Impact Factor (SJIF): 6.126 **E-ISSN: 2583-1615**

International Journal of Research in Academic World

Received: 08/March/2025 **IJRAW:** 2025; 4(4):130-135 **Accepted:** 16/April/2025

Language Trees of India

[*1]Bhumika Patel and [2]Dr. Shweta Gohel

[*1]Lecturer, GMB Polytechnic, Rajula, Gujarat, India.

[2]Dean, School of Arts and Humanities, Junagadh, Gujarat, India.

Abstract

India is home to a vast and diverse linguistic landscape, with the major language families being Indo-Aryan and Dravidian. Indo-Aryan languages are spoken by approximately 78.05% of the population, while Dravidian languages account for 19.64%. Collectively, these are known as the Indic languages. In addition to these, minor language families and language isolates make up the remaining 2.31%, including Austroasiatic, Sino-Tibetan, Tai-Kadai, and others. The number of languages spoken in India varies according to different sources; the People's Linguistic Survey of India lists 780 languages, whereas Ethnologue records 456.

Officially, Hindi in the Devanagari script is recognized as the official language under Article 343 of the Indian Constitution, with English allowed to continue for official purposes for 15 years from 1947. However, the Official Languages Act of 1963 permits the indefinite use of English alongside Hindi. The international form of Indian numerals, commonly known as Arabic numerals, is used for official purposes. Notably, the Constitution of India does not designate any national language. Instead, it recognizes 22 scheduled languages under the Eighth Schedule, which are granted recognition, status, and official encouragement. Furthermore, six languages—Kannada, Malayalam, Odia, Sanskrit, Tamil, and Telugu—have been given the status of Classical Languages.

According to the 2001 Census, India had 122 major languages and 1,599 other languages. Of these, 30 languages have over a million speakers, while 122 have more than 10,000 speakers. Historically, Persian served as the court and administrative language during the Mughal era, and English continues to play a significant role in higher education and government functions.

In terms of current linguistic dynamics, Hindi has the largest number of native speakers and functions as a lingua franca in northern and central India, though its promotion has raised concerns in non-Hindi speaking regions like Tamil Nadu, Karnataka, Maharashtra, West Bengal, Assam, and Punjab. Bengali is the second most spoken language, mainly in the eastern and northeastern regions, followed by Marathi in the southwest and Telugu in the southeast. According to the 2011 Census, Hindi is the fastest growing language, followed by Kashmiri, Meitei (Manipuri), Gujarati, and Bengali.

Ethnologue provides further classification with 148 Sino-Tibetan, 140 Indo-European, 84 Dravidian, 32 Austroasiatic, 14 Andamanese, and 5 Kra-Dai languages. This linguistic diversity illustrates the complex, multilingual character of India.

Keywords: Language family, Indian roots, classical languages, endangered languages.

Introduction

Classical Languages of India:

Indian classical languages, also known as *Shastriya Bhasha*, are languages recognized for their high antiquity and a rich, original, and distinctive literary heritage. In 2004, the Government of India established specific criteria to grant the status of "Classical Language" to languages with significant historical and literary importance. To oversee this process, the Ministry of Culture formed a Linguistic Experts' Committee to evaluate and recommend languages for this status.

As of now, six languages have been officially recognized as classical**:** Kannada, Malayalam, Odia, Sanskrit, Tamil, and Telugu. These languages are not only known for their deep literary traditions but are also referred to by unique local terms—*Sashtriya Bhasha* in Odia, *Shreshtha Bhasha* in Malayalam, and *Chemmoli* in Tamil.

In addition to the officially recognized ones, there are several other languages that scholars consider classical due to their historical and literary depth. However, these have not yet been granted classical status by the Government of India. Some of them may still enjoy recognition as official or scheduled languages but lack the formal designation of being a "classical language."

Definition: Indian classical languages, or Shastriya Bhasha, refer to languages with high antiquity and valuable, original, and distinct literary heritage.

