

Documenting India's Endangered Languages

"Our family god called *Kuladheivam* in Tamil, has a temple in our village called *polliwoj*," said the 66-year-old Kurtaz Vasamalli, who lives in the Nilgiri hills near Ooty. "We believe that the divine power comes to our villages from the temple...there is no idol worship."

The hills, which are renowned for tea and coffee plantations, are considered sacred by the Toda tribe, who have lived there for thousands of years. The pastoral community believes that their gods and goddesses lived among them once and became a part of the sacred landscape over time.



Figure 1 - Mudumalai Tiger Reserve in the Nilgiris district of Tamil Nadu

Vasamalli knows her oral traditions through her native language Toda. Toda is an endangered proto-South-Dravidian language which evolved in isolation after splitting from South Dravidian (like Kannada, Telugu and Malayalam). The Toda people number a few

thousand now. Vasamalli is one of the elders of the community striving to preserve her traditional oral heritage.



Figure 2 - A goddess temple called *teki*



Figure 3 - Toda male dance called *o:ft*

Vasamalli most recently worked with the Scheme for Protection and Preservation of Endangered Languages (SPPEL) team being implemented by the Mysuru-based Central Institute of Indian Languages (CIIL) under the Ministry of Education. SPPEL works to document and archive the country's endangered languages—languages spoken by fewer than 10,000 speakers or languages not previously studied linguistically.

SPPEL aims to preserve and revitalize endangered languages for future generations by conducting field work and documenting the grammar and words of languages, creating documentaries, bi-lingua/tri-lingual dictionaries, pictorial glossaries and ethno-linguistic profiles, and uploading the documentation work, including audio files, to online repositories for worldwide access.

Currently, SPPEL has identified 117 endangered languages, and is working toward documenting about 500 lesser-known languages in the future.

Various other schemes by the Government of India work toward the preservation and promotion of tribal and vanishing cultures. India's National Education Policy 2020 promotes multilingual studies to preserve local languages.

This local effort reflects a global concern. Endangered languages and cultures are a global phenomenon. Close to half of 7,000 languages globally are endangered globally, according to the United Nations Educational, Scientific and Cultural Organization. UNESCO has declared the period between **2022 and 2032 as the International Decade of Indigenous Languages** to draw attention to the critical situation of many indigenous languages and to mobilize stakeholders and resources for their preservation, revitalization and promotion.

The International Day of the World's Indigenous Peoples on August 9 was established by the United Nations to draw attention to the rights of indigenous peoples living in 90 countries.

AI and Indigenous Rights

This year's United Nation's International Day of the World's Indigenous Peoples (August 9) focuses on "**Indigenous Peoples and AI: Defending Rights, Shaping Futures**"—highlighting both AI's promise and perils for indigenous communities.

The Challenge

- AI development dominated by major tech companies with little indigenous representation and indigenous knowledge is being used in AI training without proper consent perpetuating colonial patterns and widening digital divides

Indigenous Innovation

- Polynesian communities using AI for reef conservation projects
- New Zealand's TeHiku Media employing AI for Māori language revitalization

AI and Technology in Indian Tribal Language Preservation

- India's Ministry of Tribal Affairs is funding *AI-based language preservation* through Tribal Research, Information, Education, Communication and Events (TRI-ECE) scheme, which includes:
 - *Bhasha Research and Publication Centre, Vadodara*: Rs. 58.70 lakh (2019-20) for studying and documenting Adivasi languages, culture and life-skills
 - *BITS Pilani with IITs and Bhashini*: Rs. 3.122 crore (2024-25) for developing AI translation tools converting English/Hindi text/speech to tribal languages and vice-versa

India's Endangered Languages

The effort for the survey and documentation of India's linguistic diversity has deep historical roots. The first linguistic survey of Indian languages was launched in 1894 by George Abraham Grierson, who also prepared the language survey report for the 1901 Census. The Linguistic Survey of India was published over a 25-year period between 1903 and 1928 and consists of 11 volumes (in 19 parts) comprising over 8,000 pages describing the languages and dialects of large parts of British India. It listed a total of 179 languages and 544 dialects.

