratives are valuable because they show that storytelling is not simply the transmission of events. The use of strategic plot structure, narrative techniques, and comparative analysis between the tales collectively adds to the artistic beauty of the tales.

Consequently, these romantic tales should be understood not merely as stories of two pairs of lovers but as linguistic and cultural

repositories through which Pashtoon communities negotiate ideas of love, honour, family, sacrifice, social obligation and identity.

References

Abbas, Z. G. (1957). *Folk tales of Pakistan*. Pakistan Publications.

Bauman, R. (1977). *Verbal art as performance*. Newbury House Publishers.

Darmesteter, J. (1888). *Chants populaires des Afghans*.

Dorson, R. M. (1978). *Folktales told around the world*. University of Chicago Press.

Hasan, M. (1976). *Famous folk tales of Pakistan*. Ferozsons.

Heston, W. L., & Nasir, M. (1989). *The bazaar of the storytellers*. Lok Virsa Publishing House.

Jan, S., et al. (2025). Storytelling tradition among Pakhtoons: A cultural heritage. *A Journal of Literary & Critical Writings, 38*(1–2), 154–160.

Jan, S., et al. (2022). Cultural heritage of Pakhtuns: A case study of Pashto speakers living in the Valley of Kashmir. In *Tribal culture, language and literature*. Romanson Print House, University of Kerala.

Lebedev, K. (1972). *Afghanskie skazki i legendy*. Nauka.

Oliver, E. E. (1890). *Across the border: Or Pathan and Biloch*. Chapman and Hall.

Omarkhel, H. (2020, February 20). Adam Khan and Durkhanai epic: Analytical look. *Taand*.

Sayad, S. (2020). National narration of Adam Khan and Durkhanai. *Research Review International Journal of Multidisciplinary, 5*(3), 67–70. https://doi.org/10.5281/zenodo.3817312

Sharif, M. M., et al. (2020). Linguistics largeness: From “Romeo and Juliet” to “Adam Khan aw Durkhanai.” *Elementary Education Online, 19*(4), 5554–556

Interdisciplinary Journal of Linguistics
Volume [18] 2025, 251-264

Verum and Exceptive Constructions in some Indian Languages

Sayantani Banerjee[i*]
Moachiba Jamir [**]
Greeshma Joseph [***]
Zahid Bashir [****]

Abstract

The paper examines verum marking and exceptive marking as two discourse-sensitive phenomena in selected South Asian languages of India. While verum marking strengthens a speaker's commitment to the truth of a proposition, exceptive marking restricts the applicability of a generalisation by excluding an individual or subset. The paper explores how both constructions are represented by context, alternatives, prosody, and pragmatic interpretation, despite serving the same semantic functions. The empirical focus includes languages such as Hindi-Urdu (Indo-Aryan), Malayalam (Dravidian), Jungli (Tibeto-Burman), Mundari (Austroasiatic), and Kashmiri (Dardic) languages. The data was collected from native speakers through questionnaires and observations. Through a cross-linguistic comparison, the study is a first attempt to investigate the marking patterns of verum and exceptive marking in some south Asian languages and whether these constructions can be understood as part of a broader domain of discourse-semantics.

Keywords: Verum Marking, Exceptive marking, South Asian Languages, Language Typology

1. Introduction

Language provides speakers with a range of grammatical and prosodic devices for expressing not only propositional content but also their stance towards what is being said. Among such devices are constructions that strengthen the speaker's commitment to the truth of a proposition and those that restrict the applicability of a generalisation. Verum marking and exceptive marking represent two such discourse-sensitive phenomena, whose interpretation is closely associated with context, alternatives, and pragmatic meaning.