Recognition

- **Official Recognition:** In 2004, the Government of India established criteria for granting the status of "Classical Language" to languages with significant literary and historical value.
- **Linguistic Experts' Committee:** Formed by the Ministry

***Corresponding Author:** Bhumika Patel

of Culture to evaluate and recommend languages for classical status.

- **Officially Recognized Classical Languages:** Kannada, Malayalam, Odia, Sanskrit, Tamil, Telugu
- **Local Names:**
  - Sashtriya Bhasha (Odia)
  - Shreshtha Bhasha (Malayalam)
  - Chemmoli (Tamil)

- **Unrecognized Classical Languages:**
- **Description:** Languages considered classical by scholars but not officially recognized by the Government of India.
- **Status:** These languages may have official or scheduled language status but lack formal "classical" designation.

Classical Languages and Literature

Indian classical languages, also known as *Shastriya Bhasha*, are those that possess great antiquity and a rich, original, and distinct literary heritage. In 2004, the Government of India laid down specific criteria for granting the status of "Classical Language" to languages with significant historical and literary value. To facilitate this process, the Ministry of Culture established a Linguistic Experts' Committee responsible for evaluating and recommending eligible languages.

Currently, six languages have been officially recognized as classical**:** Kannada, Malayalam, Odia, Sanskrit, Tamil, and Telugu. These languages are celebrated for their profound literary traditions and are known by distinctive local names—*Sashtriya Bhasha* in Odia, *Shreshtha Bhasha* in Malayalam, and *Chemmoli* in Tamil.

Apart from these, several other languages are regarded by scholars as classical due to their antiquity and literary richness. However, they have not yet received official classical status from the Government of India. While some of these may hold official or scheduled language status, they lack formal recognition as "classical languages."

Post Partition Language Trees and their Significance in India

1. Indo Aryan Language Family

The Indo-Aryan languages, a sub-branch of the Indo-European family, are spoken primarily in India, Pakistan, Bangladesh, Nepal, Sri Lanka, and the Maldives, with at least 640 million speakers according to the 1981 census. These languages, which include Hindi, Bengali, and Punjabi among others, form part of the Indo-Iranian subgroup, alongside Iranian languages such as Persian and Pashto. The Indian subcontinent serves as a dialect continuum where linguistic variations transition smoothly across regions due to minimal geographical barriers, facilitating the movement and interaction of speakers.

The Hindi Belt refers to a vast region in northern India where Hindi holds official language status. According to *Ethnologue* (1999), approximately 491 million people speak Hindi as their first language, while around 58 million speak Urdu natively. Despite Hindi's prominence, the region is marked by rich linguistic diversity, with several distinct yet related languages. G.A. Grierson's classification (1883, 1969) divides the Hindi dialect continuum into subgroups**:** Western Hindi (including Braj, Kannauji, and Haryanvi), Eastern Hindi (comprising Awadhi, Bagheli, and Chhattisgarhi), Bihari (covering Bhojpuri, Magahi, and Maithili), Central Pahari (such as Kumauni and Garhwali), and Rajasthani (including Marwari, Mewari, Harauti, and Malvi).

To the east of the Hindi Belt lie other Indo-Aryan languages such as Assamese in Assam, Bengali in West Bengal, Tripura, and Bangladesh, and Oriya in Odisha. Bishnupriya Manipuri, though spoken in the northeast, shows strong linguistic ties with Bihari languages like Bhojpuri, Magahi, and Maithili.

The broader Central Indo-Aryan group includes languages such as Eastern Punjabi, Rajasthani (with its many dialects), and Bhil languages like Bhili, Garasia, Rathawi, and Wagdi, which are geographically situated near Rajasthani, Gujarati, and Marathi regions. Gujarati, including its variant Saurashtra, also falls under this category, as does Modern Standard Hindi, the standardized form of Hindi. Beyond the Indian subcontinent, Central Indo-Aryan languages like Romani and Parya are spoken in countries such as Tajikistan and Uzbekistan.

