The languages of the Indian Subcontinent are divided into various language families, of which the Indo-Iranian and the Dravidian languages are the most widely spoken. There are also many languages belonging to unrelated language families such as Sino-Tibetan, spoken by smaller groups. Linguistic records begin with the appearance of the Brāhmī script from about the 3rd century BCE.

Indo-Aryan languages

Main article: Indo-Aryan languages

Proto-Indo-Aryan

Main article: Proto-Indo-Aryan language

Proto-Indo-Aryan is a proto-language hypothesized to have been the direct ancestor of all Indo-Aryan languages.[1] It would have had similarities to Proto-Indo-Iranian, but would ultimately use Sanskritized phonemes and morphology.

Old Indo-Aryan

Vedic Sanskrit

Devimahatmya manuscript on palm-leaf, in an early Bhujimol script, Bihar or Nepal, 11th century

Vedic Sanskrit is the language of the Vedas, a large collection of hymns, incantations, and religio-philosophical discussions which form the earliest religious texts in India and the basis for much of the Hindu religion. Modern linguists consider the metrical hymns of the Rigveda to be the earliest. The hymns preserved in the Rigveda were preserved by oral tradition alone over several centuries before the introduction of writing, the oldest among them predating the introduction of Brahmi by as much as a millennium .[*citation needed*]

The end of the Vedic period is marked by the composition of the Upanishads, which form the concluding part of the Vedic corpus in the traditional compilations, dated to roughly 500 BCE. It is around this time that Sanskrit began the transition from a first language to a second language of religion and learning, marking the beginning of the Classical period.

Classical Sanskrit

The oldest surviving Sanskrit grammar is Pāṇini's Aṣṭādhyāyī ("Eight-Chapter Grammar") dating to c. the 5th century BCE. It is essentially a prescriptive grammar, i.e., an authority that defines (rather than describes) correct Sanskrit, although it contains descriptive parts, mostly to account for Vedic forms that had already passed out of use in Pāṇini's time.

When the term arose in India, *Sanskrit* was not thought of as a specific language set apart from other languages (the people of the time regarded languages more as dialects), but rather as a particularly refined or perfected manner of speaking. Knowledge of Sanskrit was a marker of social class and educational attainment and was taught mainly to Brahmins through close analysis of Sanskrit grammarians such as Pāṇini.

Vedic Sanskrit and Classical or "Paninian" Sanskrit, while broadly similar, are separate[2] varieties, which differ in a number of points of phonology, vocabulary, and grammar.

Middle Indo-Aryan

Prakrit (Sanskrit *prākṛta* प्राकृत (from *prakṛti* प्रकृति), "original, natural, artless, normal, ordinary, usual", i.e. "vernacular", in contrast to *samskrta* "excellently made",[*citation needed*] both adjectives elliptically referring to *vak* "speech") refers to the broad family of Indic languages and dialects spoken in ancient India. Some modern scholars include all Middle Indo-Aryan languages under the rubric of "Prakrits", while others emphasise the independent development of these languages, often separated from the history of Sanskrit by wide divisions of caste, religion, and geography.

The Prakrits became literary languages, generally patronized by kings identified with the kshatriya caste. The earliest inscriptions in Prakrit are those of Ashoka, emperor of Southern India, and while the various Prakrit languages are associated with different patron dynasties, with different religions and different literary traditions.

In Sanskrit drama, kings speak in Prakrit when addressing women or servants, in contrast to the Sanskrit used in reciting more formal poetic monologues.

The three Dramatic Prakrits – Sauraseni, Magadhi, Maharashtri, as well as Jain Prakrit each represent a distinct tradition of literature within the history of India. Other Prakrits are reported in historical sources, but have no extant corpus (e.g., Paisaci).

Pali

Pali is the Middle Indo-Aryan language in which the Theravada Buddhist scriptures and commentaries are preserved. Pali is believed by the Theravada tradition to be the same language as Magadhi, but modern scholars believe this to be unlikely.[*citation needed*] Pali shows signs of development from several underlying prakrits as well as some Sanskritisation.

