
LANDMARK

Uniformity is Not Unity: A Probe into Linguistic Diversity of India

Anvita Abbi

Introduction

Indian linguistic diversity has originated due to constant contact, conflict and cohesion between migrants and locals. India represents 1,369 mother tongues from various language families (Census, 2011). The age of the language families ranges from 5000 to 700 years. One language family represented in the Andaman Islands is perhaps the oldest in the world. The members of this group are remnants of the first migration out of Africa that took place 70,000 years ago. India has served as the major corridor of human migration in the history of human evolution.

The current article unravels the genesis and the history of linguistic diversity. The diversity has created shared linguistic structures across all Indian languages and given rise to new languages, proving beyond doubt that language is not simply a means of communication but a social, regional, and cultural identity marker.

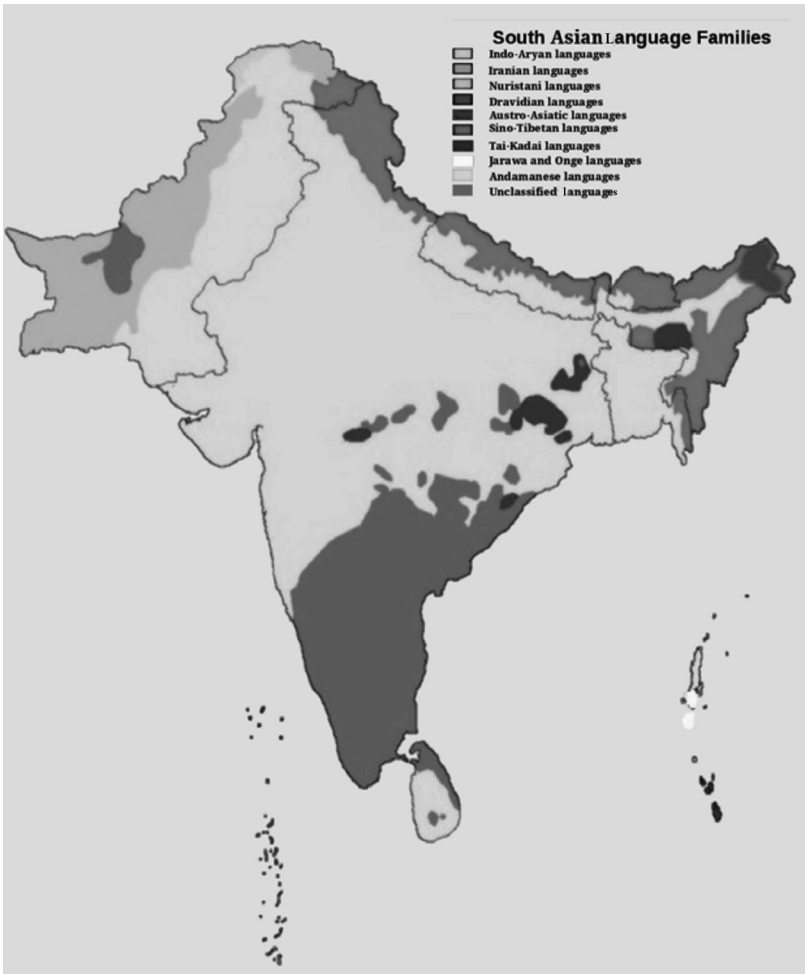
Language Families, Diversity, and Vitality of Languages

India represents six distinct language families spread over a large region and spoken by more than one billion speakers, as evident from the table. Research indicates the possible presence of the seventh language family (Abbi, 2006).

Languages subsumed under a 'family' are related to each other genealogically and have originated from a common source/s. They share several typological features that are very different from other languages of different language families. India represents the following language

families, sub-branches, and isolates spread over a large area from the Himalayas to the Andaman and Nicobar (Map).

Map: *Language Families of South Asia*



Modified version of the map found at https://en.wikipedia.org/wiki/Languages_of_India "South Asian Language Families"

The Indian Sign language (ISL) is used by more than 1.5 million signers but does not figure as a language family in any official record. However,

since ISL is very distinct from the other sign languages used in other countries, we list them here for appropriateness. Some languages are still considered isolates, meaning their affiliation to a known language family is not yet established.

The Demography

The multiplicity of languages can be visualized in the Table below.