In the northern and north-western regions, Indo-Aryan languages branch into the Pahari (hill) languages**:** Eastern Pahari (e.g., Nepali), Central Pahari (Garhwali, Kumauni), and Western Pahari (Dogri/Kangri, Jaunsari, Bilaspuri). The North-Western Indo-Aryan group encompasses the Dardic languages, including Khowar and Kalasha, as well as Sindhi, Lahnda (Hindko, Western Punjabi, Saraiki), Kashmiri, and the Kohistani and Kunar language clusters. Shina, spoken in northern Pakistan and parts of Kashmir, also falls in this category. While Dardic languages were once grouped with the Nuristani languages, since Strand's 1973 classification, Nuristani (e.g., Ashkun, Kati/Bashgali, Prasuni, Tregami, Waigali) is now treated as a sister branch to both Indo-Aryan and Iranian language families.

Surrounding this Indo-Aryan core are non-Indo-Aryan language families. The Dravidian language family, comprising Tamil, Kannada, Telugu, and Malayalam, has historically influenced Indo-Aryan languages, especially in regions like Maharashtra, Odisha, and parts of Gujarat. The Iranian language family, including Pashto, Dari, Persian, Baluchi, and Ormuri, also borders this region. Among the language isolates, Burushaski, spoken in Gilgit-Baltistan, remains unclassified.

The Austroasiatic language family is represented by two branches**:** Mon-Khmer (including Amwi, Khasi, and Pnar/Jaintia) and Munda (including Mundari, Juang, Khariya, Gorum, and Ho). Lastly, the Tibeto-Burman language family includes diverse languages like Kiranti, Newari, Meithei, Lepcha, Tangkhul, and Hmar, reflecting the rich ethnolinguistic mosaic of northeastern India.

This intricate web of languages reveals the dynamic interplay of history, migration, and cultural exchange that has shaped India's linguistic landscape over millennia.

2. Dravidian Language Family

The Dravidian language family is spoken by approximately 250 million people, primarily in southern India, northeastern Sri Lanka, and southwestern Pakistan. Significant diaspora communities also exist across Southeast Asia, the Middle East, Europe, North America, and Oceania. The earliest known evidence of Dravidian languages dates back to the 2nd century BCE, with inscriptions in the Tamil-Brahmi script found in Tamil Nadu. Among the major Dravidian languages, Telugu has the largest number of speakers, while Tamil holds the distinction of having the oldest and most extensive literary tradition. Kannada is the dominant language in Karnataka, also known for its rich literary heritage, and Malayalam, spoken in Kerala, has a well-developed and unique literary culture.

There are also smaller Dravidian languages with literary traditions, such as Tulu and Kodava, and tribal languages like

Gondi, spoken in central India. Malto and Kurukh are spoken in isolated regions of eastern India, with Kurukh extending into Nepal, Bhutan, and Bangladesh. The Brahui language, geographically isolated from other Dravidian languages, is spoken in Balochistan (Pakistan), as well as parts of Iran, Afghanistan, and Turkmenistan. Dravidian languages have exerted considerable cultural and linguistic influence across the Indian subcontinent, contributing vocabulary and grammatical features to several Indo-Aryan languages like Marathi, Gujarati, Marwari, and Sindhi. While some hypotheses suggest a migration from the Iranian plateau, linguistic evidence points toward an indigenous origin of the Dravidian family within the Indian subcontinent. No definitive relationship with other global language families has yet been established.

The Dravidian family is divided into three primary branches**:** North Dravidian, Central Dravidian, and South Dravidian. North Dravidian languages include Brahui, Kurukh, Malto, and Kurambhag Paharia, spoken across regions in Pakistan, India, Nepal, Bhutan, and Bangladesh. Central Dravidian languages such as Kolami, Duruwa, Ollari, and Naiki are mainly spoken in Maharashtra, Telangana, Odisha, Chhattisgarh, and Andhra Pradesh. The South Dravidian branch includes the most widely spoken and culturally significant languages**:** Tamil, Kannada, Malayalam, and Tulu, among others. These languages are prevalent across southern India and have spread globally due to migration.

