



Development Team

Principal Investigator:**Prof. Pramod Pandey**Centre for Linguistics / SLL&CS
Jawaharlal Nehru University, New DelhiEmail: pkspandey@yahoo.com**Paper Coordinator:****Prof. K. S. Nagaraja**Department of Linguistics, Deccan College Post-Graduate
Research Institute, Pune- 411006,
ksnagaraja2006@gmail.com**Content Writer:****Prof. K. S. Nagaraja****Content Reviewer:****Prof H. S. Ananthanarayana**Retd Prof, Department of Linguistics
Osmania University, Hyderabad 500007

Description of Module	
Subject Name	Linguistics
Paper Name	Historical and Comparative Linguistics
Module Title	Indo-Aryan Language Family
Module ID	Lings_P7_M1
Quadrant 1	E-Text

 **Pathshala**
पाठशाला
A Gateway to All Post Graduate Courses

Linguistics

Paper : Historical and Comparative Linguistics

Module : Indo-Aryan Language Family

INDO-ARYAN LANGUAGE FAMILY

The Indo-Aryan migration theory proposes that the Indo-Aryans migrated from the Central Asian steppes into South Asia during the early part of the 2nd millennium BCE, bringing with them the Indo-Aryan languages. Migration by an Indo-European people was first hypothesized in the late 18th century, following the discovery of the Indo-European language family, when similarities between Western and Indian languages had been noted. Given these similarities, a single source or origin was proposed, which was diffused by migrations from some original homeland. This linguistic argument is supported by archaeological and anthropological research. Genetic research reveals that those migrations form part of a complex genetical puzzle on the origin and spread of the various components of the Indian population. Literary research reveals similarities between various, geographically distinct, Indo-Aryan historical cultures.

The Indo-Aryan migrations started in approximately 1800 BCE, after the invention of the war chariot, and also brought Indo-Aryan languages into the Levant and possibly Inner Asia. It was part of the diffusion of Indo-European languages from the proto-Indo-European homeland at the Pontic steppe, a large area of grasslands in far Eastern Europe, which started in the 5th to 4th millennia BCE, and the Indo-European migrations out of the Eurasian steppes, which started approximately in 2000 BCE.

The theory posits that these Indo-Aryan speaking people may have been a genetically diverse group of people who were united by shared cultural norms and language, referred to as *ārya* "noble." Diffusion of this culture and language took place by patron-client systems, which allowed for the absorption and acculturation of other groups into this culture, and explains the strong influence on other cultures with which it interacted.

The Proto-Indo-Iranians, from which the Indo-Aryans developed, are identified with the Sintashta culture (2100–1800 BCE), and the Andronovo culture, which flourished ca. 1800–1400 BCE in the steppes around the Aral sea, present-day Kazakhstan, Uzbekistan and Turkmenistan. The proto-Indo-Iranians were influenced by the Bactria-Margiana Culture, south of the Andronovo culture, from which they borrowed their distinctive religious beliefs and practices. The Indo-Aryans split off around 1800-1600 BCE from the Iranians, whereafter the Indo-Aryans migrated into the Levant and north-western India.

The **Indo-European languages** are a family of several hundred related languages and dialects. There are about 445 living Indo-European languages, according to an estimate; with over two-thirds (313) of them belonging to the Indo-Iranian branch. The most widely spoken Indo-European languages by native speakers are Spanish, English, Hindi, Portuguese, Bengali, Russian, Persian and Punjabi, each

with over 100 million speakers. Today, 46% of the human population speaks an Indo-European language, by far the most of any language family. The Indo-European family includes most of the modern languages of Europe, and parts of Western, Central and South Asia. It was also predominant in ancient Anatolia (present-day Turkey), and the ancient Tarim Basin (present-day Northwest China) and most of Central Asia until the medieval Turkic migrations and Mongol invasions. With written evidence appearing since the Bronze Age in the form of the Anatolian languages and Mycenaean Greek, the Indo-European family is significant to the field of historical linguistics as possessing the second-longest recorded history, after the Afroasiatic family.

