Northeast India is a region of unusual linguistic diversity. There are at least 150 distinct languages spoken in the eight states, and although the great majority of these are Tibeto-Burman (henceforth, TB), the remainder represent three completely distinct language families – Austroasiatic, Indo-European, and Tai-Kadai – ignoring here the many languages from around India and farther abroad spoken by more recent migrants to the region. The number 150 represents a rough count of language names in Glottolog (a catalogue of world languages), but that list, like all others, represents fundamentally ethnic rather than linguistic divisions and includes both ‘languages’ which are divided into mutually unintelligible ‘dialects’ which might be considered linguistically distinct languages, and separate names for mutually intelligible varieties which could linguistically be considered to belong to a single language. For this and other reasons appropriate nomenclature for language names is a problem. Names for both language and, especially, higher-level language groupings were assigned in colonial times without concern for the ethnic sensibilities of the speakers – so that, for example, the Linguistic Survey of India (Grierson 1903–1909) designates as ‘Old Kuki’ a group of languages which include many spoken by communities that do not consider themselves ethnically Kuki and prefer not to have their languages referred to as such. This entry uses what seems at present to be the preferable name and includes other common names in brackets. Bracketed names in single quotes (‘Old Kuki’) should no longer be used but are included because they are used in older literature.

‘Northeast India’, as defined by present borders, is not a linguistically useful category. A continuum of Tibeto-Burman languages extends across the central and eastern Himalayas and down the Patkai Range, spoken in the political boundaries of Nepal, Bhutan, Bangladesh, western Myanmar, southwest China, and the northern edge of West Bengal, as well as within the boundaries of the eight states of the region. The so-called ‘seven sisters’, out of which four of them—Arunachal Pradesh, Meghalaya, Mizoram, and Nagaland—were created out of the original greater Assam, and they are centred on the Brahmaputra drainage, which is home to Khasian, Bodo-Garo, and Assamese languages. But the eastern and northern international borders run through mountain areas where we find numerous small languages and groups whose ancestral distribution included both slopes of the mountains, in Myanmar in the Chindwin and in China in the Yarlung Tsangpo drainage. The languages of Sikkim, Arunachal Pradesh, and Bhutan are closer to those of Nepal and Tibet than those to the rest of India.
There is no historical evidence for the linguistic history of the region other than scattered inscriptions in Sanskrit dating from the 6th century. It is known that the Indo-Aryans entered Northeast India from further west, and at that time there must have been both Austroasiatic and Tibeto-Burman languages present in the region, but there is no direct evidence of these until modern times. It is often assumed that the pre-Kāmarūpa population spoke one or more Austroasiatic language, probably related to the Khasian languages of Meghalaya, based on the geographical position of these languages in between their nearest cousins, the Palaung-Wa languages of northeastern Myanmar, and the Munda group of eastern India. J.H. Hutton, one of the best-known colonial administrators in the region, had speculated a lot on the Munda connection of the Khasi language.

The Brahmaputra Valley is at the edge of the Tibeto-Burman-speaking area, so the TB languages of the region must have entered the Valley from the north and/or east. The Mahābhārata describes the armies of Prāgyōtisha as ‘Kirātas and Chinese (cīnaiš)’, so when that verse was composed the armies of the Kāmarūpa state included people of East Asian appearance. That was roughly 2000 years ago, since cīn- ‘China/ese’ refers to the Qin Dynasty (221–206 BCE). These were probably speakers of one or more TB language(s), including the ancestor of the Bodo-Garo sub-branch. The Bodo-Garo languages originated and diverged in the Brahmaputra Valley along with early Prāgyōtisha/Kāmarūpa, probably as a lingua franca of what must have been a multilingual state (DeLancey 2012). The Kiranti languages of eastern Nepal and Sikkim may be of similar antiquity in the region if one accepts the popular identification of their speakers with the Kirātas of the Mahābhārata (Barua 1960), although this is unprovable.

The last language family to arrive in the region, the Tai/Ahom languages of Upper Assam, as a result of historical events, also impacted the South-Central languages. Luce (1985) argues that languages ancestral to modern South-Central languages (Kuki-Chin) and perhaps also some ‘Naga Belt’ languages were spoken in the early city-states of Chindwin Valley of western Myanmar. On this account, the present distribution of these languages results from the destruction of these states in the course of the Shan invasions, which also brought Tai languages into Northeast India.

The Tibeto-Burman Languages

The TB languages belong to the larger Trans-Himalayan (TH) or Sino-Tibetan family, which consists of the TB languages and the Sinitic (Chinese) branch. It is debated whether the TB languages constitute a branch of the family, but ‘Tibeto-Burman’ is a useful label for all of the family minus the very divergent Sinitic group. The classification of the TB languages is yet unsettled. The model adopted here recognizes three major branches: Eastern, represented mostly in Sichuan, Yunnan, and Southeast Asia; Western, in Nepal, Tibet, and Northwest India; and, most controversially, a diverse Central branch comprising the languages of the Chindwin-Irrawaddy and Brahmaputra drainages, and including most of the languages of Northeast India today.

