

# Languages of India

Languages spoken in India belong to several language families, the major ones being the Indo-Aryan languages spoken by 78.05% of Indians and the Dravidian languages spoken by 19.64% of Indians. Languages spoken by the remaining 2.31% of the population belong to the Austro-Asiatic, Sino-Tibetan, Tai-Kadai and a few other minor language families and isolates. India (780) has the world's second highest number of languages, after Papua New Guinea. According to the Census of India of 2001, India has 122 major languages and 1599 other languages. However, figures from other sources vary, primarily due to differences in definition of the terms "language" and "dialect". The 2001 Census recorded 30 languages which were spoken by more than a million native speakers and 122 which were spoken by more than 10,000 people. English continues to be an important language in India. It is used in higher education and in some areas of the Indian government. Hindi, the most commonly spoken language in India today, serves as the *lingua franca* across much of North and Central India. However, there have been concerns raised with Hindi being imposed in South India, most notably in the state of Tamil Nadu and Karnataka. Maharashtra, West Bengal, Assam, Punjab and other non-Hindi regions have also started to voice concerns about Hindi.

## Language families

Ethno linguistically, the languages of South Asia, echoing the complex history and geography of the region, form a complex patchwork of language families, language phyla and isolates. Languages spoken in India belong to several language families, the major ones being the Indo-Aryan languages spoken by 78.05% of Indians and the Dravidian languages spoken by 19.64% of Indians. The languages of India belong to several language families, the most important of which are –

### 1 Indo-Aryan language family

The largest of the language families represented in India, in terms of speakers, is the Indo-Aryan language family, a branch of the Indo-Iranian family, itself the easternmost, extant subfamily of the Indo-European language family. This language family predominates, accounting for some 1035 million speakers, or over 76.5 of the population, as per 2018 estimate. The most widely spoken languages of this group are Hindi, Bengali, Marathi, Urdu, Gujarati, Punjabi, Kashmiri, Rajasthani, Sindhi, Assamese (Asamiya) Maithili and Odia. Aside from the Indo-Aryan languages, other Indo-European languages are also spoken in India, the most prominent of which is English, as a *lingua franca*.

### 2 Dravidian language family

The second largest language family is the Dravidian language family, accounting for some 277 million speakers, or approximately 20.5% as per 2018 estimate. The Dravidian languages are spoken mainly in southern India and parts of eastern and central India as well as in parts of north-eastern Sri Lanka, Pakistan, Nepal and Bangladesh. The Dravidian languages with the most speakers

are Telugu, Tamil, Kannada and Malayalam. Besides the mainstream population, Dravidian languages are also spoken by small scheduled tribe communities, such as the Oraon and Gond tribes. Only two Dravidian languages are exclusively spoken outside India, Brahui in Pakistan and Dhangar, a dialect of Kurukh, in Nepal.

### 3 Austro-Asiatic language family

Families with smaller numbers of speakers are Austro-Asiatic and numerous small Sino-Tibetan languages, with some 10 and 6 million speakers, respectively, together 3% of the population. The Austro-Asiatic language family (*austro* meaning South) is the autochthonous language in South Asia and Southeast Asia, other language families having arrived by migration. Austro-Asiatic languages of mainland India are the Khasi and Munda languages, including Santhali. The languages of the Nicobar islands also form part of this language family. With the exceptions of Khasi and Santhali, all Austro-Asiatic languages on Indian territory are endangered.

### 4 Sino-Tibetan language family

The Sino-Tibetan language family are well represented in India. However, their interrelationships are not discernible, and the family has been described as "a patch of leaves on the forest floor" rather than with the conventional metaphor of a "family tree".

Sino-Tibetan languages are spoken across the Himalayas in the regions of Ladakh, Himachal Pradesh, Nepal, Sikkim, Bhutan, Arunachal Pradesh, and also in the Indian states of West Bengal, Assam (hills and autonomous councils), Meghalaya, Nagaland, Manipur, Tripura and Mizoram. Sino-Tibetan languages spoken in India include the scheduled languages Meitei and Bodo, the non-scheduled languages of Karbi, Lepcha, and many varieties of several related Tibetic, West Himalayish, Tani, Brahmaputran, Angami–Pochuri, Tangkhul, Zeme, Kukish language groups, amongst many others.