All Indo-European languages are descendants of a single prehistoric language, reconstructed as *Proto-Indo-European*, spoken sometime in the Neolithic era. Although no written records are extant, much of the culture and religion of the Proto-Indo-European people can also be reconstructed based on the cultures of ancient and modern Indo-European speakers who continue to live in areas to where the Proto-Indo-Europeans migrated from their original homeland. Several disputed proposals link Indo-European to other major language families.

The Indo-Iranian languages or Aryan languages, constitute the largest and easternmost extant branch of the Indo-European language family. It has more than 1 billion speakers, stretching from the Caucasus (Ossetian) and the Balkans (Romani) eastward to Xinjiang (Sarikoli) and Assam (Assamese), and south to the Maldives (Maldivian).

The common ancestor of all the languages in this family is called *Proto-Indo-Iranian*—also known as Common Aryan—which was spoken in approximately the late 3rd millennium BC. The three branches of modern Indo-Iranian languages are *Indo-Aryan*, *Iranian*, and *Nuristani*. Additionally, sometimes a fourth independent branch, *Dardic*, is posited, but recent scholarship in general places Dardic languages as archaic members of the Indo-Aryan branch.

The Indo-Iranian languages derive from a reconstructed common proto-language, called *Proto-Indo-Iranian*.

The oldest attested Indo-Iranian languages are *Vedic Sanskrit* (ancient Indo-Aryan), Older and Younger *Avestan* and *Old Persian* (ancient Iranian languages). A few words from another Indo-Aryan language (*Mitanni*) are attested in documents from the ancient *Mitanni* kingdom in northern Mesopotamia and Syria and the *Hittite* kingdom in Anatolia.

An illustration is provided here to show the closeness of Avestan (the oldest form of Iranian branch) and Vedic Sanskrit (the oldest form of Aryan branch). The closeness of Avesta and Old

Sanskrit is remarkably clear though they are quite distinct individually. Also, the differences are clearly explainable.:

Avesta: (i) tām amavantām yazatām

(2) sūrām dāmōhu səvištām

(3) miθrām yazai zaoθrdbiyō

(Jackson: Avesta Grammar)

With certain phonetic changes it is parallel to Vedic version:

(1) tām amavantām yajatām

that who is strong he sacrifices

(2) sūrām dhāmasu saviṣṭhām

wiseman abode-to the entered

(3) mitram yaje hotrābhyah

Mitra/son –to (I) offer sacrifice-item from the ladle.

Indo-Aryan family: This is a sub-branch of the larger group, Indo-Iranian, and which itself is a member of Indo-European family. The northern Indian languages evolved from Old Indic by way of the Middle Indic Prakrit languages and Apabhramśa of Middle Ages.

Indo-Aryan languages occupy the largest geographical area in the sub-continent. It stretches from Pakistan in the west to Bangladesh in the east, from Nepal in the north to Sri Lanka in the south; whereas the Iranian branch mainly occupies Iran and Afghanistan areas. The main languages of this family are listed below.

The Indo-Aryan or Indic languages are the dominant language family of the Indian subcontinent and South Asia, spoken largely by Indo-Aryan people. Indo-Aryan speakers form about one half of all Indo-European speakers (about 1.5 of 3 billion), and more than half of all Indo-European languages. This is the largest of the language families represented in India, in terms of speakers, itself the easternmost, extant subfamily of the Indo-European language family. This language family predominates, accounting for some 790 million speakers, or over 75% of the population, as per data collated during the Census of 2001. Modern north Indian languages, such as Hindi, Assamese (Asamiya), Bengali,

Gujarati, Marathi, Punjabi, Rajasthani and Odia, evolved into distinct, recognisable languages in the New Indo-Aryan Age. The most widely spoken languages of this group are Hindi, Bengali, Marathi, , Gujarati, Punjabi, Assamese, Sinhalese in Sri Lanka 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*. While the languages are primarily spoken in South Asia, pockets of Indo-Aryan languages are found to be spoken in Europe and the Middle East.

Persian or Pharsi was brought into India by the Ghaznavi and other Turko-Afghan dynasties as the court language. Persians influenced the art, history and literature of the region for more than 500 years, resulting in the Persianisation of many Indian tongues, mainly lexically. In 1837, the British replaced Persian with English for administrative purposes, and the Hindi movement of the 19th Century replaced the Persianised vocabulary for one derived from Sanskrit also replacing the use of the Perso-Arabic script for Hindi/Hindustani with Devanaagari.

