

From Documentation to Participation: Ethnographic Reflections on Tribal Language Preservation in Eastern India

Alisha Elin Indwar, Shreya Pathak
[https://doi.org/ 10.61410/had.v21i2.285](https://doi.org/10.61410/had.v21i2.285)

Abstract- Community-led initiatives in digital and cultural innovations play a pivotal role in preserving and promoting the endangered Santali, Mundari and Ho languages in Jharkhand, West Bengal and Odisha. For generations together these languages have survived through vibrant oral tradition of knowledge sharing and most of the time with no definite written script. Although festivals, music, dances, farming and other collective cultural activities have played a pivotal role in preservation of the languages yet the digital world has prominent challenges to overcome when it comes to achieving accuracy in translation and transmission of these tribal languages.

This paper seeks to explore how these three respective languages of Eastern India experience, negotiate and reshape digital and cultural innovations for language preservation within their everyday social practices. The research highlights a shift from mere documentation of tribal languages to participatory preservation. Addressal to two main questions will be made: How have these digital platforms reshape everyday linguistic practices? And what were the challenges encountered by them while incorporating such technological advancements? The current affairs on language preservation of tribal languages and government efforts through NEP 2020 has been kept as the background focal point of the research.

Keywords -Tribal Language Preservation, Ethnography, Community Participation, Indigenous Knowledge, University–Community Partnership, Community-led initiatives, Digital innovations, Cultural inclusivity, Santali, Mundari, Ho, Jharkhand, West Bengal, Odisha, Oral tradition, Written script, Language preservation, Digital world, Translation accuracy, Eastern India, Linguistic practices, Participatory preservation, Digital platforms, Technological advancements, NEP 2020.

Introduction- The linguistic landscape of the eastern tribal language- especially Santali, Ho and Mundari- have been profoundly defined by oral tradition alone for many centuries together. The early 20th century saw a colossal wave of nationalism that compelled people to embrace their roots and culture. English education act of 1835 had imposed compulsory education in English language in government managed schools with western education accentuating the need for preservation of local

-
- Designation: MA History, Institutional Affiliation: Vasanta College For Women, BHU
 - Co-AuthorProfessor, Institutional Affiliation: Vasanta College For Women, BHU
-

languages. The national movements of swadesi and boycott had ignited the will for restoration and revision of the culture rooted in ancient times in the Indian hearts. I believe these background movements had vitally initiated some milestone developments. In 1925, Pandit Raghunath Murmu, a visionary educator of Maurbhanj district, Odisha, aimed to cultivate cultural pride and improve education by developing a 30-letter alphabet set for Santali language which historically lacked a script. The script came to be known as Ol Chiki replacing the use of scripts from neighbouring languages like Odia, Bengali, Roman or Devnagiri. Ol Chiki was carefully engineered to match the poetic tone, phonetic structure and nasal tone of Santali language which the scripts of other languages couldn't provide. It was further refined and finalized in 1930s. With independence and adoption of constitution, Schedule Tribes assumed significant importance in preservation of cultural heritage, including the protection and promotion of their indigenous languages. In 1949, Rohidas Singh Nag had taken upon himself to provide a unique writing system for Mundari language. Mundari Bani (script of Mundari) was developed by him as a 27-letter alphabet set which latter was refined in the coming generations. Such initiative was winded by the will to preserve Mundari and provide it with a unique and distinct writing system which separated the script from Indo-Aryan languages of Odia and Devnagiri. Rediscovering effort of Lako Bodra came to be a success with the refined script of Ho language known as Warang Chiti (also written Varang Kshiti). Bodra invented it as an alternative to the writing systems devised by Christian missionaries. He claims that the alphabet was invented in the 13th century by Deowan Turi, and that it was rediscovered in a shamanistic vision and Bodra modernized. All the efforts appeared to have been shouldered with an objective of preserving indigenous heritage and foster a sense of pride in it

21st century has seen rapid digital development and while the other Indo-Aryan languages have kept with the digital world, it almost becomes obligatory for this generation to facilitate the advancement of the tribal language in digitalized form. The handwritten records have often failed to transmit culture as the either get manipulated or damaged in the process of transmission. Digitalization has filled in this gap, yet several challenges persists in forms of inaccuracy, unawareness and usage difficulties in reality. This paper seeks to view the problem of digital accuracy for tribal languages various angles to come up with the best outcome possible to bridge the gaps.