The prakrit of the North-western area of India known as Gāndhāra has come to be called Gāndhārī. A few documents written in the Kharoṣṭhi script survive including a version of the Dhammapada.

Apabhraṃśa/Apasabda

The Prakrits (which includes Pali) were gradually transformed into Apabhraṃśas (अपभ्रंश) which were used until about the 13th century CE. The term apabhraṃśa refers to the dialects of

Northern India before the rise of modern Northern Indian languages, and implies a corrupt or non-standard language. A significant amount of apabhraṃśa literature has been found in Jain libraries. While Amir Khusro and Kabir were writing in a language quite similar to modern Hindi-Urdu, many poets, specially in regions that were still ruled by Hindu kings, continued to write in Apabhraṃśa. Apabhraṃśa authors include Sarahapad of Kamarupa, Devasena of Dhar (9th century CE), Pushpadanta of Manikhet (9th century CE), Dhanapal, Muni Ramsimha, Hemachandra of Patan, Raighu of Gwalior (15th century CE). An early example of the use of Apabhraṃśa is in Vikramūrvashīiya of Kalidasa, when Pururava asks the animals in the forest about his beloved who had disappeared.

Modern Indo-Aryan

Hindi

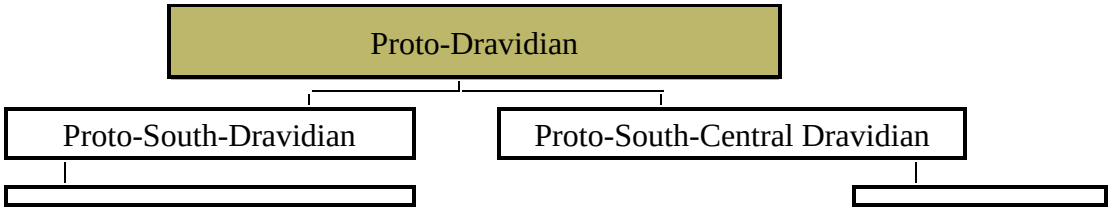
Further information: History of Hindustani

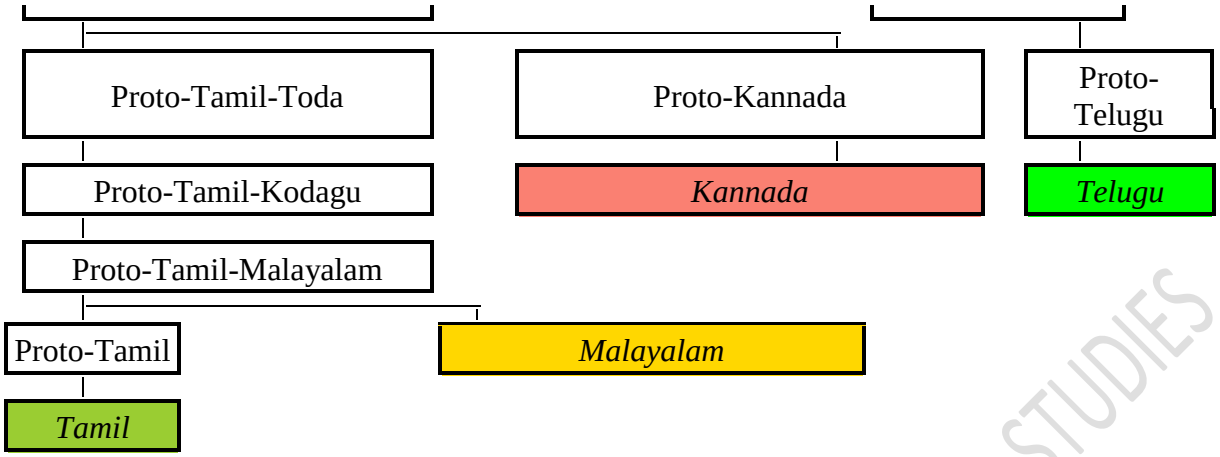
Hindi is right now the most spoken language in India and the fourth most spoken language in the world. The development of Hindi revolves around the various Hindi dialects originating mainly from Sauraseni Apabhramsha. A Jain text Shravakachar written in 933AD is considered the first Hindi book.[3] Modern Hindi is based on prestigious Khariboli dialect which started to take Persian and Arabic words too with the establishment of Delhi Sultanate however the Arabic-Persian influence was profound mainly on Urdu and lesser in Hindi, Khariboli also started to spread across North India as a vernacular form called Hindustani. Amir Khusrow wrote poems in Khariboli and Brajbhasha and referred that language as Hindavi. During Bhakti era many poems were composed in Khariboli, Brajbhasa and Awadhi. One of such classic is Ramcharitmanas written by Tulsidas in Awadhi. In 1623 Jatmal wrote a book in Khariboli with the name 'Gora Badal ki Katha'.