Numerous smaller South Dravidian languages are distributed across southern India, often among tribal communities. These include Beary, Pattapu, Irula, Kurumba, Badaga, Kodava, Jeseri, and Yerukala, among others, each localized to specific districts in states such as Karnataka, Kerala, Tamil Nadu, Andhra Pradesh, and Telangana. Other languages like Paniya, Kanikkaran, Muthuvan, Koraga, Sholaga, Toda, Kota, and Muduga are often spoken in hilly and forested regions and are associated with indigenous cultures. Several Dravidian languages, such as Kakkala, Cholanaikkan, Khirwar, and Kumbaran, remain unclassified, while extinct languages like Malaryan, Nagarchal, and Ullatan once thrived in various parts of India.

The Austroasiatic language family is widely distributed across mainland Southeast Asia, South Asia, and parts of East Asia, including Vietnam, Cambodia, Thailand, Laos, India, Myanmar, Malaysia, Bangladesh, Nepal, and southern China. With approximately 117 million speakers, it includes national languages such as Vietnamese in Vietnam, Khmer in Cambodia, Mon in Myanmar and Thailand, and Wa in Myanmar's Wa State. Santali, an Austroasiatic language, is one of India's 22 scheduled languages. Traditionally, Austroasiatic languages have been divided into Mon–Khmer and Munda branches, though newer classifications propose three groups**:** Munda, Mon-Khmer, and Khasi–Khmuic, with some scholars challenging the validity of the Mon–Khmer grouping.

These languages are considered to be indigenous to mainland Southeast Asia, predating other major language families like Sino-Tibetan and Austronesian. The term "Austroasiatic" is derived from Latin, meaning "southern Asia." Within South Asia, the branches represented are Khasic, Munda, and Nicobarese. The Munda languages, spoken by about nine million people in India and Bangladesh, are categorized into North Munda (including Santali, Mundari, Ho, Korku, Sora) and South Munda, mainly in Odisha and along the Andhra Pradesh-Odisha border. These languages are known for their grammatical complexity, with three numbers, two genders, and distinctions in pronouns, along with agglutinative and polysynthetic features.

The Khasi-Palaungic branch, recognized as distinct in recent classifications, includes the Khasic languages (Khasi, War, Lyngngam) and Palaungic languages (Palaung, Riang, Wa). Khmuic languages are primarily spoken in northern Laos, northern Vietnam, and southern Yunnan in China, with Khmu as the most prominent. The Mang languages, though lesser known, are spoken in Vietnam and China and were first documented in 1974. Pakanic languages, comprising Bolyu and Bugan, are spoken in southern China and were previously misclassified under other groups.

The Vietic languages, which include Vietnamese and Mường, are spoken in Laos and Vietnam and show significant influence from Chinese, particularly in Vietnamese. Katuic and Bahnaric languages are spoken in central and southern Vietnam, Laos, and northeastern Cambodia. The Khmer languages, spoken by the Khmer people, include several dialects and are analytic, isolating, and written in an abugida derived from Brahmi. Pearic languages, spoken by the Pear people in Cambodia and Thailand, are endangered, while Monic languages like Mon and Nyahkur trace their lineage to ancient Dvaravati. The Aslian languages are spoken by the Orang Asli in Malaysia and are classified into Jahaic, Senoic, and Southern Aslian groups. Lastly, the Nicobarese languages are spoken on the Nicobar Islands of India and include northern, central, and southern dialects. Some researchers group them with the Aslian languages based on shared features.

These detailed classifications highlight the complexity and richness of the Dravidian and Austroasiatic language families, each with its own deep history, widespread geographic presence, and intricate internal diversity.