Literature Review: One of the most important scholarly contributions in this field is '[*Austroasiatic Linguistics*](#)' edited by [K. S. Nagaraja](#) and [Kashyap Mankodi](#). The work

brings together research papers of different scholars and examine the phonological and grammatical features of various Austroasiatic languages, particularly those belonging to the family tree of proto-Munda language. The contributors to the compilation emphasize that languages such as Santali language, Mundari language, and Ho language display distinctive linguistic structures that differentiate them from Indo-Aryan and Dravidian language families. This volume also explores historical relationships and linguistic evolution, offering valuable insights into the origin and development of tribal languages. Another significant contribution, particularly to the Munda subgroup of the Austroasiatic language family, is 'The Munda Languages' edited by Gregory D. S. Anderson. This volume provides a comprehensive linguistic overview of the Munda languages spoken primarily in eastern and central India. Anderson emphasizes that these languages possess distinctive grammatical structures, including complex literary syntax and unique phonetic features, differentiates them from Indo-Aryan and Dravidian languages. Besides structural analysis, Nishaant Choksi attempts to study the sociolinguistic dimensions of the Santali language in his paper: 'Script as constellation among Munda speakers: the case of Santali'. Choksi's work also emphasizes that script choice is closely tied to identity, literacy, and institutional recognition. Press Information Bureau published an article on 'PRESERVATION OF TRIBAL LANGUAGES' by Ministry of Tribal Affairs further attempts to clear government motives on preservation and promotion of tribal languages. The article lays down the instruction which helps formulated the purpose of writing this paper more clearly.

Methodology: *The research paper calls for a dual approach method implemented simultaneously for effective analysis. Survey research methodology conducted for primary data of collection through questionnaire helped bridge the gap between assumptions and reality. The survey conducted through Google Form collected data from 35 people from eastern states of Odisha, Jharkhand and Bengal. Questions presented to them were carefully curated to have a clear implication of tribal languages in digital world. Participants were approached through social media application (i.e., WhatsApp) and consent forms were required to be signed at the end of the survey. Qualitative research methodology based on data analysis and literature review has been pursued to make a deeper understanding of the gaps between hypothesis and practicality. Moreover, the study adopts an interdisciplinary approach, drawing from linguistics, anthropology, and historical cultural studies to analyse tribal language preservation comprehensively. This methodology enabled a broader understanding of tribal languages not only as*

linguistic systems but also as carriers of indigenous knowledge, cultural identity, and community heritage.

Discussions: Based on the census of 2011, Santali takes the 3rd most populous tribal community in India followed by Munda taking its position in top ten. Origin, development and pre-history of these North-munda languages is obscure and cannot be traced back in time. According to Gregory D. S. Anderson 'at some point the ancestors of munda-speaking people must have migrated westward into the subcontinent'. How, Why and By what path still remains a topic of debate. The traditional classification of Munda is found in Zide (1969) and Zide and Stampe(1968) which is graphically represented:

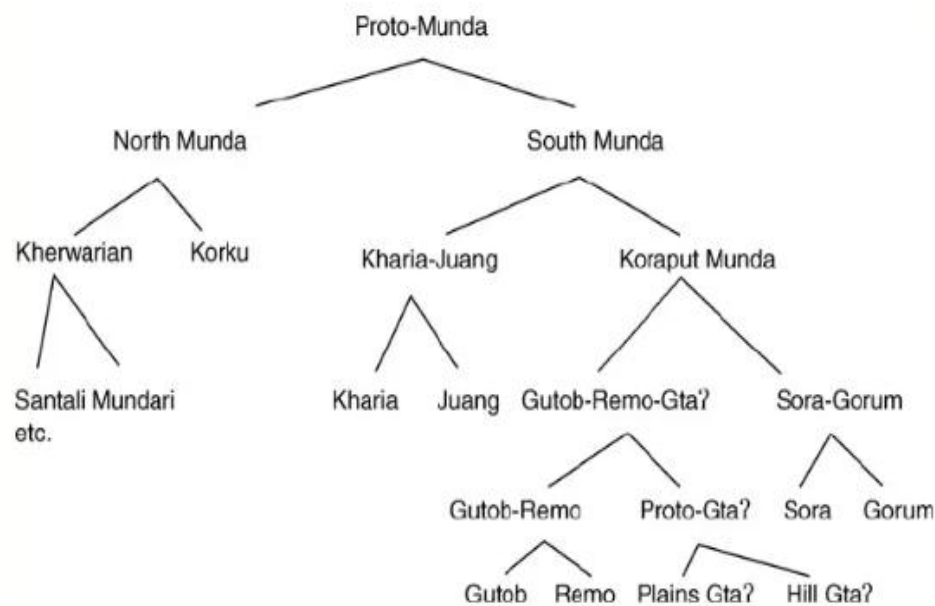


Fig: 1.1 Tree Diagram of classification of Munda languages .

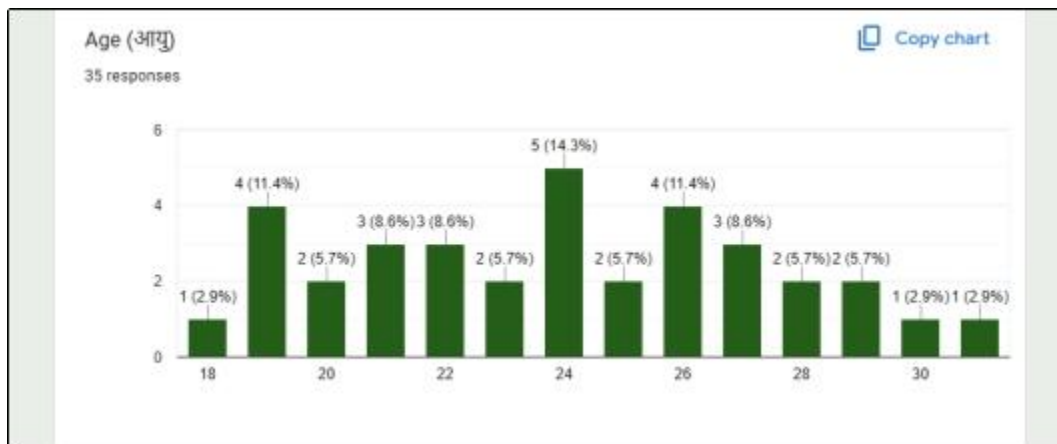
These languages have a long surviving history something that even the colonial power could not ignore. The Santal Dictionary (1932-1936) compiled by Rev. P. O. Bodding is a monumental work, providing not only descriptive linguistic data but also a rich reference of cultural information. The digitalization of the dictionary, started in 1991 as an Indo-Japanese joint research project, has been almost completed except some proof-editing work. [This collaborative research project involving scholars from University of Oslo and Tokyo University of Foreign Studies, marked a significant step toward preserving and modernizing this linguistic resource. The project aimed to](#)

convert the multi-volume printed dictionary into a searchable digital format, thereby improving accessibility for researchers, students, and native speakers. Digitalization also facilitates linguistic analysis, cross-referencing, and long-term preservation of the material. Another significant effort is the People's Linguistic Survey of India led by Ganesh N. Devy. It is a multi-volume book series documenting India's living languages, edited by linguist and literary scholar Ganesh N. Devy. Conceived as a citizen-led alternative to colonial-era linguistic surveys, it aims to record the diversity, vitality, and cultural context of Indian languages in the 21st century. PLSI was launched under the Bhasha Research and Publication Centre in 2010 to update the century-old Linguistic Survey of India. Over 3,000 scholars, activists, and community members were mobilized to catalogue linguistic diversity from a people's perspective rather than a purely administrative one.