Subsequent Indian censuses have continued to map the country's linguistic landscape. The 1961 Census of India recorded 1,652 mother tongues, of which SPPEL has identified 117 endangered languages. By 2011, this diversity had expanded significantly, with the census recording 2,843 different mother tongues. After further linguistic analysis, 1,369 were classified as recognized languages that experts could properly identify, while 1,474 remained unclassified. The census grouped all mother tongues with more than 10,000 speakers under established language categories. Languages spoken by fewer than 10,000 people are considered endangered.

A UNESCO report, developed in collaboration with the Indira Gandhi National Centre for Arts under the Ministry of Culture, emphasizes the urgent need to protect India's tribal and indigenous languages. The report highlights that when a language disappears, it takes with it irreplaceable cultural heritage and traditional knowledge systems.

India's endangered languages span the country's diverse linguistic families and are found across the country:

Zone	States/UTs Covered	Number of Languages	Some Languages
Northern	Chandigarh, Haryana, Himachal Pradesh, Jammu & Kashmir, Punjab, Uttar Pradesh, Uttarakhand	25	Spiti, Jad, Darmiya, Gahri, Kanashi
Northeast	Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura	43	Aimol, Tangam, Sherdukpen, Singpho, Tarao
East Central	Uttar Pradesh, Bihar, Jharkhand, West Bengal, Odisha, Chhattisgarh, Madhya Pradesh	15	Bhunja, Birhor, Bondo, Toto, Gorum
West Central	Gujarat, Maharashtra, Rajasthan, Daman, Diu, Dadra Nagar Haveli, Goa	5	Nihali, Baradi, Bharwad, Diwehi, Bhala
Southern	Telangana, Karnataka, Andhra	20	Toda, Soliga,

	Pradesh, Tamil Nadu, Kerala		JenuKurumba, Siddi, Urali
Andaman & Nicobar	Andaman and Nicobar Islands	9	Sentinelese, Onge, Shompen, Lamongse, Luro

Language Families in India

India's various languages belong to five language families, classified based on grammatical characteristics. The family-wise grouping of India's 22 Scheduled and 99 Non-Scheduled Languages (99) is as follows:

Language families	Number of Languages	Persons who speak it as their mother tongue	Percentage to total population
1. Indo-European	24	79,08,76,283	76.89%
2. Dravidian	17	21,41,72,874	20.82%
3. Austro-Asiatic	14	1,14,42,029	1.11%
4. Tibeto-Burmese	66	1,03,05,026	1%
5. Semito-Hamitic	1	51,728	0.01%

Languages spoken by Scheduled Tribes: Tibeto-Burman in the north and northeast, Austro-Asiatic in central/eastern regions, Dravidian in the south, and Indo-European elsewhere dominated by 21 Indo-Aryan languages, plus Iranian (2) and English.

Multilingualism and Endangered Languages

Vasamalli grew up knowing various stories in Toda, even as she learned Tamil in school and picked up English. She always wanted to learn more about her culture. "I happened to get a copy of a few books in the Toda language and found the phonetic letters."

Vasamalli has worked with various linguists who travelled to her community and were interested in learning about and preserving the Toda language. She most recently worked with the SPPEL team for the development of the Toda primer in Tamil script.



Figure 4- Kurtaz Vasamalli is a social activist and a committee member of the District Tribal Welfare Committee of the Nilgiris district. She lives in 'Karsh' Todamund village near Ooty.



Vasamalli's multilingual journey reflects a broader pattern across India. In India, out of a total population of 121.09 crore, about 89.59 crore are monolingual, 22.90 crore are bilingual, and 8.60 crore are trilingual, according to the 2011 Census.

However, multilingualism is heavily skewed toward dominant languages. The highest number of monolinguals speak Hindi (46.74 crore), followed by Bengali (7.98 crore) and Marathi (4.39 crore)—all major scheduled languages.

For endangered tribal languages like Toda, this presents a challenge: speakers must often choose between preserving their native tongue and accessing education and opportunities in dominant languages.

