

# **Foundations of Linguistics**

HS4605-S-2024

## **Team - 3**

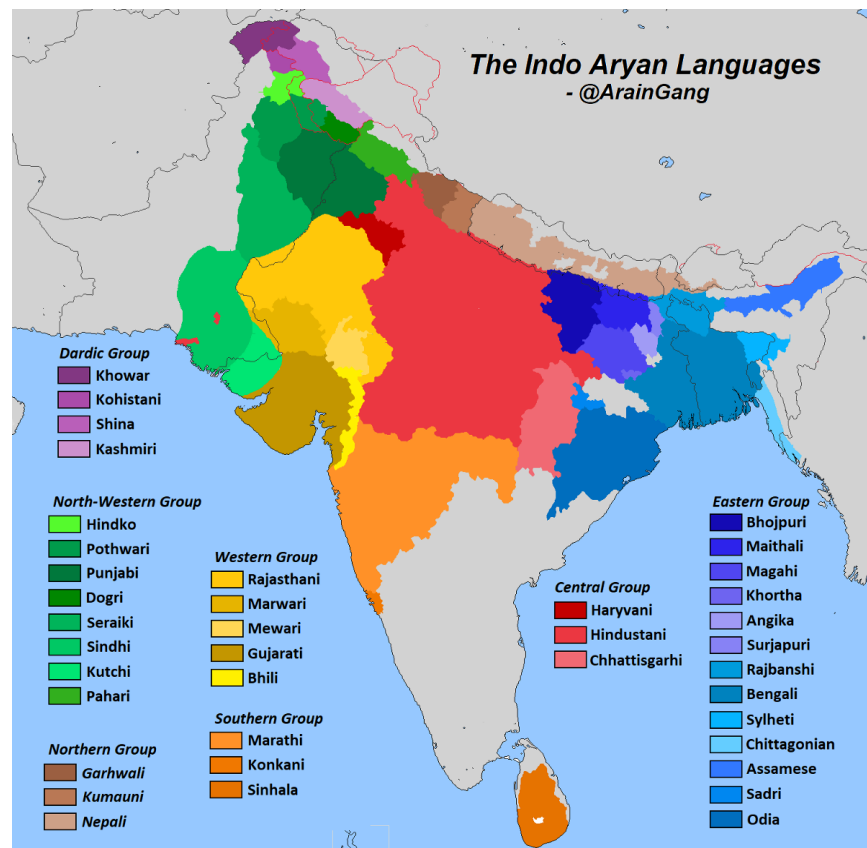
1. Nikhil S Thomas
2. BERELLI GOUTHAMI
3. DODDAPANENI UDITH
4. MULUGURU PRANITHA
5. MALOTH SONY
6. VISHNU SHREERAM M P

# Language Groups of India

India's linguistic diversity is a reflection of its rich cultural heritage, with languages serving as key markers of identity and history. The country is primarily home to four major language groups: Indo-Aryan, Dravidian, Austroasiatic, and Tibeto-Burman, alongside several regional and minor languages. These language groups not only highlight India's diversity but also reveal the complex social, cultural, and historical interactions that have shaped the nation. This report explores the major language families of India, their cultural significance, and how they interact within the social context of the country.

## Indo-Aryan Languages

The Indo-Aryan languages are among the largest and oldest language groups in South Asia, forming a branch of the Indo-Iranian language family within the broader Indo-European lineage. Known for their rich phonetic range, cultural depth, and layered vocabulary, these languages have evolved from Vedic Sanskrit to a vast family of modern languages spoken by over 900 million people.

