
South Asian Language Study Documentation

Allen Lee (Zixi Li)

May 01, 2019

| | | |
|----------|---------------------------|-----------|
| 1 | Indices and tables | 1 |
| 2 | Project Idea | 13 |
| 3 | Instructor's note | 15 |
| | Bibliography | 17 |

1.1 Introduction

1.1.1 Project Goal

South Asia is an area with a rich historical background. During its thousands of years of history, different cultures take root in this area and left their footprint here. For example, there're 22 official languages listed in Indian Constitution alone. Citizens may only speak one or two of them, but all 22 have numerous speakers across the country, not to mention minor languages.

Different characteristics of languages encode different aspects of life. From the language, we can get a brief idea of their social life. The basis of every language is that they must satisfy the need of daily communication. If some concept is being frequently or widely used, then there must be a word for it. For example, there's a word in Brazillian Portuguese: "Cafune". It means, "caress the hair of the one you love", and Sanskrit have lots of words describing woman appearance.

Languages also play a role in social classification. In most languages, different people should speak differently. Take Japanese as an example, some sentence structures are oriented only for man or women. You also need to reform the verbs when you're talking to people older than you. These characteristics reflected strong attention on gender division and respect for seniority.

In this project, I will discover different languages in South Asia and their specialties. I will focus on when and where they were spoken, what language family they belong to, what's special about each language, and so on. For convenience, I will arrange these languages according to the areas they were being spoken in.

1.1.2 Project Scope

Since the number of languages in South Asia is vast, only those with a large number of speakers, or those with significant context will be discussed.

1.2 Indo-European Family

Indo-European languages are mostly spoken in the northern part of India. All those languages belong to a subfamily of Indo-European, the Indo-Aryan branch. It includes the following languages:

1.2.1 Hindi

- *Dialects [Mas93]*
- *Devanagari Script*
- *Literature:cite:keay1920hist*
- *Early Bardic Chronicles (1150-1400)*
 - *Chand Bardai*
 - *Jagnayak*
 - *Sarang Dhar*
- *Early Bhakti Poets (1400-1550)*
- *Tulsi Das and the Rama Cult (1550-1800)*
- *The Modern Period (1800 -)*

Hindi is the most common language in India, as shown in [Languages of India](#). It is also referred to as “Language of India”. Despite that fact, it is not the national language of India - being strongly opposed by the people that do not speak Hindi, the Indian constitution did not list any language as the national language. It is also spoken in other countries, like Fiji. However, there is an ongoing plan of promoting the progressive use of Hindi, and reducing the use of English in official purposes. Yet due to opposing opinions, no restriction is imposed on using English. [Con65]

Dialects [masica1993indo]

There exist regional languages of the Hindi area, or “Hindi dialects”. Some of them are closely related to standard Hindi, some more distant. Vernacular Hindi could be roughly classified as *Eastern Hindi* and *Western Hindi*. *Western Hindi* consists of *Braj*, *Kannauji* and *Bangaru*, while *Eastern Hindi* consists of *Awadhi*, *Bagheli*, and *Chhattisgarhi*. Here’s a list of regions these languages are spoken:

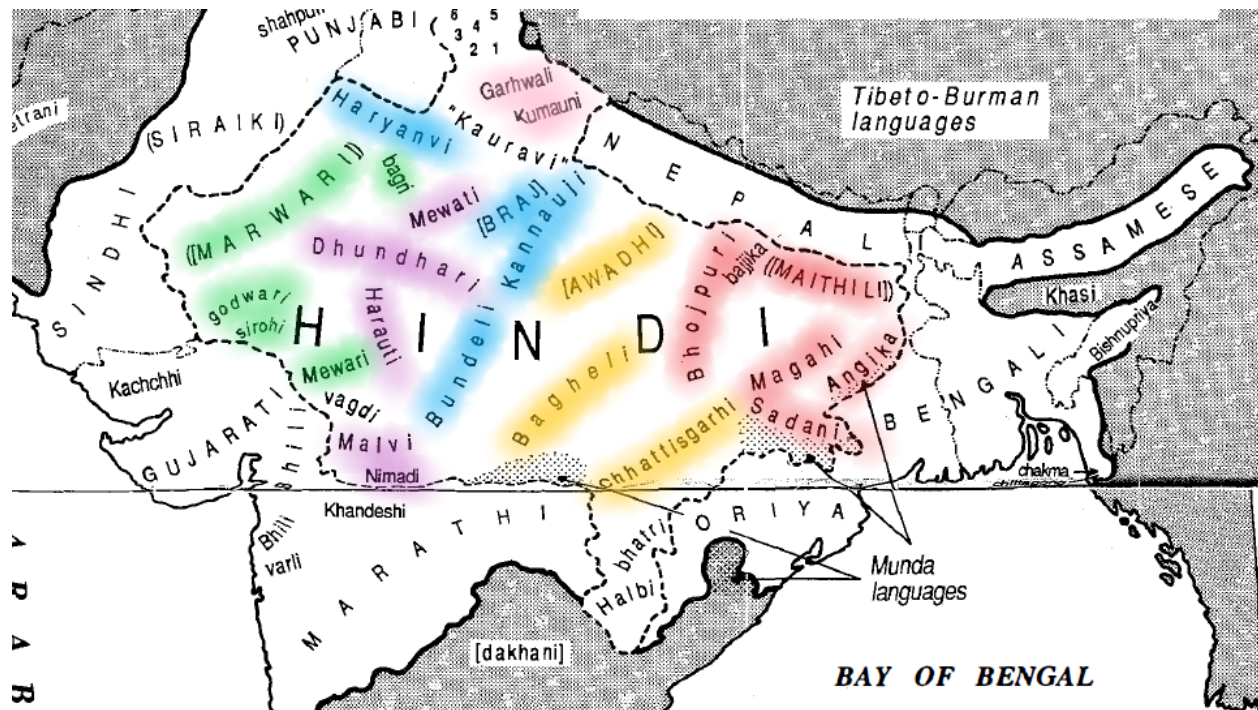
Western Hindi:

- *Braj*: An area centering Mathura, southeast of Delhi, southwest to Bareilly.
- *Bundeli*: Closely related to *Braj*.
- *Kannauji*: From Etawah and Kanpur up to Pilibhit
- *Bangaru**(also called **Haryanvi*): Haryana State and rural parts of Delhi

Eastern Hindi

- *Awadhi*: East-central Uttar Pradesh, north and south of Lucknow.
- *Bagheli*: A variety of *Awadhi*, Madhya Pradesh from Rewa to Jabalpur and Mandla
- *Chhattisgarhi*: Southeast on the border of Orissa

Here's a map from *The Indo-Aryan Languages*. Each language is marked with color, by the group of language they belong to. Blue parts are Western Hindi, and yellow parts are eastern Hindi. Other colored languages are not Hindi, but mentioned elsewhere and share the map.



Devanagari Script

Hindi is written in Devanagari script, a very complicated form of writing. Unlike most languages that have distinct characters, Devanagari symbols represent phonetic elements and are often written in compounds. There're lots of variations for one single sound when diacritic is added.

Devanagari script is a branch of Brahmi Script. More on Devanagari Script and its history, relate to [Brahmi Script](#).

Literature:cite:keay1920hist

Literature in Hindi can be divided into several periods:

Early Bardic Chronicles (1150-1400)

These are the earliest poetry written in Hindi. Developed during the conflict between Rajput clans and invading Muhammadan powers, these poets sing the bravery of soldiers. These bards must have been using local language *Prakrit* in the beginning, but then developed their modern vernacular.

Chand Bardai

Composer of *Prithviraj Raso*, in 69 books and over 100,000 verses. Consists of the history of that time and life of his patron, mixed with legends and fictions. Therefore some of its content contradict with other works of the same period. The language used in this work is a transitional form, in which some strange expressions have been obsolete for a long time. He himself was born in 1159 and died in 1192. He was the ruler of Ajmer and Delhi. He was captured during the second battle of Tarain then slain.

Jagnayak

None of his works survived, except some verses in Mahoba Khand. This poem has been handed down only by oral tradition, therefore it existed in many recensions, differing in language and subject matter. He was a contemporary of Chand Bardai.

Sarang Dhar

Author of *Hammir Rasa* and *Hammir Kavya*, chronicles of the royal house of Ranthambhor. It's a famous story about Hammir fighting against the emperor Alauddin. He is said to be the descendant of Chand Bardai.

