10.1 From Tibeto-Burman to Trans-Himalayan

Julius von Klaproth was the first scholar to assign Chinese correctly to its proper language family. In 1823, he identified the Tibeto-Burman phylum in Paris in his polyphyletic view of Asian linguistic stocks. Klaproth’s model of many distinct Asian linguistic phyla was initially controversial because many scholars in the West at the time entertained an undifferentiated view of Asian languages as all belonging to some nebulous all-encompassing language family.

His Tibeto-Burman comprised Burmese, Tibetan, Chinese and all of the languages that could be demonstrated to be related to these three. He explicitly excluded languages today known to be Kradai or Daic (e.g., Thai, Lao, Shan), Austroasiatic (e.g., Mon, Vietnamese, Nicobarese, Khmer) and Altaic (e.g., Japanese, Korean, Mongolic, Turkic). The name Tibeto-Burman gained currency in English for the language family recognized by Klaproth and was widely used by scholars in the British Isles (e.g., Hodgson 1857; Cust 1878; Forbes 1878; Houghton 1896).

Some other scholars of the day followed the Indo-Chinese theory proposed by the Scots amateur John Casper Leyden, who died at the age of thirty-five after experiencing a short but dazzling career in the British colonial administration in Asia during the Napoleonic wars. In 1807, Leyden proposed his exuberant but poorly informed Indo-Chinese theory to George Barlow, Governor General of India at Fort William, in which he claimed that all the languages in Asia and Oceania shared some “common mixed origin” (Leyden 1808).

This murky view held appeal to adherents of Biblical mythology who had been inclined to lump Chinese together with numerous other Asian languages into a grand Japhetic family, on the assumption that Chinese was one of the languages spoken by the descendants of Noah’s son Japhet, while some alternatively attempted to explain
Chinese as an antediluvian language or as one of the “confounded” forms of speech with which Yahweh had afflicted mankind after the fall of the Tower of Babel.

The Biblically inspired Japhetic was not the only pan-Asian catch-all. Wilhelm Schott wrote personally to the famous scholar of Himalayan languages Brian Houghton Hodgson to warn him against the “Turanian” theory then being propagated from Oxford. In 1856, Schott likewise published an essay warning against “Indo-Chinese.” Schott foresaw that scholars who used the label would continue to think in terms of the mistaken phylogenetic model that the label designated. Yet the Indo-Chinese model became the favourite of racist language typologists who believed that Asian languages were generally more rudimentary and that Asian peoples were more primitive than their Western counterparts.

Grammatical typology inspired language typologists such as Heymann Steinthal (1850, 1860), Ernest Renan (1858), Arthur de Gobineau (1854–1855), and John Beames (1868) to rank Chinese and Thai together on the lowest rung of the evolutionary ladder of language development based on their “monosyllabicity” and lack of inflection. These scholars argued that Chinese and Thai must be closely related and that neither was part of Tibeto-Burman. James Byrne (1885) argued that “the causes which have determined the structure of language” lay in the varying “degrees of quickness of mental excitability possessed by different races of men.” Chinese and Siamese ostensibly mediated a rudimentary, less evolved way of thinking and so were assigned to the lowest rungs of Steinthal’s ladder of language evolution. The following quote typifies this once widespread genre of scholarly discourse.

la langue chinoise, avec sa structure inorganique et incomplète, n’est-elle pas l’image de la sécheresse d’esprit et de cœur qui caractérise la race chinoise?… Suffisante pour les besoins de la vie, pour la technique des arts manuels, pour une littérature légère de petit aloi, pour une philosophie qui n’est que l’expression souvent fine, mais jamais élevée, du bon sens pratique, la langue chinoise excluait toute philosophie, toute science, toute religion, dans le sens où nous entendons ces mots. (Renan 1858:195–196)

Such reasoning was vehemently opposed by scholars following the tradition of Wilhelm von Humboldt (1822, 1825, 1836), such as August Friedrich Pott (1856) and Friedrich Max Müller (1871, 1881), who argued that the relationship between language structure and thought was not so simplistic and who stressed that biological ancestry was independent of language.

