

A short history of the official counting of India's languages (1901–2011)

Ayesha Kidwai

Jawaharlal Nehru University (New Delhi, India); ayesha@jnu.ac.in;
ORCID: 0009-0009-7527-3285

Abstract. This study investigates the persistent ambiguity surrounding the count of mother tongues in postcolonial India by tracing a century of enumeration history. It argues that the post-1971 Census shifted from identifying distinct languages to utilising “language labels”, a methodology that obscures rather than illuminates linguistic diversity. Deviating from the common perception that this practice is a colonial holdover, the paper demonstrates that the Linguistic Survey of India era censuses were actually more attuned to the country's pluralistic reality. Instead, the current impulse toward minimising is identified as a strategic instrumentalisation of Constitutional provisions.

Keywords: Linguistic Survey of India, Census Language Tables, decoloniality, India's linguistic diversity, Hindi

Acknowledgements. This article has benefited from extensive discussion with R. Karthick Narayanan, and the comments of an anonymous reviewer.

For citation: Kidwai A. A short history of the official counting of India's languages (1901–2011). *Indo-Iranian Languages*. 2026. 2 (1). P. 15–48.

DOI: 10.30842/30346800.2026.2.1.3

Краткая история официальной переписи языков Индии (1901–2011)

Айеша Кидваи

Университет Джавахарлала Неру (Нью-Дели, Индия);
ayesha@jnu.ac.in; ORCID: 0009-0009-7527-3285

Аннотация. В статье прослеживается, как неоднозначность, присущая подсчету родных языков в постколониальной Индии, отражается в вековой истории переписей. Утверждается, что в переписях, проведенных после 1971 г., наблюдается переход от выявления отдельных языков к использованию «языковых ярлыков», и такая методология затемняет, а не проясняет языковое разнообразие. В противоположность распространенному мнению, что такая практика является колониальным наследием, в статье показано, что переписи, выполненные в период создания Linguistic Survey of India (1894–1928), лучше отражали многоязыковую реальность страны. Напротив, современную ориентацию к минимизации можно связать с манипулированием конституционными положениями.

Ключевые слова: Лингвистический обзор Индии, учет языков в переписи, деколонизальность, языковое разнообразие Индии, хинди

Благодарности. Автор выражает благодарность Р. Картику Нараянану за обстоятельное обсуждение статьи, а также анонимному рецензенту за ценные комментарии.

Для цитирования: Kidwai A. A short history of the official counting of India's languages (1901–2011). *Indo-Iranian Languages*. 2026. 2(1). P. 15–48.

DOI: 10.30842/30346800.2026.2.1.3

1. Introduction

One of the enduring mysteries in postcolonial India is the actual number of languages that are spoken in the country at this point in time. While “Ethnologue” lists India as being home to 435 languages, the “People’s Linguistic Survey of India” (2010–present) (henceforth, PLSI) claims to have documented over 780 languages thus far; but the Census of India 2011 (henceforth, Census), the sole official record of languages, registers 121 languages. These three widely differing figures are driven

by considerably varying methodologies, which are in fact not obviously reconcilable with each other.

A look at the summary tables of how the three surveys present their data reveals this difference. Ethnologue names each ‘indigenous’ language by a unique code and summarises its count by linguistic genealogy, relying on the figures of the Census to determine the population of speakers, see Table 1. Since it is based on a literal head-count of the Indian population, it produces fresh estimates of the numbers of speakers every decade or so.¹

Table 1. Ethnologue’s language count by family

Language family	Number of languages
Sino-Tibetan	149
Indo-European	141
Dravidian	84
Austro-Asiatic	32
Andamanese	14
Kra-Dai	5
Creole	4
Sign Language	2
Language Isolate	1
Mixed Language	1
Pidgin	1
Unclassified	1
Total	435

The PLSI differs from the others in several respects, as it is not intended to be an “exhaustive survey of each and every language in existence in India”² or a substitute, repeat, or sequel to the Census and the colonial Linguistic Survey of India (1896–1928; henceforth, LSI). As its goals are, generally stated, to build networks between diverse language communities and to foster a rights-based advocacy on matters of language, it does not

¹ <https://www.ethnologue.com/country/IN/> (retrieved on 10.10.2025).

² <https://www.bhasharesearch.org/plsi/language-archive> (retrieved on 10.10.2025).

actually list or classify the languages it does study by any other metric than provinces. Nevertheless, a four-fold classification can be inferred from the manner in which the archive is proposed to be organised into, as shown in Table 2. This classification is not linguistic or genealogically determined, but rather uses an admixture of sociological (*adivasi*, i.e. tribal) and Census (Scheduled vs. Non-Scheduled) categories with a purely function-based category of ‘code’ language thrown in. The few entries that have been uploaded on the website so far contain a mix of linguistic, historical, and cultural information, with references both to the Census (particularly for speaker strength) and the LSI as its source.

Table 2. The PLSI language classification

PLSI	Number of languages
Languages of Adivasis	32
Scheduled Languages	30
Non-Scheduled Languages	–
Code Languages	1

The Census, on the other hand, offers two classifications in its chapter on languages, and then groups them in terms of their inclusion in the Eighth Schedule of the Indian Constitution (Table 3). It then presents various statements examining the distribution of population for these two groups of languages by different parameters, and closes with a genealogical classification of the languages of the two groups (Table 4). This is in contrast to the LSI’s counts, which follows only a genealogical grouping (Table 5).

Table 3. The Census’s language count (2011)

Census year 2011	Number of languages
Scheduled Languages	22
Non-Scheduled Languages	99 + others
Total	121 (+ others)

Table 4. The Census's language count by family (2011)

Language family		Number of languages
Indo-European	Indo-Aryan	21
	Iranian	1
	Germanic	1
Dravidian		17
Austro-Asiatic		14
Tibeto-Burman		66
Semitic-Hamitic		1
Total		121

Table 5. LSI's language/dialect count by family for India

Language family	Number of languages/dialects
Austro-Asiatic	45
Tibeto-Burman	344
Dravidian	45
Indo-Aryan	289
Unclassified	20
Total	743³

In fact, the enumerations in the Census Reports have never yielded the same number of languages/dialects ever postulated by the LSI (Table 5), even in 1901–1921 (Table 6), when Grierson was still loosely associated with the Census operations.⁴

Table 6. India languages in the Census 1901–1921

Census year	Number of languages
1901	147
1911	220
1921	222

³ This figure is approximate as the LSI covered undivided India.

⁴ The Census counts in Table 6 are inclusive of foreign languages spoken within the territory of India.

In this paper, I present a short history of the methodology of the Census regarding the treatment of ‘mother tongue’ language returns, in order to make the general point that the post-1971 period represents a trend similar to the methodologies of the early colonial censuses. My main argument runs as follows: The methodology of the Census in evaluating language returns does not, whatever be the allusions it makes, have anything to do with the LSI or the 20th-century colonial censuses. I will argue that the treatment of language returns in postcolonial India has had very little to do with the LSI, barring the efforts in the 1961 Census, where the Census Commissioners and officials sought to take a decolonial approach to the question of language returns. I suggest that the Census’s reference to the LSI must be seen as merely a superficial appropriation of some features of an LSI-inspired Census in order to advance an agenda that minimises and undercounts India’s minority languages and ‘promotes’ Hindi and the other 21 scheduled languages.

To this end, the paper is structured as follows. Section 2 traces the Census’s methodology in the postcolonial period 1971–2011 and shows that the Census can be argued to trace the colonial roots of many of its processes to the 1881 and 1891 Censuses. Section 3 discusses the 1901 Census report on language and the LSI framework that underlay it. Section 4 turns its attention to the Census in 1961 and describes how its critical engagement with the LSI signalled a period of decolonial thought in measuring India’s linguistic diversity. Section 5 concludes the essay with a discussion of the political developments that ensured a decolonial Census could not be maintained.