This paper lists two kinds of semantic markers, truth semantic markers and exceptive markers. These markers in the sentence do not

[i] Equal Contribution
[*] University of Kashmir; banerjeesayantani01@gmail.com
[**] University of Kashmir; jamirmoachiba@gmail.com
[***] IIT, Delhi; josephgreeshma97@gmail.com
[****] University of Kashmir; zahidbashir367@gmail.com

interact with agreement structure semantics and are usually related to pragmatics or context in the conversations. Speakers utilise these prosodic resources to negotiate what counts as true and which members of a relevant set are to be excluded from a generalisation. Two constructions that seem to particularly exemplify these dimensions are verum marking and exceptive marking. At first glance, they appear to be in quite different domains. Verum marking concerns the speaker's commitment to the truth of a proposition, whereas exceptive constructions express the exclusion of an individual or subset from a generalisation. However, both constructions are sensitive to the interpretation of alternatives, and to the discourse conditions under which a proposition or generalisation is evaluated. This paper addresses this potential link from a cross-linguistic perspective, asking if verum and exceptive marking can be seen as part of a larger domain of discourse-semantic and alternative-sensitive marking, without assuming that they form single grammatical category.

In this paper, the empirical focus is on South Asian Languages (SAL) in India. We consider representative language from each of the major SAL language families present in India- Hindi-Urdu from the Indo-Aryan language family, Malayalam from the Dravidian language family, Jungli from the Tibeto-Burman language family, Mundari from the Austroasiatic language family and Kashmiri from the Dardic language family.

2. Review of Related Literature

2.1 Verum Marking

The notion of verum stems from the literature on verum focus, most prominently associated with Höhle (1992). The canonical form for verum marking is to indicate the truth of a proposition, not of an individual constituent of a proposition. Consider a discourse like:

1. A. I heard Rina didn't come yesterday.
   B: Rina DID come yesterday.

The 'did' here does not simply contribute the assertion that Rina came. Instead, the utterance is made in a discourse situation in which the truth of the proposition Rina came yesterday is in question. Thus, verum marking is usually linked to correction, contradiction, confirmation, or the resolution of an explicitly or implicitly posed question concerning the truth of a proposition.

This discourse-sensitive reading has led to a variety of analyses. For example, Gutzmann and Castroviejo Miró (2011) treat verum as a use-conditional expressive operator on the question under discussion: verum roughly marks the truth of p. More recent cross-linguistic work has questioned the assumption that verum is just a special case of ordinary focus. Gutzmann, Hartmann, and Matthewson (2020) contrast the Focus Accent Thesis, where verum is a regular focus on truth value, with a Lexical Operator Thesis, where apparently focal marking realises a dedicated verum operator. Their investigation of Gitksan,

Bura, and South Marghi is especially important because these languages employ strategies for verum that do not pattern straightforwardly with their ordinary focus constructions.

The cross-linguistic picture has thus become central to the study of verum. Recent work indicates that marking truth can be accomplished by prosody, dedicated particles, polarity-sensitive morphology, or indirect syntactic strategies. For example, Kerr and Van der Wal (2023) discuss Bantu languages, where truth marking can be achieved through backgrounding, while other languages use specific markers. The empirical base in an effort to identify which properties of these constructions are genuinely shared across languages. A key question in modern verum research is thus not so much whether a language 'has verum' but how a language grammatically or prosodically marks the speaker's commitment to the truth of a proposition and under what discourse conditions that marking is licensed.

Exceptive constructions deal with a different semantic problem. It is explained in the following subsection

2.2 Exceptive marking

In exceptive constructions, the speaker generalises over a set but explicitly points out an element that does not belong to this generalisation. For example, in a sentence like (2), Everyone left but Kim.

Description wise, the construction can be divided into an associative (everyone), an exceptive phrase (except) and an exception (Kim). This tripartite characterisation has been a useful starting point for cross-linguistic research.

von Fintel (1993) distinguishes between connected but-phrases and free except for constructions:

2. The meeting was attended by all but John.
3. Excepting John, all students were present at the meeting.

In this respect, the exceptive restricts the universe of discourse of the relevant quantificational statement. Nevertheless, von Fintel shows that connected but-phrases and free exceptives are not semantically equivalent: the former require a stronger lexical condition, while free exceptives behave more like domain subtractors.