---

**Figure 10.1** Julius von Klaproth’s Tibeto-Burman family.
At first, Indo-Chinese encompassed Asian languages from the Caspian Sea to Polynesia. This untenable construct embodied numerous misguided phylogenetic conjectures and so came to be whittled down in successive stages. After Philipp von Siebold (1832) and Anton Boller (1857) presented their case for a distinct Altaic phylum, Ernst Kuhn (1883, 1889) began to fix what was still wrong with the Indo-Chinese model by correcting the erroneous inclusion of Austroasiatic, but the resulting model still represented a false family tree. Yet some scholars and several notable sinologists adopted the Indo-Chinese name and the false Indo-Chinese phylogeny (e.g., von der Gabelentz 1881; Forchhammer 1882; Conrady 1896; Laufer 1916; Wulff 1934).

In 1924, the French orientalist Jean Przyluski coined sino-tibétain as the French term for Indo-Chinese in the English and German sense. This French term entered English in 1931 when Jean Przyluski and Gordon Luce co-authored an article on the root for the numeral hundred in “Sino-Tibetan.” The new term did not catch on at once, but during the Great Depression in 1935 the American president Franklin Roosevelt instituted the employment scheme called the Works Progress Administration. Through this programme, the famous Berkeley anthropologist Alfred Kroeber, inspired by the enthusiasm of Robert Shafer, managed to raise funding for his Sino-Tibetan Philology project. Changing the name of the model of linguistic relationship to the new Gallic label helped to deflect the widespread criticism against Indo-Chinese.

Shafer effectively ran the project for Kroeber but saw two things fundamentally wrong with “Sino-Tibetan.” In 1938, Shafer proposed to remove Kradai or Daic from the language family, but in the end he was not allowed to do so (Shafer 1955:97–98). Shafer also put Sinitic on par with other divisions in the family. The two operations would effectively have heralded a return to Julius von Klaproth’s original Tibeto-Burman model. After Paul Benedict came to Berkeley in the winter of 1938–1939 to join the project, he traded in the name Indo-Chinese for “Sino-Tibetan.” Moreover, after the conclusion of the project in 1940, he took credit for removing Daic (1942). Benedict (1972) also restored Sino-Tibetan to its original Indo-Chinese shape, again isolating Chinese as the odd man out.

Ironically, after the Cultural Revolution, Chinese scholars adopted as orthodoxy the Indo-Chinese model as it had been repackaged in America. Sino-Tibetan became 汉藏语系 Hàn-Zàng yǔxì, notwithstanding its empirically unsupported phylogeny and

![Figure 10.2](image_url) The Indo-Chinese or Sino-Tibetan theory: Kradai or Daic has been excluded since the Second World War.
its racist legacy. Historically, Sino-Tibetan is rooted in the fact that morphosyntactic typology had perplexed less enlightened linguists of the 19th century into believing that Chinese and Thai represented an inferior developmental stage on a Steinhthal’s ladder of language evolution. This view relied on the assumption that Sinitic languages had never evolved and that Chinese had remained typologically unchanged and “without inflection, without agglutination” for millennia (e.g., Chalmers 1866).

By contrast, the informed historical linguistic view represented quite a different understanding of Chinese. Carl Richard Lepsius (1861:492–496) proposed that Chinese tones had arisen from the merger of initials and the loss of finals based on correspondences between Chinese and Tibetan. He argued that entire syllables had been lost in Chinese and that Chinese ideograms once represented words that may often have contained more than just the root syllables whose reflexes survive in the modern pronunciations. The view of Chinese promulgated by Lepsius later inspired Bernhard Karlgren (1920, 1957) to conceive of Old Chinese as a “langue flexionelle” and to undertake the reconstruction of Old Chinese in accordance with the principles of the comparative method.

Two models of phylogenetic relationship sought to defy the Sino-Tibetan paradigm propagated from Berkeley, that is, Sino-Himalayan (Bodman 1973, 1980) and Sino-Kiranti (Starostin 1994). Although neither proposal gained acceptance, these salients made the crucial point that to date no evidence has ever been adduced in support of the Sino-Tibetan phylogenetic model, defined by its truncated “Tibeto-Burman” taxon encompassing all non-Sinitic languages. Methodologically, attempts to define all non-Sinitic languages negatively in terms of Sinitic innovations that other languages lack or to invoke the argument of gross word order for Karen and Sinitic, as Benedict (1976) once did, are known to be phylogenetically meaningless. All comparative evidence amassed to date supports Julius von Klaproth’s 1823 minimalist Tibeto-Burman tree, which epistemologically therefore continues to represent the default model.