2. The Census (1971–2011) methodology in comparison with early colonial censuses (1881–1891)

For the discussion in this section, I will lay out the process observed in the Census 2011 in the treatment of language returns and record the similarities from the previous postcolonial censuses in comparison, making the connections to the 1881 and 1891 censuses where relevant.

2.1. Enumeration

The beginning of any census exercise requires a definition of the return to be elicited by the enumerators. The Census 2011 does not elicit the names

of ‘languages’ directly, instead using the notion of ‘mother tongue’ (MT) as defined in the Indian Constitution in Art. 350A.⁵ The prefatory note to the paper on Language defines ‘mother tongue’ [Census 2011: 3] as follows:

Mother tongue is the language spoken in childhood by the person’s mother to the person. If the mother died in infancy, the language mainly spoken in the person’s home in childhood will be the mother tongue. In the case of infants and deaf mutes, the language usually spoken by the mother should be recorded. In case of doubt, the language mainly spoken in the household may be recorded.

This definition of MT has remained constant since 1971, and is at least partly an inheritance from the colonial censuses of 1931 and 1941, which instructed enumerators to record a child’s first language. The source of this, what Tiery [2024] calls the ‘primordialist’ definition of the unit of enumeration, is most likely the 1881 and 1891 censuses, which were defined characterising ‘language’ as ‘spoken by the individual from the cradle’, and as ‘spoken in childhood by the person’s mother to the person’, respectively. This definition was used again in the 1941 and 1951 censuses as well. Only in the 1901, 1911, 1921, and 1961 censuses was “mother-tongue” not defined literally as the respondent’s mother’s language.

In addition, enumerators were instructed to record the MT in full as stated, and to desist from linking it to the respondents’ religion or ethnicity. In case the enumerator has grounds to suspect the truth of the MT, (s)he is instructed to nevertheless record the return given and follow it up with a supervisory report to their supervising officer for verification. These instructions are intended not only to minimise enumerator bias and the politicisation of the Census process, but also to collect data in a manner that is amenable to post-enumeration rationalisation, which we discuss in the next section.

2.2. Rationalisation

In the Census 2011, raw returns are subjected to a process of “thorough linguistic scrutiny, edit, and rationalisation” [Census 2011: 4]. These pro-

⁵ 350A provides for instruction in the mother tongue at the primary stage of education to children belonging to linguistic minority groups. This provision was not part of the Constitution of India, 1950, and was inserted by the Constitution (Seventh Amendment) Act, 1956.

cesses, the Census states, are necessary because “mother-tongues as returned in the census are basically the designations provided by the respondents of the linguistic mediums in which the respondents think they communicate; they need not be identical with the actual linguistic mediums” [Ibid.]. For the 2011 Census, a total of 19,569 raw returns were received.

Although Census 2011 does not make its detailed criteria explicit in any of the documents it has made public since 1971, the Language Census Data website⁶ provides us with a reasonable impression of the steps that follow immediately post-enumeration:

- Editing involves standardisation of spellings in the return and correction of spelling and other language errors.
- Rationalisation involves the elimination of spurious returns.

In Census 2011, the total number of rationalised mother tongues was 2,843, i.e. 14.53 % of the original raw returns.

The institution of this step of rationalisation has been an important one, as the colonial Census overall was dogged by this problem of spurious returns. For example, the 1891 Census notes in dismay [Census 1891: 130]:

[T]he first impulse, in many cases, is to return the name of the caste as that of the language. For example, the potter gives “potterish”, the tanner “tannerish” or the weaver “weaverish” as his mother tongue, especially if he be either a member of a large caste or a stranger to the locality where he is being enumerated. In other instances, the name of the District or State will be returned. (...) [But] (...) there are cases, too, where the tribal designation indicates a real distinction. But a further difficulty arises when the name mentioned happens to be not in accordance with the enumerator’s notion of what is fit, or where official views on matters philologic may not have filtered down to him. In that way a good deal of valuable information may have been lost in the course of the census; as where, for instance, a general term, such as Hindi, was substituted for some name locally recognised.

This process of scrutiny and rationalisation was introduced since the 1961 Census. All names that turn out to be cases of misreporting, caste names, place names, names of religious faiths, motivated names, etc. are weeded out at this stage.

⁶ <https://language.census.gov.in/showHomePage> (accessed on 10.10.2025).

All these measures notwithstanding, it is therefore somewhat surprising that the number of raw returns has only been increasing in every Census since 1971; as Table 7 shows, the number of rationalised MTs has fluctuated widely.

Table 7. Rationalised MTs Census Language Reports

Census year	Raw returns	Rationalised MTs
1971	3,000	–
1981	<3,000	–
1991	10,400	3,372 (32.4 %)
2001	6,661	3,592 (53.9 %)
2011	19,569	2,843 (14.5 %)

It is impossible to ascertain in particular as to why the list of rationalised MTs does not stay within approximately the same range in every Census, especially if the information in either the LSI or the 1931 Census, both of which published classified lists of language names, is used as a basis for the rationalisation.

2.3. Classification

The process of classification is then applied to rationalised mother tongues. It involves the “arrangement of speeches into coherent typological systems so that broad language areas could be delimited and each speech form could be appropriated to the place it belongs in the system on the basis of comparability of its linguistic features.”⁷ Effectively, this statement signifies that first, the rationalised MT returns are classified into language families and subgroups that were proposed by the LSI [Grierson 1927], and possibly also subsequent Censuses. Census 2011 was found to contain only 1,369 classified MTs, i.e. just 7 % of the raw MT returns and approximately 48 % of rationalised MTs.

⁷ <https://language.census.gov.in/showConceptsDefinitions> (accessed on 10.10.2025).

The remainder of the 2,843 non-spurious MT returns, i.e. 1,474 rationalised MTs, are considered to be ‘unclassified’ languages. These unclassified languages are not subjected to any further processes and are reported under ‘Others’ in the language tables. In the 2011 Census, this means that the non-spurious MT returns of 1,875,542 Indians are presented in the Census reports as not having language names that the state recognises.

The second step of classification is also a step of a further elision of language names, as the Census, since 1991, chooses not to name those classified MTs that have less than 10,000 speakers. These, too, are listed as undifferentiated ‘Others’. As Karthick Narayanan [2025: 9] points out, this means that in the Census 2011, both types of ‘Others’ languages appear to total 1,099 classified MTs, i.e. over 80 % of classified MTs, as the Census publication [Census 2011] states that 270 classified MTs have been identified at the All-India level.⁸

The third and final step of classification maintains the language names and returns of the 270 MTs that have survived thus far, but with the additional step of using selected language names as labels, rather than signifying an actual MT. For example, consider the first page of the ‘Abstract of Speaker Strength of Languages and Mother Tongues-2011’ [Census 2011: 6].⁹

The items in bold in Figure 1 are the use of language names as, what I call, ‘language labels’. While there are languages like Assamese and Dogri that have no other classified MT from the 1,369 classified MTs, there are also language labels like Hindi, which have 56 classified MTs grouped under it. In all cases, the ‘Others’ category contained under individual language labels signifies classified MTs which are spoken by less than 10,000 speakers.

The Census documents are exceedingly vague about how such decisions for classification are made. Although reference is made to the LSI, and the Census held during the pendency of the LSI, are as providing some sort of foundation for this classification, this is an untruth on several levels.

⁸ The Language Census Data website, however, cites 373 languages as being excluded by the 10,000 speakers strength benchmark. For the purposes of the description in this paper, I rely on the official Census papers, but this confusion is informative insofar as it indicates that the published figures and the data contained on the Language Census data website do not always agree.

⁹ There is another ‘Others’ section, which records the numbers of unclassified MTs and is listed independently, on par with language labels.