The two domains, then, are different in regard to their primary semantic contribution: verum is about the truth status of a proposition, whilst exceptives are about the exclusion status of an individual or subset from a generalisation. However, they have an important methodological and possibly theoretical property in common. Both are sensitive to the alternatives. In a verum construction, the relevant contrast can be represented schematically as {p, ¬p}, with the speaker marking p as the proposition holding in the discourse. In an exceptive construction, the interpretation requires a larger domain or set, and the

exclusion of one or more members from the relevant generalisation. In this sense, both constructions constrain the structuring and interpretation of alternatives in discourse.

In recent years the verum literature has been very explicit in its arguments against identifying verum with ordinary focus. The notion of generalisations and exceptions is considered as a different semantic relation in the exceptive literature. A more cautious umbrella is discourse-semantic marking, or more precisely, discourse-functional marking sensitive to alternatives and to propositional content.

This point of view is also a good methodological basis for the study of Indian languages. Rather than starting with pre-established grammatical labels like verum particle or exceptive marker, the investigation can begin with discourse functions and work its way to their formal realisation. For example, contexts of correction, contradiction, confirmation, and disputed truth show how languages encode commitment to the truth of a proposition. For exceptives, the contexts of exclusion from universal and non-universal generalisations can show the markers that are available, their syntactic position, the difference between connected and free constructions and the morphosyntactic properties of the exception. The bigger goal is to construct a typological picture of exceptive constructions in genetically and structurally diverse languages.

The present study concerns verum and exceptive marking as two distinct yet potentially comparable domains of discourse-semantic marking. The aim is not to argue that they instantiate one grammatical phenomenon, but to ask whether their cross-linguistic realisation reveals a broader tendency for languages to grammatically encode the management of truth, exclusion and alternatives. Such an approach may also serve to identify areas of convergence between information structure, quantification, discourse semantics and morphosyntax that are obscured when the two phenomena are studied in isolation.

3. Research gap

While there is a huge amount of theoretical literature on verum and exceptive constructions, their distribution and grammatical realisation across Indian languages are still largely undocumented from a comparative perspective. While existing studies of verum have mostly focused on a single language or on cross-linguistic samples not from South Asia, the literature on exceptives has likewise been dominated by semantic analyses based on English and a relatively small number of other languages.

In this context, the present study is a first systematic attempt to document and investigate verum and exceptive marking in Indian languages in a comparative perspective. Instead of treating the English categories of verum and except as directly translatable grammatical categories, the present study takes a function-to-form approach, investigating the ways in which languages encode the discourse

function of affirming the truth of a proposition and the semantic function of excluding an individual or subset from a generalisation. This is particularly important in the Indian context, where languages belonging to several genealogical families and displaying considerable morphological and syntactic diversity may use very different strategies to express these functions.

Consequently, the goal of the study is not only to find lexical equivalents of verum and except, but also to establish their distribution, morphosyntactic realisation, discourse conditions, and interaction with related categories such as focus. For verum, the investigation asks whether prosody, particles or verbal markers express truth-related meanings. For the exceptives, it examines the nature and position of the exceptive marker, the types of associates and exceptions it allows, and the distinction between connected and free constructions. Such a comparative study will provide an empirical basis for determining whether the two domains show patterns that recur across Indian languages, and whether their formal overlap (where it exists) can be understood as part of a larger system of discourse-semantic and alternative-sensitive marking.

In this respect, the study attempts to fill an important descriptive gap. This work is meant to be the first step toward a systematic typology of verum and exceptive marking in Indian languages, providing new empirical documentation and a basis for evaluating existing theoretical claims against data from a linguistically diverse region that has so far received limited attention in this area.