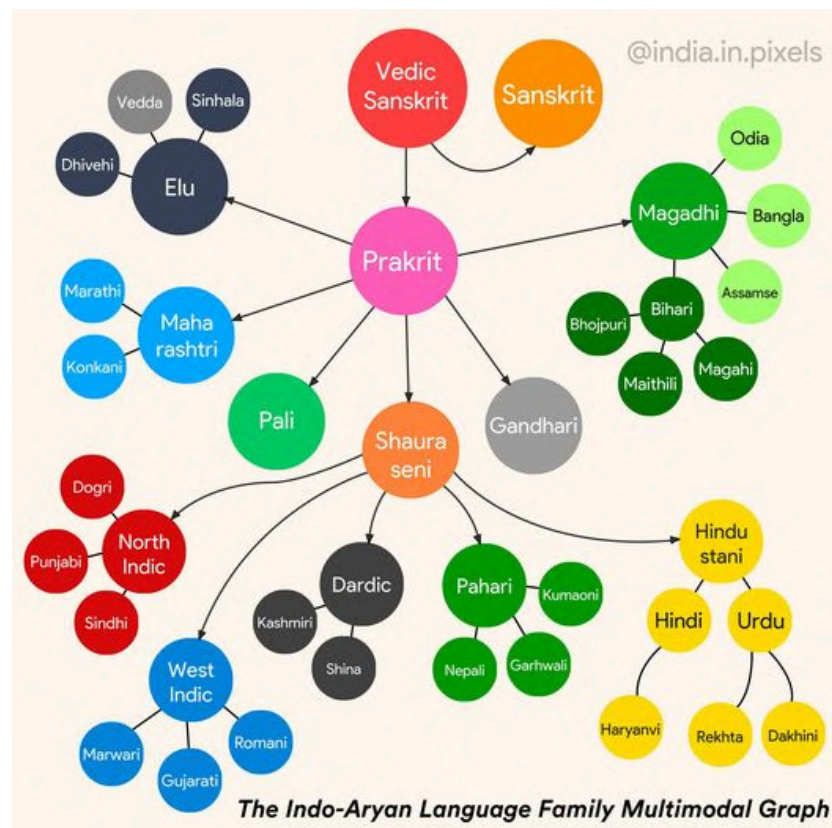


## Vedic and Classical Sanskrit Foundations

Indo-Aryan began with Vedic Sanskrit (1500 BCE), the language of the Rigveda and ancient Indian spiritual texts. Its transition into Classical Sanskrit introduced formal grammar codified by Panini in works like the *Ashtadhyayi*, establishing a language system of such precision that it resembles programming syntax. This linguistic structure facilitated the development of philosophical, scientific, and literary works, seen in epics like the *Ramayana* and *Mahabharata*, and created vocabulary layers with words that convey meanings difficult to capture with single English equivalents. This layered quality lends depth to Indo-Aryan languages, making their terms often evocative of multiple concepts.

## Prakrits and Cultural Spread

Over time, Sanskrit branched into simpler, widely spoken Prakrit languages that permeated India's cultural and religious life. Key Prakrits include Pali (Buddhist canon), Maharashtri (poetry and drama), Shauraseni (classical drama), Magadhi (Jain and Buddhist texts), Gandhari (Silk Road communities), and Elu (early Sri Lankan language). These accessible languages connected with local traditions and influenced major religious scriptures, helping Indo-Aryan reach diverse communities.



## Apabhraṃśa and the Rise of Modern Languages

Between 1000–1500 CE, Prakrit languages transitioned into Apabhraṃśa forms, bridging Prakrits to modern Indo-Aryan languages. The Bhakti movement played a role in this evolution, with poets like Kabir and Mirabai making the language accessible and spiritual. By 1500–1800 CE, Apabhraṃśa evolved into distinct languages, including Hindi, Bengali, Marathi, Punjabi, and Gujarati, each forming its own literature and identity while retaining roots in classical heritage.

## Phonetic Range, Script Diversity, and Cross-Cultural Influence

Indo-Aryan languages are noted for their extensive phonetic range, a legacy of Sanskrit’s organized sound structure. These languages are written in various scripts that enhance regional identity: Devanagari (Hindi, Marathi, Nepali), Bengali (Bengali, Assamese), Gurmukhi (Punjabi in India, Shahmukhi in Pakistan), Sinhalese (Sri Lanka), and Odia. Indo-Aryan languages, especially Sanskrit and Pali, also influenced Southeast Asian languages like Thai, Khmer, Javanese, and Balinese through cultural exchange, underscoring Indo-Aryan’s far-reaching impact. Indo-Aryan languages feature nuanced terms for concepts, shaped by historical influences from Sanskrit, Persian, Arabic, and English, as well as regional cultural diversity, enabling subtle distinctions in meaning.

Devanagari	अ आ इ ई उ ऊ ऋ ॠ ए ऐ ओ औ क ख ग घ ङ च छ ज झ
Bengali	অ আ ই ঐ উ ঊ ঋ ৠ এ ঐ ও ঔ ক খ গ ঘ ঙ চ ছ জ ঝ ঞ ট ঠ ড
Gurmukhi	ਅ ਆ ਇ ਈ ਉ ਊ ਋ ਐ ਓ ਔ ਕ ਖ ਗ ਘ ਙ ਚ ਛ ਜ ਝ ਞ ਟ ਠ ਡ ਤ ਥ
Gujarati	અ આ ઇ ઈ ઉ ઊ ઋ એ એ ઔ ઓ ઔ ક ખ ગ ઘ ઙ છ જ ઝ ઞ ટ ઠ
Oriya	ଅ ଆ ଇ ଈ ଊ ଋ ଓ ଐ ଓ କ ଖ ଗ ଘ ଙ ଛ ଜ ଝ ଞ ଟ ଠ ଡ ଣ

## Conclusion

From ancient Vedic hymns to modern languages, Indo-Aryan languages embody continuity, cultural depth, and adaptive evolution. With their layered vocabulary, phonetic richness, and widespread usage, they are a cornerstone of South Asian heritage and linguistic diversity.

# **The Dravidian Languages**

Dravidian languages are a family of languages spoken in South Asia, including India, Pakistan, and Sri Lanka. Family of around 26 languages spoken by over 300 million people. The Dravidian languages form a rich and historically significant language family, primarily spoken in South Asia. This includes some of the oldest and most well-preserved languages in the world, especially Classical Tamil, known for its extensive literary and grammatical traditions.