Table: Multiplicity of Languages

Language Families	Speech Area	Number of Languages	Percentage/Number of Speakers
Indo-Aryan	North of India	574	78.05%,
Dravidian	South of India	153	19.64%,
Austroasiatic		65	1.11%,
Mon-Khmer	Northeast		
Nico-Monic	Nicobar Islands		
Munda	Central, East		
Tibeto-Burman ¹	Himalayan ranges	226	1%
Tai-Kadai	Assam and Arunachal Pradesh	6	
Great Andamanic/ Great Andamanese	Andaman Islands	1 (Jero Language)	3 Speakers
Angan	Onge-Jarawa (Andaman)		
Indian Sign language	all over India		<1.5 million signers
Isolates		3	
Shom Pen	Nicobar Islands		
Nahali/Nihali	Maharashtra		

According to the Census (2011), there are 1,369 rationalized mother tongues, but reported mother tongues [whether classified or not] are 19,569. More than 10,000 speakers speak a total of 270 identifiable mother tongues (excluding English). Out of these identified mother tongues, only

121 are *classified languages*; Non-Scheduled 99 and Scheduled 22. There are around 156 languages spoken by those whose strength is less than 10,000. The largest number of languages, 56 in all, are spoken by less than 10,000 speakers in Jammu and Kashmir alone². The Constitution of India recognizes only twenty-two languages termed “Scheduled languages”. These are listed under the Eighth Schedule (called ‘Languages’), Articles 343-351 of the Constitution, and the rest as ‘unscheduled’. The former are official languages of the Indian states.

Then there are ‘tribal languages’. The term tribal is used for administrative, judicial, and political purposes and has no pejorative sense. The tribal languages are spoken natively by the tribal communities, scheduled as per Article 342 of the Constitution by the President and Parliament. Neither a linguistic definition of a tribal language nor any specific characteristic linguistic feature defines a tribal language. Out of 500 tribal languages, only two are included in the Eighth Schedule — Santali and Bodo.

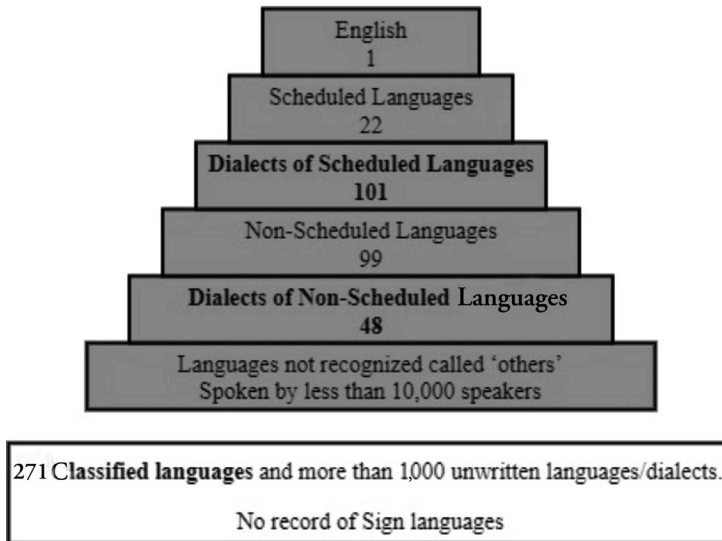
There are more unwritten languages in the country than written ones. The history of our spoken languages is older than the written ones. These languages constantly evolve with time and give rise to new languages and their varieties. They, by any standard, should not be considered inferior to the written ones, as they are the storehouse of the indigenous knowledge-system, ecological information, and the living oral documents of our history, civilization, human culture, human evolution, and human migration.

The enumeration of languages is not easy. It is messy because of how the term ‘mother tongue’ has been defined in various Census and how people generally change their language loyalty while reporting.

Language Vitality and Language Endangerment

Not all languages are considered equal in their prestige and use in society. A hierarchy of prestige can be diagrammatically expressed in a pyramid structure.

Diagram: *Hierarchical Status of Indian Languages and Dialects Based on Census 2011*



English, our official language [but outside the purview of the ‘Scheduled Language’], occupies the highest position. Single language domination in a plurilingual society pushed many languages to a minority status even though some had more speakers than those on the throne of dominance. About 197 languages are endangered according to UNESCO (Moseley, 2010). Linguistic imperialism and linguistic marginalization are two points of the same spectrum. The artificial demarcation between major and minor languages has adversely affected the vibrancy of these languages.