4. Empirical Domain

The present study is based on a genetically diverse sample of five languages representing the major language families spoken in the country. The sample includes Hindi-Urdu (Indo-Aryan), Malayalam (Dravidian), Jungli (Ao) (Tibeto-Burman branch of Sino-Tibetan), Mundari (Austroasiatic), and Kashmiri (Dardic group of Indo-Aryan). The choice is made to provide a broad genealogical and structural basis for the investigation of whether similar discourse-semantic functions are achieved by similar or different grammatical strategies.

The sample contains one of the most widely spoken language varieties in the Indian subcontinent, Hindi-Urdu, which is the Indo-Aryan component. It is an SOV language with a complex system of verbal morphology, postpositions, split ergative case marking and information-structural strategies. The present study takes Hindi-Urdu as an important testbed, because the discourse-related meanings can be expressed by a combination of particles, prosody, verbal morphology, and syntactic configurations.

Malayalam is a South Dravidian language of the Dravidian family spoken primarily in Kerala and the neighbouring regions of South India. Typologically, Malayalam is essentially an agglutinating and suffixing language, grammatical information being expressed in the

main by readily identifiable morphemes attached to the stems. Morphologically, Malayalam is a language with a productive case-marking system, in which the grammatical relations are often expressed by suffixes, including the accusative, genitive, dative, instrumental, locative and other cases. Nouns mark singular and plural, and plural is often marked by suffixes like *-kaḷ*; the pronominal system encodes distinctions in person and number. The verbal system is also mainly suffixing and expresses categories like tense, aspect, mood, negation and besides person, number and gender. One particularly important feature of Malayalam is the prevalence of non-finite and participial forms of verbs in subordinate and relative-clause constructions. Syntactically, Malayalam is basically an SOV language and is predominantly head final. Objects precede verbs, possessors precede possessed nouns and modifiers generally precede their heads. Relative clauses are usually prenominal and can be formed by participial constructions, e.g. a structure corresponding to 'the book that he read' in which the relative clause precedes the noun. Malayalam makes extensive use of non-finite constructions to express temporal, causal and other subordinate relations. SOV is the basic constituent order, but Malayalam allows some flexibility in word order, partly due to morphological case marking helping to identify grammatical roles. Adpositions are mainly postpositional and case suffixes play an important role in encoding spatial, grammatical and semantic relations. Negation can be expressed through negative verbal morphology and auxiliary-like constructions. The interaction of negation with the verb and other constituents results in a number of possible constituent-order patterns. Typologically, Malayalam can be described as a language that is predominantly agglutinative, suffixing, case-marking, head-final and SOV with a syntax that makes extensive use of non-finite and prenominal relative-clause constructions.

Jungli (previously written as Chungli), is a Tibeto-Burman language spoken by the Ao tribe, who predominantly live in the Northeast Indian state of Nagaland. It belongs to the larger Central Naga (or Ao group) of languages. Jungli is an SOV language. It is a tonal language with a three-tone system: Low, Mid, and High tones marked on the vowels. It is largely an agglutinating language and contains both prefixes. It has postpositions and head-initial nouns phrases. Jungli has a rich case system with overt case marking for ablative "*nungi*" agentive *-i*, allative "dangi or *i*", benefactive "*atema*", comitative "*den*", instrumental "*agi*", dative "*nem*" and locative case "*nung*". The existence of Jungli provides the opportunity to test how verum and exceptive meanings are encoded by dedicated particles in Tibeto-Burman languages.

Mundari language family is one of the major non-Indo-Aryan language families of India. It is part of the Munda group of the Austro-Asiatic family. The Mundari language, which is mainly spoken in Jharkhand and neighbouring areas, presents a rich interaction between verbal morphology, argument structure and grammatical processes related to discourse. This language is included to broaden the

genealogical scope of the study beyond the Indo-Aryan and Dravidian families.