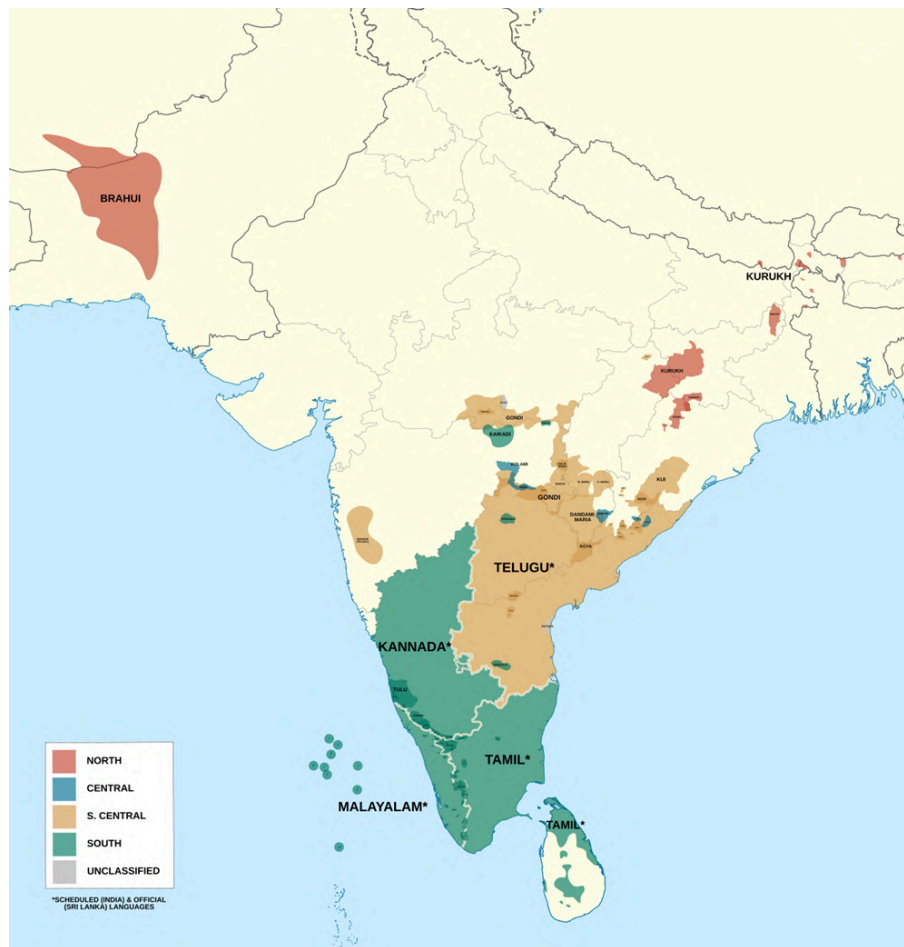
## **Origins and the Term "Dravidian"**

The term "Dravidian" was first introduced by the British missionary and linguist Robert A. Caldwell. He derived it from the Sanskrit word Dravida. the term refers to a family of languages spoken across South India, parts of Central India, and even Baluchistan on the Pakistan- Afghanistan border, where Brahui is spoken.

## **Classification of Dravidian Languages**

Dravidian languages are generally classified into three main groups: Northern, Central, and Southern.

- Northern Group: This is the smallest group, containing Brahui, Malto, and Kudukh.
- Central Group: This includes a more diverse set of languages that are more widely distributed, such as Gondi, Konda, Kui, Manda, Parji, Gadaba, Kolami, Pengo, Naiki, Kuvi, and Telugu. Telugu, among these, is particularly notable for its large number of speakers and significant literary contributions.
- Southern Group: This group includes Tulu, Kannada, Kodagu, Toda, Kota, Malayalam, and Tamil. Each of these languages possesses its own distinct script and literary history, with Tamil, Malayalam, and Kannada having a well-documented presence dating back to the pre-Christian era.



## Major Dravidian Languages

- Tamil: One of the oldest classical languages with rich literature and grammar traditions.
- Telugu: Known for its lyrical quality, often called the "Italian of the East."
- Kannada: Has a deep literary heritage with ancient inscriptions and classical works.
- Malayalam: Known for its complex poetry and grammar, with a strong independent literary tradition.

## Minor Dravidian Languages

While the four major languages provide a well-known face to Dravidian culture, the lesser-known Dravidian languages, such as Tulu, Kodagu, Toda, Kota, and more, bring regional stories, songs, and folklore to life. These languages, though smaller in the number of speakers, carry invaluable oral traditions and stories of the indigenous communities.