PART-A
LANGUAGES SPECIFIED IN THE EIGHTH SCHEDULE (SCHEDULED LANGUAGES)

Name of Language & mother tongue(s) grouped under each language	Number of persons who returned the language (and the mother tongues grouped under each) as their mother tongue)	Name of Language & mother tongue(s) grouped under each language	Number of persons who returned the language (and the mother tongues grouped under each) as their mother tongue)
1	2	1	2
1 ASSAMESE	1,53,11,351	Gawari	19,062
Assamese	1,48,16,414	Gojri/Gujjari/Gujar	12,27,901
Others	4,94,937	Handuri	47,803
2 BENGALI	9,72,37,669	Hara/Harauti	29,44,356
Bengali	9,61,77,835	Haryanvi	98,06,519
Chakma	2,28,281	Hindi	32,22,30,097
Hajjong/Hajong	71,792	Jaunpuri/Jaunsari	1,36,779
Rajbangsi	4,75,861	Kangri	11,17,342
Others	2,83,900	Khari Boli	50,195
3 BODO	14,82,929	Khorthal/Khotta	80,38,735
Bodo	14,54,547	Kulvi	1,96,295
Kachari	15,984	Kumauni	20,81,057
Mech/Mechhia	11,546	Kurmali Thar	3,11,175
Others	852	Lamani/Lambadi/Labani	32,76,548
4 DOGRI	25,96,767	Laria	89,876
Dogri	25,96,763	Lodhi	1,39,180
Others	4	Magadhii/Magahi	1,27,06,825
5 GUJARATI	5,54,92,554	Malvi	52,12,617
Gujarati	5,50,36,204	Mandeali	6,22,590
Gujrao/Gujrau	15,431	Marwari	78,31,749
Pattani	16,510	Mewari	42,12,282
Ponchi	13,812	Mewati	8,56,643
Saurashtra/Saurashtri	2,47,702	Nagpuria	7,63,014
Others	1,62,895	Nimadi	23,09,265
6 HINDI	52,83,47,193	Padari	17,279
Awadhi	38,50,906	Pahari	32,53,889
Baghati/Baghati Pahari	15,835	Palmuha	23,579
Bagheli/Baghel Khandi	26,79,129	Panch Pargania	2,44,914
Bagri Rajasthanhi	2,34,227	Pando/Pandwani	15,595
Banjari	15,81,271	Pangwali	18,668
Bhadrawahi	98,806	Pawari/Powari	3,25,772
Bhagoria	20,924	Puran/Puran Bhasha	12,375
Bharmauri/Gaddi	1,81,069	Rajasthani	2,58,06,344
Bhojpuri	5,05,79,447	Sadan/Sadri	43,45,677
Bishnoi	12,079	Sirmauri	1,07,401
Brajbhasha	15,56,314	Sondwari	2,29,788
Bundeli/Bundel khandi	56,26,356	Sugali	1,70,987
Chambeali/Chamralli	1,25,746	Surjuia	17,38,256
Chhattisgarhi	1,62,45,190	Surjapuri	22,56,228
Churahi	75,552	Others	1,67,11,170
Dhundhari	14,76,446	7 KANNADA	4,37,06,512
Garhwali	24,82,089	Badaga	1,33,550
		Kannada	4,35,06,272
		Kuruba/Kurumba	24,189
		Prakritha/Prakritha Bhasha	12,257
		Others	30,244

Fig. 1. Languages grouped under Hindi by Census 2011

First of all, the number of MTs in none of the Censuses since 1971 corresponds to the number of language varieties that the LSI identified: a total of 743 languages, from which 179 ‘languages’ and 544 ‘dialects’ were identifiable. Assuming that the Censuses 1971–2011 also follow the Census

1961 decision to abandon the colonial distinction of language vs. dialect (see Section 3), the fact is that the closest we have ever got to 723 MTs is in Census 2011, and even that is off by about 500 languages.¹⁰ As Table 8 shows, the MT-identification rate of the Census has never crossed 20% of even the classified MTs!

Table 8. India languages in the Census General Reports

Census year	Number of classified MTs	Number of MTs > 10,000 speakers
1971	–	–
1981	–	106
1991	1576	216 (13.7% of classified MTs)
2001	1635	234 (14.3% of classified MTs)
2011	1369	270 (19.7% of classified MTs)

Second, the LSI did not ever make a classification that puts all the 56 varieties listed under the language label ‘Hindi’ as being included in the same sub-group or branch. For example, Awadhi and Bagheli are to be found in the Eastern Hindi subgroup of the Central sub-branch of Grierson’s classification of the Indo-Aryan subfamily, but Braj and Hindi are in the Western Hindi sub-group. Bhojpuri is classified as an outer-branch Eastern Indo-Aryan language, and all the Pahari languages in the list are part of the outer North-Western branch of Indo-Aryan, as is Gojri. No linguistic justification is given for this classification by the Census, and it is exceedingly difficult to take as principled in any way.

Moreover, the classification of various MTs under the language label of Hindi has not remained constant either. As Table 9 records, the language label ‘Hindi’ has had varying different numbers of MTs grouped under it.

¹⁰ Quite a different picture emerges, however, if we include the othered languages into the MT count, using the formula [Classified MTs] — [No. of MTs > 10,000 speakers]. Then, Census 1981 isolates 1,360 MTs, Census 1991 1,401 and 1,099. These figures are much closer to the only Census that actually engaged with Grierson’s LSI and the Census of 1931, but as the question that was asked at the point of enumeration was also very different, too much cannot be made of the apparent closeness to the Census figures.

Table 9. MTs listed under the language label ‘Hindi’ in Census Reports

Census year	Number of MTs
1971	48
1981	50
1991	48
2001	49
2011	56

In fact, we may infer that the Census has, since 1971, been engaged in bowdlerising the LSI and the 1931 lists to inflate figures for the language label of Hindi. All the languages of the LSI/Census 1931 subgroups named Bihari, Rajasthani, Central Indo-Aryan languages, Eastern Hindi, and West, East, and Central Pahari languages, have been successively incorporated into the current ‘Hindi’ group. There is no escaping the extreme irony of this ‘promotion’ of Hindi by the Census operations, particularly given that many of the languages grouped so are also the ones which, on the ground, are under intense pressure from Hindi in their home regions, in that the domains in which these languages are used are shrinking rapidly, with Hindi being used in those domains. Nevertheless, resisting that pressure, the speakers of these languages hold out and assert their distinct linguistic identity to the Census enumerators, only for it to be substantially erased by this policy of using language labels.

Rather than the LSI, the current Census table most closely resembles the kind of language information and groupings that were undertaken in the 1891 Census, which also groups together languages under language labels. In that Census, language names like ‘Hindi’ are explicitly admitted to be language labels which, the report says, “predominates” through Hindustan and Bihar [Census 1891: 138–139] (emphasis mine):

This title [Hindi] is admittedly a comprehensive one, and includes all the varieties of Aryan speech between the Jamna and Rajmahal. It has thus absorbed such distinctions as Vraj, Baiswari and Bhojpuri, as well as Maithili and other Bihar varieties of Hindi. <...> [T]**herefore, all such sporadic entries have been absorbed into the general title, lest the separate recognition of such small and scattered numbers should give a misleading notion of the actual prevalence of the dialects in question.**

Just as the 1891 Census collapsed several independent mother tongues under a single language label of ‘Hindi’, lest there be “separate recognition of such small and scattered numbers”, the postcolonial Census since 1971 does exactly the same. Why the 1881 and 1891 Censuses did so is understandable, as the early colonial state was deeply invested, as Goswami [2004: 64] has argued, in the creation of sustainable state-generated classificatory schemes that were constructed by a “rule of difference” that mandated the “continual performance of the nonidentity between the state and its colonised subjects”. Building on her arguments, Majeed [2019] has argued that the colonial state’s attempts in the early Censuses were to develop ‘language’ to be a parameter by which it could classify social constructs like ‘race’ and ‘nationality’. But why should the postcolonial Census wish to replicate the intentions of the colonial one? Before I discuss the reasons for this apparent coincidence of strategies in Section 4, I first consider the impact that the LSI and Grierson’s work on the 1901 (and subsequent Censuses) had for postcolonial India, and particularly the 1961 Census.