The **vitality index** of Indian language families differs from family to family. While Indo-Aryan, Dravidian, and to some extent, Tibeto-Burman language families are not threatened, other language families are not that fortunate. Excluding Santali, many Munda languages are threatened. The Great Andamanic is on the verge of extinction. Languages of the Tai-Kadai group, except for Tai-Khamti, also face speedy extinction due to language shift to Assamese. Angan languages and isolates such as Shom Pen in the Nicobar Islands and Nahali in Maharashtra also face endangerment. English has been a killer language as it made inroads into formal education distanced from the oral tradition—our centuries-old strong tradition—giving rise to linguistic apartheid.

The History of Indian Languages

The population structure in the subcontinent has been shaped by multiple historical migrations and culturally mediated patterns of gene flow as per the geneticists. Research on DNA samples has established the following migration routes and the dates of human dispersal to South Asia.

- Homo Sapiens from Africa came around 70,000 ybp (years before the present) and inhabited the Andaman Islands. These are known as Great Andamanese. There seem to be two migrations from different regions at two different times, the former of the Great Andamanese and the latter of the Onge and Jarawa. They are the First Indians or First Voices of India. The migrations imply that India has served as a major corridor for dispersing modern humans that started from Africa 1,00,000 years ago.
- The second migration comprised the Austroasiatic population around 20,000 ybp from the Southeast after the Last Glacial Maximum. There seem to be two migrations from different regions at two different times.
- Dravidian speakers came around 6,000 ybp from the southern part of Asia, below Asia Minor/Mideast (disputed).
- Tibeto-Burmans came around 6,000-5,000 ybp from the Northeast.
- Speakers of the Indo-European languages came in several waves from Western or Central Asia around 4,000 ybp.
- Speakers of Tai-Kadai languages came from the Southeast around 700 ybp.

Great Andamanese was recently identified as an independent language family (for details see Abbi, 2018), distinct from the other group of indigenous languages spoken in the Andaman Islands. The geneticists corroborated the research.

What Do Our Historical Records Tell Us?

The earliest written texts, the Ṛg-Veda, written around 1500-1200 BC, can be found in the Indo-Aryan language family. Vedic and Classical Sanskrit marks the Old Indo-Aryan (OIA) period with *Aśṭādhyāyī* written by Panini around 500 BC. Various spoken languages (e.g. Prakrits) are found in Ashokan inscriptions in Middle Indo-Aryan (MIA). *Apbhramsa*

originated around 500 BC to 1000 AD giving rise to New Indo-Aryan (NIA) or Modern Indo-Aryan languages. Thus, it is a misnomer to consider that all Indo-Aryan languages are derived from Sanskrit. The so-called Aryan community was diglossic, with women speaking Prakrit and men speaking Sanskrit. So, the mother tongues of growing up children were Prakrits of various regions that gave birth to MIA languages. Kalidasa's *Abhigyaan Shankuntlam* is the best example where men speak in Sanskrit and women in Prakrit. There are essentially no compositions by women in the R̥gveda as women were prohibited from learning Sanskrit. Sanskrit cast a profound influence on MIA languages and later Dravidian languages.

The earliest written records of Dravidian languages, 'Sangam' literature written in Brahmi script, date back to the 3rd century BC. Tamil is, thus, older than Sanskrit. Traces of the Dravidian language have been found in Vedic literature to prove the two communities were in contact.

Although no written records are available in Austroasiatic languages, their existence predates the Aryan invasion and the origin of Indo-Aryan languages in India. The artefact of a dancing figurine in the Indus Valley civilization shows traces of the Austroasiatic presence.

Languages of OIA have words of Austroasiatic origin, proving that Aryans intermingled with this group.

The Geographical Spread of Languages

As seen in the Map, **Indo-Aryan** languages occupy the largest geographical area in the country. They are spread in the vast area covering India's north, west, east, and central parts. There are north-western languages such as Sindhi, Lahnda, and Punjabi; central languages such as Hindi, Urdu, Pahari, Rajasthani, and Bhojpuri; eastern languages such as Bangla, Assamese, Odia, Maithili, and Bihari as well as south-western languages such as Gujarati, Marathi, and Konkani. Although state boundaries bind most languages, no state is monolingual and houses several languages not necessarily from the same language family. A chain of mutual intelligibility runs through Indo-Aryan languages from west to east and north to central India, with two points at the end of the chain unintelligible. This chain is not applicable for either Dravidian or Tibeto-Burman languages.

