

August 5, 2021

India's Linguistic Diversity

How the Census obscures it

By: Shivakumar Jolad, Aayush Agarwal

The Indian census' classificatory practices obscures the full extent of language diversity in the country. Including what the census calls "mother tongues" presents a fuller picture.

India is one of the most linguistically diverse countries, [ranking 4th in terms of the number of language](#) spoken, according to the Ethnologue Language Catalogue of the world. Yet, enumerating its languages has been contentious, given the implications of legitimising and delegitimising linguistic identities. The Indian Census' way of classifying of languages occludes this diversity. We present here a truer picture by including in our calculation of diversity what the census calls “mother tongues”.

Is it a language? Is it a mother tongue?

Counting and classifying the languages and dialects in India has been challenging for both linguists and state agencies from colonial times to the present. In many North Indian states, for instance, people distinguish between *Bhasha* (language) and *Boli* (spoken language). They might Hindi as their language in official documents, while their language may actually be, say, Bhojpuri. This is roughly in line with *granthika*, the literary variety, and *vyavaharika*, the colloquial variety.

The [Linguistic Survey of India conducted between 1894-1903](#), led by Sir George Abraham Grierson, reported 179 languages and 544 dialects spoken in India. The [People's Linguistic Survey of India \(2010-2012\)](#) reported the existence of 780 languages (without making a distinction between language and dialect). The [Ethnologue classified 447 languages in India](#).

Forms which vary regionally or are used exclusively by lower classes or castes are relegated to the status of dialects.

Disagreements about the classification of languages and dialects leads to these divergent estimates of languages in India. Linguists generally distinguish between the two on the basis of mutual comprehension. But often, the distinction between language and dialect is more political than a linguistic question. The power and prestige of certain selected forms, widely used across large geographic regions, often acquire the status of a language, whereas forms which vary regionally or are used exclusively by lower classes or castes are relegated to the status of dialects. If mutual intelligibility is the criteria, Hindi, Urdu, and Punjabi should have been the same language, whereas Rajasthani and Kumauni should be different.

The [Census of India in 1951](#) attempted to steer clear of controversy and counted only what people declare as their “mother tongue”, and reported the existence of at least 783 mother tongues. After 1971, the censuses stopped listing languages spoken by less than 10,000 people. These languages are lumped into the “other languages” category. Similarly, the mother tongues spoken by less than 10,000 are placed under the “other mother tongue” category. The process of rationalisation of language data, and the “other” categorisation leads to the invisibilisation of languages spoken by ‘minority’ people. Census 2011 reported 19,569 raw mother tongue returns, [which were ‘rationalised’ into 1,369 mother tongues](#), and then regrouped into 270 mother tongues (each spoken by more than 10,000 people) and 121 languages. The languages are further grouped into 22 Scheduled (listed under Schedule VIII of the Indian Constitution) languages — which receiving state patronage — and 99 non-Scheduled languages.

The Census' process of enumeration does not “represent the full linguistic diversity of India; in fact, it minimises it.”

Collectively, these “other languages” are spoken by 1.87 million people, and “other mother tongues” by 18.6 million people ([Census of India 2018](#)). Under the “Hindi” language alone, there are 16.7 million speakers who speak ‘other mother tongues.’ The population of some of the ‘minoritised’ languages, such as Bhili (with 10.4 million speakers) and Gondi (with 2.98 million speakers, based on Census 2011 figures), exceeds that of many small countries, including the country with the most languages spoken— Papua New Guinea. The

mother tongues under Hindi Bhojpuri (50.6 million speakers), Rajasthani (25.8 million), and Chhattisgarhi (16.2 million) have speakers comparable to Spanish speakers in Spain, Polish in Poland, Dutch in the Netherlands, and yet, they do not have the status of ‘language’ according to the Indian Census.

As is evident from these stark discrepancies, the Census’ process of enumeration does not “represent the full linguistic diversity of India; in fact, it minimises it” (Kidwai 2019).

An occluded diversity

Despite all its limitations in classification, undercounting, and exclusion, the census classification still shows enormous linguistic diversity in India and within Indian states. The languages broadly belong to the following families: Indo-European (Indo-Aryan, Germanic (English), Iranian (Persian, Afghani)), Dravidian, Austro-Asiatic, and Tibeto-Burmese. These, along with Semito-Hamitic (Arabic), constitute the main languages in India. The most diverse language group is the Tibeto-Burmese group with 66 languages. The most spoken languages belong to the Indo-Aryan branch with 78% of the population, Dravidian 20%, Austro-Asiatic languages 1.1%, and Tibeto-Burmese just 1%, spoken in North-Eastern India (see Table 1).