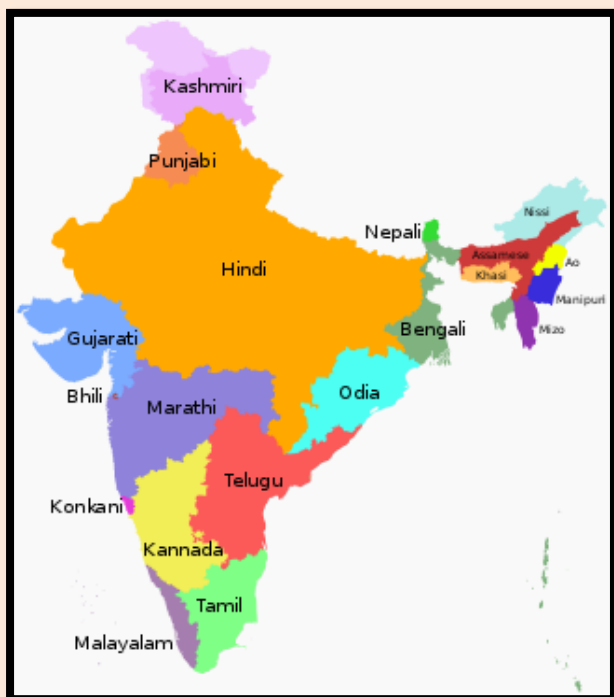
### 5 Tai-Kadai language family

Ahom language, a South western Tai language, had been once the dominant language of the Ahom Kingdom in modern-day Assam, but was later replaced by the Assamese language (known as Kamrupi in ancient era which is the pre-form of the Kamrupi dialect of today). Nowadays, small Tai communities and their languages remain in Assam and Arunachal Pradesh together with Sino-Tibetans, e.g. Tai Phake, Tai Aiton and Tai Khamti, which are similar to the Shan language of Shan State, Myanmar; the Dai language of Yunnan, China; the Lao language of Laos; the Thai language of Thailand; and the Zhuang language in Guangxi, China

## 6 Great Andamanese language family

The languages of the Andaman Islands form a group:

- the Great Andamanese languages, comprising a number of extinct, and one highly endangered language
- the Ongan family of the southern Andaman Islands, comprising two extant languages, Önge and Jarawa, and one extinct language, Jangil.



Language	Family
Assamese	<b>Indo-Aryan</b>
Bengali (Bangla)	
Dogri	
Gujarati	
Hindi	
Kashmiri	
Konkani	
Maithili	
Marathi	
Nepali	
Odia	
Punjabi	
Sanskrit	
Urdu	
Sindhi	
Meitei	<b>Sino-Tibetan</b>
Bodo	
Santali	<b>Austro-Asiatic</b>
Tamil	<b>Dravidian</b>
Telugu	<b>Dravidian</b>
Malayalam	<b>Dravidian</b>
Kannada	<b>Dravidian</b>

## **Hindi**

Hindi, written in Devanagari script, is the most prominent language spoken in the country. In the 2001 census, 422 million (422,048,642) people in India reported Hindi to be their native language. This figure not only included Hindi speakers of Hindustani, but also people who identify as native speakers of related languages who consider their speech to be a dialect of Hindi, the Hindi belt. Hindi (or Hindustani) is the native language of most people living in Delhi, Uttar Pradesh, Uttarakhand, Chhattisgarh, Himachal Pradesh, Chandigarh, Bihar, Jharkhand, Madhya Pradesh, Haryana, and Rajasthan.

"Modern Standard Hindi", a standardised language is one of the official languages of the Union of India. In addition, it is one of only two languages used for business in Parliament however the Rajya Sabha now allows all 22 official languages on the Eighth Schedule to be spoken.