However, the history of the field has left us with an unfortunate nomenclatural legacy. Whereas Tibeto-Burmanists in Klaproth’s tradition used the name “Tibeto-Burman” for the family as a whole, Sino-Tibetanists have used the term “Tibeto-Burman” to denote all non-Sinitic languages as comprising a single taxon. In an attempt to escape this terminological morass, in 2004 the alternative name “Trans-Himalayan” was proposed for the linguistic phylum because the world’s second most populous language family straddles the great Himalayan range along both its northern and southern flanks (van Driem 2007:226).

This neutral geographical term is analogous to “Indo-European” and “Afro-Asiatic” in reflecting the geographical distribution of the language family. The term “Afro-Asiatic” was coined in 1914 and replaced the earlier “Hamito-Semitic” for similar reasons. Hamitic was shown not to be a valid subgroup, just as Sino-Tibetan defined by its unitary non-Sinitic taxon likewise denotes a false tree. The linguistic phylum is genuinely Trans-Himalayan in distribution in that by far most of the roughly 300 different Tibeto-Burman languages and three-fourths of the major Trans-Himalayan subgroups are situated along the southern flanks of the Himalayas (Figure 10.3), while by far most speakers of Trans-Himalayan languages live to the north and east of the great Himalayan divide (Figure 10.4).
Figure 10.3 Geographical distribution of the major Trans-Himalayan subgroups. Each dot represents not just one language but the putative historical geographical centre of each of forty-two major linguistic subgroups.

Figure 10.4 Geographical distribution of Trans-Himalayan languages.
10.2 Subgroups and Geography

Much more is known about the Tibeto-Burman language family today than in the days of Klaproth. Today we can identify forty-two subgroups for which there appears to be evidence and about which there is some degree of consensus. The 2012 version model of the Fallen Leaves model, shown in Figure 10.5, contains a number of groups not mentioned when this model was first presented (van Driem 2001). The rGyalrongic subgroup was proposed and validated by Jackson Sun (2000a, 2000b). The Nàic subgroup, comprising Nàmhì and Shìnxìng and the closely related Nàish languages (i.e., Nàxi [naJi], Na [naJ] and Laze [la4ze]) has been proposed by Jacques and Michaud (2011). Evidence for an Ěrsùish subgroup has been presented by Yu (2011).

Post and Blench (2011) presented evidence for Siangic, a group comprising Milang and Koro. At one level, Post and Blench envisage Siangic not as a Tibeto-Burman subgroup but as an altogether non-Tibeto-Burman phylum that has left vestiges in Koro and Milang. A more conservative stance would be to treat Koro and Milang together as a Tibeto-Burman subgroup in their own right. In a similar vein, many scholars have recently publicly aired the view that Puroik (or Sulung), normally deemed to be a member of the Kho-Bwa cluster of languages, is not a Tibeto-Burman language at all. Despite the apparently aberrant nature of some of the lexicon, Puroik, Koro, and Milang all exhibit a good

---

**Figure 10.5** The 2012 version of the agnostic Fallen Leaves model. Thirty out of the forty-two Tibeto-Burman subgroups lie south of the great Himalayan divide, seven to the north and east of the Himalayas, and five (i.e., Tshangla, Bodish, Nungish, Lolo-Burmese, and Kachinic) on both sides of the Himalayas.
share of Tibeto-Burman vocabulary. The history of Indo-European is instructive in this regard.

French shows a smidgen of Celtic lexicon that can be viewed as substrate, while the language itself is indisputably a Romance dialect. Words borrowed from the substrate language do not determine the linguistic affinity of a language. Until Ritter von Xylander (1835), Albanian was held to be a language isolate in Europe just like Basque. It is sobering to reflect that less is known today about Tibeto-Burman historical grammar than was known in 1835 about Indo-European historical grammar. The Gongduk language in Bhutan is analogous to Albanian or, for that matter, much like Koro, Milang and Puroik, in exhibiting much vocabulary that appears outlandish from a Tibeto-Burman perspective. Yet our perspective on Tibeto-Burman has been changing rapidly in recent years, as more becomes known about the less well documented languages of the phylum. Our understanding of what Starostin called “Tibeto-Burman in the narrow sense” is broadening to encompass a more informed and fine-mesh view.