## Distinctive Features of Dravidian Languages

- **Retroflex Sounds:** Dravidian languages have sounds where the tongue curls back. In Tamil, for instance, "ṭa" (ṭ) is a retroflex sound different from "ta" (t). You create this sound by curling your tongue back slightly. Retroflex ḍ: The word "koḍi" (S&) means "hen" or "chicken," where "ḍ" (ḍ) has a retroflex sound. Regular d: Meanwhile, "padi" (ā&) means "ten," where "d" (&) is a regular, non-retroflex sound.
- **Agglutinative Structure:** Words are built by adding prefixes and suffixes. In Kannada, the word "ooduvudakke" (ಓದುವುದಕ್ಕೆ) means "for reading":
  - "oodu" (ಓದು) = read (root word)
  - "vudu" (ವುದು) = noun suffix
  - "kke" (ಕ್ಕೆ) = for
- **Gender and Number:** Dravidian languages have three genders (masculine, feminine, neuter) and distinct plural forms. In Telugu, "pillalu" (పిల్లలు) means "children" with "lu" added to show plural, while "pilla" (పిల్లా) is singular.
- **Subject-Object-Verb (SOV) Order:** The sentence structure usually follows SOV. In Malayalam, "avan saapikkunnu" (അവൻ കഴിക്കുന്നു) means "he eats" where:
  - "avan" (he) is the subject,
  - "saapikkunnu" (eats) is the verb.

## Scripts

- **Tamil Script:** One of the oldest and most distinctive, known for its straight lines and simplicity. It is linear, with distinct straight lines and curves.
- **Telugu Script:** Characterized by rounded, flowing shapes, Telugu script is round and more curved compared to others. It also has 16 vowels and 41 consonants, with characters often resembling circles and loops.
- **Kannada Script:** Kannada script is similar to Telugu but has sharper edges and a more angular design.
- **Malayalam Script:** Malayalam script is known for its flowing, curvy lines and circular shapes, making it visually distinct from the other scripts.

Tamil	ஷிவோ ரகஷது கீர்வாணபாஷாரஸாஸ்வாததத்பராந்
Telugu	శివో రక్షతు గీర్వాణభాషారసాస్వాదతత్పరాన్
Kannada	ಶಿವೋ ರಕ್ಷತು ಗೀರ್ವಾಣಭಾಷಾರಸಾಸ್ವಾದತತ್ಪರಾನ್
Malayalam	ശിവോ രക്ഷതു ഗീർവാണഭാഷാരസാസ്വാദതതൃരാൻ

## Cultural and Literary Contributions of Dravidian Languages

Dravidian languages have contributed significantly to India's literary, religious, and philosophical traditions. From Tamil Sangam literature to Kannada Vachanas, with classical works like Ramacharita in Malayalam and Telugu Bhakti poetry, the richness of their vocabulary and the distinctiveness of their phonetic systems have allowed them to adapt and evolve through centuries, fostering a deep cultural connection for millions.

## Influences and Borrowings

- Dravidian languages have borrowed many words from Sanskrit (especially in religious and philosophical contexts) and have influenced Indo-Aryan languages.
- They have also influenced languages outside India, like Sinhalese (in Sri Lanka) and Burushaski (in parts of Pakistan).

## Conclusion

Dravidian languages embody deep cultural and linguistic history, with diverse phonetics and scripts, shaping South Asian heritage. They continue to evolve, maintaining their significance in modern times.



## **Austic Group of Languages**

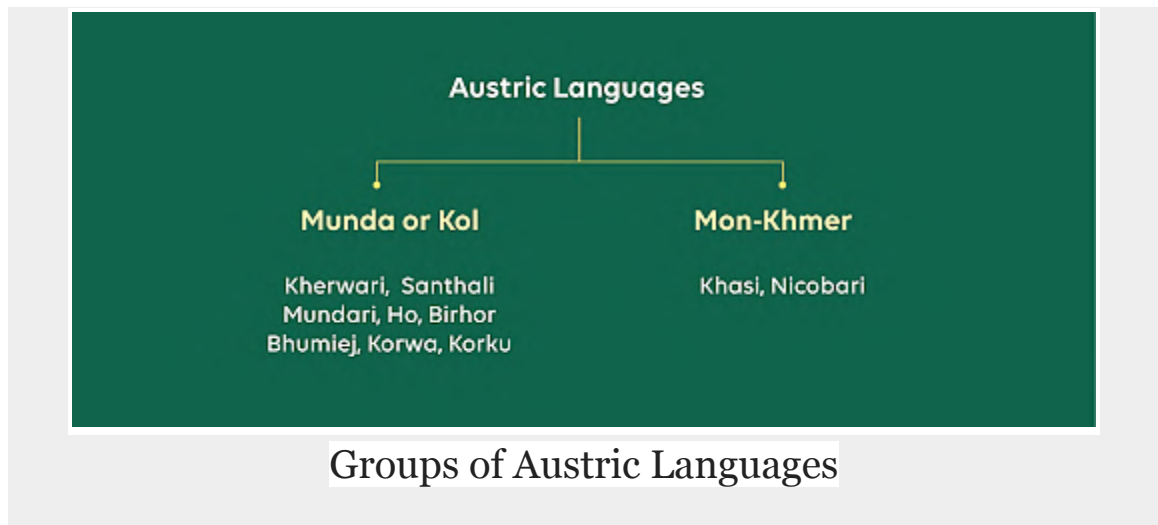
The Austic group of languages of India belong to the Austro-Asiatic sub-family, represented by languages of Munda or Kol group spoken in central, eastern, and northeastern India. It also includes the languages of Mon-Khmer groups like Khasi and Nicobarese. These languages existed long before the arrival of the Aryans and were referred to as Nisadas in ancient Sanskrit literature.