A landmark initiative towards the tribal empowerment, development and linguistic preservation by The Ministry of Tribal Affairs was the launch of Beta Version of Adi Vaani, India's first AI-powered translation platform for tribal languages. Adi Vaani has been developed by a national consortium led by IIT Delhi, in collaboration with BITS Pilani, IIIT Hyderabad, IIIT Nava Raipur, and the Tribal Research Institutes of Jharkhand, Odisha, Madhya Pradesh, Chhattisgarh, and Meghalaya. Launched in 2022 under the National Language Translation Mission, Bhashini becomes another AI powered application that aims at breaking language barrier by translating real time conversations voice to voice, speech to text and text to text for numerous Indian languages including Ho and Santali. Developed by Ministry of Electronics and IT, it aims for seamless communication for the generation. Sanchika, a digital repository for Indian languages aggregates dictionaries, primers, storybooks, and multimedia resources for Scheduled and tribal languages. Managed by the Central Institution of Indian Languages, this centralized archive is a vital data source for training language models, developing translation system and preserving cultural narrative. The platform provides digitally organized textual, audio and visual resources supporting academic researches and language learning. These comprehensive collections further serves as a foundational data for developing artificial intelligence and 'Natural Language Processing' applications. Anuvadini again an AI-based multilingual translation platform developed by the All India Council for Technical

Education (AICTE), claims to be a 'small but powerful translation tool' for quick and accurate language conversion. Anuvadini enables practical use cases such as: expanding access to higher education and technical knowledge for students in regional languages, supporting skilling and capacity building by making learning resources available in local languages and strengthening Indian languages as mediums of knowledge, not only communication. Many such government initiatives has come up in the last decade that has significantly affected the way languages are preserved.

However, the extent to which these opportunities are accessible and effective for tribal language speakers remains a matter of concern. Santali being the only tribal language recognized in the 8th schedule of Indian constitution stands as a proof alone. To get a deeper understanding of such gaps and to ultimately find the answers to the questions mentioned in the abstract, survey was carried with 35 respondents to the questionnaire. Each of them have given their assent to use their



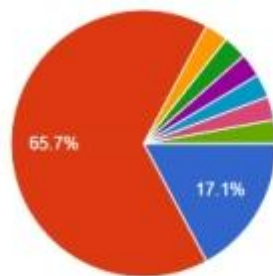
response for academic purpose.

Fifteen females and twenty males were the respondents belonging to the age group of 18 to 30. 23 respondents from Odisha, 10 belonged to Jharkhand and 2 from West Bengal. While 64.71 percent of the respondents were student, while the other were employees of government jobs, teachers and private sector. All respondents identified themselves as native speakers of Santali, Ho, and Mundari languages.