With the establishment of British rule in India, saw the clear division of Hindi and Urdu languages. This period also saw the rise of modern Hindi literature starting with Bharatendu Harishchandra. This period also shows further Sanskritization of Hindi language in literature. Hindi is right now official languages in nine states of India- Uttar Pradesh, Bihar, Rajasthan, Jharkhand, Madhya Pradesh, Chhattisgarh, Uttarakhand, Haryana and Himachal Pradesh and the national Capital Territory of Delhi. Post independence Hindi became the official language of Central Government of India along with English.

Further information: Bengali Language

Dravidian languages





This tree diagram depicts the genealogy of the primary Dravidian languages spoken in South India.

The Dravidian family of languages includes approximately 73 languages[4] that are mainly spoken in southern India and northeastern Sri Lanka, as well as certain areas in Pakistan, Nepal, Bangladesh, and eastern and central India, as well as in parts of Afghanistan, and overseas in other countries such as the United Kingdom, United States, Canada, Malaysia and Singapore.

The origins of the Dravidian languages, as well as their subsequent development and the period of their differentiation, are unclear, and the situation is not helped by the lack of comparative linguistic research into the Dravidian languages. Inconclusive attempts have also been made to link the family with the Japonic languages[*citation needed*] and with the extinct Elamite language (by Elamo-Dravidian hypothesis).

Many linguists, however, tend to favour the theory that speakers of Dravidian languages spread southwards and eastwards through the Indian subcontinent, based on the fact that the southern Dravidian languages show some signs of contact with linguistic groups which the northern Dravidian languages do not. Proto-Dravidian is thought to have differentiated into Proto-North Dravidian, Proto-Central Dravidian and Proto-South Dravidian around 1500 BCE, although some linguists have argued that the degree of differentiation between the sub-families points to an earlier split.

It was not until 1856 that Robert Caldwell published his *Comparative grammar of the Dravidian or South-Indian family of languages*, which considerably expanded the Dravidian umbrella and established it as one of the major language groups of the world. Caldwell coined the term "Dravidian" from the Sanskrit *drāvida*, related to the word ‘Tamil’ or ‘Tamilan’, which is seen in such forms as into ‘Dramila’, ‘Dramiˇa’, ‘Dramida’ and ‘Dravida’ which was used in a 7th-century text to refer to the languages of the southern India. The publication of the *Dravidian Etymological Dictionary* by T. Burrow and M. B. Emeneau was a landmark event in Dravidian linguistics.

Languages of other families in India

Sino-Tibetan languages

Main article: Sino-Tibetan languages

Sino-Tibetan languages are spoken in the western Himalayas (Himachal Pradesh) and in the highlands of Northeast India. The Sino-Tibetan family includes such languages as Meithei, Bodo, Naga and Garo. Some of the languages traditionally included in Sino-Tibetan may actually be language isolates or part of small independent language families.[*citation needed*]

Austroasiatic languages

Main article: Austroasiatic languages

The Austroasiatic family is thought to be the first to be spoken in ancient India. Austroasiatic languages include the Santal and Munda languages of eastern India, Nepal, and Bangladesh, and the Mon–Khmer languages spoken by the Khasi and Nicobarese in India and in Burma, Thailand, Laos, Cambodia, Vietnam, and southern China. The Austroasiatic languages are thought to have been spoken throughout the Indian subcontinent by hunter-gatherers who were later assimilated first by the agriculturalist Dravidian settlers and later by the Indo-Aryan peoples arriving from northwestern India.