Language Family and Speakers in India– 2011

Census 2011: Language Family of Scheduled and Non-Scheduled Languages

Sl no	Language Family	Sub family	Number of Languages	Number of speakers	Percentage of Speakers
1	Indo-European	Indo-Aryan	21	945,052,555	78.05%
		Iranian	1	21,677	0%
		Germanic	1	259,678	0.02%
2	Dravidian		17	237,840,116	19.64%
3	Austro-Asiatic		14	13,493,080	1.11%
4	Tibeto-Burmese		66	12,257,382	1.01%
5	Semito-Hamitic		1	54,947	0%

Table: Shivakumar Jolad • Source: Census 2011 • Created with Datawrapper

Table 1: Language families and speakers in India

“Many ‘mother tongues’ so defined would be considered a language rather than a dialect by linguistic standards” (Bhattacharya 2002). Accordingly, we would expect the linguistic diversity of the mother tongue to be much higher than the language diversity based on Census Data. We compute India’s linguistic diversity at the national and state levels; at the two hierarchical levels of languages provided by Census 2011 — language and mother tongue — and for Scheduled languages and non-Scheduled languages.

For this, we use the [Greenberg’s Diversity Index](#), a validated instrument that measures linguistic diversity. The Greenberg’s Linguistic Diversity Index (LDI) is a number between 0 and 1. It is 0 when every person in the region under consideration speaks the same language and reaches the maximum value (close to 1) when every language has an equal number of speakers. The Linguistic Diversity of India (LDI): Language and Mother Tongue plot for India and its states is shown in Figure 1.