The growing awareness in the field that the Tibeto-Burman analogues of Armenian, Hittite, and Albanian all appear to be found within the eastern Himalayas highlights the fact that the language family’s centre of phylogenetic diversity lies squarely within the eastern Himalayas. The lexical diversity observed in many subgroups of the eastern Himalayas is just one residue of a complex and many-layered ethnolinguistic prehistory in a region of ancient human habitation.

The whereabouts and the names of the languages in the forty-two leaves that have fallen from the Trans-Himalayan tree are listed in the following. The most obvious disambiguations are indicated with the symbol ≠ with additional elucidation. Realities on the ground are far more complex than any short list can show. Related but entirely distinct and mutually unintelligible languages sometimes go by the same name (e.g., Magar, Limbu, Chinese). So the roughly 280 language labels in this non-exhaustive list obscure a great deal of dialectal and linguistic diversity. Terms in brackets represent alternative names by which the languages are also known.

Sometimes the ethnic designation and the mother tongue do not match, as when a community, for example, considers itself Jingpō but speaks the Lolo-Burmese language Zaiwa or considers itself Tibetan but speaks a rGyalrongic language. Some languages are extinct (e.g., Pyu, Dura), believed to be extinct (e.g., the Sak languages) or moribund (e.g., Barām). In fact, most Tibeto-Burman languages are endangered with imminent extinction. A more detailed account can be found in the ethnolinguistic handbook *Languages of the Himalayas* (van Driem 2001) and in the literature referenced therein.

*Angami-Pochuri* (southern Nagaland, northern Manipur, neighbouring portions of Burma and Assam): Angami, Chokri [Chakri], Kheza, Mao [Sopvoma], Pochuri, Ntenyi, Maluri [Meluri], Sema, Rengma, Kezhama, Senkadong

*Ao* (central Nagaland and neighbouring portions of Burma): Yacham, Ao Chungli, Ao Mongsen, Yimchungrü [Yachumi], Sangtam [Thukumi], Yacham, Tengsa, Lotha [Lhota]
Bái (the area around Dàlì in Yúnnán province): Bái

Black Mountain Mönpa (the Black Mountains of Bhutan): ’Olekha, Riti, Jangbi, ’Wangling

Bodish (Tibet, Pakistan, India, Nepal, Sikkim, Bhutan): Balti, Purik, Ladakh, Zanskar, Lahul, Central Tibetan (dBus and Tsang), Sherpa, Ölmo Sherpa, Lhomì, Jirel, Kagate, Mustang, Limirong, Mugu, Northern Kham, Eastern Tibetan, Amdo Tibetan, Brokpa, Dzongkha, Lakha, Dränjøke, Cho-ca-nga-ca-kha, Bumthang, Kheng, Mangde, Kūrtōp, Chali, Dzala, Dakpa

Brahmaputran [i.e., Bodo-Koch plus Northern Naga] (West Bengal, Assam, Meghalaya, Arunachal Pradesh, northern Nagaland and adjacent portions of Burma): Chutiya, Kokborok, Tiwa, Dimasa [Hills Kachāḍī], Bodo, Plains Kachāḍī, Meche, Garo, Atong, Pani Koch, Ruga, Rabha, Tāngsa, Nocte, Wāncho, Kuwa, Haimi, Htangan, Konyak, Pono, Phom, Chang, Wēlam, Nokaw

Chepangic (central Nepal): Chepang, Bhujeli

Dhimalish (eastern Nepalese Terai, western Bhutanese duars): Dhimal, Toto

Digarish [Northern Mishmi] (Dibang River valley, Lohit district, Arunachal Pradesh): Idu, Tāraon [Digaro]

Dura (central Nepal's Lamjung district): Dura

Ěrsūish (southern Sichuān, northern Yúnnán): Ėrsū, Tosu, Lizu

Gongduk (south-central Bhutan): Gongduk

rGyalrongic (southern Sichuān): Situ, Japhug, Tsobdun, Zbu, Lavrung (including Thurje Chenmo, and nDzorogs), Horpa (including rTau and Stod-sde)

Hrusish (western Arunachal Pradesh): Hruso [Aka], Dhimma [Miji], Levai [Bangru]

