



The Historical Evolution of Linguistics Survey in India - A Study

N. Ramya¹ & Dr. Priya Premlatha²

¹Ph. D Research Scholar, PG & Research Department of History, Nirmala College for Women (A), Coimbatore
Affiliated to Bharathiar University, Coimbatore

²Research Supervisor, Assistant Professor & Head, PG & Research Department of History
Nirmala College for Women (A) Coimbatore
Affiliated to Bharathiar University, Coimbatore



Open Access

Manuscript ID:
BIJ-SPL2-NOV25-MD-125

Subject: History

Received : 23.09.2025

Accepted : 24.10.2025

Published : 14.11.2025

DOI: 10.64938/bijsi.v10si2.25.Nov125

Copy Right:



This work is licensed under
a Creative Commons Attribution-
ShareAlike 4.0 International License.

Abstract

This study provides a comprehensive overview of the historical evolution of linguistic surveys in India, from pre-colonial times to the post-independence era. It begins by exploring early linguistic observations by figures like Amir Khusrau, who informally classified Indian languages, and how these laid the groundwork for more systematic efforts during British colonial rule. The most significant development was the Linguistic Survey of India (LSI), initiated in 1894 under Sir George Abraham Grierson, which documented 179 languages and 544 dialects over several decades. The study also highlights the impact of linguistic data on critical policy decisions, such as the States Reorganization Act of 1956, which redefined state boundaries based on linguistic identities. Post-independence, institutions like the Linguistic Society of India and the Central Institute of Indian Languages (CIIL) played a key role in documenting and preserving languages. In conclusion, the study advocates for inclusive census practices, expanded mother tongue education, and greater institutional support to protect India's rich linguistic heritage.

Keywords: linguistic surveys, amir khusrau, sir george abraham grierson, states reorganization act, central institute of indian languages

Introduction

A linguistic survey is a comprehensive study of languages and dialects spoken in a particular region or by a specific community. Its purpose is to document, analyze, and understand the linguistic landscape, including language distribution, vitality, and changes over time. The linguistic survey helps to preserving cultural heritage, formulating effective language-related policies in education and governance, empowering linguistic minorities by recognizing their languages, informing sociolinguistic research, and actively working to document and prevent the extinction of endangered

languages. India is home to thousands of languages and dialects, representing several major language families, including Indo-Aryan, Dravidian, Austro-Asiatic, and Tibeto-Burman. The Office of the Registrar General's Language Division has conducted the Linguistic Survey of India (LSI). The survey captures the diversity and complexity of India's linguistic environment with a systematic examination of several languages and mother tongues. This documentation helps in understanding the social and cultural processes related to language and helps to preserve and protect India's linguistic legacy for future generations.



Methodology

The Present study is based on both Quantitative and Qualitative in nature and both Primary and Secondary Sources were used.

Pre-Colonial Period

Amir Khusrau a renowned Sufi poet from the Delhi Sultanate, he is credited with an early, informal, "linguistic census". In his work *Nuh Sibhir*, he classified the languages of India into Bhashas. The languages like Sindhi, Lahori, Kashmiri, Gujarati, Tamil, Telugu, Bengali, Awadhi, Hindavi and Sanskrit. He noted the uniqueness of each and their differences from foreign languages. Various travelers accounts, including Ibn Battuta, also included comments on the linguistic diversity encountered across the subcontinent. These early linguistic observations were not scientifically quantified.

The History of Linguistic Survey in Colonial Period

The arrival of European in 1600, particularly the British East India Company, gave a new impetus to linguistic documentation in India. The British realized that to rule India effectively, they needed a clear understanding of the many languages and dialects spoken across the country. Knowing these languages made it easier for them to communicate with local people, collect taxes, deal with legal matters, and run military operations smoothly. As a result, language study became a key part of colonial governance. The first census of India was conducted in 1872, but it did not cover the entire British territory and did not include a focus on linguistic data. In 1878, a committee was formed to consider proposals for the general Census of India scheduled for 1881. This 1881 census was the first synchronous census of British India and marked a deliberate attempt to collect information on "mother tongues." The 1891 Census continued this effort and expressed satisfaction that the information on mother tongues, aligned with scientific classifications, had successfully mapped broad language regions across India.