## **Bengali**

Native to the Bengal region, comprising the nation of Bangladesh and the states of West Bengal, Tripura and Barak Valley region of Assam. Bengali is the sixth most spoken language in the world. After the partition of India (1947), refugees from East Pakistan were settled in Tripura, and Jharkhand and the union territory of Andaman and Nicobar Islands. There is also a large number of Bengali-speaking people in Maharashtra and Gujarat where they work as artisans in jewellery industries. Bengali developed from Abahatta, a derivative of Apabhramsha, itself derived from Magadhi Prakrit. The modern Bengali vocabulary contains the vocabulary base from Magadhi Prakrit and Pali, also borrowings & reborrowings from Sanskrit and other major borrowings from Persian, Arabic, Austro-Asiatic languages and other languages in contact with. Like most Indian languages, Bengali has a number of dialects. It exhibits diglossia, with the literary and standard form differing greatly from the colloquial speech of the regions that identify with the language. Bengali language has developed a rich cultural base spanning art, music, literature and religion. There have been many movements in defence of this language and in 1999 UNESCO declared 21 Feb as the International Mother Language Day in commemoration of the Bengali Language Movement in 1952.

## **Marathi**

Marathi is an Indo-Aryan language. It is the official language and co-official language in Maharashtra and Goa states of Western India respectively, and is one of the official languages of India. There were 83 million speakers in 2011 and 72 million speakers in 2001. Marathi has the third largest number of native speakers in India. Marathi has some of the oldest literature of all modern Indo-Aryan languages, dating from about 1200 AD (Mukundraj's *Vivek Sindhu* from the close of the 12th century). The major dialects of Marathi are Standard Marathi and the Varhadi dialect. There are other related languages such as Khandeshi, Dangi, Vadvali and Samavedi. Malvani Konkani has been heavily influenced by Marathi

varieties. Marathi is one of several languages that descend from Maharashtri Prakrit. Further change led to the Apabhraṃśa languages like Old Marathi.

Marathi is the official language of Maharashtra and co-official language in the union territories of Daman and Diu and Dadra and Nagar Haveli. In Goa, Konkani is the sole official language; however, Marathi may also be used for all official purposes.

### **Telugu**

Telugu is the most widely spoken Dravidian language in India and around the world. Telugu is an official language in Andhra Pradesh, Telangana and Yanam, making it one of the few languages (along with Hindi, Bengali, and Urdu) with official status in more than one state. It is also spoken by a significant number of people in the Andaman and Nicobar Islands, Chhattisgarh, Karnataka, Maharashtra, Odisha, Tamil Nadu, Gujarat and by the Sri Lankan Gypsy people. It is one of six languages with classical status in India. Telugu ranks fourth by the number of native speakers in India (81 million in the 2011 Census), fifteenth in the *Ethnologue* list of most-spoken languages worldwide and is the most widely spoken Dravidian language.

### **Tamil**

Tamil is a Dravidian language predominantly spoken in Tamil Nadu, Puduchery and many parts of Sri Lanka. It is also spoken by large minorities in the Andaman and Nicobar Islands, Kerala, Karnataka, Andhra Pradesh, Malaysia, Singapore, Mauritius and throughout the world. Tamil ranks fifth by the number of native speakers in India (61 million in the 2001 Census) and ranks 20th in the list of most spoken languages. It is one of the 22 scheduled languages of India and was the first Indian language to be declared a classical language by the Government of India in 2004. Tamil is one of the longest surviving classical languages in the world. It has been described as "the only language of contemporary India which is recognisably continuous with a classical past." The two earliest manuscripts from India, acknowledged and registered by UNESCO Memory of the World register in 1997 and 2005, are in Tamil. Tamil is an official language of Tamil Nadu, Puducherry, Andaman and Nicobar Islands, Sri Lanka and Singapore. It is also recognized as minority language in Canada, Malaysia, Mauritius and South Africa.

### **Urdu**

After independence, Modern Standard Urdu, the Persianised register of Hindustani became the national language of Pakistan. During British colonial times, a knowledge of Hindustani or Urdu was a must for officials. Hindustani was made the second language of British Indian Empire after English and considered as the language of administration. The British introduced the use of Roman script for Hindustani as well as other languages. Urdu had 70 million speakers in India (as per the Census of 2001), and, along with Hindi, is one of the 22 officially recognised regional languages of India and also an official language in the Indian states of Jammu and Kashmir, Delhi, Uttar Pradesh, Bihar and Telangana that have significant Muslim populations.