3. Tibetian – Burman Language Family

The Tibeto-Burman (TB) languages form a diverse and widespread language family spoken across the Himalayan region and parts of Southeast Asia. These languages are found from Kashmir in the west, stretching through Tibet and the sub-Himalayan regions of India, Nepal, Bhutan, Bangladesh, and China, and further extending into Southeast Asia, including Burma (Myanmar), Thailand, Laos, and Vietnam. With several hundred languages, many still underdocumented, the TB family represents a vast linguistic landscape. Scholars have attempted various classifications; Shafer (1974) proposed four main groups—Bodic, Baric, Burmic, and Karenic—while Benedict (1972) initially listed eight subgroups, later reclassifying them into three in 1976.

The North-Eastern India Group includes Shafer's Baric or Benedict's Bodo-Garo-Konyak group (now Sal group), featuring languages like Jinghpaw and the Sak/Luish group. These languages typically show lexical peculiarities and simple tonal systems. The Western Group, corresponding to Shafer's Bodic and Benedict's Tibetan-Kanauri Group, includes subgroups like Tibetan Proper, Western and Eastern Bodish, Kanauri (West Himalayish), Kiranti (e.g., Rai), and the Newar-Chepang cluster. The Central North-Eastern or Qiang Subgroup, including Nungish and Naxi, is characterized by shared verb morphology and tonal systems. The South-Eastern Group (Shafer's Kukish and Benedict's Kuki-Chin-(Southern) Naga) includes the Burmese-Lolo and Karen subgroups. These languages often have complex tonal systems, minimal morphology, and SVO word order, especially in Karen and Bai, likely due to Chinese influence.

Western Tibeto-Burman (Bodic) includes Bodish languages

such as Tibetan Proper and its regional forms like West Bodish (Gurung, Tamang), East Bodish (Tshangla), and Central Tibetan (Lhasa). Tibetan varieties include Central Bodish (e.g., Lhasa dialect), spoken across Tibet, Ladakh, Bhutan, and Nepal with around 4.9 million speakers, and Western Tibetan languages like Balti, Ladakhi, and Zangskar. These are found in Baltistan, Ladakh, and the Zangskar Valley, respectively. Other Western Himalayan varieties include Lahuli dialects like Tod and Ranglo, spoken in Lahul, and Piti (Spiti) dialects like Bhar and Sham. Additional lesser-known varieties include mNyam, mNyamskad, and Jad (Garhwal Bhotia), as well as the mNgahris subgroup found in Ngari Prefecture, western Tibet, with around 40,000 speakers.

The Gtsang subgroup covers dialects in northern Nepal (e.g., Dolpo, Reng Pungmo, Tichurong) and Tibetan areas like Shigatse and Gyantse, totaling nearly half a million speakers. Prominent within this group is Sherpa, with over 70,000 speakers in Nepal, and related languages like Kagate, Jirel, Lhomi, and Halung. The Dbus (Central Tibetan) variety includes Lhasa Tibetan, which serves as the standard language and is spoken by approximately 450,000 people, plus about 150,000 Tibetan refugees worldwide. Southern Tibetan varieties such as Tromowa (Gromo), Danjong (spoken in Sikkim by over 70,000 people), and Dzongkha (spoken by 225,000 and the national language of Bhutan) show further diversity. The Amdo region preserves archaic Tibetan features across varieties like Golok and Braghgo, spoken by over 800,000 people. Khams Tibetan is split into Western, Eastern, Northern, Southern, Cone, and bBrugchu varieties, collectively spoken by over a million people.

The West Bodish group includes lesser-known but significant languages such as Kutang Ghale, Ghale (15,000 speakers), Kaike (Tarali Kham), Gurung (about 150,000 speakers), Thakali (Tapaang), Chantel, Rohani, Manang, and Tamang (approximately 600,000 speakers). These languages often show Tibetan influence and exhibit substantial dialectal variation.