In 1895, the British government formally initiated the Linguistic Survey of India (LSI). George Abraham Grierson was appointed as the Director of the Survey and later wrote the language report for the 1901 Census. The 1901 Census Report became a major milestone in Indian linguistic history, largely due to Grierson's detailed chapter on languages. The LSI documented a total of 179 languages and 544 dialects across India. The last volume of the LSI was published in 1927. The 1911 Census introduced new instructions that required the recording of the language spoken at home, and in the case of infants or deaf-mutes, the mother's language. The language-related questions in the 1901, 1911, and 1921 censuses were limited to determining respondents' knowledge of English. A significant change occurred in the 1931 Census with the introduction of a bilingual question, aimed at identifying any language commonly used in addition to the mother tongue. This bilingual question was repeated in the 1941 and 1951 censuses based on the 1931 definition. There appear to be no published language reports for either the 1941 or 1951 censuses.

Grierson Contribution to LSI from 1894-1927

Under the leadership Sir George Abraham Grierson, a linguistic survey was conducted in India between 1894 and 1927. An Irish Indologist and Indian government officer. At the Seventh International Oriental Congress in Vienna in 1886, he proposed the first proposal for a systematic linguistic survey in India. The project was delayed owing to funding issues, but the survey was approved in 1891. Field operations started in 1894 and lasted until 1927. He was in charge of the entire project, did not personally visit each location. He used his extensive network of district collectors and other government officials to collect data. Between 1903 and 1928, the LSI results were published in I I volumes (19 sections), more than 8,000 pages. These volumes examined the phonology, morphology, syntax, and vocabulary of several languages and dialects. The LSI identified 179 languages and 544 dialects spoken in British India, including an estimated 224 million people. It classified languages into families like Indo Aryan,



Dravidian, Tibeto-Burman, and Austroasiatic and sub-groups and revealing fundamental information about their distribution and relationships. It specifically omitted the areas like Madras Presidency, the princely realms of Hyderabad and Mysore, and Burma.

Language family	No. of Language	Percentage
Indo-Aryan	21	78.05
Dravidian	17	19.64
Austro-Asiatic	14	1.11
Tibeto-Burman	66	1.01

Source: Census of India 2011(Language), Paper I of 2018, States and Union Territories (Table C-16)

LSI in Post Independence India

The growth of linguistic surveys in post-independence India started from 1951 onwards. The 1951 Census presented language data only in the form of raw, alphabetical listings of mother tongues along with speaker counts. It offered minimal analytical insight into India's vast linguistic diversity. By the 1961 Census, it was recognized that the full name of the mother tongue including dialects should be recorded exactly as stated by the respondents. In cases where the mother had died early, the language predominantly spoken in the childhood home was noted. For infants and deaf-mutes, the mother's language continued to be recorded. The 1961 Census was particularly significant for its detailed enumeration of languages, identifying over 1,554 distinct "mother tongues." This provided a foundational baseline for post-independence linguistic planning in India. To streamline data processing, the 1971 Census adopted a new policy: languages spoken by fewer than 10,000 individuals were excluded from detailed publication.

By the time the 1991 Census was conducted, people across India reported over 10,400 different language names, many of these turned out to be repeated entries, dialects, or simply spelling variations. After careful analyzed by experts were

able to identify 1,576 unique mother tongues. 1,796 language names couldn't be clearly identified and were classified as "unclassified." Using linguistic methods, the 1,576 recognized mother tongues were further grouped into 114 major languages. These 114 languages were further classified into two categories like

- Part A— Scheduled Languages, which included the 18 official languages recognized by the Indian Constitution. Initially 15 in 1981, Konkani, Manipuri, and Nepali language were added in 1992.
- Part B — Non-Scheduled Languages, consisting of 96 other languages. Some of these had fewer than 10,000 speakers or were not clearly classified.

In total, 216 mother tongues were spoken by more than 10,000 people 85 of them falling under Scheduled Languages and 131 under Non-Scheduled Languages. Languages with fewer than 10,000 speakers were grouped under the category "others". The Census revealed that 96.29% of India's population spoke Scheduled Languages, while 3.71% spoke Non-Scheduled or unclassified languages. According to the 2011 linguistic census, published in 2018, more than 19,500 languages or dialects are spoken in India as mother tongues. These were grouped into 121 main mother tongues for better classification. Among them, Hindi is the most widely spoken, with 52.8 crore people (or 43.6% of the population) identifying it as their mother tongue. The second most spoken language is Bengali, with about 9.7 crore speakers across the country.