Great Andamanese & Ongan languages

On the Andaman Islands, language from at least two families are spoken: the Great Andamanese languages and the Ongan languages. The Sentinelese language is spoken on North Sentinel Island but contact has not been made with the Sentinelis, so its language affiliation is unknown. While Joseph Greenberg considered the Great Andamanese languages to be part of a larger Indo-acific family, not established through the comparative method and considered spurious by historical linguists, Stephen Wurm suggests similarities with Trans-New Guinea languages and others are due to a linguistic substrate.[69] Juliette Blevins has suggested that the Ongan languages are the sister branch to the Austronesian languages in an Austronesian-Ongan family based on sound correspondences between protolanguages.[70]

Isolates

The Nihali language is a language isolate spoken in Madhya Pradesh and Maharashtra. Affiliations have been suggested to the Munda languages but they have yet to be demonstrated.

Evolution of scripts

Indus script

Spread of scripts in Asia.
Main article: Indus script

The Indus script is the short strings of symbols associated with the Harappan civilization of ancient India (most of the Indus sites are distributed in present-day Pakistan and northwest India) used between 2600–1900 BCE, which evolved from an early Indus script attested from around 3500–3300 BCE. Found in at least a dozen types of context, the symbols are most commonly associated with flat, rectangular stone tablets called seals. The first publication of a Harappan seal was a drawing by Alexander Cunningham in 1875. Since then, well over 4000 symbol-bearing objects have been discovered, some as far afield as Mesopotamia. After 1500 BCE, coinciding with the final stage of Harappan civilization, use of the symbols ends. There are over 400 distinct signs, but many are thought to be slight modifications or combinations of perhaps 200 'basic' signs. The symbols remain undeciphered (in spite of numerous attempts that did not find favour with the academic community), and some scholars classify them as proto-writing rather than writing proper.

Brāhmī script

Main article: Brāhmī script

The best known inscriptions in Brāhmī are the rock-cut edicts of Ashoka, dating to the 3rd century BCE. These were long considered the earliest examples of Brāhmī writing, but recent archaeological evidence in Sri Lanka and Tamil Nadu suggest the dates for the earliest use of Brāhmī to be around the 6th century BCE, dated using radiocarbon and thermoluminescence dating methods.

This script is ancestral to the Brahmic family of scripts, most of which are used in South and Southeast Asia, but which have wider historical use elsewhere, even as far as Mongolia and perhaps even Korea, according to one theory of the origin of Hangul. The Brāhmī numeral system is the ancestor of the Hindu-Arabic numerals, which are now used worldwide.

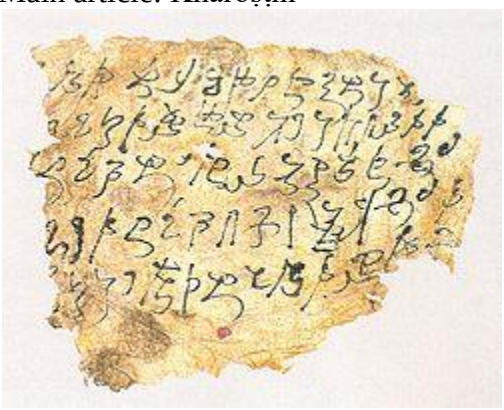
Brāhmī is generally believed to be derived from a Semitic script such as the Imperial Aramaic alphabet, as was clearly the case for the contemporary Kharosthi alphabet that arose in a part of northwest Indian under the control of the Achaemenid Empire. Rhys Davids suggests that writing may have been introduced to India from the Middle East by traders. Another possibility is with the Achaemenid conquest in the late 6th century BCE. It was often assumed that it was a planned invention under Ashoka as a prerequisite for the his edicts. Compare the much better documented parallel of the Hangul script.

Older examples of the Brahmi script appear to be on fragments of pottery from the trading town of Anuradhapura in Sri Lanka, which have been dated to the early 5th century BCE. Even earlier evidence of the Brahmi script has been discovered on pieces of pottery in Adichanallur, Tamil Nadu. Radio-carbon dating has established that they belonged to the 6th century BCE.[71]

A minority position holds that Brāhmī was a purely indigenous development, perhaps with the Indus script as its predecessor; these include the English scholars G.R. Hunter and F. Raymond Allchin.