The Central languages, whether or not they form a genealogical unit, are indisputably areally connected. Their distribution along the Irrawaddy-Chindwin and Brahmaputra valleys suggests a centre of dispersal in what is now the Myanmar-Tibet borderland, which is thus the presumed location of the language ancestral to most of the languages of Northeast India.

There is one language of the Eastern branch spoken in the region: Lisu, a Burmese-Ngwi (Lolo-Burmese) language which has in recent times expanded from a centre in Yunnan through northern Myanmar and more recently into far eastern Arunachal. The languages of the westernmost Arunachal and Bhutan are probably Western, as are all the languages of Sikkim, with the probable exception of Lepcha. All the other TB languages of the region are traditionally
The Sal Languages

The Sal group (named for an innovative lexical form *sal for ‘sun’, e.g. Bodo san) consists of three sub-groups. The Jinghpaw-Asakian languages are spoken primarily in Myanmar, but a few are spoken in Northeast India as well. Singpho (Morey 2010), in Upper Assam, is a variety of Jinghpaw, which is spoken across northern Myanmar. The small Asakian (Sak, Luish) group includes Andro, Chairel, and Sengmai, three small or extinct languages of Manipur. The Northern Naga subgroup includes an undetermined number of languages including Chang, Konyak, and Phom in Nagaland, and Wancho, Nocte, Tangsa, Tutsa, and others in Upper Assam and Arunachal. Nocte, Tangsa, and perhaps Tutsa are umbrella ethnic designations which include speakers of very distinct linguistic varieties (Morey 2019). There has been a recent explosion of documentation of these languages, including several comprehensive grammars, mostly still unpublished (Nagaraja 2010; Bano 2017; Boro 2017; Mulder 2020).

The most important branch of the history of the region is Bodo-Garo, a dozen languages spoken in and around the Brahmaputra Valley, including Bodo with over a million speakers in Assam. Other languages include Dimasa, Tiwa (Lalung), Kokborok (Tripura), Barman, Deuri (Deori Chutiya), Moran, Koch, Rabha, Ruga, Atong, and Garo. These languages have received increasing scholarly attention in recent years, with several comprehensive grammars (Burling 2004; Jacquesson 2005; Joseph 2007; van Breugel 2014; Bhattacharya 1977 also remains useful) and substantial comparative reconstruction (Joseph and Burling 2006; Debnath 2014).

The Kuki-Naga Languages

Other than the Northern Naga languages of Nagaland, the languages of the Myanmar border states belong either to the demonstrated South Central or Kuki-Chin branch, or to a group of typologically similar languages of Nagaland and Manipur, which is referred to here as the ‘Naga Belt’ languages. These two sets of languages have long been considered to comprise a Kuki-Naga branch; this idea has fallen into some disfavour but is probably correct. Another probably Kuki-Naga language is Karbi, with nearly a million speakers in and near Karbi Anglong district in Assam. It is not known when Karbi came to its present location in Assam; linguistic connections, as well as local traditions suggest that they came from the south/southeast several centuries ago.

South Central Languages

The largest demonstrated subgroup within the Kuki-Naga branch is South Central (Kuki-Chin), consisting of dozens of varieties spoken throughout the Chin State in Myanmar and Chittagong in Bangladesh as well as Mizoram, Manipur, and Assam. The clade, or a group of closely related languages, divides into several sub-branches (Peterson 2017). The languages of the Northwestern (‘Old Kuki’) sub-branch are spoken in southern Manipur and the Barak River Valley in Assam and Tripura. The Manipur languages include Aimol, Anal, Chiru, Chhothe, Koieng, Kom, Lamkang, Monsang, Moyon, Purum, Tarao, and others; the Barak Valley group includes Biate, Chorei, Hallam, Hmar, Hrangkhol, Ranglong, Saihriem, Sakachep, and others. The Central subgroup includes Mizo, the state language of Mizoram, and other languages.
of Chin State, Mizoram, and Chittagong, including Bawm, Lai, Mara, Mizo (Lushai, Lushei), Pangkhua, Zahao, and more. The Northeastern or Northern Chin languages of Chin State and southern Manipur include Thado, Gangte, Paite, Ralte, Zomi, and others. Languages of the Southern (Southern Chin) group are spoken mainly in Myanmar and Bangladesh. VanBik’s (2009) is a phonological and lexical reconstruction of South Central. Comprehensive grammars include the ones written by So-Hartman (2009) and Zakaria (2017).

The ‘Naga Belt’ Languages

Earlier classifications sometimes lumped Ao, Angami-Pochuri, Zeme, and Tangkhulic into a Naga branch, but the current opinion is that this is a purely ethnic category and that these groups do not constitute a genealogical branch distinct from South Central. However, these languages plus Meitei or Manipuri, the language of the Imphal Valley and the state language of Manipur, show similar grammatical and lexical patterns, probably as a result of centuries of contact.