### **Features of Austic Languages**

- The most important language in this group is Santhali, which is spoken by Santhal tribals in Jharkhand, Bihar, and Bengal.
- All Austro-asiatic languages on Indian territory are endangered, with the exceptions of Khasi and Santhali.
- There are no tones in Austro-Asian languages, but there is a wide range of vowels.
- Under the influence of Chinese and Indo-Aryan languages, the sound systems of the Mon-Khmer and Munda branches have diverged significantly.
- Austronesian languages are also included in this group.
- These are spoken in Pacific Islands, Southeast Asia, Madagascar.
- This group also includes Austroasiatic languages.
- These are spoken in mainland South Asia and Southeast Asia.
- This group of languages was proposed in 1906 by Wilhelm Schmidt.

## Groups of Austric Languages

This group of languages is subdivided into following sub-groups:



### Munda or Kol Language

- They form the largest of the Austric group of languages.
- A group of fourteen tribal languages are included in this group.
- Its major group is known as Kherwari .
- Kherwari is spoken in Chota Nagpur, Orissa, Chhattisgarh and West Bengal.
- Kherwari also includes Santhali, Mundari, Ho, Birhor, Korwa and Korku (or Kurku).

### Mon-Khmer Language

- It is divided into two sub-groups: Khasi and Nicobari.
- Khasi speakers include Khasi tribal people of Meghalaya.
- Nicobari speakers include tribal people of the Nicobar Islands.
- Khasi was earlier written in Bengali- Assamese script about a century ago.
- Due to the influence of Welsh Methodist missionaries, Roman alphabets were modified for Khasi to produce literary works.

## Conclusion:

Language constitutes a means of exploration, exchange and expression of ideas and knowledge. Austric languages are a group of languages that are mainly spoken in South and Southeast Asia. Due to contact with various cultures it consists of typological changes and basic-vocabulary borrowings.



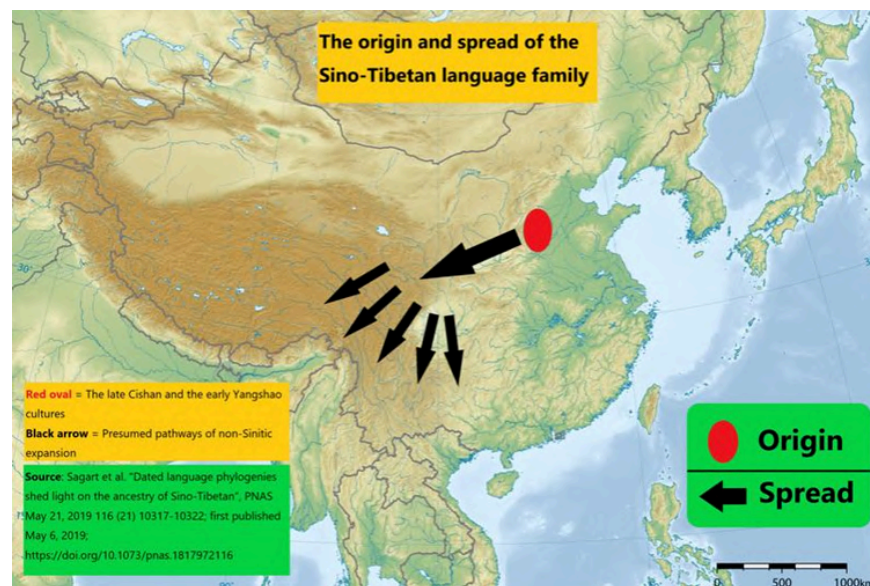
Countries where Austric Languages are spoken

## Tibeto-Burman languages

Tibeto-Burman languages, part of the Sino-Tibetan family, include over 400 languages spoken across Southeast Asia, East Asia, and South Asia, with around 60 million speakers. The name comes from Burmese and Tibetic languages, which have rich literary traditions dating back to the 12th and 7th centuries. Most other languages in this group are spoken by smaller communities and are less documented.

While Sino-Tibetan is often split into Sinitic and Tibeto-Burman groups, some linguists dispute this, arguing that non-Sinitic languages lack shared features in phonology or morphology to form a single branch.