3. The LSI and its impact on the Census (1901–1931)

The beginning of the 20th century saw a radical change in the colonial Census of India methodology because of the initiation of the Linguistic Survey of India in 1896. Although an initiative independent of the Census throughout its long existence spanning more than three decades (the last volume of the LSI was published in 1928), the LSI and its Superintendent, George Abraham Grierson, played quite a crucial role in the Censuses of 1901, 1911, and 1921. In Kidwai [in press], I show that the proposal to conduct the LSI, which was a study of the modern vernacular languages of India, was originally conceived as a scholarly exercise, which represented a consensus between philologists meeting at the Seventh International Oriental Congress, Vienna, 27–29 September 1886. It was only upon the denial of governmental sponsorship to the original scholarly exercise that led to the LSI’s opportunistic alliance with the Registrar General’s office.

As Majeed [2019: 56–60] has shown, even as there was a close connection between the Census and the LSI, Grierson quite clearly enforced a separation between the two, cf. [Ibid.: 58]:

From the outset, he was anxious about how his engagement with the Census might hinder the Survey's progress, since the LSI's work could not be 'dropped altogether for a lengthened period, and then taken on again, without injury and considerable expense'. (...) It was therefore in Grierson's interest to make sure that his work for the Census could be utilised in the LSI, while also ensuring that the Survey came into its own as a project in its own right. At the same time, it could ride on the back of the Census for some of its information.

One of the crucial ways in which Grierson did this was by introducing into the Census the genealogical classifications and groupings arrived at by the LSI following its own independent methodology, thereby ensuring that Census observations, starting from the 1901 Census itself, were at least broadly in line with the classifications being arrived at by the LSI on philological/linguistic grounds. This was doubly ensured by Grierson by having circulated¹¹ to all the Provincial Census Commissioners the rough lists of languages spoken in each district that the LSI had covered¹² and by agreeing to author the chapter on 'Languages' in the 1901 Census report,¹³ in such a way that the languages recorded by the Census enumerations are classified according to the groupings being developed by the ongoing LSI. As Grierson says in his opening paragraph of the chapter [Census 1901: 247] (emphasis mine):

The survey is not yet completed (...) but the greater portion of the work has been done, and the present chapter may be considered as an abstract of its results so far as they are available. (...) **Speculations regarding Indian languages must wait till the survey is concluded, and all the facts are presented in a convenient form. Till then, even the classification adopted in the following pages must be taken as provisional.** It represents the arrangement which at the present moment, to my mind, agrees best with the facts which have been hitherto examined, but the discovery of new facts may induce me to modify it.

¹¹ H. H. Risley to G. A. Grierson dated 24.10.1900 (Mss Eur E223/260, Asian and African Studies, British Library).

¹² The LSI did not cover all of British India, as the entire Madras Presidency and some princely states had refused to participate in it at all.

¹³ G. A. Grierson dated 19.09.1900 to the Census Commissioner, H. H. Risley (Mss Eur E223/260, Asian and African Studies, British Library).

The provision of the LSI's Rough Lists to the Census Commissioners can be seen as a rationalisation measure by Grierson, as the offer was made in the context of Risley's early suggestion¹⁴ that Grierson's help was needed in particular to determine the 'doubtful names' of languages. Why Grierson felt the LSI's Rough Lists were a good guide for rationalisation was because of the procedure he had thus far employed in the LSI in compiling them.

Between April and July 1896, Grierson had sent a circular with a form to all District-level officials (Assistant/Deputy Collectors and Magistrates) requesting them to list the names of "genuine local dialects".¹⁵ These were defined as "home-language, spoken in the country away from the cities". Respondents were instructed to provide all the names if multiple ones were used. If a local dialect had no special name but was known by a well-known general term, it was to be entered. It was emphasised that no language census was being requested, and that at this stage, all that was required was that officials provide a return for the names of speech varieties and short remarks about prior mention of these languages/dialects in government reports such as the Census, Settlement Reports, Gazetteers, as well as scholarly journals and books. A facsimile of the form that accompanied this circular is to be found in Figure 2, which Grierson requested district officials to have filled up "by some intelligent persons with the local modes of speech in the area under your charge."

The letter also informed officials that in this first stage of the survey, their responses would be compiled into "a rough preliminary catalogue of all the names by which every form of speech used in the area of the survey is known." This would then be followed by a second stage, which would consist of "collecting specimens of all the various forms of speech so catalogued; and of revising the first rough catalogue by the help of these specimens." This second stage was therefore one in which linguistic phenomena (mostly morphosyntactic) were used to confirm whether the names supplied actually identified distinctive language varieties. By October 1900, many of the specimens received had indeed been analysed, but since quite a few specimens still remained to be received, Grierson presents

¹⁴ H. H. Risley to G. A. Grierson dated 17.10.1900 (Mss Eur E223/260, Asian and African Studies, British Library).

¹⁵ Grierson to the Deputy Commissioner, Central Provinces and the Political Agent, Raipur, dated 04.04.1896 (Progs., Nos.V/11, 1896–1903, National Archives, Delhi).

No.	<i>Name of District, or other local area</i>	1	2	3	4
		Local name of the dialect or form of speech. To be written in the local Vernacular character and in the Roman character,	Name of language of which it is a dialect, if known. To be written in the Roman character.	Approximate number of persons speaking the dialect in the locality to which the return refers.	REMARKS.

DEPUTY COMMISSIONER'S OFFICE
1 APR 1896
SYLHET.

Fig. 2. Form accompanying the April 1896 circular

the classification he uses as provisional in his chapter in the 1901 Census report, which was the only time he ever wrote for a Census publication. In this chapter, he resolves the anxieties that Majeed in the quote above has noted about the Census's vis-à-vis by instituting a scheme in which the LSI was insulated from the pressure of strict compliance with the Census findings and did not have to consider the findings of the Census as a validation of its conclusions. Instead, it was the Census's responsibility to ensure that its findings were framed within the LSI's classifications and findings.

Such was the success of this stratagem that all three subsequent census reports — 1911, 1921, and 1931 — were careful to respect the LSI classifications; indeed, the Census seemed to believe that with the LSI making the advances in understanding that it was, there was nothing much left for it to do. By 1911, Census Commissioners were noting that “the question of language and dialect has now fully been dealt with”¹⁶ and by 1920, had declared that it was “doubtful” whether there was any scope left for “a detailed treatment of language” by it.¹⁷ The 1931 Census notes that post the publication of the LSI's important introductory volume in 1927, “the little that remains for a census to do with regard to the main language of the country beyond recording their corresponding increase and decrease, since the area of any given tongue is hardly ever stable” [Census 1933: 348].

The LSI archive is a valuable source to understand the process by which the LSI's imprint was stamped on the Census. Early on in 1900, Grierson supplied the then Commissioner, Herbert Hope Risley, with a paper detailing how the linguistic classification of Central Indo-Aryan (prominent amongst them, Hindi) was undertaken. Once Risley read it, he was driven to admit that “the entries in the language column of the Census schedule will be worth much”¹⁸, as the information the Census had collected about language, dialect, and region was not in any way superior to the LSI. In fact, the “statistical part of the Chapter is the least important as it is no case to be more than approximate.” He therefore recommends to Grierson that he:

¹⁶ Note for Report. Ch. IX. Language, p. 1, attached to letter from E. A. Gait to Grierson, dated 06.07.1911 (IOR/S/1/14/2, Asian and African Studies, British Library).