Early *Bhakti* Poets (1400-1550)

Note: This is exactly the devotional poets we read in the lectures.

Tulsi Das and the Rama Cult (1550-1800)

Tulsi Das is the author of Ramayana, the legendary literature that is well-known around the world. Originally named Rambola, he took the name of Tulsi Das after becoming a devotee. He lived a normal, insignificant life apart from the legend. Many legends are told about him, but none of them is reliable. *Nabha Das*, the author of the *Bhaktamaldy* is said to have been his friend, and *Sur Das* is also supposed to have visited him. The language he used was Awadhi, a dialect that belongs to Eastern Hindi. However, he also incorporated lots of words from other dialects, especially *Braj*.

The Modern Period (1800 -)

At the beginning of the 18th century, India got influenced by the western diaspora. The East India Company took over the government of vast areas of India. The introduction of printing facilities (presses) helped the distribution of literature, and technologies changed the life of local people in many ways. The Company encouraged vernacular Hindi literature.

His real name is Dhanpat Rai Shrivastava. He is one of the most celebrated writers of India. Before him, Hindi literature was mostly about fairy tales, epics, entertainment and religion. He is the pioneer of bringing realism to Hindi prose literature.

1.2.2 Urdu

Urdu is one of the 22 official languages in India, and the official national language of Pakistan. It is also a registered regional language of Nepal.

Script

Urdu is written in [Perso-Arabic](#) Script. More on that in the linked page.

Relation with Hindi

Urdu is mutually intelligible with Standard Hindi, which means the speaker of either can easily understand another language. They're structurally the same language. Hindi is written in Devanagari script, while Urdu is written in Persian Alphabet. Both Urdu and Hindi have a large number of Sanskrit, Arabic and Persian words. Hindi have more special words from Sanskrit, and Urdu have more special words from Persian. The separation of this two languages are mostly religious and political reasons, after India and Pakistan separate from each other.

Urdu Literature

Urdu Literature started quite late, with the earliest found around 14th century, and flourishing around 16th century. However, no real prose literature developed before the 19th century. Urdu literatures lean heavily towards poetry and short stories. Long prosas are restricted to an ancient form called Dastan, which dealt with magic and superficial events, and they're called Dastaangois.

Ghazal

Ghazal is a type of poetry, which is well developed in Urdu. It have a fixed format:

- It consists of rhyming couplets, called *Sher* or *Bayt*.
- A Ghazal have at least 5 *sher*s. Most have 7 to 12 *sher*s. Almost all of them have less than 15 *sher*s.
- Couplets end with the same rhyming pattern and have the same meter.

The most strict form must follow five rules, listed below. The uniqueness of Ghazals come from *qaafiyaa*(rhyme) and *radif*(refrain).

- Matlaa: The first *sher*. Set the tone of the ghazal, and determine its rhyming and refrain pattern.
- Radif: Refrain. Both lines of the matlaa and the second lines of all *sher*s must end in the same rhyme.
- Qaafiyaa. Following the radif, words with the same end rhyme pattern.
- Maqtaa. Last couplet of the ghazal. The most personal part, reflects creativity.
- Beher. Each line of a ghazal must have the same metrical pattern (same number of words).

1.2.3 Bihari

Bihari is a group of Eastern Indo-Aryan languages. Here's a list of languages, and their areas being spoken:

- *Magahi*: central Bihar
- *Maithili*: north of the Ganga. This language have a long literary tradition.
- *Sadani*: South Bihar centering on Ranchi
- *Angika*: eastern Bihar
- *Bajjika*: Muzaffarpur and part of Champaran Districts in northwest Bihar. and Nepal
- *Bhojpuri*: Spoken in Assam state; Bihar state: Champaran, Saran, and Shahabad districts; Delhi; Jharkhand state: Palamau and Ranchi districts; Madhya Pradesh state; Uttar Pradesh state: Azamgarh, Ballia, Basti, Deoria, Ghazipur, Gorakhpur, Mirzapur, and Varanasi districts; West Bengal state. [Eeds19]
- etc. (Kudmali, Majhi, Musasa, Panchpargania, Nagpuri, Surjapuri)

Among these languages, only Maithili is constitutionally recognised in India in 2003. Both Maithili and Bhojpuri is recognised in Nepal.