Language and Mother Tongue Diversity of India

Greenberg's Linguistic Diversity Index

Language Mother Tongue

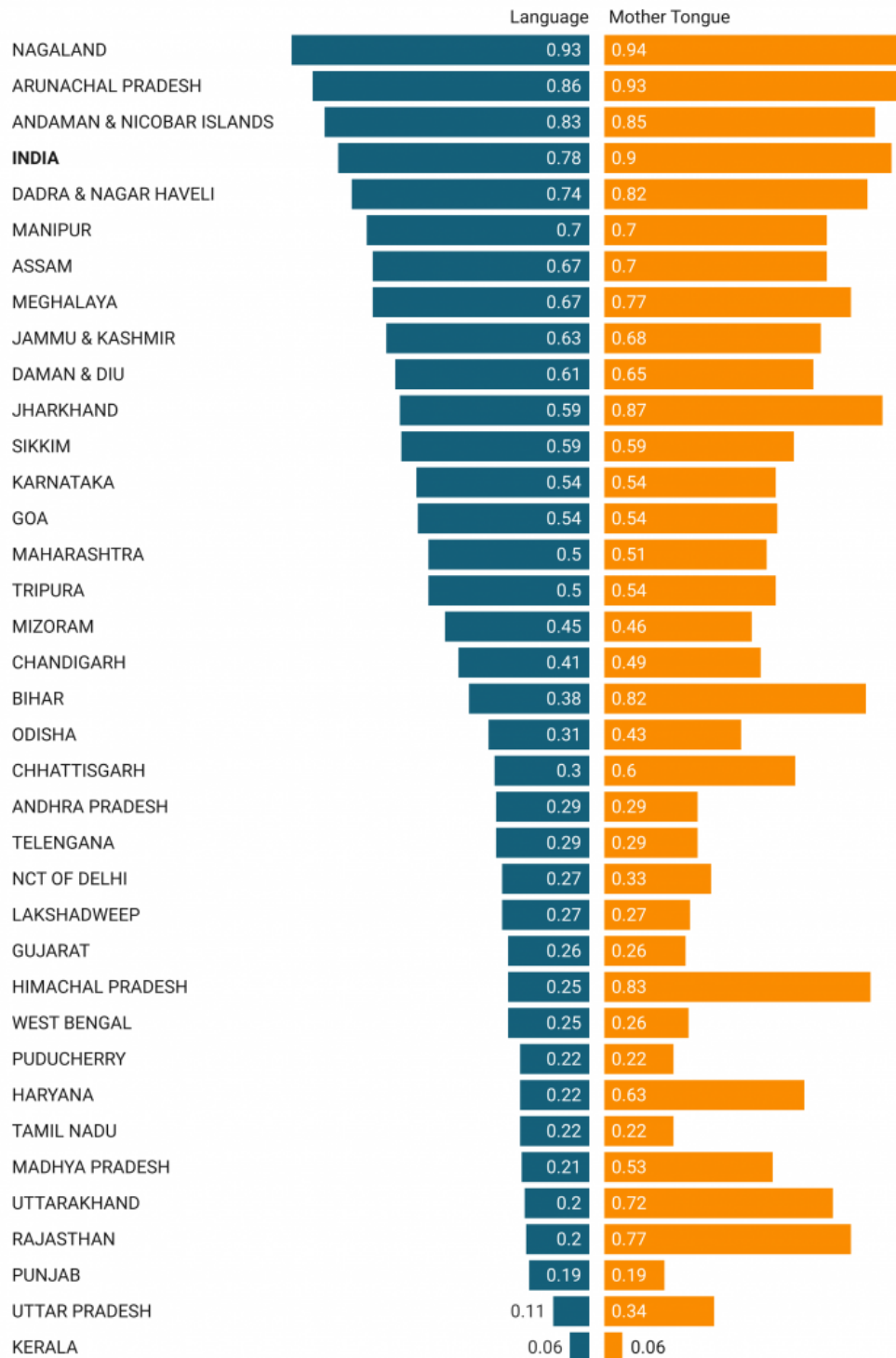


Chart: Shivakumar Jolad • Source: From: Census - C16 Languages of India • Created with Datawrapper

Figure 1: Linguistic Diversity of Indian States (left) LDI: Languages (right) LDI: Mother Tongue, computed using state-level data from Indian Census 2011 language tables.

Linguistic Diversity of India

Greenberg Linguistic Diversity (Language and Mother Tongue)