¹⁷ J. T. Marten to Grierson dated 11.03.1920 (IOR/S/1/14/1, Asian and African Studies, British Library).

¹⁸ H. H. Risley to G. A. Grierson dated 24.3.1901 (Mss Eur E223/260, Asian and African Studies, British Library).

⟨...⟩ should start writing on the basis of the first totals at once and make ...[the]... chapter a forecast in general terms of the conclusions which the Linguistic Survey will set forth in full philological detail. For a Census report, we want something that is generally understood, and which will be read by people who do not run to philology except in a general way. I should be inclined to start by an exposition of the general principles on which classification is based ⟨...⟩ Then, I would treat each group singly, bringing out their group character, and the characters distinguishing the languages, which make up the group. You would in fact treat each group much on the lines of the article which you sent me. ⟨...⟩ I will endeavour to fit in the ethnological data (forming of course part of a separate chapter) with your philological conclusions.

This is exactly the chapter that Grierson wrote for Census 1901, a text that was eventually to serve as a draft for his introductory volume of the LSI published in 1927 [Grierson 1927]. This chapter is a remarkable inclusion in a Census that is widely acknowledged for its application of ‘scientific racism’ in the way that it classifies Indians by racial types and argued for a racial origin of the Indian caste system, because it argues specifically, in its third paragraph [Census 1901: 249], against any attempt to view language in India as amenable to ethnological conclusions:

Nowhere are there presented stronger warnings against basing ethnological theories upon linguistic facts than in India. The “unholy alliance” between the two sciences has long been condemned, and has now fallen into disrepute, and I have, hence, in the following pages refrained so far as was possible from discussing questions of racial origin. When I have done so, it has only been to bring forward theories regarding the origin of nationalities which have been previously suggested by professed ethnologists, and to attempt to throw light upon them when they are confirmed by philology.

Grierson’s chapter [Census 1901: 250ff.] which proposes a tentative classification of languages and dialects based on philological grounds alone, is thus an important counterpoint to the overall racialist project in Census 1901.¹⁹ Although Risley’s anthropometric methods and con-

¹⁹ It is important to note here that the archive shows that it was Risley himself, in his letter dated 09.04.1902, who suggested to Grierson that he begin his discussion with

clusions have long been discredited, his racial interpretation of caste had a lasting and profound impact on how the British colonial government viewed and administered Indian society [Bates 1995]; indeed, many would argue that it has had even a longer afterlife, stretching into postcolonial India.²⁰

The manner in which Grierson discussed the language returns in the 1901 Census chapter marked the major change of listing languages by family that is to be seen in the Indian Census today. Figure 3 is an image of Table A of the 1891 census for Parent Tongue [Census 1891: 136]. This is to be contrasted with the 1901 overall table in Figure 4 [Grierson 1903: 2], in which there is, instead of a semi-random assortment of language names and population numbers, a structured list of language families provided. But it is not only in this table that the difference between the two censuses lies. Following the format suggested by Risley, Grierson then proceeded to discuss each group of languages, revealing in each section the LSI's tentative sub-groupings down to the dialect level along with speaker strength statistics provided to him by the provincial census commissioners. Figure 5 presents a few of these subsidiary tables to point out some general characteristics of Grierson's discussion.

The most important thing to note from these samples of subsidiary tables is that the framework for classification is provided by the LSI and

a paragraph on this relationship. The next few letters then cordially discuss the content of this discussion, with Grierson frankly expressing the exact same views one finds in the chapter. This suggests that whatever commitments Risley had towards racialised scientism, he did not harbour an opinion that it had to be necessarily applied as the sole explanation for all differences.

²⁰ Majeed [2019: 21–22] presents a picture of Grierson's politics that portrays his unquestioning acceptance of racial types and 'scientific' racism. I think this position does not place adequate weight on Grierson's oft-repeated pronouncements, both in private correspondence as well as in published work like the LSI and the Census chapter quoted above, where he explicitly rejects the use of ethnology, anthropology, and their anthropometric methodologies. In my view, while it is indeed true that Grierson seems to never have directly confronted the positive endorsements that many of his interlocutors gave to anthropometric methodology — in general, he was a conservative person, always committed to maintaining civility even in disagreement and careful not to cause affront to men who held high public office — the fact is that neither in the LSI or in the work he did for the Census did he 'racialise' his philological classifications (contra [Bates 1995]).

TABLE A.—Showing the GENERAL LINGUISTIC DISTRIBUTION of 10,000 of the POPULATION.

Language and Class.	Number per 10,000.	Language and Class.	Number per 10,000.
A. Indic-Aryan	7,460	F. Thibeto-Burman—cont.	
Hindi	3,269	Kathé (<i>Manipuri</i>)	3
Bengáli	1,578	Miri-Abor	1
Maráthi	721	Kúki	1
Panjábi	676	Lálung	1
Gujaráthi	405	Lúshai	1
Úriya	344	Thibetan	1
Úrdú	140	<i>Minor Languages</i>	1
Sindhí	99		
Western Pahári	58	G. Môn-Annam	9
Assamese	55	Môn (<i>Talaing</i>)	9
Central Pahári	44		
Márwádi	44	H. Shán (Tái)	7
Kachhi	17	Shán (<i>unspecified</i>)	7
Halabi	6		
Kashméri	1	J. Malayan	—
Goanese, &c.	1		
Eastern Pahári	1	K. Sinitic	27
<i>Minor Languages</i>	1	Karén	26
		Chinese	1
B. Dravidian	2,021	L. Japanese	—
Telugu	759		
Tamil	581	M. Eránic-Aryan	51
Kanarese	371	Pakhtú	41
Malayálam	207	Balóch	9
Gónd	53	Persian	1
Túlú	19		
Oráon	14	N. Semitic	2
Kandh	12	Arabic	2
Kodagu	2		
Mal-Pahádia	1	O. Turánic (<i>Úgrian</i>)	—
Bráhui	1		
<i>Minor Languages</i>	1	P. European-Aryan	9
		English	9
C. Kolárian	113		
Santhál	65	Q. Basque	—
Múndá and Kól	25	Ha mitic	1
Kórwá and Kúr	7	<i>Not distinguishable or not returned</i>	—
Bhíl	6		
Sáwara	4		
Kharría	3		
Baiga, Bhinja, &c.	2		
Gadaba	1		
D. Gipsy Dialects	15		
E. Khási Dialects	7		
F. Thibeto-Burman	278		
Burmese	212		
Arakanese	14		
Kachári	8		
Nipáli dialects	7		
Gáro	6		
Tippera	5		
Khyin	5		
Nága dialects	4		
Míkir	4		
Méch	4		

Fig. 3. The 1891 Census: Languages of India

		Number of languages spoken.	Number of speakers.
A. VERNACULARS OF INDIA.			
Malayo-Polynesian Family—			
Malay Group	2	7,831*
Indo-Chinese Family—			
Mōn-Khm̄r Sub-family	427,760	4	
Tibeto-Burman Sub-family	9,560,454	79	11,712,299
Siamese-Chinese Sub-family	1,724,085	9	
Dravido-Mundā Family—			
Mundā Sub-family	3,179,275	10	59,693,799
Dravidian Sub-family	56,514,524	14	
Indo-European Family, Aryan Sub-family—			
Eranian Branch	1,377,023†	3	221,157,673
Indo-Aryan Branch	219,780,650	22	
Semitic Family			
Hamitic Family	1	42,881
Unclassed Languages			
Anđamanese	1,882	1	5,530
Gipsy Languages	344,143	2‡	346,150
Others	125		
TOTAL VERNACULARS OF INDIA		147	292,966,163
B. LANGUAGES OF OTHER ASIATIC COUNTRIES, §			
AFRICA, AUSTRALIA	1	76,673
C. EUROPEAN LANGUAGES			
Language not returned	269,997
Language not identified, traced, etc	947,164
			101,059
GRAND TOTAL—INDIA		147	294,861,056