Post-Independence Linguistic Reorganization of States

After independence, India's administrative structure began to undergo significant changes to support national planning and governance. The Indian Independence Act of 1947, British India was divided into two Dominions like India and Pakistan. The Act recognized two types of territories within India: (a) British provinces and (b) Princely states. When the Constitution of India came into effect in 1950, the country was administratively divided into categories



of 'A', 'B', and 'C' states. The boundaries of Madras State were adjusted on a linguistic basis through the Andhra State Act, 1953, which transferred Telugu-speaking areas to form the newly created state of Andhra. In 1954 Himachal Pradesh and Bilaspur were merged to form the new state of Himachal Pradesh. In 1953 with the establishment of the States Reorganization Commission, headed by Fazal Ali. Based on recommendations States reorganization Commission, the States reorganizations Act was implemented on November 1, 1956. As a result, India's internal boundaries, reducing the number of states from 27 to 14 and creating 6 Union Territories. Andhra Pradesh was expanded to include the Telugu-speaking regions of Hyderabad.

States and its Language

Hindi speakers rank first among all the languages being spoken in India and is followed by the speakers of Bengali, Telugu, Marathi, Tamil and Urdu languages. The Indian census has included Bodo, Dogri, Maithili, Santali, Konkani, Manipuri, and Nepali as Scheduled languages since 1991. The percentage of speakers of each language for 1981 was calculated based on the total population of India, excluding Assam, where the 1981 Census was not conducted due to disturbed conditions. The 1991 Census was conducted excluding Jammu & Kashmir due to unfavorable conditions.

Language	State	Percentage
Hindi	Bihar, Haryana, Himachal Pradesh, Madhya Pradesh, Rajasthan, Uttar Pradesh, Chandigarh and Delhi mainly located in northern and central India	43.63
Bengali	Tripura and West Ben al.	8.03
Marathi	Maharashtra,	6.86
Telugu	Andhra Pradesh,	6.70
Tamil	Tamilnadu, Srilekha	5.70
Gujarati	Gujarat	4.56
Urdu	Jammu & Kashmir	4.19

Kannada	Karnataka	3.61
Odia	Orisha	3.10
Malayalam	Kerala	2.88
Punjabi	Pun ab	2.74
Assamese	Assam	1.26
Maithili	Bihar	1.12
Santali	Jharkhand, Odisha, West Bengal, Bihar	0.61
Kashmiri	Jammu and Kashmir	0.56
Nepali	Sikkim, West Bengal	0.24
Sindhi	Maharashtra, Rajasthan, Gujarat, Madhya Pradesh, and Delhi	0.23
Dogri	Jammu and Kashmir, Himachal Pradesh	0.21
Konkani	Goa, Karnataka, Maharashtra	0.19
Manipuri	Manipur, Assam, Tripura.	0.15
Bodo	Assam, Arunachal Pradesh, Nagaland, and West Ben al	0.12

Source: Census of India 2011 (Language), Paper I of 2018, States and Union Territories (Table C-16)

The Linguistic Society of India

The Linguistic Society of India (LSI) was founded in 1928 at the fifth Oriental Conference in Lahore, with its publications starting in 1931. Its main goal is to promote Indian linguistics and the scientific study of Indian languages. The Society was registered in 1954 under the Act XXI of 1960 in Calcutta and again in 1966 under the Bombay Public Trust Act. Its office is based at Deccan College Postgraduate & Research Institute, Pune. Its inaugural meeting, held on April 1, 1928, was attended by linguists from Punjab and presided over by

Dr. Taraporewala. The meeting coincided with the award of the Order of Merit to Sir George

Grierson, who had led the first Linguistic Survey of India. In 1937, the office moved from Lahore to Calcutta, where Professors S.K. Chatterji and



Sukumar Sen took charge. In 1955, the Indian Philological Association merged with the LSI, and the office shifted to Deccan College, Pune. The Linguistic Society of India (LSI) organizes the All-India Conference of Linguists (AICL) annually, starting in 1970. The latest, the 33rd AICL, was held in Chandigarh in 2011.