Each of the northern Indian languages had different influences. For example, Hindustani was strongly influenced by Sanskrit, Persian, and Arabic, leading to the emergence of Modern Standard Hindi and Modern Standard Urdu as registers of the Hindustani language. Modern Standard Hindi is recognised as the official language of India while Urdu is a scheduled language.

Distribution of major IA languages (native speakers) given below:

Language	Geographical location	Population, Census 2001
Hindi/ <u>Hindustani</u> (Hindi-Urdu, about 250 million),	UP, MP, Rajasthan, Haryana	422,048,642
Bengali	WB, Bangladesh	83,369,769
Punjabi	Punjab, and other places	29,102,477
Marathi	Maharashtra	71,936,894
Gujarati	Gujarat	46,091,617
Konkani	Goa	2,420,140
Odia	Orissa	33,017,446
Saraiki (<u>Doabi</u> , <u>Lahnda</u>)	Kashmir	20 million
Bhojpuri	Bihar	33099497
Rajastani	Rajasthan	50 million
Sindhi	Delhi, Maharashtra	2,535,485
Nepali	Nepal	1,835,485
Assamese	Assam	1,3168,484

Sinhalese	Sri Lanka	15 million
Maithili	Bihar, WB	12,178,673
<u>Bhili/Bhilodi</u>	Rajasthan, MP, Gujarat, Maharashtra	9,582,957
Kashmiri	Jammu & Kashmir	5,362,349
Tharu	Nepal, India	1.53 M. In Nepal; 400,000-India
Harauti	Hadoti region of Rajasthan	2,462,867
Dogri	Jammu, Himachal Pradesh, Punjab	2,282,547
Garhwali	Uttaranchal	2,267,314
Nimadi	Nimar in MP	2,148,146
Sadri/Nagpuria/Jhaari	Chota Nagpur belt	20,44,776
Kumaoni	Kumaon, Uttarakhand, Dharchula dt.	2003786
Dhundhari	Dhundhar region of Rajasthan	1871130
Ahirani, Khandeshi	Khandesh, Maharashtra	1865813
Maldivian	Maldives, Minicoy Island	3,40,000

List of languages by number of native speakers in India



Linguists generally recognize *three major divisions/stages* of Indo-Aryan languages: Old, Middle, and New (or Modern) Indo-Aryan. These divisions are primarily linguistic and are named in the order in which they initially appeared, with later divisions coexisting with rather than completely replacing earlier ones.

- I. Old Indo-Aryan (ca. 1500–300 BCE)
 - (a) early Old Indo-Aryan: Vedic Sanskrit (1500 to 500 BCE)
 - (b) late Old Indo-Aryan: Epic Sanskrit, Classical Sanskrit (500 to 300 BCE)
- II. Middle Indo-Aryan or Prakrits, Old Odia? (ca. 300 BCE to 1500 CE)
- III. Early Modern Indo-Aryan (Mughal period, 1500 to 1800)

(a) early Dakkhini (*Kalmitul-hakayat* 1580)

(b) emergence of *Khariboli* (*Gora-badal ki katha*, 1620s)

Old Indo-Aryan: Consists of huge literature from the earliest times, which is represented by Vedic literature.

Old Indo-Aryan

The earliest evidence of the group is from Vedic and Mitanni-Aryan. Vedic has been used in the ancient preserved religious hymns, the foundational canon of Hinduism known as the Vedas. Mitanni-Aryan is of similar age to the language of the Rgveda, but the only evidence of it is a few proper names and specialized loanwords. Vedic Sanskrit is only marginally different from, Proto-Indo-Aryan the proto-language of the Indo-Aryan languages.

In about the 4th century BCE, an artificial language based on Vedic, called "Classical Sanskrit" by convention, was codified and standardized by the grammarian Panini.

The knowledge embodied in Sanskrit texts can be classified into two groups, viz. Vedic texts and Classical texts. In the former we have the four major works of Veda, Rgveda, Yajurveda, Samaveda and Atharvaveda. The first one is purely in verse form; while the second one contains both verse and prose texts; while the third one contains prayers, etc.