The Ministry of Tribal Affairs addresses this challenge through bilingual dictionaries and trilingual education modules, supporting the National Education Policy 2020's three-language formula to help tribal students maintain their heritage while gaining broader linguistic skills. SPPEL has documented the Toda language, which lacks a native script, and is about to publish a Toda primer—an introductory book for beginners—in the Tamil script for dissemination among children.

The importance of such written documentation is something Vasamalli understands. "The elders are not teaching us and others like they used to and there is a lot more to the language. It is better to have written records." She sees that many Toda youth are interested in learning about their culture and history, giving her hope for the language's future. "We need to create situations that bring them back to their roots and back to the community, while also helping them engage with the outside world."

"Language is what makes us human. When peoples' freedom to use their language is not guaranteed, this limits their freedom of thought, of opinion and expression, as well as their access to rights and public services. This Decade must accelerate the mobilization of the international community to safeguard Indigenous languages in the long term."

- Audrey Azoulay, Director-General of UNESCO

Documenting Endangered Languages

“We believe there is divine power. If we protect the land, trees, and nature, we will receive blessings from God. That is the basic principle of the Toda community,” Vasamalli said. She said that a mountain is called *kotajen* and the Toda believe that their god by the same name lives there.

Toda's linguistic heritage is being promoted through SPPEL's extensive work, which involves the following documentation and promotion steps, which are as follows:

- *Recording phase*: Capturing words, sentences, songs and stories.
- *Transcription and analysis*: Converting recordings into the written form; examining the language's grammar, sound systems, and word-formation process; and building lexicons.
- *Grammar construction*: Writing the grammar by explaining the rules of the language including that of sentence formation.
- *Cultural documentation*: Recording livelihood practices, documenting rituals, capturing festivals and community traditions.
- *Digital archiving*: Digitally preserving the records of the language by creating metadata for all recorded materials and uploading them to repositories for worldwide access, and long-term preservation.
- *Revitalizing*: Producing community-driven primers to support early childhood education and encourage literacy in the language.

Use of Technology to Preserve and Promote Languages

The CIIL has published 8 digital dictionaries thus far for various endangered languages. For its documentation work, SPPEL uses various kinds of advanced technology tools, such as:

- High-end recording facilities for capturing audio
- Video-recording equipment for visual documentation
- Specialized software for creating metadata
- Various linguistic analysis software

The digital infrastructure includes digital repositories for storing and organizing data, and metadata creation systems that catalogue recorded materials. The website Sanchika: <https://sanchika.ciil.org/home>, launched July 17, 2025, features SPPEL's documentation work, including audio files for the elicitation of various words. Sanchika contains hundreds of language samples and audio files of various endangered languages.

SPPEL also creates short and long-form documentaries, such as "Panuha Not: The Pig Festival Chowra" about the pig festival celebrated among the Sanenyo tribe who live on Chowra Island in the Andaman and Nicobar Islands.

Other schemes preserving tribal cultures and languages

Various other schemes by other ministries also work toward preserving and promoting tribal culture and languages.

Ministry/Organization	Program/Initiative	Focus Area
Ministry of Culture	Folk, Tribal Arts Documentation	Documentation and promotion of folk and tribal arts through research projects and publications
	National Manuscripts Mission (NMM)	Digitization and conservation of manuscripts, rare books, and archival materials under the National Manuscripts Mission
	National Mission for Cultural Mapping (NMCM)	Implementation of the National Mission for Cultural Mapping involving cultural asset mapping of artists, art forms, and heritage practices in respect of 4.5 lakhs villages of India
	JanapadaSampada Division	The JanapadaSampada Division continues extensive fieldwork and audio-visual documentation of vanishing traditions, rituals, performing arts, and oral histories
	RashtriyaSanskritiMahotsavs	Ministry of Culture also organizes RashtriyaSanskritiMahotsavs at the national level where many folk & tribal artists from all over India are engaged to showcase their talents to spread awareness about rich culture amongst the masses/youth of the country
	IGNCA Programs	THE IGNCA, under the ministry, also organizes various workshops and regional cultural events to promote traditional practices and languages