### HISTORY



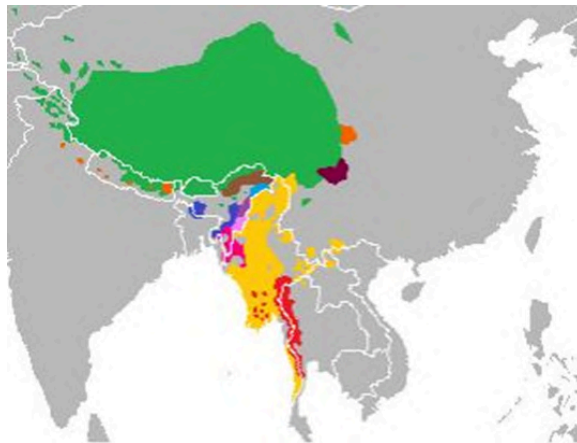
In the 18th century, scholars noticed similarities between Tibetan and Burmese, both with rich literary traditions. In the 19th century, Brian Houghton Hodgson gathered data on non-literary languages in the Himalayas and northeast India, linking them to Tibetan and Burmese. The term "Tibeto-Burman" was introduced by James Logan in 1856, who later added the Karen languages. Charles Forbes saw it as part of a broader Turanian family, excluding Semitic, Indo-European, and Chinese languages.

Julius Klaproth observed in 1823 that Burmese, Tibetan, and Chinese shared basic vocabulary, unlike Thai, Mon, and Vietnamese. By the late 19th century, linguists like

Ernst Kuhn and August Conrady classified these languages into "Indo-Chinese" with two branches: Tibeto-Burman and Chinese-Siamese. The term "Sino-Tibetan" was later coined by Jean Przyluski in 1924.

After World War II, Western scholars generally excluded Tai languages from Sino-Tibetan, though some Chinese linguists still include them. Most linguists today agree on a link between Tibeto-Burman and Chinese, but the classification into separate Chinese and Tibeto-Burman branches remains debated, despite support from scholars like Paul Benedict and James Matisoff.

## OVERVIEW



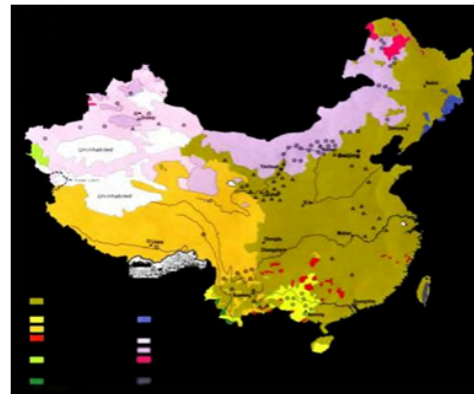
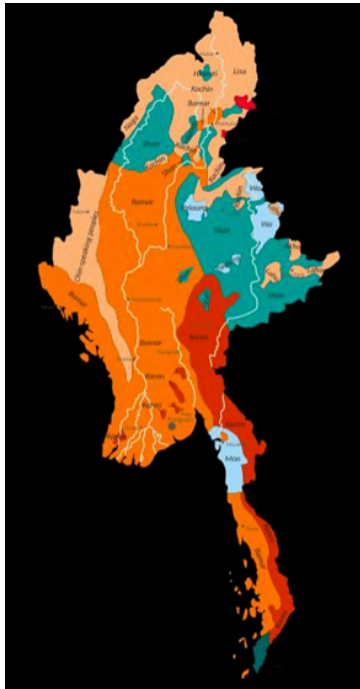
Tibeto-Burman languages are spoken in isolated mountainous regions, making them difficult to study, and many have no written form. While identifying a language as Tibeto-Burman is often straightforward, pinpointing its exact relation to others in the group is challenging. There are several dozen established subgroups, ranging from well-researched groups with millions of speakers to rare isolates, some only recently discovered and at risk of extinction. These subgroups are typically categorized by their geographic location.

## Southeast Asia and southwest China

The Karen languages, spoken by 3 million people along the Burma-Thailand border, are unique among Tibeto-Burman languages (except Bai) for using subject–verb–object word order due to influence from Tai-Kadai and Austroasiatic languages.

Burmese, the most spoken Tibeto-Burman language with over 32 million speakers, is the national language of Myanmar. It belongs to the Lolo-Burmese group, which includes about 100 languages spoken across Myanmar, Thailand, Laos, Vietnam, and

southwest China, like Loloish, Akha, and Hani, showing Austroasiatic influences.



In southwest China, Tibeto-Burman languages have been influenced by Chinese, complicating their classification. The Bai language's relation to Chinese is debated, while Naxi is often grouped under Lolo-Burmese. Sichuan's Qiangic and Rgyalrongic languages preserve archaic traits, with Tujia being the easternmost Tibeto-Burman language.