In the sample, another Indo-Aryan comparison is Kashmiri, the representative of the Dardic group. It is a language spoken predominantly in the Kashmir Valley and is typologically distinct within the Indo-Aryan domain with a range of morphosyntactic properties distinguishing it from languages such as Hindi-Urdu. Its word order is flexible and exhibits a split-ergative alignment system**.** The alignment is sensitive particularly to aspect/tense subjects of transitive clauses may receive ergative marking in the perfective, while nominative-accusative patterns occur in other environments. Kashmiri has a rich case system, including nominative**,** ergative, dative, ablative, locative, and instrumental functions. This is especially helpful in distinguishing genealogical inheritance from grammatical innovations that are peculiar to a language.

This empirical design allows us to move beyond identifying direct lexical equivalents of English did or except. But the study raises a wider question: how do genetically and typologically diverse Indian languages grammatically encode discourse functions of affirming truth and expressing exclusion? This comparison is expected to reveal recurrent and language-specific strategies and to contribute to the development of a broader typology of discourse-semantic marking in Indian languages.

4.1 Verum particles in Indian languages

This subsection describes different types of verum morphology seen in Indian languages. In Jungli, for context, where A says 'Ravi did not go to Delhi yesterday' and B knows this is false and corrects A by saying 'Ravi did go to Delhi', the verum construction is seen in (4).

5. *ravi yashi* *Delhi* *ao-a* *ta*
   Ravi yesterday delhi go.PST-DEC V
   'Ravi DID go to Delhi yesterday'

However, without any prior discourse context, the sentence will be as seen in (5).

6. *ravi yashi* *Delhi* *ao-a*
   Ravi yesterday delhi go.PST-DEC
   'Ravi DID go to Delhi'

For context, everyone expected Ravi not to come to the meeting, but he came. Now A says, Ravi did come to the meeting

7. *ravi* *meeting-i* *aru-a* *tu*
   Ravi meeting-ALL come-PST-DEC V
   'Ravi DID come to the meeting'

However, in the case of surprising information and confirmation, the *ta* marking disappears, as seen in (7).

8. *ravi* *shiitake-ba* *yashi* *aru-a* *na*
   ravi *true-NOM* *yesterday* *come.PST-DEC INT*
   'Did Ravi REALLY come yesterday?'

Jungli uses various particles to encode truth semantics. Example (4) shows *ta* which gives a sense of that the speaker has been saying this truth repeatedly and the listener is wrong, and it is in fact true that Ravi went to Delhi yesterday. Example (5) has particle *tu*, which is used when there is a sense of 'contrary to what was expected', when it is in fact true that Ravi came to the meeting'. Particle *ja* is also used for two cases- emphasis and correction. For example, see (8).

9. *Masu, ravi delhi ao-a ja*
   no, ravi delhi-ALL go.PST-DEC V
   'No, Ravi went to to Delhi'

Now we move to Malayalam data

10. *i. Ravi innale (serikum) delhi-kky po-i*
    Ravi yesterday really Delhi-DAT go.PST
    'Ravi DID go to Delhi yesterday'

    ii. *ravi innale delhi-il pov-uka tanne cheyth-u*
    Ravi yesterday Delhi-LOC go-NMLZ EMPH do-PST
    'Ravi DID go to Delhi yesterday'

11. *ravi innale serikum vann-irunn-o*
    Ravi yesterday really come-be.PST-Q
    Did Ravi really come yesterday?

Another strategy of verum marking is also seen in Malayalam, where the copula is used. For example

12. *Illa, ravi delhi-kky-aanu poyat-u*
    no Ravi Delhi-DAT-COP go.PST-NMLZ
    'No, Ravi went to DELHI*'*
    *(context: A: Ravi went to Srinagar; B corrects only the destination and says '*No, Ravi went to DELHI*')*

Malayalam uses two particles for verum marking- *serikum* and *tanne.* However, (9 (ii)) is rarely used, though it is grammatical.