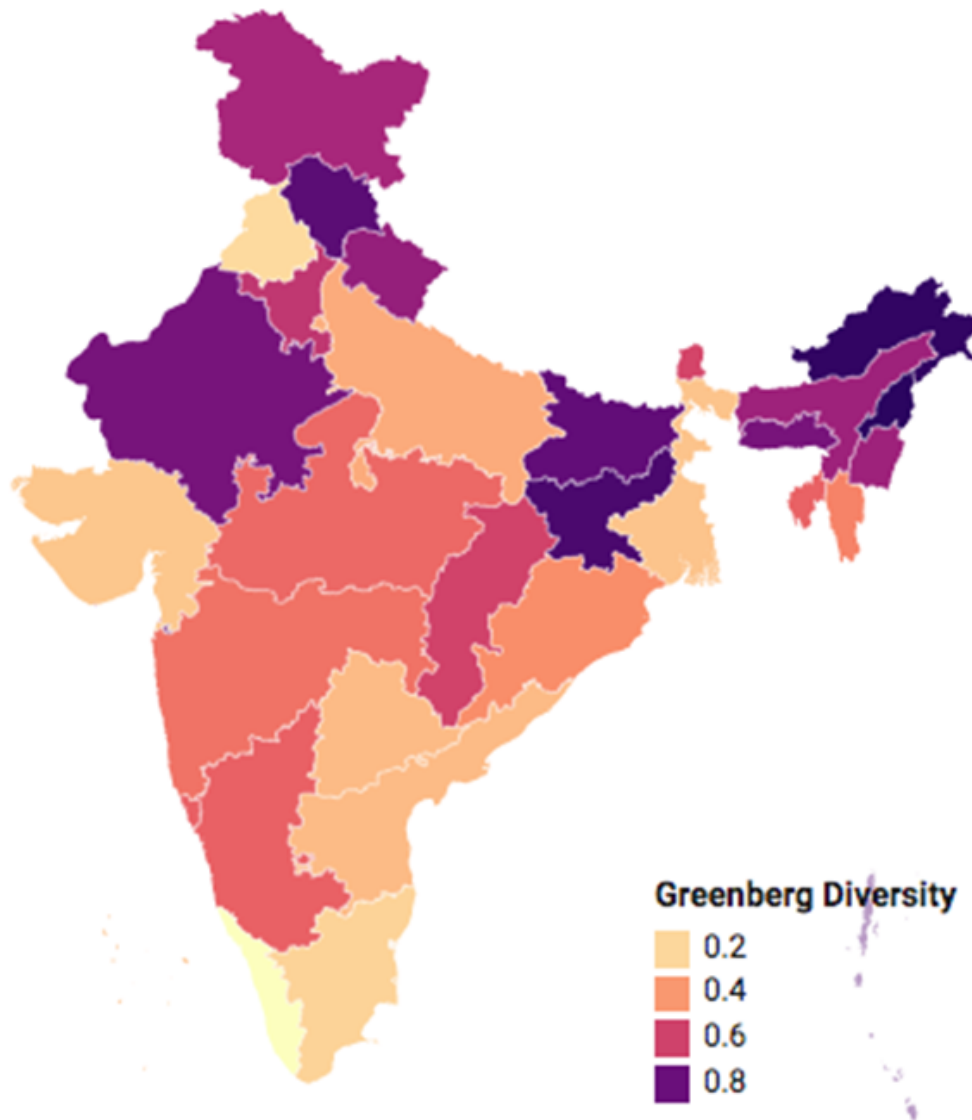


Figure 2: Map of Linguistic Diversity of India. Based on Greenberg's Index

The LDI of India, considering the restrictive 121 language classification of Census 2011, is 0.78. If we include mother tongues, diversity rises to 0.9. The latter is closer to the [diversity index of 0.93 calculated by UNESCO](#) using Ethnologue's catalogue of 425 languages (UNESCO 2009), which includes more languages — for instance, many of the 197 endangered languages spoken by less than 10,000 people.

The most linguistically diverse states in India are Nagaland and Arunachal Pradesh, followed by tribal areas and islands like Andaman and Nicobar. Small states with distinct ethnic and tribal groups show greater linguistic diversity than large, densely populated states like Uttar Pradesh and Kerala, where there is homogenisation induced by diffusion of people and culture over time. In Kerala, 97% of the population speaks one language: Malayalam.