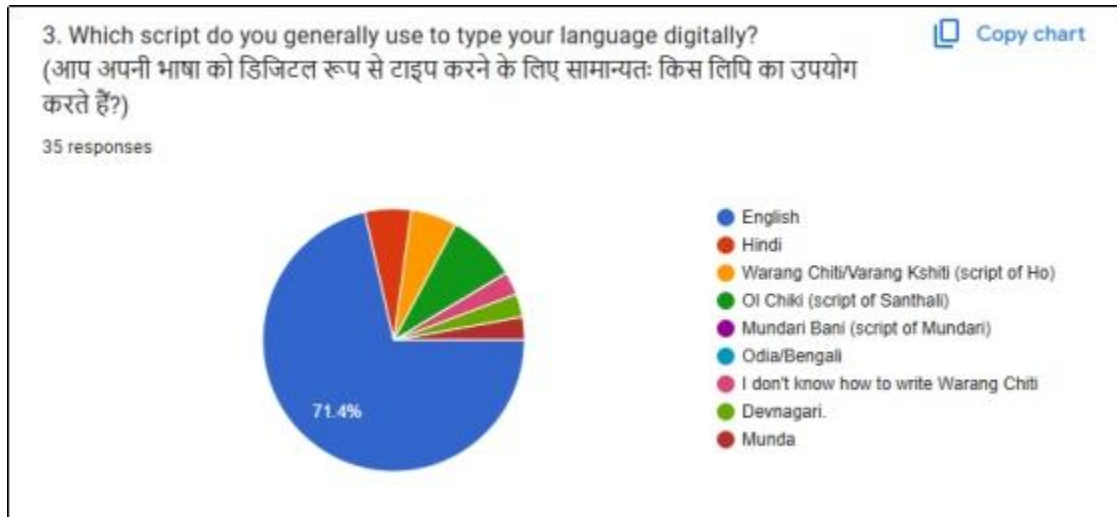
Analysing the responses indicate that most participants continue to use tribal languages in informal settings such as family interactions and community

meetings. However, the usage of tribal languages decreases significantly in formal domains, particularly in education, professional communication, and digital platforms. Respondents have expressed significant cultural and emotional attachment towards their tribal language yet there is limited usage of tribal languages in digital communication. This observation imparts that tribal language still continues to exist primarily as oral language with limited digitalization. Responses reveal varying levels of awareness regarding digital platforms supporting tribal languages. While some respondents reported knowledge of keyboards, fonts, and social media usage, a significant portion indicated limited awareness of available digital tools. 65.7 percent of the responses says they don't find it easy to write their language digitally while some admit to use their language in digital social media platforms with scripts of Odia, Devnagiri and Roman. Respondents admits that borrowing of neighbouring languages' script has effect on the original pronunciation and sometimes meaning of the appropriate word. This ultimately affects the accuracy of writing and effective communication. Further what makes the situation grave is the youth turning towards the use of dominant languages of Hindi or English in the digital spaces, which makes it harder to promote or propagate the language. The accuracy of digital scripts is yet to be achieved as seen from the responses.

35 responses



- Yes (ହଁ)
- No (ନାହିଁ)
- May be
- It's not even available on Google Keyboard.
- There is a digital keyboard for Ol Chiki(Santali), but it is still not very co...
- usually through odia or hindi keyboards
- Don't know about the script and letters...
- No, it is not very easy to write my lang...



While all the participants have shown moderate awareness towards the government-led initiatives for tribal language preservation, many still have not involved themselves into any such schemes or became a participant of these programmes. This forms a clear indication that there is still a gap to be filled with continuous efforts of approach. NEP 2020 according to some have become a major milestone for achieving success for the filling the gap mentioned above. Some Board of Secondary Education schools in Odisha have been reported to incorporate Santali, Ho and Munda languages in their studies, while the script they use to teach these languages are still obscure. 98 percent of the respondents are in favour of university-community initiatives for achieving accuracy in preservation of tribal scripts.

Coming back to the two questions I assured to seek an answer for - how have these digital platforms reshape everyday linguistic practices? And what were the challenges encountered by them while incorporating such technological advancements? - The survey responses presents a nuanced and layered reality. The findings suggests that digital platform have gradually begun to reshape everyday linguistic practices by encouraging informally written or oral communication in tribal languages. Many respondents have reported using native languages in messaging apps, social media comments and community based digital groups. This heralds a new era of shifting from traditional oral usage towards written communications digitally, particularly among youth. Digital keyboards, Unicode fonts, and localized content have enabled users to experiment with writing their languages more frequently. In some cases, respondents also mentioned using tribal

languages to share cultural content, folk stories, songs, and community announcements, thereby strengthening linguistic identity in digital spaces. Nevertheless the transition feels partial and uneven. The vast borrowing of scripts from dominate for easy typing, wider readability and familiarity suggests that usage of tribal languages have not yet been normalized in everyday digital communication. The most commonly reported lack was absent of hardware gadgets like tribal language keyboard and an environment for these languages to shine. As the population of tribal people is pretty less in formal settings, frequent communication even with their own people becomes rare. Thus organizing of tribal community programmes in formal setting can encourage the tribals to come forward and uphold their culture while ensuring survival of these language.