Kachinic [Jinghpaw] (northeastern India, northern Burma, southern Yúnnán): The various Kachin, Singpho, Jīngpō, or Jinghpaw languages and the Sak [Luish] languages Sak, Kadu, Andro, Sengmai, Chairel

Karbī [Mikir] (Mikir Hills or Karbī Anglón, neighbouring districts of Assam): Karbī [Mikir]

Karenic (lower Burma, the Tenasserim, and adjacent Thailand coastal regions): Pāo, Pwo, Sgaw, Kayah, Brek [Bwe], Bghai

Kho-Bwa (western Arunachal Pradesh): Khowa [Bugun], Sherdukpen, Puroik [Sulung], Lishpa

Kiranti (eastern Nepal): Pāṅchṭare Limbu, Phedāppe Limbu, Tamarkhole Limbu, Chathare Limbu, Yakkha, Chiling, Āṭhpahariyā (including Belhare), Lohorung, Yamphu, Mewahang, Kulung, Nachiring, Sampeng, Sam, Chamling, Puma, Bantawa, Chintang, Dungmali, Thulung, Jero, Wambule, Tilung, Dumi, Khaling, Kohi, Bahing, Sunwar, Hayu

Chinbok, Laizo, Lakher, Ashö, Khumi Chin, Hmar, Anal, Lakher [Mara], Falam, Vaiphei, Lamgang, Simte

**Lepcha** (Sikkim, Darjeeling, Kalimpong): Lepcha

**Lhokpu** (southwestern Bhutan): Lhokpu [Doya]

**Lolo-Burmese** (southwestern China, Burma, Southeast Asia): Burmese, Zaiwa (≠ Midzuish Zaiwa) [Atsi], Lāshi, Māru (≠ Mru in the Chittagong), Maingtha [Achang a.k.a. Ngachang], Hpon [Hpun], Dānu, Taungyo [Tāru (≠ Danaw)], Phunoi, Akha, Lahu, Lisu, mBisu, Ahsi, various Yi languages

**Magaric** (central Nepal): Syāngjā Magar, Tanahū Magar, Pālpā Magar, Khām Magar [Kham (≠ Tibetan Kham)]

**Meithei** (Manipur): Meithei [Manipuri]

**Pyu** (extinct language of pre-Burmese epigraphy in Burma): Pyu

**Midzuish** [Southern Mishmi] (Lohit drainage, Lohit district, Arunachal Pradesh): Kaman [Miju a.k.a Mijhu], Zaiwa (spoken by the Meyöl clan near Walong ≠ Burmic Zaiwa)

**Mru** (in the Chittagong of Bangladesh): Mru [Māru] (≠ the Shan State Māru in Burma)

**Nāic** (southern Sichuān, northern Yūnnán): Nàmöyi, Shǐxing, Nàxi, Na, Laze

**Newaric** (central Nepal): Kathmandu Newar, Pahari Newar, Badikhel Newar, Chitlang Newar, Dolakha Newar, Barām, Thangmi

**Nungish** (Yūnnán province, northern Burma): Trung, Álóng, Rāwang, Róuruò, Nung including Nūsū and Ānū (≠ the Daic Nung in northern Vietnam)

**Qiāngic** (southern Sichuān, northern Yūnnán): Southern Qiāngic, Northern Qiāngic, Mi-ńag (Mùyà), Prinmi (Pūmī), Choyo (Quèyù), Tangut (Xiāxià), Zhābā, Ėrgōng, Guiqióng

**Raji-Raute** (western Nepal, Uttarakhand): Raji, Raute

**Siangic** (Arunachal Pradesh): Koro, Milang

**Sinitic** (China): Mandarin, Cantonese, Wū, Gān, Xiāng, Hakka [Kējiā], Southern Min (including Hokkien), Eastern Min, Northern Min, Central Min

**Tamangic** (central Nepal): Tamang, Gurung, Thakali, Chantyal, Ghale, Kaire, 'Narpa, Manangba

**Tangkhul** (northeastern Manipur, neighbouring parts of Burma): Tangkhul, Maring

**Tani** [Abor-Miri-Dafla] (Arunachal Pradesh, neighbouring portions of Assam): Apatani, Nyisu, Bengni, Nishing, Tagin, Yano, Sarak [Hill Miri], Galo, Bokar, Ramo, Ashing, Pailibo [Libo, Damu, Bori, Mishing [Plains Miri], Padam, Shimong, Pasi, Panggi, Tangam, Karko, Minyong