1.2.4 General Indo-European

Indo-European is the language family with the most speakers. [Cam08] It is also the most popular language family in India, with 78.05% of its population speaking some kind of Indo-European language [Bri17], and 19.65% speaking Dravidian language.

1.2.5 The Indo-Aryan branch

Indo-Aryan languages are mostly related to Iranian languages to the west, and are collectively called *Indo-Iranian* family, a major branch of *Indo-European* Family. They have similar features to Greek, but are more likely to be common innovations, rather than preserved features from the Indo-European dispersal.

History

Indo-Aryan and Iranian speakers split no later than 2000 B.C. Rather than the geographical distribution nowadays, it was a north/south division rather than a east/west one. The southern branch was Proto-Aryans, moving south via Central Asia, reaching south Asia. They entered northwest India through Afghanistan and Bactria. The Bactria part is then occupied by Proto-Iranians.

1.2.6 References

1.3 Languages of India

India has as much as 22 official languages, each with a decent amount of speakers (except Sanskrit). Since the number is many, I'll only focus on some of them. Here's the list of languages covered:

Below are statistics of all languages spoken in India, and some comments on them.

1.3.1 Speakers of 22 official languages

From this piechart, we can see that India does not have a "central" language. Hindi has a dominance, but not by a large margin. That's also why Indian government documents have to be distributed in several languages, and we often see multilingual traffic signs. One language to notice is Sanskrit. Modern Sanskrit is almost a dead language [Nel17]. It is an old language that carried lots of great literature, but few people use it as the first language. Although Sanskrit has few speakers nowadays (so few that it is shown as 0.00% in the piechart), Hindi is actually a register of Sanskrit. Lots of its vocabulary came from Sanskrit.

Also from this chart, we can see that there's little difference between the ratio of languages used by men or women. Traditionally Sanskrit are only spoken by noble men, and women must speak dialects no matter how prestigious they're. But in the census data from 2011, the number of men Sanskrit speaker is 8,189, while women have 5,946 Sanskrit speakers. Other languages show a similar indifference of speaker by gender. This is a demonstration of ongoing equivalence between men and women.

1.4 Brahmi Scripts

There're dozens of forms of writing in the Indian area, and many of them have a common root to the Brahmi script, which could be dated back to the third century B.C.

- *History*
- *Evolution and Diversification*
- *Devanagari*

1.4.1 History

The origin of Brahmi script is controversial. Most Western Scholars consider it a derivation of North Semitic, while a few think it's from South Semitic or a progenitor of both. However, both would agree that this script got again refined and improved a lot in India, especially on phonetics which contributed a lot to the rhyme of the verses.

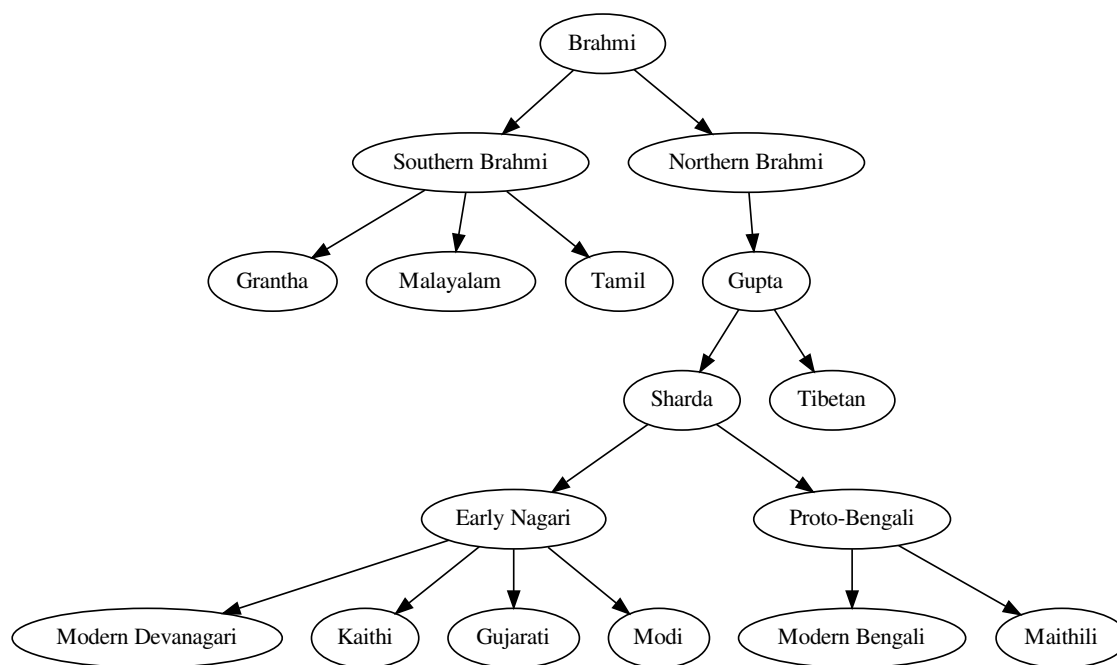
At the beginning, Brahmi was being used for Prakrit, then being used by Sanskrit about four centuries later. At first, it is only being used for administrative, literary and scientific purposes. This also helped in settling those oral sacred texts on writing.