The particle 'serikam' is used to represent verum semantics in (9(i)). The emphatic marker *tanne* can also be used as a verum marker.

Next, we show verum strategies in Mundari below.

13. *Ravi hola Delhi seno[ʔ] ke-dan*
    Ravi yesterday delhi go be.PST
    'Ravi DID go to Delhi'

14. *Ravi hola hiju[ʔ] ta[ʔ] ke-dan*
    Ravi yesterday c ome Q be.PST
    Did Ravi really come yesterday?

Mundari has no dedicated verum marker. However, we see a rising tone to denote verum semantics. A similar pattern is used in Kashmiri, as seen below.

15. *ravi gov dili ra:th*
    ravi.NOM go.PST Delhi yesterday
    'Ravi DID go to Delhi yesterday'

16. *ravi* *ava:* *pazej* *rath*
Ravi come-PSR-3 INT yesterday

Hindi-Urdu uses a dual strategy to denote verum marking. They are particle *to* and rising prosody.

17. i. *ravi* *to kal* *dili* *gaye*
Ravi V yesterday Delhi go-PST
ii. ravi *kal* *dili* *gaye*
Ravi yesterday Delhi go-PST
'Ravi DID go to Delhi yesterday'

The marker *to* is multifunctional (See Bhalla, 2022 for a discussion). However, its role as verum marker is yet unexplored.

In examples, such as (16), *sach mei* in rising tone can also be used.

18. *kya* *ravi* *sach* *mei* *kal* *aaye*
Q ravi true LOC yesterday go-PST
'Did Ravi REALLY come yesterday'

We summarise the findings in a schematic format in the table below.

Language	Family	Verum strategy	Marker/ strategy	Core discourse function observed
Jungli	Tibeto-Burman	Particle-based (multiple)	*ta*	Correction / rejection of a false assertion; speaker insists that *p* is true
			tu	Contrary to expectation
			ja	Emphasis + correction
			ta/tu absent in some contexts	Confirmation / surprise
Malayalam	Dravidian	Particle + emphatic + copular strategies	*serikum*	Affirmation / confirmation of truth
			tanne	Emphasis / verum
			-āṇu (copula)	Corrective focus on a constituent
			serikum / interrogative strategy	Confirmation / surprise
Mundari	Austroasiatic	Prosodic	Rising intonation	Verum / confirmation
Kashmiri	Dardic	Prosodic	Rising tone	Verum / emphatic affirmation
			Rising intonation in questions	Confirmation
Hindi-Urdu	Indo-Aryan	Particle + prosodic	*to*	Correction / emphatic affirmation
			Rising prosody	Verum / emphasis
			sach mein + rising tone	Confirmation / surprise

Table 1. Summary of Observations: Verum Marking

4.2 Exceptive marking in Indian languages

Phrasal exceptive marking is seen in two ways in Hindi-Urdu. Those are the use of phrasal *chhodkar* and *bagair* and clausal *nehi aye* (with or without the connectors but or and) as seen in (18-19). The rare use of the complex predicate *reh gaye* is also observed.

19. *ali* *ke bagair/chodkar* *koyi* *nehi* *aya*
    Ali GEN without noone NEG come-PST.3
    'Nobody came except Ali'

20. *Sab* *aa gaye* *sirf* *ali* *nahi aya* *everyone*come
    go-PST only ali remain go-PST
    'Everyone came, except that Ali did not come'

Jungli uses *alaka* in exceptive constructions. In clausal exceptions, it uses the connector 'but', as seen below.

21. *ali* *alaka* *kaketshir-tem* *kar* *aru*
    Ali except students-PL some come.PST
    'Some students came except Ali'

22. *Ajak* *aru* *saka* *Ali* *m-aru*
    All come.PST but Ali NEG-come.PST
    'Everyone came, except that Ali did not come'

Kashmiri uses *vərəːj* in both phrasal exceptive constructions and uses *khali in*, as seen below.