The East Bodish and Himalayan groups offer another layer of linguistic richness, particularly in northeast India, Bhutan, and Tibet. Dzalakha is spoken in northeastern Bhutan, Tawang District (India), and Tibet's Cuona region, with around 52,000 speakers in total. East Monpa is spoken in Arunachal Pradesh and adjacent Tibetan areas, while Sherdukpen, found in West Kameng District, includes Bot and Lish subgroups. Lhokpu (Lhobikha) and Gongduk (Gongdubikha), spoken in southwestern and south-central Bhutan respectively, are considered remnants of pre-Tibetan indigenous communities. Tshangla, also called Sharchop, functions as a lingua franca among Monpa communities in southeastern Bhutan and Arunachal Pradesh.

This comprehensive overview of the Tibeto-Burman languages illustrates their incredible diversity and complexity, spanning an enormous geographic and cultural expanse. Each subgroup offers a window into the intricate interplay between language, history, and identity across the Himalayan and Southeast Asian worlds.

The Central Himalayan and North-East Indian regions are home to a remarkable diversity of languages, reflecting a blend of historical, geographical, and cultural factors. The West Himalayish languages, spoken along the Chandra and Bhaga Rivers in northern Himachal Pradesh, include a variety of smaller languages, such as Pattani, with around 20,000 speakers, and Tinan, spoken by approximately 2,500 people. Bunan, with about 5,000 speakers, is found along the Bhaga River in western Lahul. Kanauri, spoken in the Kinnaur area, has around 60,000 speakers and includes several dialects, including Lower and Upper Kanauri, Thebor, and Chitkbuli. The Kanashi language, spoken by about 1,100 people in Malana village near Kulu, also contributes to the region's linguistic diversity. Rangpa, with approximately 7,500 speakers in north-central Garhwal, and Thami, spoken by 14,000 people in eastern Nepal, further highlight the complexity of the region's languages.

In Nepal, the Central Himalayan languages, such as Magar (with 290,000 speakers), Raute, and Raji, illustrate the historical shifts from dominant languages, as many speakers have adopted languages like Kham or Nepali. The Rai languages, including Sunwar, Yakkha, Limbu, and Bantawa, are significant in the eastern part of Nepal but face endangerment, with only about 400,000 speakers across various subgroups. The Newari language, with around 600,000 speakers in the Kathmandu Valley, has a rich literary tradition but is also experiencing a decline in use. The Kiranti or Rai languages are highly diverse and include many dialects such as Chamling, Belhare, and Chintang, with some facing the risk of extinction.

In North-East India, the Sal languages, including Boro, Garo, Kokborok, and Dimasa, are notable for their innovative etyma and widespread use, particularly in Assam, Meghalaya, and Tripura. The Naga languages, such as Ao, Chang, and Konyak, represent a highly diverse linguistic group with distinct dialects, and these languages have substantial populations across India and Burma. Similarly, the Kuki-Chin group, spoken across Nagaland, Mizoram, and Manipur, includes languages like Thado, Zeliangrong, and Meithei, with some communities transitioning to the Meithei language. In addition to the Tibeto-Burman languages, the region also includes languages like Jinghpaw and the moribund Sak languages, which highlight the historical and cultural interactions across Burma and neighboring regions.

The South-Eastern Tibeto-Burman languages include the Burmish and Loloish subgroups, with Burmese as a major representative language, and Loloish languages spoken across southern China and northern Southeast Asia. The Ugong language, with only about 300 speakers in Thailand, is critically endangered. Mru, a language with around 40,000 speakers in Myanmar and Bangladesh, is also of uncertain classification but shows influence from nearby Kuki-Chin and Burmese languages. The complexity of these languages, their historical migrations, and ongoing shifts underscore the rich, multifaceted linguistic landscape of the Central Himalayan and North-East Indian regions.