1.4.2 Evolution and Diversification

Brahmi went through some serious diversification in India, eventually turning into several mutually intelligible forms. This diversification is notable because no other scripts undergo such dramatic change during the same time period.

Such a dramatic change might be related to politics. The political unity India attained under the Mauryas did not approach again for a whole nineteen hundred years. Therefore, no one centralized writing word or one sacred written text was maintained. The Brahmins themselves are also regionally divided. When they maintain the Sanskrit language, they wrote it in various regional scripts.

Its detailed evolution path is very complicated. So here's a simplified tree structure, containing several branches we care about most. Note that this diagram is largely simplified. Some intermediate scripts are neglected. Shown as direct ancestor in this graph does not mean they're direct ancestors: there might be several generations between them.



Here's a graph comparing Brahmi with several modern scripts derivated from it.

| | <i>Brahmi</i> | <i>Gurmukhi</i> (= <i>Punjabi</i>) | <i>Takri</i> (<i>Chameali</i>) | <i>Gujarati</i> | <i>Devanagari</i> | <i>Bengali</i> (+ <i>Assamese</i>) | <i>Oriya</i> | <i>Sinhalese</i> |
|----|---------------|--|-------------------------------------|-----------------|-------------------|--|--------------|------------------|
| A | 𑀅 | ਅ | 𑀓 | અ | अ ऌ | অ | ଅ | අ |
| Ā | 𑀆 | ਆ | 𑀔 | આ | आ ऌ | আ | ଆ | ආ |
| I | 𑀇 | ਇ | 𑀕 | ઇ | इ | ই | ଈ | ඈ |
| Ī | | ਈ | 𑀖 | ઈ | ई | ঈ | ଈ | ඉ |
| U | 𑀈 | ਉ | 𑀗 | ઉ | उ | উ | ଉ | උ |
| Ū | | ਊ | 𑀘 | ઊ | ऊ | ঊ | ଊ | ඌ |
| R | | | | ૨ | ऋ ॠ | ঋ | ଋ | ඬ |
| E | 𑀉 | ਏ | 𑀙 | એ | ए | এ | ଏ | ඒ |
| AI | | ਐ | 𑀚 | ઐ | ऐ | ঐ | ଐ | ඹ |
| O | 𑀊 | ਓ | 𑀛 | ઓ | औ ओ | ও | ଓ | ඔ |
| AU | | ਔ | 𑀜 | ઔ | औ औ | ঔ | ଔ | ඵ |
| * | | | | | | | | |
| KA | 𑀋 | ਕ | 𑀝 | ક | क | ক | କ | ක |

Fig. 1: Excerpt of symbol table for several scripts using Brahmi-derived scripts. Adopted from [Mas93]

It is obvious that the evolved scripts looked nothing like the original Brahmi, and they're also very different from each other. They have been evolving individually for a long time, leading to a very different use of drawing elements.

Also, all the derived scripts have added symbols to represent additional diacritics or new sounds compared to Brahmi. Something not represented in this extracted page is that the derived scripts also have a difference in symbols. Some scripts have symbols that do not exist in other scripts.