Fig. 4. The 1901 Census: Languages of India

Main Group.	Population returning it.
Non-Sanskritic . . .	54,425
Sanskritic . . .	219,726,225
TOTAL, Indo-Aryan Branch . . .	219,780,650

(A)

Group.	Population returning it.
Shinā-Khōwār . . .	54,425
Kāfir
Kalāshā-Pashai
TOTAL, Non-Sanskritic languages	54,425

(C)

Dialect.	Population returning it.	Dialect.	Population returning it.
Lahndā	350	<i>Dialects north of the Salt Range,—</i>	
Jatkī	220,836	contd.	
Dērāwāl	526,577	Sawāīnī	42,291
Chināwar	165	Jandāli	39,016
<i>Dialects north of the Salt Range.</i>		Khātri	34,980
Awānkāri	621	<i>Dialects south of the Salt Range</i>	
Dhannī	11,711	Kāchhrī	16
Ghēbī	74,082	Khētrānī	96
Hindkī	661,283	Mūltānī	961,997
Pōthwāri	226,542	Bahāwalpuri	580,036
Tināoli	2	Ṭhalōchhri	1,555
Rēshī	3,011	Ubhē-dī boli	1,924

(B)

Fig. 5 (A-C). The 1901 Census: Samples of subsidiary language tables

Of the above Jatkī is the language of the Jat population of Dera Ghazi Khan. Dērāwāl is found in Dera Ismail Khan. Chināwari is spoken along the Chenab. Nearly all the speakers of Awānkāri, Dhannī, Ghēbī, and Pōthwāri, belong to Rawalpindi. Hindkī is principally found in Hazara and Rawalpindi. Rēshī (valley of the Rēeh), Sawāīnī (valley of the Soan), Jandāli and Khātri (both north of Pindi Gheb) all belong to Rawalpindi. Bahāwalpuri is the Lahndā of Bahawalpur. Ubhē-dī boli is the Ubhēchi of Sindh.

Mārṣwārī.	
Where spoken.	Number of speakers.
Ajmer-Merwara	333,401
Assam	7,205
Bengal	10,663
Berar	41,521
Bombay	253,503
Central India	125,498
Central Provinces	30,941
Hyderabad	57,777
Punjab (excluding Bāgrī)	196,026
Rajputana (excluding Bāgrī)	3,712,262
United Provinces	8,050
Elsewhere	5,144
TOTAL, Mārṣwārī (excluding Bāgrī)	4,781,991
Bāgrī.	
Punjab	281,491
Rajputana	564,252
TOTAL, Bāgrī	845,743
TOTAL, Mārṣwārī (including Bāgrī)	5,627,734

(D)

Language.	Approximate Population speaking it.
Eastern Hindi	22,136,358

(E)

Language.	Population returning it.
Tableng	198
Tamlu	1,545
Mojung	152
Banpurā
Mutoniā
Mohongiā
Namsangiā
Mōshāng
Shānggō
TOTAL, Eastern Nāgā group	1,895

(F)

Fig. 5 (D–F). The 1901 Census: Samples of subsidiary language tables

not by the actual returns received by the Census. This leads to the grouping of speaker strength under purely theoretical parameters, based on the LSI's philological conjectures, concepts like that in (A), i.e. Sanskritic vs. Non-Sanskritic Indo-Aryan, rather than actual speaker returns noted down by the enumerators. Similar abstractions are denoted by the names 'Eastern Hindi' in (E), and 'Eastern Naga' Group in (F). Note also that in (F), even those dialect names are included which do not figure in the language returns. (D) represents the case of what the LSI considers to be two closely related languages, Marwari and Bagri. It is interesting to see that even though the language name Bagri is recorded by a significant number and reflected in the Census counts, Marwari is used as a language label, presumably to signify the closely related nature of these languages.

It will be obvious to the reader that the 1901 Census represents quite a different enterprise than both the Censuses that preceded it and the 2011 Census on several counts when it comes to recording MTs. While the summary table of Census 1901 (given in Figure 4 above), does not name and count every ‘dialect’, the subsidiary tables included in the chapter name for many if not all of them, including those that cannot be classified. Speaker strength was never a criterion that was employed to determine listing or naming of India’s languages.

As already mentioned, the fact that the 1911 and 1921 Censuses acknowledged that the LSI was the authority on the philological classification of Indian languages and dialects, and tried to implement its findings and classification in their reports. However, as Majeed shows [2019: 59–60], more often than not, the Census officials struggled to translate the LSI’s findings into their own system of enumeration, which prefers to count ‘languages’ rather than ‘dialects’, especially given that Grierson’s use of the distinction was solely determined by extralinguistic factors such as ‘possession of a literature’, ‘social prestige’ and the like. A consideration of the weaknesses of the language chapters of these Censuses would take me too further afield at this point, so I shall put that discussion aside, merely underscoring the point that Grierson’s import of the LSI framework into the 1901 Census did not actually build sufficient capacity in the Census organisation to analyse and classify language returns.

Furthermore, this establishment of the LSI’s supremacy also eclipsed two important changes in the Census methodology. The first important innovation that was initiated in the 1901 Census was the collection of information about bilingualism, a question which has continued to feature in the Census to this day. Enumerators were asked to elicit information about ‘Other Indian languages in common use’ as well as ‘whether literate in English’. A discussion of this measure is beyond the scope of this paper, but the fact that Grierson in the 1901 Census report does not discuss these returns at all, is indicative of how narrow his interest in the Census returns actually was.

The second innovation was that unlike the previous censuses, the 1901 Census asked the enumeration question as ‘language ordinarily used’ rather than parent-tongue or mother’s language as it had been in the 1881, 1891 Censuses. It was clarified further in 1911 that the language to be returned was that which each person ordinarily spoke in his own home. Only in the case of infants and deaf mutes was the mother’s language to be

entered. As Tieri [2024: 449] has pointed out (talking here about the 1961 Census which also used exactly this definition), this shift of language “is a subtle but profound change: differently from ‘mother’s language’, ‘language mainly spoken in the household’ implies a non-deterministic approach to linguistic identity. This definition locates mother tongue in language use (‘spoken in the household’), which is not an ascriptive feature but a deliberate action, and hence can be altered at will.”

Grierson does not appear to have been involved in effecting this change, as the archive shows him being approached by Risley only after enumeration had already been completed. But while there may not have been any direct connection that can explain this change, the fact is that the LSI had for the past few years been using that very definition of language cannot be a mere coincidence. In his April 1896 circular (see footnote 15 on page 30), Grierson had defined genuine local dialects in terms of use in the home domain:

I hope that it will be clearly understood that the names which I want entered are the names of the genuine local dialects. In some parts of India, people are more or less bilingual. **A man will speak one language in his own house and amongst his familiar friends and another language, a kind of court language, when talking in the city, bazaars, in a public court, or to a superior. It is the former language which I wish to have recorded. The other may be recorded too, provided its use is carefully stated, but the important one is the home language spoken in the country, away from the cities.** To take a concrete example, in Allahabad, for instance, the language spoken in the courts, in the bazaars and to European gentlemen is no doubt Hindustani, but the language or languages of the surrounding district differ widely from Hindustani. According to one excellent authority, no less than six different dialects are current in that single district. It is the local name of these dialects which I wish to have entered in the form. Hindustani should only be entered if it is the home language of any particular class or group of people, for example, of the Musalman population of any town or tract.

It is therefore not unreasonable to conjecture that the Census’s redefinition of the term as the ‘language ordinarily used in the home’ was influenced by the LSI. However, while for the LSI, the definition was crucial in enabling the officials to identify the linguistic object they were to report

on, it is not at all obvious exactly what impact the change in definition had on the actual Census returns, because there is no proof that the enumerators actually framed the language question in the use-based manner that was intended. Further, the use of LSI classification as a framework naturally constrained the results.