Middle Indo-Aryan (Prakrits)

Middle Indo-Aryan (Prakrits) Outside the learned sphere of Sanskrit, vernacular dialects (Prakrits) continued to evolve. The oldest attested Prakrits are the Buddhist and Jain canonical languages Pali and Ardha Magadhi, respectively. By medieval times, the Prakrits had diversified into various Middle Indo-Aryan dialects. "Apabhramsa" is the conventional cover term for transitional dialects connecting late Middle Indo-Aryan with early Modern Indo-Aryan, spanning roughly the 6th to 13th centuries. Some of these dialects showed considerable literary production; the Sravakachar of Devasena (dated to the 930s) is now considered to be the first Hindi book.

The next major milestone occurred with the Muslim conquests on the Indian subcontinent in the 13th–16th centuries. Under the flourishing Turco-Mongol Mughal empire, Persian became very influential as the language of prestige of the Islamic courts due to adoption of the foreign language by the Mughal emperors. However, Persian was soon displaced by Hindustani. This Indo-Aryan language is a combination with of Persian, Arabic, and Turkic elements in its vocabulary, with the grammar of the local dialects.

The two largest languages that formed from Apabhramsa were Bengali and Hindustani; others include Gujarati, Odia, Marathi, and Punjabi. Assamese, Konkani, Nepali, Romani, Sinhalese...

New Indo-Aryan

Dialect continuum

The Indo-Aryan languages of Northern India and Pakistan form a dialect continuum. What is called "Hindi" in India is frequently Standard Hindi, the Sanskrit-ized version of the colloquial Hindustani spoken in the Delhi area since the Mughals. However, the term Hindi is also used for most of the central Indic dialects from Bihar to Rajasthan.

The Indo-Aryan Prakrits also gave rise to languages like Gujarati, Assamese, Bengali, Odia, Nepali, Marathi, and Punjabi, which are not considered to be part of this dialect continuum.

Proto-Indo-Aryan, or sometimes Proto-Indic, is the reconstructed proto-language of the Indo-Aryan languages. It is intended to reconstruct the language of the Proto-Indo-Aryans. Proto-Indo-Aryan is meant to be the predecessor of Old Indo-Aryan (1500–300 BCE) which is directly attested as Vedic and Mitanni-Aryan. Despite the great archaicity of Vedic, however, the other Indo-Aryan languages preserve a small number of archaic features lost in Vedic.

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 Rgveda to be the earliest. The hymns preserved in the Rgveda 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.

The end of the Vedic period is marked by the coſmposition 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.

The oldest surviving Sanskrit grammar is Pāṇini's Aṣṭādhyāyī ("Eight-Chapter Grammar") dating to c. the 5th century BCE. It is essentially a Descriptive grammar, i.e., it describes the usage of people as the adage goes – prayogaśaraṇāḥ vaiyākaraṇāḥ), although it contains descriptive parts, 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 mark 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 "Pāṇinian" Sanskrit, while broadly similar, are separate varieties, which differ in a number of points of phonology, vocabulary, and grammar.

Prakrit (Sanskrit *prākṛta* प्राकृत (from *prakṛti* प्रकृति), "original, natural, artless, normal, ordinary, usual", i.e. "vernacular", in contrast to *samskrta* "excellently made", both adjectives elliptically referring to *vaak* "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 – Śauraseni, Maagadhi, Maahaaraashtri, 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., Paiśaaci).

Pāli language

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 Maagadhi, but modern scholars believe this to be unlikely. 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 Kharosthi script survive including a version of the Dhammapada.

Middle Indo-Aryan and Apabhramśa

The Prakrits (which includes Pali) were gradually transformed into Apabhramśas (अपभ्रंश) which were used until about the 13th century CE. The term **apabhramśa** 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 apabhramśa 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 Apabhramśa. Apabhramśa authors include Sarahapad of Kamarupa, Devasena of Dharand 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 Apabhramśa is in Vikramorvashīya of Kalidasa, when Pururava asks the animals in the forest about his beloved who had disappeared.