NAGALAND- Languages spoken by the people

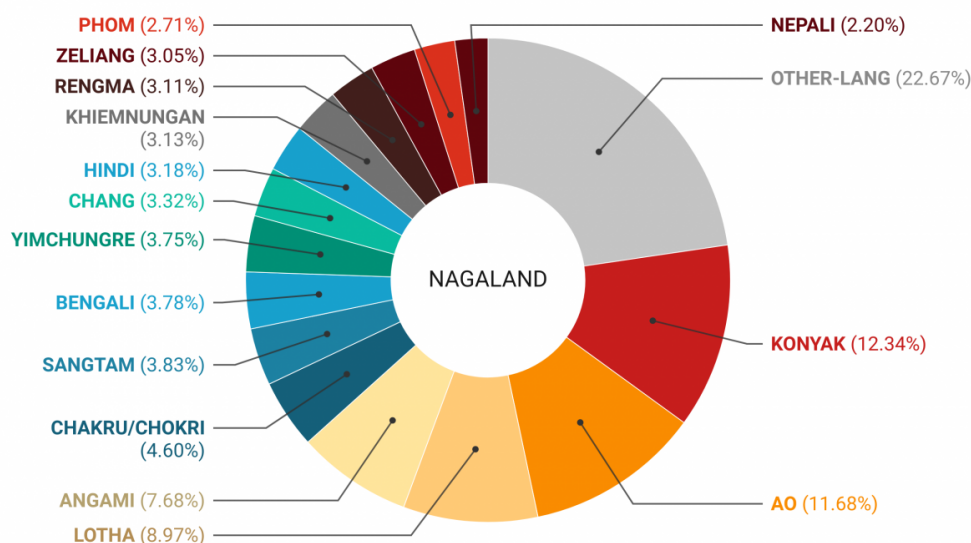


Chart: Shivakumar Jolad • Source: Census 2011 • Created with Datawrapper

KERALA- Languages spoken by the people

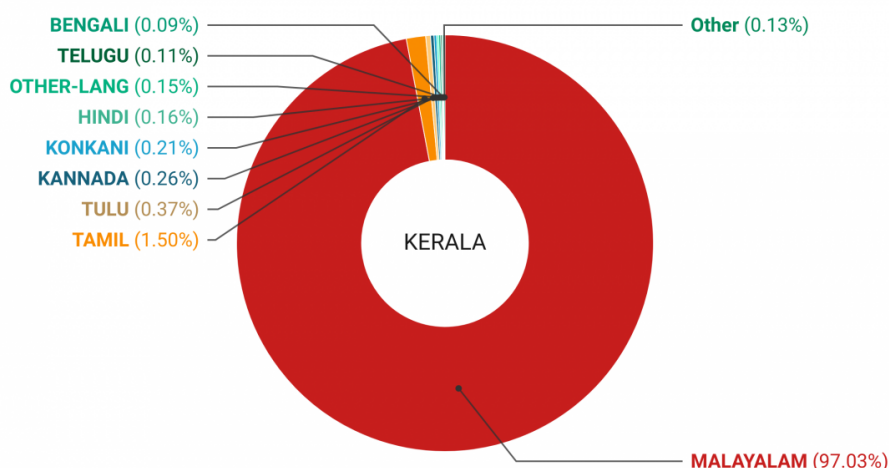


Chart: Shivakumar Jolad • Source: Census 2011 • Created with Datawrapper

Figure 3: Languages spoken in the most diverse (Nagaland) and the least diverse state (Kerala) of India

Homogenisation of the dialects of Hindi

The gap between LDI-language and LDI-mother tongues is particularly stark for states in the ‘Hindi’ belt of north, central, and western India, which have several mother tongues that have been grouped under the rubric of ‘Hindi’. The LDI in these states is low primarily because 54 languages are treated as mother tongues under Hindi. Of the top 10 states ranked according to highest mother tongue diversity, nine have Hindi as the official state language.

Take Uttar Pradesh, the largest state of India, which has 94% of its population speaking Hindi as the dominant language, followed by Urdu at 5.4%. It has a low LDI of 0.11. But the state has a significant proportion of speakers of mother tongues classed under Hindi: Bhojpuri (10.8%), Awadhi (1.9%), Bundeli/Bundelkhandi (0.65%), and Brajbhasha (0.36%), among others. If all these mother tongues are included, UP’s LDI rises by three times to 0.34.

Hindi States and India: Difference between Language and Mother Tongue diversity

Greenberg's Linguistic Diversity Index

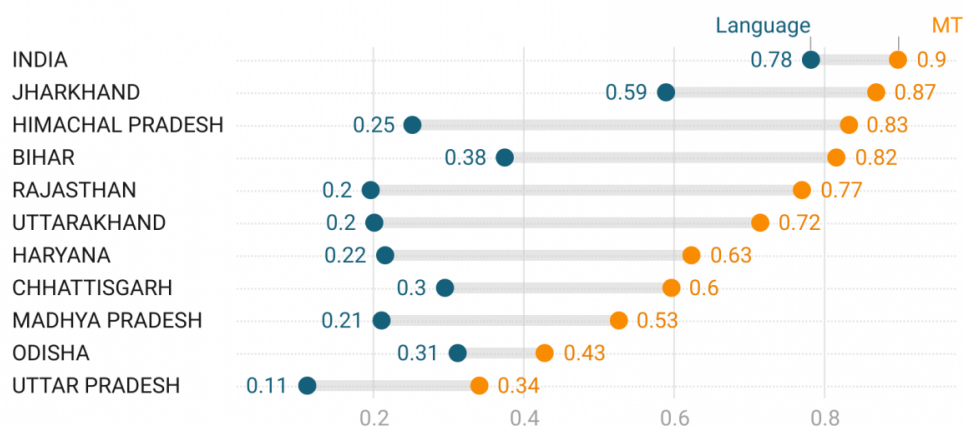


Chart: Shivakumar Jolad • Source: From: Census -C16 Languages of India • Created with Datawrapper

Figure 4: Hindi States - Language and Mother Tongue Diversity

In Himachal Pradesh, Hindi as a ‘language’ has 85.89% of the population as speakers. But Hindi as a ‘mother tongue’ (15.68%) comes third after Pahari (31.9%) and Kangri (16.25%). Both the latter languages are clubbed under Hindi, along with Mandeali, Kulvi, Bharmouri/Gaddi, Chambeali/Chameli, and Sirmauri. Himachal’s LDI-mother tongue (0.83) is more than three times larger than LDI-language (0.25).