Kharoṣṭhī script

Main article: Kharoṣṭhī



Paper strip with writing in Kharoṣṭhī. 2nd to 5th century CE, Yingpan, Eastern Tarim Basin, Xinjiang Museum.

The Kharoṣṭhī script, also known as the Gāndhārī script, is an ancient abugida (a kind of alphabetic script) used by the Gandhara culture of ancient northwest India to write the Gāndhārī and Sanskrit languages. It was in use from the 4th century BCE until it died out in its homeland around the 3rd century CE. It was also in use along the Silk Road where there is some evidence it may have survived until the 7th century CE in the remote way stations of Khotan and Niya.

Scholars are not in agreement as to whether the Kharoṣṭhī script evolved gradually, or was the work of a mindful inventor. An analysis of the script forms shows a clear dependency on the Aramaic alphabet but with extensive modifications to support the sounds found in Indian languages. One model is that the Aramaic script arrived with the Achaemenid conquest of the region in 500 BCE and evolved over the next 200+ years to reach its final form by the 3rd century BCE. However, no Aramaic documents of any kind have survived from this period. Also intermediate forms have yet been found to confirm this evolutionary model, and rock and coins inscriptions from the 3rd century BCE onward show a unified and mature form.

The study of the Kharoṣṭhī script was recently invigorated by the discovery of the Gandharan Buddhist Texts, a set of birch-bark manuscripts written in Kharoṣṭhī, discovered near the Afghan city of Haḍḍā (compare Panjabi HAḌḌ ਹੱਡ s. m. "A bone, especially a big bone of dead cattle" referring to the famous mortuary grounds if the area): just west of the Khyber Pass. The manuscripts were donated to the British Library in 1994. The entire set of manuscripts are dated to the 1st century CE making them the oldest Buddhist manuscripts in existence.

Gupta script

Main article: Gupta script

The Gupta script was used for writing Sanskrit and is associated with the Gupta Empire of India which was a period of material prosperity and great religious and scientific developments. The Gupta script was descended from Brahmi and gave rise to the Siddham script and then Bengali script.

Siddhaṃ script

Main article: Siddham

A replica of Uṣṇīṣa Vijaya Dhāraṇī Sūtra manuscript in Siddham on palm-leaf in 609 CE. Hōryū-ji, Japan. The last line is a complete Sanskrit syllabary in Siddhaṃ script

Siddhaṃ (Sanskrit, accomplished or perfected), descended from the Brahmi script via the Gupta script, which also gave rise to the Devanāgarī script as well as a number of other Asian scripts such as Tibetan script.

Siddhaṃ is an abugida or alphasyllabary rather than an alphabet because each character indicates a syllable. If no other mark occurs then the short 'a' is assumed. Diacritic marks indicate the other vowels, the pure nasal (anusvara), and the aspirated vowel (visarga). A special mark (virama), can be used to indicate that the letter stands alone with no vowel which sometimes happens at the end of Sanskrit words. See links below for examples.

The writing of mantras and copying of Sutras using the Siddham script is still practiced in Shingon Buddhism in Japan but has died out in other places. It was Kūkai who introduced the Siddham script to Japan when he returned from China in 806, where he studied Sanskrit with Nalanda trained monks including one known as Prajñā. Sutras that were taken to China from India were written in a variety of scripts, but Siddham was one of the most important. By the time Kūkai learned this script the trading and pilgrimage routes over land to India, part of the Silk Road, were closed by the expanding Islamic empire of the Abbasids. Then in the middle of the 9th century there were a series of purges of "foreign religions" in China. This meant that Japan was cut off from the sources of Siddham texts. In time other scripts, particularly Devanagari replaced it in India, and so Japan was left as the only place where Siddham was preserved, although it was, and is only used for writing mantras and copying sutras.

Siddhaṃ was influential in the development of the Kana writing system, which is also associated with Kūkai – while the Kana shapes derive from Chinese characters, the princlple of a syllable-based script and their systematic ordering was taken over from Siddham.