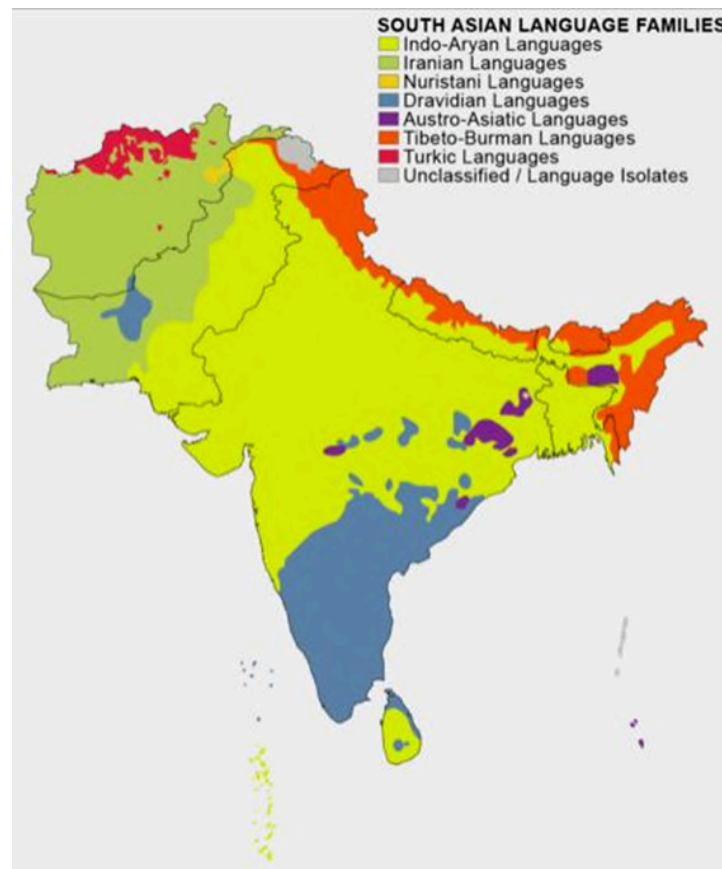
Historically, the Pyu and Tangut languages are linked to Tibeto-Burman, but their exact classification remains unclear.

## Tibet and South Asia

Over 8 million people in the Tibetan Plateau and nearby regions (e.g., Nepal, Bhutan) speak Tibetic languages, with a rich literary tradition since the 8th century. These, along with East Bodish languages, form the Bodish group.

The southern Himalayas host diverse Tibeto-Burman languages. Key groups include West Himalayish, Tamangic (e.g., Tamang with 1 million speakers), and Kiranti. Newar in central Nepal also has a literary tradition, and Magaric languages have nearly a million speakers. Smaller groups include Lepcha and Chepangic.

In Bhutan, most languages are Bodish, but there are isolates like 'Ole and Gongduk. Arunachal Pradesh primarily speaks Tani languages, with smaller diverse groups like Kho-Bwa.



Northern Myanmar to northeast India features Nungish, Jingpho–Luish (e.g., Jingpho with nearly a million speakers), and Brahmaputran languages like Boro–Garo. In Manipur and Nagaland, languages include Meitei (1.4 million speakers) and Kuki-Chin languages. Mru is spoken in the Chittagong Hill Tracts between Bangladesh and Myanmar.

## Conclusion

The Tibeto-Burman languages are a diverse group spoken across the Himalayan region, Southeast Asia, and parts of East Asia. These languages, including well-known ones like Burmese and Tibetan, have rich literary traditions, but many smaller languages remain under-researched and endangered. While some groups have been heavily influenced by neighboring languages, they all share historical and cultural connections within the Sino-Tibetan family.

## Language Isolates: Nihali and Burushaski

Language isolates are unique languages that have no known affiliations with other languages, making them particularly interesting from a linguistic perspective. Two notable examples of language isolates in South Asia are Nihali and Burushaski. This report provides an overview of these languages, their unique features, and their uncertain origins.

### **Nihali**



Nihali is spoken by a small community along the border of Madhya Pradesh and Maharashtra. As of 1991, it was estimated that only around 2000 people spoke the language, placing Nihali in an endangered status. The language has borrowed extensively from neighboring languages, with approximately 25% of its vocabulary derived from Korku, an Austroasiatic language. Additionally, Nihali contains vocabulary from Dravidian languages and Marathi.

Despite these borrowings, many core words in Nihali remain native and cannot be traced to other languages. Words representing basic concepts, such as numerals and terms like “blood” and “egg,” are intrinsic to the language. Interestingly, no monolingual speakers of Nihali exist today, indicating that all speakers are bilingual, at the very least.

Linguist Franciscus Kuiper was the first to propose that Nihali might be a language isolate. He even suggested that it could have originated as an argot—a secret language used by specific groups. Like many other languages in the region, Nihali follows a Subject-Object-Verb (SOV) word order.