Indo-Aryan languages are sister languages of Indo-Iranian languages

and thus form a subfamily of the large Indo-European language family. As said earlier, MIA languages are derived from Prakrit, the spoken form of the OIA and are influenced by Sanskrit, the language of the educated elite in 500 BC.

Hindi, in its Devanagari script, is the official language. If the two mutually intelligible Hindi and its sister language Urdu (inclusive of all its varieties) are combined, they represent more than 50 per cent of the Indian population, with 57,91,19,824 speakers using it as their first language. Hindi-Urdu also has the largest number of speakers using it as a second language. Our languages are written in multiple scripts, such as Konkani is written in five different scripts depending on the state where it is used.

Dravidian languages occupy the southern part of India; however, Kurux and Malto, the two languages of the same family, are spoken in central India and are known as north Dravidian languages. The term is reserved to include Brahui, the only Dravidian language spoken in northwest Pakistan. These languages are proof of the migration route of Dravidians. Kurux and Malto are spoken in the vicinity of the Munda languages incorporating many of their features. The major languages of the Dravidian family are Tamil, Telugu, Kannada, and Malayalam. These are literary languages and have been given the status of Classical languages by the Government of India. However, all four languages have several regional and society-bound dialects as well.

It is to be noted that all the languages of the Tibeto-Burman, barring Meitei in Manipur and Austroasiatic language family, are considered 'tribal.'

Austroasiatic languages are known to be the oldest, predating any other language family except those represented in the Andaman Islands. They consist of two major groups—Munda represented in the central and east of India and Mon-Khmer languages spoken in the northeast in Meghalaya hills and its sub-branch Mon-Nico in the Nicobar Islands in the Indian Ocean. Santali is the major Munda language listed in the Eighth Schedule, while Khasi occupies the prominent position among the Mon-Khmer languages. Nicobarese languages with five varieties³ belong to the Mon group.

Tibeto-Burman languages belong to the Sino-Tibetan language family. They are spoken in the Himalayan ranges stretching from northwest

India to northeast India and extending to Nepal, Burma, and China. This family has the most numerous and highly differentiated languages in its fold. The languages of this family mark India for its linguistic diversity. Most of the languages of this family are mutually unintelligible. In post-independent India, varieties of contact Hindi have emerged in Meghalaya, Arunachal Pradesh, and Manipur, bringing various tribes closer. Interestingly, each language of this group is written down, unlike the case with other tribal languages of the country.

Tai-Kadai, a branch of the major Tai group of languages, is spoken in Arunachal Pradesh and Assam. Out of the seven languages in its fold, Ahom and Nora have become extinct; The number of speaker in Aiton (1,500), Khamiyang (50), Phakial (5,000), and Phake (2,000) make them endangered, leaving only Tai Khamti spoken in Arunachal Pradesh as a very vibrant language. However, most of the speakers of this family are shifting to Assamese, the Indo-Aryan language. Khamti, found in the Lohit district of Arunachal Pradesh, is the largest group (12,890) among all the Tai languages in India and is used in education.

Indigenous Languages Spoken in the Andaman Islands

Languages spoken in the Andaman Islands belong to the hunter and gatherers community and thus, are a rich repository of ecological knowledge. The living Andamanese tribes can be classified into four major groups, the Great Andamanese, the Jarawa, the Onge, and the Sentinelese. Of the 14 known Andamanese languages, only four Onge, Jarawa, Sentinelese, and the Koiné form of present-day Great Andamanese are breathing their last. Most languages are in danger of extinction due to a low population base and/or language replacement by Hindi.

Great Andamanic language family had ten languages in its fold by the end of the 19th century: Aka-Sare, Aka-Khora, Aka-Jero, Aka-Bo, Aka-Puchikwar, Aka-Kede, Aka-Kol, Aka-Juwai, Âkà-Bêa and Aka-Bale; all of these except Aka-Jero are now extinct. This language has only three speakers.

The Angan Language Family

The Angan family of languages consists of Jarawa (400), Onge (100), Sentinelese (exact number not known), and Jangil (extinct), all

represented in the Andaman Islands. Except for Jangil, which became extinct by 1925, all the other languages are spoken by small populations, transmitted to the younger generation, and used in all relevant domains of the hunter and gatherer society. However, recent contact with outsiders may endanger Onge and Jarawa.