4. Decolonial thinking in the 1961 Census

The Second World War brought an abrupt halt to the collection and classification of language and mother-tongue data in the 1941 Census, which was the last Census conducted by the British, as India won its independence from colonial rule in August 1947. The large-scale social tensions, upheavals, and migrations triggered by the Partition of India and Pakistan entailed that the data for the 1951 Census was collected in a time of great disturbance.

The 1951 Census of India recorded approximately 783 mother tongues. The data for mother tongues were generally tabulated by large administrative units called “tracts” (usually consisting of several tehsils or police stations) rather than by individual villages, although special sorting was later done for certain multilingual border areas. By the mid-1950s, however, as the country gradually stabilised, confidence grew in the provisions of the new Constitution and in the commitment to realise its promises of universal education and social and cultural equality. In this climate, the Census was identified as a key instrument for gathering knowledge about the ‘real’ India.

To my mind, the 1961 Census vis-à-vis at least the language census data represents India’s first and perhaps only Census that represents decolonial thinking [Mignolo 2021], as it seeks to eliminate “coloniality” from the way that Indians and the state that they had recently created could come to understand and count the languages spoken within its area. In a mammoth effort that spanned over nearly a decade, a team of the Census Commissioner Asok Mitra and the linguist R. C. Nigam, advised by the great linguist Suniti Kumar Chatterjee, thoughtfully tried to free language data collection and interpretation from the two distinct strands of epistemology that were its inheritance from the colonial period. The scheme that they came up with ensured the counting and naming of 1652 MTs returned, including those that were spoken by small numbers of people.

Considerations of space prevent me from dwelling on the 1961 Census methodology at length, so I shall focus on just two aspects of why I think the Census of 1961 is an inspiring moment of decolonial thinking, however short-lived, in nascent India. The first represents the manner in which it embraces an epistemic delinking from both the LSI and the Census practices of the colonial past. In an outstanding critique of the colonial distinctions between ‘language’ and ‘dialect’, Nigam [1971: xxiv–xxx] argues that the latter term has no scientific validity, either as an object of Census enumeration or of linguistic analysis. Discarding popular definitions that treat dialects as inferior, substandard, illiterate, scriptless, or ungrammatical forms of a language, Nigam points out that even though the LSI shows awareness of the status-based nature of this distinction and laid down criteria to control for this, the LSI still contains many lapses, as “uniformity of following his own laid-down criteria was never a virtue with the great Editor of the Linguistic Survey Volumes.” He therefore recommends abandoning this distinction altogether, as “pursuing the line of ‘language/dialect’ discussions is not likely to yield any concrete results (<...> [as to distinguish the two] (...)) there have always got to be considerations taken into account which fall outside of strictly linguistically structural or grammatical considerations” [Ibid.: xxx].

The importance of this move for the 1961 Census needs to be underscored, as the 1961 Census was committed to the LSI system of analysis and only then classification as a means of identifying distinct MTs [Nigam 1971: xv]. Rather than a simplistic rejection of the classification method in linguistics, this rejection of the language/dialect distinction creatively **relinks** the LSI to the Census, as providing historically relevant information about MT returns. Whereas the 1901, and subsequent Censuses, imposed the LSI framework onto the Censuses and rendered the actual MT returns invisible, the 1961 Census treats returns of MTs listed in the LSI at par with other MT returns given by the people of India. As a consequence, the final report names and lists the statistics of all 1652 mother tongues.

Perhaps the most striking decolonial move of all is how the 1961 Census, invoking the LSI, envisions a new role for the Census of India. Figure 6 [Census 1964: ccxxxiv] reproduces an importable table from the final Census report, which specifies the results obtained from a scrutiny of MT statistics when the LSI information is integrated with the other MT returns.

	No. of tongue	Total No. of speakers
1. No. of mother tongue returns	1,652	438,936,918
2. No. of mother tongue attested, in Grierson's Linguistic Survey of India classifications	572	436,224,545
3. No. of mother tongues not traced in the L.S.I. and <i>tentatively classified</i>	400	425,076
4. No. of mother tongues attested in the Linguistic Survey of India but <i>tentatively reclassified</i> .	50	1,908,399
5. No. of mother tongues considered unclassifiable	527	62,432
6. Foreign mother tongues	103	315,466

Fig. 6. The 1961 Census: Mother tongue statistics

Seen in the light of the comment made by the Census Commissioner, Asok Mitra's introductory note to the 1961 Census, which makes it clear that the Census officials planned to undertake "a more extended long-term survey, in the manner of Grierson's LSI (...) in the intercensal period and would bring in a richer harvest to 1971 and 1981" [Census 1964: iv], it is obvious that Census 1961 saw its role in the next twenty years to be one of orchestrating a new linguistic survey of India, presumably through the Language Division that was instituted in 1961. Such a survey was warranted because Mitra recognised that the 1961 report "stands for work in progress and is no more than a tentative presentation", and also that the Census had still to convince the government: "I must accept responsibility for my share in the statements made and the conclusions drawn, even in the Linguist's Note, which do not necessarily represent the views of the Government, insofar as misstatements, errors or misplacements of emphasis are concerned" [Ibid: v].

Clearly, however, efforts to convince the government were not successful, as even though Nigam's [1971] invaluable "Language handbook on mother tongues in Census" was published by the Office of the Registrar-General, which detailed notes on all the 1652 returns, when the 1971 Census report on Languages did finally appear, it was in the form described

in Section 2 of this essay, rather than as a continuation of the work initiated by the 1961 Census. Although it is far from certain that the Census and its Language Division, or for that matter any consortium of linguists, could have succeeded in this endeavour, it is worth closing this essay with some speculations from recent politics as to why the direction given by the 1961 Census was not pursued.

5. Decoding the politics of the change in the Census methodology

Why was India's decolonial moment so short-lived? Why is it that despite the fact that the 1961 Census revealed that so many Indian mother tongues lay outside the classifications provided by the LSI, why is it, since 1971, cited as the sole document that the Census of India relies on? The answer, I believe, lies in both policies as well as the way that politics over the last five decades has instrumentalised certain key provisions in the Indian Constitution.

In a fine essay, Agnihotri [2015] has shown how the various tensions in the Indian Constituent Assembly debates ensured that most of the linguistic decisions taken by the Constituent Assembly ended up in either ignoring, or in some cases even weakening "the multilingual and multi-cultural ethos that is constitutive of Indian society" [Ibid.: 47]. Although several important provisions for the protection of linguistic minorities were pushed through, the compromises that had to be made in the processes of winning them have ensured the survival of other articles that sufficiently diminish the actual protection given to the rights of linguistic minorities.

Consider first the apparent protections. The Indian Constitution protects linguistic minorities through Articles 29 and 30, which guarantee the right to conserve their distinct language, script, and culture and to establish and administer educational institutions of their choice. Article 347 allows the President to recognise a language used by a substantial portion of a state's population as an official language. Additionally, Article 350A mandates the provision of facilities for instruction in the mother tongue at the primary stage for all linguistic minorities, and Article 350B provides for a Special Officer to report on the measures for the protection of linguistic minorities.

It will be obvious to the reader that for these protections to be delivered, there has to be an actual count of the number of linguistic minorities. As the Constitution does not name or list them, the only means by which such minorities can be enumerated is the Census, which, as we have seen, has, since 1971, developed a number of ruses to mask minority status. The purposeful undercounting of Indian MTs by the post-1971 Census is at least partially determined by the Indian state's panic-stricken response to a decolonial Census actually enumerating all the minorities that exist. Hence, the decision of refusal to count below the 10,000 speaker strength threshold and to discard those varieties not covered by the LSI.