4. Tai Kadai/Kra-Dai/DaicKadai and Language Group

The Kra-Dai language family, also known as Tai-Kadai or Daic, is one of the major language families in Southeast Asia and Southern China, encompassing nearly 100 languages spoken by approximately 100 million people across regions like China, Thailand, Laos, Myanmar, India, and Vietnam. This family is divided into two primary branches**:** Tai (Kam-Tai) and Kadai. The Tai branch includes prominent languages such as Thai, the national language of Thailand, and Lao, the national language of Laos, along with related languages like Zhuang, Kam, and Sui. Thai and Lao use Indic-based scripts, while Zhuang and Kam-Sui use Chinese characters or Romanized scripts. The Kadai branch is made up of lesser-known languages such as Gelao, Lachi, and Laqua, with many Kadai languages being primarily oral and using Chinese characters or Romanized scripts in some cases.

The historical and genetic affiliations of Kra-Dai languages

have been the subject of extensive research and debate. Early theories, such as the Tai-Austronesian hypothesis proposed by Paul Benedict, suggested a connection between Tai-Kadai and Austronesian languages. Other historical theories include the proposition by James R. Chamberlain that the Tai-Kadai family may have originated as early as the 12th century BCE in Southern China, and the suggestion by Weera Ostapirat that Kra-Dai and Austronesian could be sister language families. Laurent Sagart also proposed that Kra-Dai might be a branch of Austronesian that migrated back to mainland China. Current theories include the Sino-Tai Hypothesis, which proposes a connection between Tai and Sino-Tibetan languages, and the Austro-Tai Hypothesis, suggesting a genetic link between Tai and Austronesian languages. Additionally, the Sino-Tibetan-Austronesian Hypothesis postulates a larger language phylum that includes Tai-Kadai, Sino-Tibetan, Miao-Yao, and Austronesian.

Kra-Dai languages are categorized into several internal branches**:** Tai, spoken in Southern China and Southeast Asia; Kra, found in Southern China and Northern Vietnam; Kam-Sui, spoken in Guizhou and Guangxi in China; Be, located in Hainan, possibly including Jizhao of Guangdong; and Hlai, also spoken in Hainan. There are also unclassified languages such as Lakkia, Biao, Jiamao, and Jizhao, as well as mixed-origin languages like Hezhang Buyi (a mix of Northern Tai and Kra), E (a mix of Northern Tai and Pinghua Chinese), Caolan (a blend of Northern Tai and Central Tai), Sanqiao (a mix of Kam-Sui, Hmongic, and Chinese), and Jiamao (a blend of Hlai and other unknown elements).

The writing systems used by Kra-Dai languages are diverse. Tai languages, such as Thai and Lao, typically use Indic-based scripts, while languages like Zhuang and Kam-Sui employ Chinese characters or Romanized scripts. Many Kadai languages, however, lack their own writing systems, and some use Chinese characters or Romanized scripts. The genetic affiliation of Kra-Dai remains a topic of ongoing research, with scholars continuing to explore its relationships with other language families and its internal classification. The complexity and diversity of Kra-Dai languages highlight their rich linguistic history and evolving theories surrounding their origins and connections.

5. Andamanese Language Family

The Andamanese languages are spoken by the indigenous peoples of the Andaman Islands in the Indian Ocean, and are grouped into two known families**:** Great Andamanese and Ongan. There are also two unattested languages, Sentinelese and Jangil, which have added to the complexity and mystery of the region's linguistic diversity.

The Great Andamanese family historically included ten distinct languages, such as Aka-Bea, Akar-Bale, A-Pucikwar, Oko-Juwoi, and others. However, many of these languages have become extinct, and by the early 21st century, only a small number of people spoke a modified version called "Present Great Andamanese" or "Jero," which is primarily based on Aka-Jeru. In recent times, Hindi has increasingly replaced the indigenous languages as the primary language among the descendants. The Ongan family consists of the Jarawa and Onge languages, which are still spoken by approximately 670 people. These languages have survived due to the relative isolation of their speakers and their resistance to outside contact.

Two unattested languages, Sentinelese and Jangil, remain of great interest to linguists. The Sentinelese people refuse all contact with outsiders, meaning their language remains unknown, although estimates suggest there are between 50 to 250 speakers. Some scholars, such as Anvita Abbi, have speculated that Sentinelese might be related to Ongan. Jangil, on the other hand, became extinct in the 1920s, with the last known speakers having died off. Maurice Vidal Portman noted similarities between Jangil and Jarawa, suggesting a potential relationship, but this has not been confirmed.