23. *səːriː* *sabak* *ʃurʲ* *aːjɨ* *alijas* *vərəːj*
    All Students come.PST. 3Pl. Ali-DAT except
    'Some students came except Ali'

24. *pretʰ* *kanh* *aːv ,* *kʰeːliː* *aːv nɨ* *aliː joat.*
    Everyone come.PST. 3Pl except Ali
    'Everyone came, except that Ali did not come'

In Malayalam, we see use of both *ozhike*, the exception denoting particle and also the use of *only*, in (24-25), respectively.

25. *Ali* *ozhike* *ellavarum* *vann-u*
    ali except everyone come-PST
    'Everyone came except Ali'

26. *ellavarum* *vann-u* *pakshe Ali* *matram*
    everyone-ADD come-PSTbut Ali only *vann-illa*
    come-NEG.PST
    'Everyone came, except that Ali did not come'

In Mundari, *bageke* is used for exceptive constructions

27. *Ali do rabi bageke sobe* *hiju kedan*
    ali and ravi except everyone here be.PST
    'Everyone is here except Ali and Ravi'

We summarise our findings regarding exceptive constructions below.

Language	Type of exceptive	Marker / strategy	Main observation
Hindi-Urdu	Phrasal	*chhodkar, bagair*	Uses dedicated phrasal expressions for exclusion. *bagair* is a 'without' expression, while *chhodkar* has an exceptive/exclusion use.
	Clausal	*nahi aya*, with *sirf* and optionally a connector such as *but/and*	Clausal exception is expressed through a negative clause, with the exception interpreted as the one entity excluded from the generalization.
	Complex predicate	*reh gaye*	A relatively rare strategy for expressing exception, suggesting that exceptive meaning can also arise through a complex predicate construction.
Jungli	Phrasal	*alaka*	*alaka* functions as an overt exceptive marker introducing the exception. Notably, it can occur with a non-universal associate ('some students').
	Clausal	*saka* 'but' + negative clause	Clausal exceptions are formed with the connector *saka* 'but', followed by a negative clause expressing the exception.
Kashmiri	Phrasal	*vərəːj*	*vərəːj* functions as the overt phrasal exceptive marker and introduces the exception.
	Clausal	*khali* / negative-clausal strategy	Clausal exception is expressed through a construction involving a negative clause; the exact status of *khali* requires further investigation.
Malayalam	Phrasal	*ozhike*	*ozhike* is a dedicated exceptive marker meaning roughly 'except'. It directly introduces the excluded member.
	Clausal	*pakshe* 'but' + *matram* 'only' + negative clause	Exception is expressed through a separate negative clause. The combination of *but* and *only* produces an exceptive interpretation.
Mundari	Phrasal and Clausal	*bageke*	Exception is expressed through phrase *bageke*

Table 2. Summary of Observations: Exceptive Marking

5. Key Findings

The preliminary data indicate that there is substantial cross-linguistic variation in the formal realisation of both verum and exceptive meanings across the five languages. As for verum, the languages seem to have various strategies rather than a single dedicated grammatical category. Jungli presents the richest particle-based system, where *ta*, *tu*, and *ja* are associated with distinct discourse conditions such as correction, counter-expectation and emphasis. Malayalam has overt particles (*serikum*, *tanne*) and a copular construction that may mark corrective focus. Hindi-Urdu also combines the multifunctional particle to with prosodic marking, whereas Mundari and Kashmiri, based on the present data, seem to predominantly use prosody, especially rising intonation, in expressing verum-like interpretations. These patterns suggest that verum meanings may be spread along a continuum from overt lexical or particle-based marking to prosodic and constructional strategies. Crucially, the data also suggest that discourse context is an important factor in the realisation of verum: correction, counter-expectation, emphatic affirmation, and confirmation/surprise do not necessarily call for the same formal strategy, as is especially apparent in Jungli and Malayalam.