Language Contact, Bilingualism, and Areal Features

Migrant speakers of various language families intermingled with the local population and created new languages. The OIA speakers did not come with their women, and subsequent intermixing with the local population created NIA languages. Each family tree grew on this soil with its branches extending wide and ultimately touching the branches of other trees. Pollination of this contact gave rise to many similar linguistic structures known as Areal features in the discipline of linguistics. For instance,

Duplicating a word twice is common in all Indian languages.

- Hindi: *jate jate* 'while going'
- Punjabi: *munde munde* 'boys boys'

Using two verbs at the end of a sentence where the second verb loses its original meaning.

- Malayalam: 'the bottle broke'
- Hindi: *botal tuut garii* 'the bottle broke'

Expressive morphology where duplication of syllables emotes several kinds of sensations and feelings.

- Hindi: *dhak dhak* 'throbbing of heart', *jhan jhan* 'tinkling of bells'
- Nepali: *chul chul* 'unsteady'
- Tai Khamti: *syen4 -sok2-sok2* 'very beautiful'

Echo formations as

- Hindi: *chay vay* 'tea etc'
- Tamil: *puli gili* 'tiger etc'
- Telugu: *puvu givu* 'flower etc'

These grammatical structures are shared across all the languages of mainland India except those in the Andaman and Nicobar Islands.

From the areal point of view, the shared semantic field/s represented by similar structures indicate shared underlying cognitive ability, a single coherent conceptual-semantic space and common interpretation across the speakers of diverse languages of the country.

Enduring contact between the speakers of different language groups results in robust bilingualism and the generation of new languages such as Halbi in Chhattisgarh, Sadari in Jharkhand, Chakesang [derived from the conglomeration of Kheza, Chokri and Sangtam], Tenyidie, or the language of Zeliangrong [a group of three to four tribes] in the northeast, and varieties of contact Hindi (Abbi 2021) across cosmopolitan cities, northeast, Andaman, and Nicobar.

Conclusions

No language is 'pure' or 'impure'. Admixture of various races and communities gave rise to 'mixed languages' responsible for the genesis of modern Indian languages. Linguistic diversity is part and parcel of the diversity of life in nature and culture. Any loss in linguistic diversity is a loss in the vitality and resilience of the whole web of life.

The high linguistic diversity is because of several human migrations from various parts of the world and adapting and adopting the local population and its languages. India thus has a legacy of bi/multilingualism. Indian linguistic diversity has been sustained because of its functional bilingualism, which allows each language to maintain its identity and thrive at its own pace while interacting in a complex social network. The main feature of Indian bilingualism is that our languages are not competitive but complementary. Children are already equipped with this latent knowledge when they enter school.

This multiplicity of languages has taught us social cohesion and tolerance for others. Diversity is the *mantra* for human evolution and sustenance. Uniformity is not unity.

Notes

1. Unfortunately, the Indian Census is silent on the other two language families. These are enumerated under the category of 'less than 10,000 speakers'.
2. Inclusive of the territories of Ladakh.
3. Multilingual dictionaries of two languages of the Austroasiatic family are available online. Sanenyo (Abbi & Pachori, 2021) <https://www.sppel.org/>

sanenyodictionary.aspx and Luro (Abbi & Vysakh 2021) <https://sppel.org/lurodictionary.aspx>.

References

- Abbi, A. (2006). *Endangered languages of the Andaman Islands*. GMBH.
- Abbi, A. (2018). 'A Sixth language family of India: Great Andamanese, its historical status and salient present-day features. In R. Mesthrie & D. Bradley (Eds.), *Dynamics of language. Plenary and focus lectures from the 20th International Congress of Linguists* (pp. 134-152). International Congress of Linguists. UCT Press.
- Abbi, A. (2021). Social cohesion and emerging standards of Hindi in a multi-lingual context. In W. Ayres-Bennett & J. Bellamy (Eds.), *The Cambridge Handbook of Language Standardization* (pp. 115-138). Cambridge University Press.
- Census of India. (2011). *Data on language and mother tongue*. Government of India. https://censusindia.gov.in/2011Census/Language_MTs.html
- Moseley, Christopher (Ed.). (2010). *UNESCO atlas of the world's languages in danger* (3rd edn). <http://www.unesco.org/culture/en/ endangeredlanguages/atlas>

Anvita Abbi, a Padma Shree Awardee, retired as Professor of Linguistics from Jawaharlal Nehru University, Delhi and currently holds the B.B. Borkar Chair of the Comparative Literature, University of Goa. She identified and described the Great Andamanese, the sixth language family of India. She is an advisor to UNESCO on language issues.
anvitaabbi@gmail.com