Such homogenisation of mother tongues under dominant languages diminishes the linguistic rights of their speakers. In the Hindi belt states, the administrative language and medium of instruction in schools remain Hindi, undermining those who speak languages like Bhojpuri, Pahari, Maithili, Rajasthani, and Marwari.

Many of India’s excluded languages and mother tongues fall under UNESCO’s endangered languages list, and are mostly spoken by people in the North East and in the tribal states of India. For example, Aka/Hruso and Koro languages in Arunachal Pradesh, Asur in Chota Nagpur belt, Nahali in Madhya Pradesh fall under the ‘definitely endangered’ group, Atong in Meghalaya, and Geta in Odisha fall under the ‘severely endangered’ languages. The excluded ‘other’ minority and foreign languages spoken within India include Portuguese, French, and German.

State recognition

The promotion of certain selected languages through the Constitution and state machinery undermines the rights of ‘minoritised’ linguistic communities throughout India. The State Reorganisation Committee’s (SRC) recommendations in 1953 on the linguistic reorganisation of states stirred fears about the suppression of rights of linguistic minorities. The linguistic reorganisation of the states in 1956 aimed to create states with homogenous linguistic identity even though multilingual groups co-existed in every administrative region in India.

|| Linguistic reorganisation brought most of the scheduled languages “home” under a state umbrella.

With the exceptions of some states adopting non-scheduled languages as their official language, like Kokborok in Tripura and Mizo in Mizoram, linguistic reorganisation brought most of the scheduled languages “home” under a state umbrella (Pandey 1997, 81). Although the SRC sought to secure political and economic justice to linguistic groups, the language so secured was the dominant language at the state-level and not the diverse set of languages spoken within the state. States continue to promote their politically dominant languages in political and administrative affairs, largely ignoring linguistic minorities.

Inclusion under the Eighth Schedule, which obligates the government to use and promote the language, has largely occurred through political considerations and lobbying. In 1950, there were 14 languages under the schedule. This has now expanded to 22 through successive constitutional amendments. The 21st amendment of 1967 added Sindhi; the 71st Amendment of 1992 added Nepali, Manipuri, and Konkani; and the 92nd constitutional amendment of 2004 added Maithili, Dogri, Santali, and Bodo. India’s Scheduled Languages are mostly from the two dominant language families. Only Sanatali from the Austro-Asiatic family and Bodo and Manipuri (Maitei) from the Tibeto-Burmese family are on the list.

A dendrogram of Scheduled Languages and Mother Tongues, scaled to the population (logarithmic scale) is shown below (see Figure 5).



Figure 5: Scheduled Languages and their Associated Mother Tongues in India, based on Census 2011 Classification (The height of the bar represents the number of speakers on a logarithmic scale). Data Source: Census 2011, C-16 Series.

Non-scheduled languages

There is a demand for 38 languages to be included in the 8th schedule. But there is no “right to be scheduled” based on sheer numbers of speakers, even when the 2011 Census shows that, 10 'non-scheduled languages' have more than one million native speakers, and 31 have more than 100,000 speakers. In all, census 2011 lists 99 non-Scheduled languages (which includes English) and 147 mother tongues spoken by more than 10,000 people. The largest of these are Bhili/Bhilodi with 10.4 million native speakers, Gondi (3 million), Kurukh (2 million), and Khandeshi with 1.86 million speakers (see Figure 6), exceeding the population of certain European countries. Yet Sanskrit is in the official schedule even though only 24,821 people claim it as their language.

Non-Scheduled Languages in India (Top 25)