**Tshangla** [Shâchop] (eastern Bhutan, enclaves in Arunachal Pradesh and Tibet): Tshangla [Shâchop or loconyms], Bjokapakha, Dirang Tshangla

**Tūjiā** (Hūnán, Hūbèi and Guìzhōu provinces): Tūjiā
West Himalayish (Himachal Pradesh, Uttarakhand): Manchad, Tinan, Bunad [Gari], Kanashi, Rangpo, Darma, Byangsi, Rangkas, Zhangzhung


Some of the subgroups in this list of forty-two fallen leaves represent tentative subgrouping hypotheses that have yet to be subjected to closer scrutiny (e.g., Newaric, Qiangic). By the same token, questions arise such as whether Bodish should include East Bodish as well as Bodish proper and how East Bodish should otherwise be renamed, or whether Brahmaputran should encompass both the Bodo-Koch as well as the Northern Naga languages. In historical linguistics, it is preferable to work from the bottom up, that is, starting with the tangible leaves that have fallen from the tips of the branches and then moving upward to gain an understanding of the nodes in the tree. Yet many Tibeto-Burman languages are still poorly documented and scantily described.

10.3 Internal Phylogeny and Higher-Order Subgrouping

The Fallen Leaves model is no definitive phylogeny by definition. Though agnostic about higher-order subgrouping, the model does not deny that there is a family tree whose structure must be ascertained by historical linguistic methods. The continuing identification of subgroups presents a challenge to the current generation and to future generations of historical linguists to reconstruct the internal phylogeny of Tibeto-Burman on the basis of reliable data and regular sound laws and not to accept false family trees that we inherit from our predecessors or find in the literature without the support of historical comparative evidence. Two of Shafer’s (1966–1974) old “divisions” continue to lead robust lives of their own as higher-order albeit vaguely delineated subgrouping proposals (i.e., Bodic and Burmic).

Recently, Jacques and Michaud (2011) have proposed a higher-order subgroup called Burmo-Qiāngic, comprising Lolo-Burmese and a subgroup newly christened Nā-Qiāngic. Nā-Qiāngic essentially represents the same catch-all that used to be called “Qiāngic.” This constellation of subgroups has now been rendered less nebulous, however, by Sun (2000a, 2000b), Yu (2011), and Jacques and Michaud (2011), who have validated the rGyalrongic, Ėrsūish, and Nāic subgroups, respectively. In addition to these three subgroups, Nā-Qiāngic also contains Mi-ňag (Mùyà), Prinmi (Pûmì), Choyo (Quèyù), Tangut (Xixià), Zhābā, Qiāngic sensu stricto, and perhaps Ėrgōng and Guiqióng. The internal phylogeny of the latter medley of subgroups still has to be worked out, and the higher-order subgrouping hypotheses Nā-Qiāngic and Burmo-Qiāngic likewise require validation.
Another higher-order subgrouping hypothesis, Sino-Bodic, has a long history. Julius von Klaproth (1823) observed that Tibetan and Chinese appeared to be more closely related to each other than either were to Burmese. Simon (1927, 1928, 1929) and Forrest (1956, 1962) adduced lexical evidence that suggested a closer relationship between Chinese and Tibetan within the family. Although Shafer criticized Simon’s work, Shafer (1955) too observed that a closer genetic affinity obtained between Sinitic and Bodic than between any other two divisions. Later Bodman (1973, 1980) too adduced evidence indicating a closer relationship between Sinitic and Bodic. The name “Sino-Bodic” was proposed for the hypothesis, and additional lexical evidence for this affinity was presented (van Driem 1997). Matisoff (2000) protested, but most of the Sino-Bodic evidence still stands (van Driem 2005). Possible new evidence for Sino-Bodic has been adduced by Nathan Hill (2011) and Zhèngzhāng (2011). Future research will determine whether any of these supergroups will survive the test of time.

Notes
1. Hodgson’s correspondence is kept at the Royal Asiatic Society in London.
2. *tibeto-birmanskij v uzkom smysle*.

References


Tibeto-Burman

147