### **Gujarati**

Gujarati is an Indo-Aryan language. It is native to the west Indian region of Gujarat. Gujarati is part of the greater Indo-European language family. Gujarati is descended from Old Gujarati (c. 1100 – 1500 CE), the same source as that of Rajasthani. Gujarati is the chief language in the Indian state of Gujarat. It is also an official language in the union territories of Daman and Diu and Dadra and Nagar Haveli. According to the Central Intelligence Agency (CIA), 4.5% of population of India (1.21 billion according to 2011 census) speaks Gujarati. This amounts to 54.6 million speakers in India.

### **Kannada**

Kannada language is a Dravidian language which branched off from Kannada-Tamil sub group around 500 B.C.E according to the Dravidian scholar Zvelebil. According to the Dravidian scholars Steever and Krishnamurthy, the study of Kannada language is usually divided into three linguistic phases: Old (450–1200 CE), Middle (1200–1700 CE) and Modern (1700–present). The earliest written records are from the 5th century, and the earliest available literature in rich manuscript (*Kavirajamarga*) is from c. 850. Kannada language has the second oldest written tradition of all vernacular languages of India. Current estimates of the total number of epigraph present in Karnataka range from 25,000 by the scholar Sheldon Pollock to over 30,000 by the Sahitya Akademi, making Karnataka state "one of the most densely inscribed pieces of real estate in the world". According to Garg and Shipely, more than a thousand notable writers have contributed to the wealth of the language.

### **Malayalam**

Malayalam has official language status in the state of Kerala and in the union territories of Lakshadweep and Puducherry. It belongs to the Dravidian family of languages and is spoken by some 38 million people. Malayalam is also spoken in the neighboring states of Tamil Nadu and Karnataka; with some speakers in the Nilgiris, Kanyakumari and Coimbatore districts of Tamil Nadu, and the Dakshina Kannada and the Kodagu district of Karnataka. Malayalam originated from Middle Tamil (Sen-Tamil) in the 7th century. As Malayalam began to freely borrow words as well as the rules of grammar from Sanskrit, the Grantha alphabet was adopted for writing and came to be known as *Arya Eluttu*. This developed into the modern Malayalam script

### **Odia**

Odia (formerly spelled *Oriya*) is the only modern language officially recognized as a classical language from the Indo-Aryan group. Odia is primarily spoken in the Indian state of Odisha and has over 40 million speakers. It was declared as a classical language of India in 2014. Native speakers comprise 91.85% of the population in Odisha. Odia is thought to have originated from Udra Prakrit similar to Udra Magadhi, a language spoken in eastern India over 2,500 years ago. The history of Odia language can be divided to Old Odia (3rd century BC –1200 century AD), Early Middle Odia (1200–1400), Middle Odia (1400–1700), Late

Middle Odia (1700–1870) and Modern Odia (1870 till present day). The National Manuscripts Mission of India have found around 213,000 unearthed and preserved manuscripts written in Odia.

### **Punjabi**

Punjabi, written in the Gurmukhi script in India, is one of the prominent languages of India with about 32 million speakers. In Pakistan it is spoken by over 80 million people and is written in the Shahmukhi alphabet. It is mainly spoken in Punjab but also in neighboring areas. It is an official language of Delhi and Punjab.

### **Assamese**

Asamiya or Assamese language is most popular in the state of Assam and Brahmaputra Valley. It's an Eastern Indo-Aryan language having more than 15 million speakers as per world estimates by Encarta.

### **Maithili**

Maithili is an Indo-Aryan language spoken in the Bihar and Jharkhand states of India. In Nepal, it is spoken in the eastern Terai, and is the second most prevalent language of Nepal. Tirhuta was formerly the primary script for written Maithili. Less commonly, it was also written in the local variant of Kaithi. Today it is written in the Devanagari script. In 2003, Maithili was included in the Eighth Schedule of the Indian Constitution as a recognised regional language of India, which allows it to be used in education, government, and other official contexts.