### **Sample Vocabulary in Nihali**

Some common words in Nihali include:

- **Kalen:** Egg
- **Joppo:** Water
- **Cakni:** Hear



- **Nan:** What
- **Nani:** Who
- **Poe:** Bird

## Burushaski



Burushaski is spoken in the northwestern region of Ladakh, near the borders of Pakistan, China, and Afghanistan. The language has intrigued linguists for decades, and various theories have attempted to link it to other language groups. However, no proposal has been widely accepted. Some scholars have placed Burushaski within the hypothetical Dené–Caucasian macrofamily, along-side languages such as Basque (spoken in Spain), North Caucasian, Yeniseian (a central Siberian language), Sino-Tibetan, and Na-Dene (a North American language). Another theory links Burushaski with Yeniseian through a proposed family known as Karasuk. Additionally, some researchers have speculated on connections between Burushaski and ancient Anatolian languages such as Hittite and Phrygian.

## Unique Characteristics of Burushaski

Burushaski exhibits some distinctive linguistic features. Unlike most languages, which typically have 2 or 3 noun classes, Burushaski categorizes nouns into 4 classes:

- The first class includes male human beings, gods, or spirits.
- The second class includes female human beings, gods, or spirits.
- The third class encompasses animals and countable objects.
- The fourth class is reserved for abstract concepts, fluids, and uncountable nouns.

Burushaski also has a base-20 numeric system, reminiscent of the old French counting system. This characteristic is relatively rare among world languages and adds to the unique nature of Burushaski. Like Nihali, Burushaski follows an SOV word order.

## **Sample Vocabulary in Burushaski**

Some common words in Burushaski include:

- **Ja:** I
- **Un:** You
- **Nana / Nani:** Mother
- **Tata / Tati:** Father
- **Hir:** Man
- **Gus:** Woman
- **Men:** To drink
- **Thoo:** To spit
- **Chrome:** To come
- **Han:** One
- **Alto:** Two
- **Isko:** Three
- **Toorumo:** Ten
- **Tha:** Hundred

## **Conclusion**

Both Nihali and Burushaski exemplify the diversity and complexity of language isolates. Each language possesses unique features, such as Nihali's status as a potentially secret language and Burushaski's extensive noun classification and base-20 numeric system. The origins of both languages remain subjects of scholarly debate, underscoring the mystery and fascination surrounding language isolates.

# Social Context of India's Linguistic Diversity

## *Why does Social context matter?*

India's linguistic landscape is not only about diversity in communication but also intertwined with social identity, hierarchy, and region-based autonomy.

1. **Language as Identity:** For many communities, language is a primary marker of ethnic identity, often linked to specific regions, social customs, and spiritual practices. Linguistic identity can sometimes become a basis for political movements, as seen in the Gorkhaland movement for Nepali speakers or the Bodo people's quest for autonomy in Assam. Language identity also influences **social interactions, marital choices, and community cohesion**.
2. **Linguistic Hierarchy and Social Stigma:** Certain languages are historically associated with **higher status**. For instance, Sanskrit, and by extension its descendants in the Indo-Aryan family, has traditionally been seen as languages of **scholarship and religion**. In contrast, Indigenous languages spoken by tribal communities, particularly in the Austroasiatic and Tibeto-Burman families, have been marginalised or dismissed as "tribal dialects." This stigmatisation has contributed to unequal access to education, social mobility, and representation for speakers of these languages.
3. **Language in Politics and Religion:** Although a complex and often controversial topic, languages are frequently—and mistakenly—linked with religious identities, leading to misunderstandings. Language has been a significant tool in political strategies, shaping communities and influencing social dynamics. From the reorganization of states to large-scale political movements, language has consistently played a pivotal role in shaping the political landscape.

## Understanding the threats to language diversity in India

1. **Urbanisation and Migration:** Economic migration to urban centres where dominant languages like Hindi or regional languages like Kannada (in Bengaluru) and Marathi (in Mumbai) are used has led to a decrease in the transmission of minority languages. As people shift to cities and adopt more widely spoken languages, there is less incentive or opportunity to pass down indigenous languages within families.
2. **Status of Power, Authority, and Opportunity:** Many Indian families aspire for their children to learn and become proficient in languages like English, often as a means of showcasing status. This reflects a widespread societal obsession with the English language. Another factor is the desire for social acceptance, which leads individuals to gradually drift away from their native languages, adopting a

dominant language—whether due to demographics, perceived power, or authority.

3. **Politics:** Politics has, in various ways, contributed to the decline of certain languages. Despite the ideal of “Unity in Diversity,” the government has sometimes, intentionally or inadvertently, favored certain languages over others—whether to ensure national stability, maintain political control, or pursue specific agendas. This can create an impression of linguistic imperialism, where certain languages are promoted at the expense of others.

## **Conclusion**

In summary, India’s richness of languages is more than just a means of communication; it is a foundation of identity, culture, and history for millions. However, the forces of modernisation, social hierarchies, and political pressures are eroding this invaluable heritage. By recognizing the importance of each language, not only as a cultural asset but also as a vital part of individual and community identity, we can strive to create a future that respects and preserves linguistic diversity. Let us remember that protecting our languages is not merely about saving words—it's about honouring the voices, traditions, and lives that each language embodies.