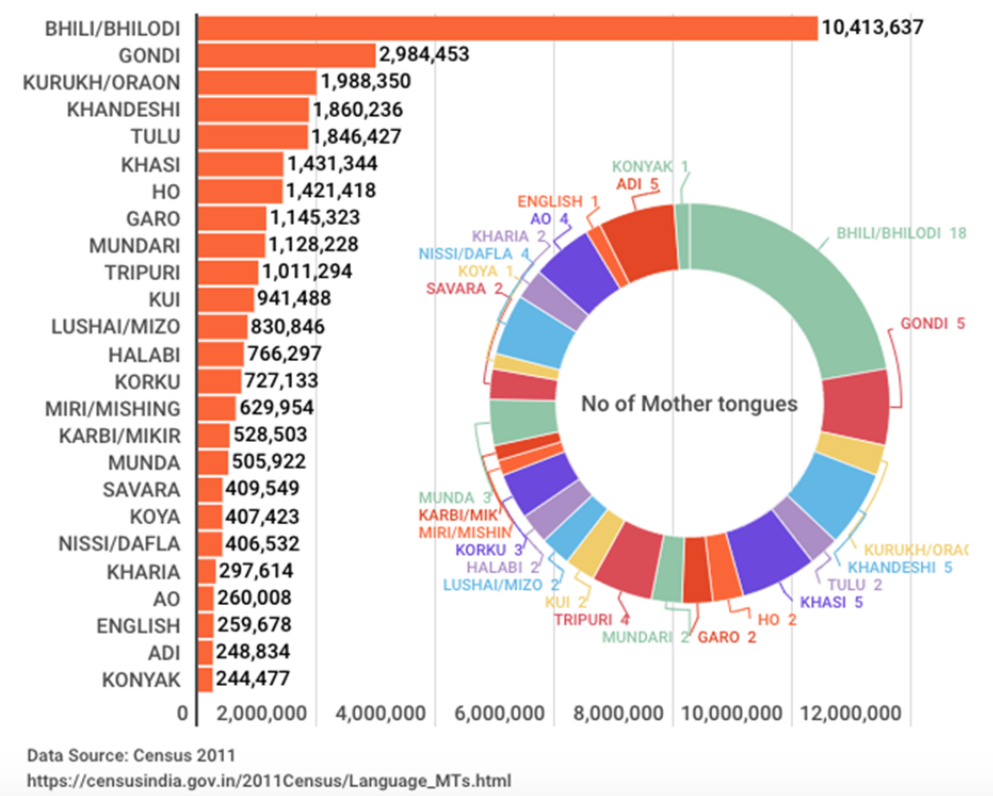


Figure 6: Top 25 Non-Scheduled Language Speakers in India, and their Associated Number of Mother tongues. (Data Source: Census 2011, C-16 Series)

According to [Thomas Benedikter](#), “Linguistic rights need official recognition, need instruments, infrastructures and funds of application, need a secure and clear legal framework, need validation and support from the State, which exercises sovereignty and public power”. The ‘minoritised’ non-scheduled languages are underrepresented in public education, media, and the economy and business, violating the linguistic rights of its speakers.

Despite the constitutional directive of Article 350A for states to provide adequate facilities for instruction in the mother tongue at the primary stage, the languages of linguistic minorities continue to be underrepresented as media of instruction in schools. An analysis of [DISE](#) (District Information System of Education, 2015-16) reveals that only 28 languages are used as the main media of instruction in schools throughout the nation. This poses [challenges](#) for children who are unfamiliar with the language used in their schools, especially during their foundational learning years.

The current ruling party strongly advocates the notion of ‘[One Nation, One Language](#)’, with Hindi to be used across India, administrative and otherwise, thus promoting monolingualism, rather than valuing multilingualism and multilingual practice. The Census classification and grouping of languages and mother tongues through ‘rationalisation’ does not acknowledge the languages of millions of people. The ongoing pandemic has delayed the decennial Census of India house-listing, [in all likelihood pushing the Census 2021 enumeration to 2022](#). This has provided extra leg space to mobilise reforms. We advocate reforms to the Census enumeration and classification of languages process to improve transparency, make it inclusive of all languages, and reflect the true linguistic diversity of India.

References:

Benedikter, Thomas. 2013. *Minority Languages in India: An appraisal of the linguistic rights of minorities in India*. Bolzano: European Academy of Bolzano/Bozen (EURAC), Institute for Minority Rights.

Bhattacharya, S.S. 2002. “Languages in India: Their Status and Function.” In *Linguistic Landscape in India*, edited by N.H. Itagi and Shailendra Kumar Singh. Mysore: Central Institute of Indian Languages and Mahatma Gandhi International Hindi University.

Kidwai, Ayesha. 2019. “The People’s Linguistic Survey of India Volumes: Neither Linguistics, Nor a Successor to Grierson’s LSI, but still a Point of Reference.” *Social Change* 49, no 1: 154–159.

Rajendra, P. 1997. *Minorities in India: Protection and Welfare*. New Delhi: APH Publishing.