1.4.3 Devanagari

Devanagari is a composition of “deva” and “nagari”. “Nagari” means “city” or “metropolitan”, which is the original name of the language. “deva” means “divine”. There’s also another branch of Nagari, called Nandinagari. The Devanagari is being used for not only Hindi, but also Sanskrit, Marathi, Nepali, Maithili, Sindhi, etc.

This picture shows how Devanagari for Sanskrit evolved from Brahmi over time.

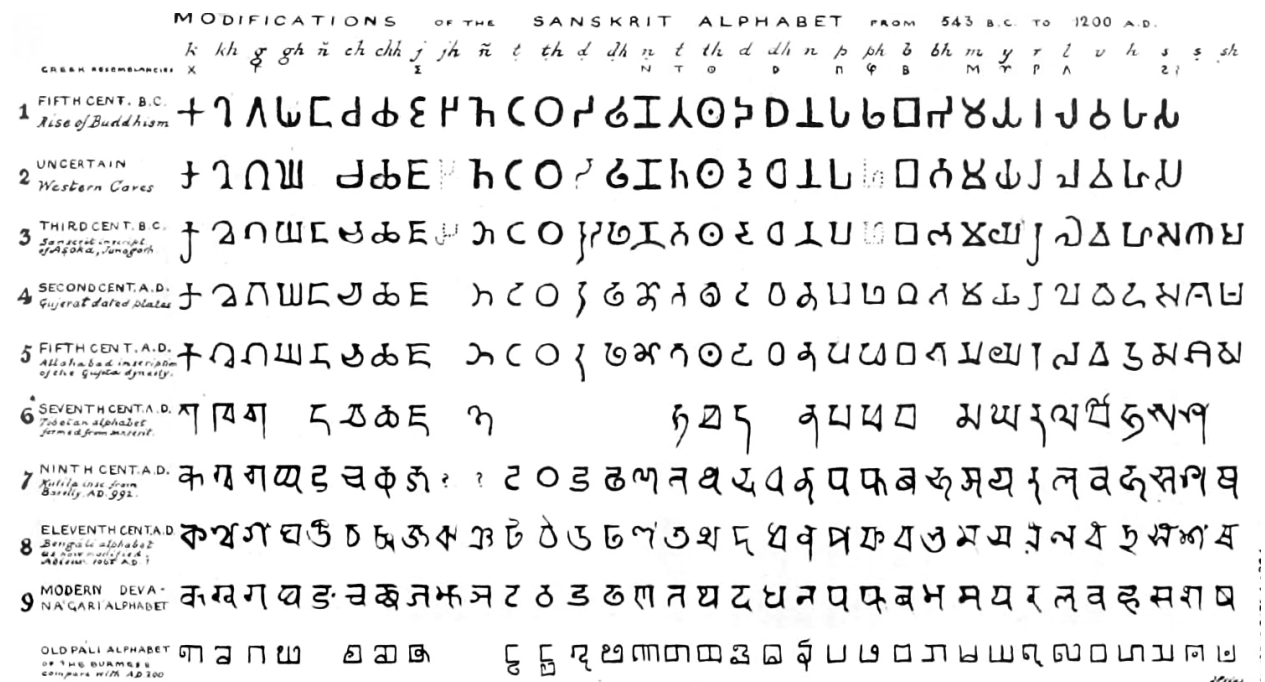


Fig. 2: Evolution of Brahmi into Devanagari. By James Prinsep, 1838. [Pri38]

1.5 Perso-Arabic Scripts

Perso-Arabic scripts are not designed for Indo-Aryan languages. To fit Indo-Aryan languages, diacritics is added to the script. Its evolution to fit the modern languages a much simpler and easily done than with [Brahmi](#). The redundant symbols for Indo-Aryan family are preserved to represent foreign lexicon from Arabic.

Urdu and Sindhi employ this script, with only minor differences on several elements.

There’s also the Maldivian script, called Tana or Thanna, that looks similar to Arabic script. It employed certain Arabic diacritics and numerals, and it’s also written from right to left. However, this script is a completely original invention, not a modification of any Arabic script.

1.6 Referenced Articles

Here is a list of articles that were referenced in my project, and was originally open to the public for free download.