A constant generational differences were observed where younger participants were more open to experimenting with digital tools that older ones. A prominent gap discovered during the survey was that many older generation people have problem writing a standardized script of their own language. As the scripts developed by individual efforts were made during colonial era, much of the uneducated class, which constituted almost of the tribal population, remained unaware of such developments. The script was preserved and carried on by only some educated tribals. Thus such a limitation becomes difficult to address.

This paper considers it necessary to develop digital hardware like keyboard for the Santali, Ho and Mundari languages while also ensuring community-based awareness and training initiatives. Without the development of digital resources such as keyboards, fonts, educational materials, and translation tools, preservation and communication becomes fragmented and inconsistent, limiting the effective use of tribal languages in digital spaces. Once keyboards make typing of these languages easy, further digitalization of oral traditions, folklores, songs and traditional knowledge can also be initiated, thereby preserving cultural heritage for the future generation to come. Additionally, such efforts will encourage inclusion of tribal language in social media platform, digital apps and academic resources. The difference can be brought about by the university-community collaboration towards achieving accuracy in the process. Awareness campaigns, community workshops and digital literacy programmes can become the next step for generational gap identified during the survey.

Reference :-

1. Gregory D.S. Anderson. The Munda Languages. Routledge, 2015, p. 3.

2. Nagaraja, K. S., and Kashyap Mankodi. *Austroasiatic Linguistics*. Central Institution of Indian Languages, 2010, p. 260,
<https://archive.org/details/dli.language.0100/page/n3/mode/2up>.
3. Nagaraja, K. S., and Kashyap Mankodi. *Austroasiatic Linguistics*. Central Institution of Indian Languages, 2010, p. 260,
<https://archive.org/details/dli.language.0100/page/n3/mode/2up>.
4. Devy, G. N., editor. *The Being of Bhasha: General Introduction to the People's Linguistic Survey of India*. Vol. 1, Orient BlackSwan, 2014.
5. Devy, G. N., editor. *The Being of Bhasha: General Introduction to the People's Linguistic Survey of India*. Vol. 1, Orient BlackSwan, 2014.
6. "Ministry of Tribal Affairs Launches Beta Version of Adi Vaani – India's 1st AI-Powered Translator for Tribal Languages Leading to Large Language Model." Pib.Gov.In, 2025,
<https://www.pib.gov.in/pressreleasepage.aspx?prid=2162846>.
7. "Digital India BHASHINI Division, Current AI and Kalpa Impact Launch VYOMA Innovation Challenge to Advance Multilingual Edge AI Innovation." Pib.Gov.In, 2026,
<https://www.pib.gov.in/pressreleasepage.aspx?prid=2268112®=3&lang=1>.
8. "22 Languages, Digitally Reimagined." Pib.Gov.In, 2017,
<https://www.pib.gov.in/pressreleasedetail.aspx?prid=2182427®=48&lang=2>.
9. "22 Languages, Digitally Reimagined." Pib.Gov.In, 2017,
<https://www.pib.gov.in/PressReleaseDetailm.aspx?PRID=2182427®=3&lang=2>.
10. "Artificial Intelligence for Culture and Languages." Pib.Gov.In, 2026,
<https://www.pib.gov.in/PressReleasePage.aspx?PRID=2225474®=3&lang=2>

**Alisha Elin Indwar , Email ID: alisha2indwar@gmail.com
Contact Number: 7978127273**