Furthermore, although Article 350B is procedurally implemented in the appointment of a Special Officer and although it does define linguistic minorities — “any group or groups of people whose mother tongues are different from the principal language of the State, and at the district and taluka/tehsil levels, different from the principal language of the district or taluka/tehsil concerned” — it leaves it to the States / Union Territories to define them. No list of minority languages is ever compiled by this office using the information from the Censuses, nor does it exercise the power to demand the inclusion of minority languages in the State / Union Territories. A consideration of this office's last annual report in 2014–2015 available online²¹ shows that it has no justiciable objection to most States / Union Territories, barring a handful of, mainly identifying other Scheduled Languages, as the minority languages of their regions, paying scant attention to even the minorities that the Census does identify. Again, the post-1971 Census's strategy of grouping MTs under language labels helps in this endeavour, particularly if the MT is grouped under a Scheduled Language.

As we have seen, there is no other language label that is a beneficiary of the post-1971 strategy of grouping under a language label than ‘Hindi’. The steady inflation of the number of MTs it labels is the result of the weaponisation of the most problematic article of all, Article 351 (emphasis mine):

It shall be the duty of the Union **to promote the spread of the Hindi language**, to develop it so that it may serve as a medium of expression for all the elements of the composite culture of India and to secure its enrich-

²¹ http://164.100.166.181/annualreport/52ndReport_CLM_English.pdf (accessed on 10.10.2025).

ment by assimilating without interfering with its genius, the forms, style and expressions used in Hindustani and in the other languages of India specified in the Eighth Schedule, and by drawing, wherever necessary or desirable, for its vocabulary, primarily on Sanskrit and secondarily on other languages.

Article 351 was originally intended to achieve linguistic harmony and national integration by developing Hindi's connections to the originally 14 languages listed in the Eighth Schedule, but as the years have progressed, language politics in India has changed the meaning of inclusion in the Eighth Schedule. The schedule has been expanded through amendments to now total 22 languages, such as Sindhi (1967), Konkani, Manipuri, and Nepali (1992), and Bodo, Dogri, Maithili, and Santhali (2004). Languages in the Schedule are eligible for government support in development, use in secondary education, and for official purposes, but none of these 'advantages' invest such overwhelming power that there should be 38 languages queueing up for inclusion in the Schedule. The real reason for this aspiration is the sad fact that, since 1971, the only way a citizen can get to be recognised as a speaker of a 'language' is if its name is included in the Eighth Schedule. Because then, and only then, the state's Census methodology tells her, can she truly count.

Abbi, Gupta, and Aggarwal [Abbi et al. 1995] point out that the Eighth Schedule is best characterised as the state's elevation of a select group of Indian languages into an elite tier onto which it confers legitimacy. In the process, they argue that it constructs a hierarchy where languages outside the Schedule are relegated to the status of dialects. This has installed, what Abbi [1995] calls, a system of 'linguistic apartheid', where the 10,000-speaker threshold serves as a gatekeeper to political and social visibility. For a speaker of a minority or tribal mother tongue, the Census methodology ensures that their linguistic identity is either othered or forcibly subsumed under a dominant "language label" like Hindi. In this framework, the Eighth Schedule is not a tool for preservation but a mechanism of functional displacement, forcing citizens to abandon their native tongues for the prestige of a Scheduled language in order to be recognised by the state. It is this system of structural exclusion that the Census seeks to reproduce every decade through its practice of counting language labels rather than languages.

6. Conclusion

This article has examined the history of the enduring problem of the undercounting of India's minority languages and the ensuing confusion as to exactly how many languages are spoken in the country. It argues that this undercounting is not the result of an inheritance of colonial language policy, by showing how the LSI disrupted colonial categories of thought, and actually forced the counting of all minority languages for over three Censuses. It shows that barring the 1961 Census, where the aspiration to strengthen the LSI's legacy into a truly decolonial one was first showcased, the role adopted by the Census is to play handmaiden to the state. It does not therefore count the 'mother tongue' at all, insofar as it is a native speaker's self-identification, but only invented language labels.

Abbreviations

LSI — Linguistic Survey of India; MT — mother tongue; PLSI — People's Linguistic Survey of India.

References

- Abbi 1995 — A. Abbi. Language contact and language restructuring: A case study of tribal languages in Central India. *International Journal of the Sociology of Language*. 1995. Vol. 116. No. 1. P. 175–185. DOI: <https://doi.org/10.1515/ijsl.1995.116.175>.
- Abbi et al. 1995 — A. Abbi, R. S. Gupta, K. S. Aggarwal (eds.). *Language and the State: Perspectives on the Eighth Schedule*. New Delhi: Creative Books, 1995.
- Agnihotri 2015 — R. K. Agnihotri. Constituent Assembly debates on language. *Economic and Political Weekly*. 2015. Vol. 50. No. 80. P. 47–56.
- Bates 1995 — C. Bates. Race, caste and tribe in Central India: The early origins of Indian anthropometry. P. Robb (ed.). *The concept of race in South Asia*. Delhi: Oxford University Press, 1995. P. 219–259.
- Goswami 2004 — M. Goswami. *Producing India: From colonial economy to national space*. Chicago: Chicago University Press, 2004. DOI: <https://doi.org/10.7208/9780226305103>.
- Grierson 1903 — G. A. Grierson. *The languages of India: Being a reprint of the chapter on languages*. Calcutta: Superintendent of Government Printing, 1903.
- Karthick Narayanan 2025 — R. Karthick Narayanan. *Multilinguality, vitality, and endangerment: Insights from the lived experiences of multilingual India*. Cambridge: Cambridge Scholars Publishing, 2025.
- Kidwai in press — A. Kidwai. The Linguistic Survey of India: The prequel. R. K. Agnihotri (ed.). *Census, survey, policy*. Delhi: Orient BlackSwan. In press.

- Majeed 2019 — J. Majeed. *Colonialism and knowledge in Grierson's Linguistic Survey of India*. London; New York: Routledge, 2019. DOI: <https://doi.org/10.4324/9780429439230>.
- Mignolo 2021 — W. Mignolo. *The politics of decolonial investigations*. Durham, NC: Duke University Press, 2021. DOI: <https://doi.org/10.1215/9781478002574>.
- Nigam 1971 — R. C. Nigam. *Language handbook on mother tongues in census*. Delhi: Manager of Publications, Government of India, 1971.
- Tieri 2024 — S. Tieri. The census amid language politics: The production of language data and the politics of mother tongue in the Census of India, Punjab, 1941–71. *South Asia: Journal of South Asian Studies*. 2024. Vol. 47. No. 2. P. 434–454. DOI: <https://doi.org/10.1080/00856401.2023.2275444>.

Sources

- Census 1891 — *General report on Census of India, 1891*. London: Eyre and Spottiswoode, 1893.
- Census 1901 — *Census of India, 1901*. Vol. 1. *India*. Pt. 1. *Report*. Calcutta: Superintendent of Government Printing, 1903.
- Census 1933 — Hutton J. H. *Census of India, 1931*. Vol. I — *India*. Pt. I — *Report*. Delhi: Manager of Publications, 1933.
- Census 1964 — *Census of India 1961*. Vol. I. Pt. II-C(ii): *Language tables*. Delhi: Government of India, 1964.
- Census 2011 — *Census of India 2011*. Paper 1 of 2018: *Language*. Delhi: Government of India, 2018.
- Ethnologue — D. M. Eberhard, G. F. Simons, Ch. D. Fennig. *Ethnologue: Languages of the world*. 28th edn. Dallas: SIL International, 2025. Available at: <https://www.ethnologue.com/> (accessed on 10.10.2025).
- Grierson 1927 — G. A. Grierson. *Linguistic Survey of India*. Vol. 1. Pt. 1: *Introductory*. Calcutta: Government of India, 1927.

Получено/received 17.11.2025

Принято/accepted 29.12.2025