Two clades are located primarily in Nagaland. The Ao (Central Naga) group consists of very closely related Ao, Lotha, Sangtam, Yimchungri, and others. Bruhn (2014) is a phonological/lexical reconstruction, and Coupe’s (2007) is a comprehensive grammar of a variety of Ao. Angami-Pochuri includes very closely related Ntenyi, Pochuri, Rengma, and Sumi as the Pochuri subgroup, and Angami, Chokri, Kheza, Mao (Sopvoma), and Poula in the Agami subgroup.

The other groups are centered in Manipur. Zeme (Zeliangrong) is a dialect chain with named varieties Liangmai, Rongmei, Zeme, spoken in and around Tamenglong District, Manipur. Tangkhulic languages are found throughout eastern Manipur. The Ukhrul district of Manipur probably has the same kind of linguistic diversity as Chandel, but for ethnopolitical reasons the different varieties are locally considered to be dialects of Tangkhul rather than distinct languages. Tangkhul, as spoken around Ukhrul town, serves as a lingua franca and is spoken and understood by most inhabitants of the district. The Tangkhulic clade includes the Tangkhul varieties in Ukhrul and Kamjong districts of Manipur, as well as Maring and Uipo (or ‘Khoibu’) in Chandel and Tengnoupal districts.

Meitei or Manipuri is the language of the Imphal Valley of Manipur and the dominant language and lingua franca of Manipur. Chelliah’s (1997) is a grammatical description of the Meitei language. The Meitei royal chronicle, the Cheitharol Kumbaba, dating from the 15th century, recounts a tradition of the origins of the Manipur state through the amalgamation of several distinct tribes 2000 years ago. Linguistically Meitei has lexical connections with many of its neighbors and shows layers of outside lexical influence (Singh 2002).

Central Languages of Arunachal

The languages spoken along the eastern Himalayas, in Arunachal Pradesh and adjacent districts of China, are not yet very well-known, and linguists are at a very preliminary stage in classifying them. In the east are several poorly understood languages like Meyor in Anjaw District, K’man (Kaman, ‘Miju’), and two very closely related languages, viz. Kera’a (Idu) and Tawrā (Taruang, Taraon, Digaru). Speakers of K’man, Kera’a, and Tawrā share an ethnic identity as Mishmi, and these languages have previously been considered to form a group on that basis. Linguistically, however, K’man is quite distinct from Kera’a-Tawrā, and is probably more closely related to Meyor.

The best-studied Arunachal language group is Tani (Abor-Miri-Dafla), including Adi, Apatani, Bangni, Bokar, Galo, Mising, Nyishi, Tagin, and other varieties spoken in central and
eastern Arunachal Pradesh and adjacent districts in Assam and China. Sun (1993) provides a phonological and lexical reconstruction, and Post (2007), a comprehensive description of Tani languages. Closely related but probably not Tani proper is Milang in the upper Siang drainage on which Modi (2017) provides a comprehensive description.

Farther west are several small languages or groups, most barely documented. Koro (Koro Aka), in East Kameng district was once incorrectly considered a variety of Hrusso; Hrusso (Aka) and Miji-Bangru form the Hrussish clade and Kho-Bwa (Bugunish, Kamengic), includes Bugun (Khowa), Puroik (‘Sulung’), Sherdupken, and Sartang (Butpa Monpa), spoken in western Arunachal.

One can only speculate as regards the history of these languages. Post (2015) notes some linguistic and ethnographic traits which suggest a Southeast Asian origin for some of these languages and peoples. If these languages are in fact Central, i.e. part of a genealogical unit along with Jinghpaw and South Central, then they too originated not very far to the east of their present location.

**Western Languages of Eastern Arunachal and Sikkim**

From far eastern Arunachal across Bhutan and Sikkim and into Nepal we find languages belonging to the Western branch of the Tibeto-Burman. Lepcha is the most characteristic TB language of Sikkim, as it is spoken there or thereabouts in the Ilam district of Nepal and the Kalimpong district of West Bengal, both the places being geographically contiguous to Sikkim. It is clearly a TB language, possibly belonging to the Central branch. All the other TB languages belong to one or another sub-branch of the Western branch. Tshangla is centered in Bhutan, with speakers also in Arunachal; its higher affiliation is unclear. Andvik’s (2010) is a comprehensive grammar. The East Bodish language Dakpa is spoken in Bhutan and Tawang District of Arunachal. Like several other languages in the area, including Tshangla, Dakpa is sometimes called ‘Monpa’. In Sikkim there are several Tibetan languages, particularly Denjongke (Sikkimese, Bhutia), also Tibetan, Sherpa, and Dzongka, which is the national language of Bhutan, and Tamangic languages including Tamang and Gurung. Several Kiranti languages are represented in Sikkim and probably in Nepali settlements in other Northeast Indian states, including Limbu, Yakkha, Kaling, Kulung, Chamling, and Bantawa. There are also Magar and Thami communities, representing respectively the Central Himalayan and Newaric sub-branches.

**References**