The historical context of the Andamanese languages reveals the impact of external influences on the indigenous populations. Before the late 18th century, the indigenous peoples of the Andaman Islands had minimal contact with outside powers. However, when the British established a colonial presence, significant disruptions followed. The establishment of a penal colony in the 1860s brought mainland settlers and indentured laborers, which contributed to a severe population decline and cultural disruptions among the Andamanese communities. Early linguistic documentation was initiated by Alexander John Ellis, who studied South Andamanese languages and shared his findings with the Philological Society, laying the groundwork for the understanding of Andamanese languages.

In contemporary times, the Great Andamanese languages have largely disappeared due to historical disruptions, with only a modified version and Hindi being used by the remaining descendants. The Ongan languages, Jarawa and Onge, have managed to survive largely because of the isolation of their speakers. The languages of Sentinelese and Jangil, however, remain poorly understood due to a lack of direct contact and the extinction of Jangil.

Regarding linguistic relationships, there is no conclusive evidence to suggest that the Great Andamanese and Ongan languages are genealogically related, despite some typological similarities. Scholars such as Anvita Abbi and Radcliffe-Brown have debated potential connections, but there is a general consensus that the languages are unrelated. The relationship of Sentinelese with other Andamanese languages is still speculative, with some scholars proposing a potential link to Ongan based on linguistic features.

The Andamanese languages, as a whole, are more of a geographic label than a genetic one, as the languages within this category are not closely related. Historical contact, population decline, and cultural disruptions have all significantly impacted the preservation and diversity of Andamanese languages. Ongoing research and exploration are necessary to better understand these languages and their relationships, particularly those that remain unattested.

Conclusion

India is home to a diverse linguistic landscape, with several major language families. The Indo-Aryan family, spoken across northern, central, and western India, includes languages such as Hindi, Bengali, Marathi, and Punjabi. The Dravidian family, primarily found in southern India, encompasses languages like Tamil, Telugu, Kannada, and Malayalam. The Tibeto-Burman family is spoken in the northeastern states and includes languages like Bodo and Meitei, while the Austroasiatic family, with languages like Santali and Khasi, is found in central and eastern India. The Austronesian family, spoken in the Andaman and Nicobar Islands, includes languages such as Nicobarese and Great Andamanese. There are also smaller groups like the Iranian languages (e.g., Kashmiri) and Turkic languages (e.g., Uighur), along with Semitic languages (e.g., Arabic). Additionally, India is home to various isolated languages and smaller families, contributing to the nation's rich linguistic

diversity. This diversity is a result of India's complex history, culture, and geography, with ongoing efforts to preserve and recognize the country's many languages.

References

1. "50th Report of the Commissioner for Linguistic Minorities in India (July 2012 to June 2013)" (PDF). Commissioner for Linguistic Minorities, Ministry of Minority Affairs, Government of India. Archived from the original (PDF) on 26 December 2014. Retrieved 17 September 2016.
2. "Constitution of India". Archived from the original on 2 April 2012. Retrieved 21 March 2012.
3. "Official Language-The Union-Profile-Know India: National Portal of India". Archive.india.gov.in. Retrieved 28 December 2017.
4. "Official Language Act | Government of India, Ministry of Electronics and Information Technology". meity.gov.in. Retrieved 24 January 2017.
5. Salzmann, Zdenek; Stanlaw, James; Adachi, Nobuko (8 July 2014). Language, Culture, and Society: An Introduction to Linguistic Anthropology. Westview Press. ISBN 9780813349558-via Google Books.
6. "Indo-Aryan languages". Encyclopædia Britannica. Retrieved 10 December 2014.
7. "Hindi languages". Encyclopædia Britannica. Retrieved 10 December 2014.