The expression of exceptive meanings is similarly variable. The languages use both phrasal strategies, where an overt expression directly introduces the exception, and clausal or coordination-based strategies, where the exception is expressed through a separate negative clause. In Hindi-Urdu, chhodkar and bagair are used in phrasal constructions, and a construction involving sirf 'only' and a negative clause provides a clausal strategy. . In phrasal constructions, Jungli uses *alaka* as an overt exceptive marker while in clausal constructions, it uses *saka* 'but' and negation. In Malayalam phrasal constructions, the dedicated exceptive marker is *ozhike*, while the corresponding clausal meaning is expressed by a combination of *pakshe* 'but', *maathram* 'only', and negation. Kashmiri also has an overt phrasal marker, vərəːj, but its clausal strategy remains to be investigated. One particularly interesting finding is that a non-universal associate occurs in the Jungli data, which means that exceptive constructions are not necessarily limited to universal quantificational contexts. These findings are promising, showing that both domains display a considerable amount of form-function variation, with languages recruiting dedicated particles, multifunctional markers, prosody, copular constructions, coordination, negation and other grammatical resources. These patterns support a function-to-form approach to the comparative study of verum and exceptive marking, and raise the broader question of whether apparently similar discourse-semantic functions are grammaticalized

6. Conclusion

The initial comparison of Hindi-Urdu, Malayalam, Jungli, Mundari and Kashmiri reveals substantial cross-linguistic variation in the expression of verum and exceptive meanings. Verum is expressed through the use of particles, prosody, emphatic constructions and copular strategies, while exceptive meanings are expressed through phrasal markers and also through clausal and coordination-based constructions. The data also suggest that these meanings are not uniformly encoded by dedicated grammatical markers, but are strongly conditioned by discourse context. Particularly interesting are the many discourse-sensitive particles in Jungli, the interplay of emphasis and verum in Malayalam and Hindi-Urdu, and the apparent dependence on prosody in Mundari and Kashmiri. To summarise, the study provides initial empirical documentation of two relatively understudied domains in Indian languages and demonstrates the potential of a function-to-form approach for revealing how genetically diverse languages encode truth, exclusion, and alternatives. The findings also provide a basis for a larger typological investigation of discourse-semantic marking across Indian languages.

References

Bhalla, O. (2022). *'To'-marked XPs in Hindi* [Doctoral dissertation, Jawaharlal Nehru University].

Gutzmann, D., & Castroviejo Miró, E. (2011). The dimensions of verum. In O. Bonami & P. Cabredo Hofherr (Eds.), *Empirical issues in syntax and semantics 8* (pp. 143–165). CSSP.

Gutzmann, D., Hartmann, K., & Matthewson, L. (2020). Verum focus is verum, not focus: Cross-linguistic evidence. *Glossa: A Journal of General Linguistics*, *5*(1), 51, 1–48. https://doi.org/10.5 334/gjgl.347

Höhle, T. N. (1992). Über Verum-Fokus im Deutschen. In J. Jacobs (Ed.), *Informationsstruktur und Grammatik* (pp. 112–141). Westdeutscher Verlag.

Jordanoska, I., Kocher, A., & Bendezú-Araujo, R. (2023). Introduction special issue: Marking the truth: A cross-linguistic approach to verum. *Zeitschrift für Sprachwiss enschaft*, *42*(3), 429–442. https://doi.org/10.1515/zfs-2023-2012

Kerr, E. J., & Van der Wal, J. (2023). Indirect truth marking via backgrounding: Evidence from Bantu. *Zeitschrift für Sprachwissenschaft, 42*(3), 443–492. https://doi.org/10.1515/zfs- 2023-2010

von Fintel, K. (1993). Exceptive constructions. *Natural Language Semantics, 1*(2), 123–148. https://doi.org/10.10 07/BF0037 2560