Table 6.3 *NIA Consonants in Perso-Arabic script: Urdu vs. Sindhi*

| | Urdu | Sindhi | | Urdu | Sindhi |
|-----------|------|--------|-----------|------|--------|
| <i>p</i> | پ | پ | <i>jh</i> | جھ | جھ |
| <i>t</i> | ت | ت | <i>gh</i> | گھ | گھ |
| <i>ʈ</i> | ٹ | ٹ | | | |
| <i>č</i> | چ | چ | <i>b</i> | | ب |
| <i>k</i> | ک | ڪ | <i>ɖ</i> | | ڌ |
| | | | <i>f</i> | | ف |
| <i>b</i> | ب | ب | <i>ɡ</i> | | گپ |
| <i>d</i> | د | د | <i>m</i> | م | م |
| <i>ɖ</i> | ڌ | ڊ | <i>n</i> | ن | ن |
| <i>j</i> | ج | ج | <i>ɳ</i> | | ڻ |
| <i>ɡ</i> | گ | گ | <i>ṇ</i> | | ڇ |
| | | | <i>ŋ</i> | | ڱ |
| <i>ph</i> | ڦ | ڦ | | | |
| <i>th</i> | ٿ | ٺ | <i>y</i> | ی | ی |
| <i>ʈh</i> | ٺ | ٺ | <i>r</i> | ر | ر |
| <i>čh</i> | چھ | چھ | <i>r̥</i> | ڙ | ڙ |
| <i>kh</i> | کھ | ک | <i>l</i> | ل | ل |
| | | | <i>ʋ</i> | و | و |
| <i>bh</i> | بھ | پ | | | |
| <i>dh</i> | دھ | ڌ | <i>s</i> | س | س |
| <i>ɖh</i> | ڌھ | ڊ | <i>h</i> | ھ | ھ |

Note: This does not include all of the letters of the Perso-Arabic alphabet as used in Urdu and Sindhi, or in the proper order. Other letters are found only in Perso-Arabic loanwords. The vowel signs are also omitted.

Matthew Nelson, Life in a Dead Language: Modern Sanskrit as an Ultraminor Literature., 2017 [\[PDF\]](#) Indian Census in 2001, [\[XLSX\]](#)

- [genindex](#)

CHAPTER 2

Project Idea

Languages appear and die over time, and there're many reasons for a language to disappear: Natural disaster causing the death of all speakers, invaded by another tribe, becoming unsuitable for new concepts, did not evolve over time, become obsolete and no one wants to speak it, being considered a low-level language(language for lower people) and avoided, etc.

I want to discover the specialties of each language(and dialects) that appeared in history, what aspect of life do they mainly focus on (have a large vocabulary in). We can also infer their daily life by looking at their vocabulary: if there's a word for something, namely "pot", then they must have invented pots at that time.

CHAPTER 3

Instructor's note

Your idea sounds great; be sure to use proper academic sources for all the information you use. When you have gathered all your research, I can create a wiki page for you where you can upload it, with images etc., if you have any.

Bibliography

- [Con65] Indian Constitution. The official languages act. *Indian Constitution*, 01 1965.
- [Mas93] Colin P Masica. *The Indo-Aryan languages*. Cambridge University Press, 1993.
- [Eeds19] Gary F. Simons Eberhard, David M. and Charles D. Fennig (eds.). Ethnologue: languages of the world. twenty-second edition. *World Language Ethnologue*, 2019.
- [Bri17] Encyclop?dia Britannica. *Indo-Aryan languages*. Encyclop?dia Britannica, inc., 2017.
- [Cam08] Lyle Campbell. Ethnologue: languages of the world. 2008.
- [Nel17] Matthew Nelson. Life in a dead language: modern sanskrit as an ultraminor literature. *Journal of World Literature*, 2:411–432, 01 2017. doi:[10.1163/24056480-00204004](https://doi.org/10.1163/24056480-00204004).
- [Mas93] C.P. Masica. *The Indo-Aryan Languages*. Cambridge Language Surveys. Cambridge University Press, 1993. ISBN 9780521299442.
- [Pri38] James Prinsep. Inscription in the old character on the rocks of girnar in gujerat, and dhauri in cuttack. *Journal of the Asiatic Society of Bengal*, 1838.
- [Mas93] C.P. Masica. *The Indo-Aryan Languages*. Cambridge Language Surveys. Cambridge University Press, 1993. ISBN 9780521299442.