SahityaAkademi	International Day of Indigenous Peoples	The Akademi celebrates the International Day of the World's Indigenous Peoples on 9 August each year by organizing an All-India Tribal Writers' Meet
	Dedicated Tribal Centres	To further preserve and promote oral and tribal literature, it has established dedicated centres in Agartala and Delhi
	UNMESHA Literature Festival	In the first edition of UNMESHA – the International Literature Festival (Shimla, 16-18 June 2022), around 30-35 tribal writers representing more than 25 tribal languages participated. The second edition (Bhopal, 3 to 6 August 2023), inaugurated by the President of India, DroupadiMurmu, featured over 575 writers from India and foreign, took part in 85 literary programs, and almost 103 languages were represented
Ministry of Tribal Affairs	Support of Tribal Research Institute	
		Research studies/publication of books/documentation including audio visual documentaries on promotion of rich tribal cultural heritage
		Research and documentation of Indigenous practices by tribal healers and medicinal plants, Adivasi Languages, agriculture system, dances and paintings etc
		Tribal cultural exchange programmes
		To acknowledge the heroic and patriotic deeds of tribal people, Ministry has sanctioned setting up 10 Tribal Freedom Fighters

		Museum. These museums will also exhibit rich tribal cultural heritage of the region
		Enhancement of learning achievement level amongst the Scheduled Tribe Students

SK/RK

References

- International Day of the World's Indigenous Peoples 9 August:
<https://www.un.org/en/observances/indigenous-day>
- Protection of Tribal Languages:
<https://www.pib.gov.in/PressReleasePage.aspx?PRID=2150806>
- Preservation and Promotion of India's Tangible and Intangible Cultural Heritage:
<https://www.pib.gov.in/PressReleasePage.aspx?PRID=2153597>
- PRESERVING TRIBAL CULTURAL HERITAGE:
<https://www.pib.gov.in/PressReleasePage.aspx?PRID=2147865>
- Uthagamandalam Municipality:
<https://www.tnurbantree.tn.gov.in/uthagamandalam/>
- National Education Policy 2020:
https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
- SPPEL: <https://www.sppel.org/default.aspx>
- UNESCO launches Report on India's Tribal and Indigenous Languages:
<https://www.unesco.org/en/articles/unesco-launches-report-indias-tribal-and-indigenous-languages>
- Indigenous Languages Decade (2022-2032):
<https://www.unesco.org/en/decades/indigenous-languages>
- "Indigenous Peoples and AI: Defending Rights, Shaping Futures":
https://social.desa.un.org/sites/default/files/IDWIP%202025%20Concept%20Note%20FINAL.pdf?_gl=1*12vn9nb*_ga*OTQ2OTg5ODMzLjE3NTE4NjQ2MjI.*_ga_TK9BOL5X7Z*czE3NTQ3NDI1NjQkbzkkZzAkdDE3NTQ3NDI1NjQkajYwJGwwJGgw
- PRESERVATION OF TRIBAL LANGUAGES:
<https://www.pib.gov.in/PressReleasePage.aspx?PRID=2042993>
- Ministry of Education, Higher Education: <https://www.education.gov.in/en/language-education>
- Languages Atlas of India 2011:
https://language.census.gov.in/eLanguageDivision_VirtualPath/Atlas/pdf/2011.pdf
- Linguistic Survey of India by Sir G.A. Grierson:
<https://language.census.gov.in/map/data/showLSIGrierson>
- Linguistic Survey of India Vol. 1:
<https://www.indianculture.gov.in/rarebooks/linguistic-survey-india-vol-i->

0#:~:text=survey%20of%20India.-

This%20book%20was%20compiled%20and%20edited%20by%20G.A.,published%20in%20the%20year%201927.&text=v.&text=dc.description-

This%20is%20the%20third%20volume%20of%20the%20linguistic%20Survey%20of,of%20the%20survey%20of%20India.

- THE LANGUAGES OF INDIA, Census 1901:
<https://censusindia.gov.in/nada/index.php/catalog/31294>
- The Nilgiris District: <https://nilgiris.nic.in/gallery/glenmorgan/>