Protecting Language Diversity in India

The Constitution of India lists 22 official languages among the 22 languages in the Eighth Schedule, including Hindi, Tamil, Bengali, and others and Sanskrit, Tamil, and Kannada are designated as classical languages due to their ancient written and oral traditions spanning over 1,000 years. The Constitution of India not only recognizes scheduled and classical languages but also protects minority languages as a fundamental right. The language policy ensures the protection of linguistic minorities and appoints a Special Officer to safeguard their interests. Post-independence, the Central Institute of Indian Languages (CIIL) in Mysore was tasked with conducting a comprehensive language survey, but it remains incomplete. The 1991 Census identified 1,576 mother tongues with distinct grammatical structures and 1,796 speech varieties. The Indian Constitution ensures the right of children from linguistic minority groups to receive primary education in their mother tongue, the United Nations declared International Mother Language Day on February 21.

The Language Bureau, under the Ministry of Human Resource Development, oversees this policy. Supporting language diversity online, Union Minister Ravi Shankar Prasad emphasized that the internet should not be limited to English and called for the inclusion of local languages. The Government of India's Digital India initiative mandates that all mobile phones sold from July 2017 must support Indian languages, aiming to bridge the digital divide and empower non-English speakers. Minority languages are still at risk of extinction, such as the Bo language in Andaman and Nicobar Islands, which became extinct with the death of its last speaker,

Boa. The efforts to preserve language diversity, many communities are shifting to major languages, abandoning their mother tongues.

Conclusion

The historical evolution of linguistic surveys in India reflects the country's rich and complex linguistic diversity and its vital role in shaping governance, education, and cultural identity. From the early observations of scholars like Amir Khusrau to the comprehensive colonial-era surveys led by Sir George Grierson, and later the post-independence census initiatives, India has made remarkable progress in documenting its vast array of languages. Key milestones such as the 1961 Census and the States reorganization Act of 1956 highlight how language has significantly influenced administrative decisions and the redrawing of state boundaries. Organizations like the Central Institute of Indian Languages (CIIL) and the Linguistic Society of India (LSI), along with modern digital efforts, continue to contribute to language preservation. Protecting India's linguistic heritage calls for more representative data collection, inclusive educational policies, and active community participation to ensure that every language regardless of how widely it is spoken is recognized, preserved, and respected.

References

1. Language Division Report, Registrar General & Census Commissioner of India, Ministry of Home Affairs New Delhi, 2011, p. 5.
2. Grierson, G. A, Linguistic Survey of India, Central Publication Branch, Calcutta, 1927, p.85.
3. Chatterji, S.K. Languages and Literatures of Modern India, Bengal Publishers, Calcutta, 1963, p. 44.
4. Annamalai, E. Language Movements in India, Central Institute of Indian Languages, Mysore, 2005. pp. 61,65.
5. Grierson, India: A Linguistic Introduction. Government Printing Office, Calcutta, 1904,



6. Zide, Norman H. *Studies in Comparative Dravidian Phonology*, University of Pennsylvania Press, Philadelphia, 1965, p. 9.
7. Grierson, *Linguistic Survey of India*, Superintendent of Government Printing, Calcutta, vol. I-XI, 1903, p.21
8. Singh, *India: A Regional Geography*, National Geographical Society of India, Varanasi, 1971, p. 435
9. Nilakanta Sastri, *The Pandyan Kingdom: From the Earliest Times to the Sixteenth Century*, Swathi Publications, Madras, 1972, p. 4.
10. Sivaramamurti, *Art of India*, Publications Division, Ministry of Information and Broadcasting, New Delhi, 1974, p. 134.
11. Menon, *The Story of the Integration of the Indian States*. Orient Longman publication, New Delhi, 1956, p. 342
12. I I) Brass, Paul R. *Language, Religion and Politics in North India*, Cambridge University Press, Cambridge, 1974, p. 27
13. Government of Tamil Nadu. *Tamil Nadu Official Language Act, 1956*, Government Press, 1974.
14. Office of the Registrar General & Census Commissioner of India, *Mother Tongue Survey of India: Project Report*, New Delhi, 2007.
15. Chandramouli, C. *Linguistic Survey of India: A Perspective*, Office of the Registrar General, India, New Delhi, 2011, p. 14.
16. Central Institute of Indian Languages, *Annual Report 2011—12*, Mysore, CIIL, 2012, p. 18.
17. *The Constitution of India*, Article 350A and 350B, Ministry of Law and Justice, 1950,
18. p. 190.
19. Government Order. Ms. No 136, Dated 12- 06- 2006
20. Government. Order. Ms. No. 133, dated 26-12- 2021
21. Tamil University, *Tamil University Act*, Tamil University Publication Wing, Thanjavur, 1982, p. 2.