Tibeto-Burman subgroups and historical grammar

George van Driem
Universität Bern

Abstract

Several distinct strains of thought on subgrouping, presented in memory of David Watters and Michael Noonan, are united by a golden thread. Tamangic consists of Tamangish and maybe something else, just as Shafer would have wanted it. Tamangic may represent a wave of peopling which washed over the Himalayas after Magaric and Kiranti but before Bodish. There is no such language family as Sino-Tibetan. The term 'trans-Himalayan' for the phylum merits consideration. A residue of Tibeto-Burman conjugational morphology shared between Kiranti and Tibetan does not go unnoticed, at least twice. Black Mountain Mönpa is not an East Bodish language, and this too does not go unnoticed.

Keywords

Tibeto-Burman, subgrouping, Tamangic, Tamangish, Limbu, Tibetan, Black, Mountain Mönpa, Michael Noonan, David Watters, Richard Keith Sprigg
Tibeto–Burman subgroups and historical grammar

George van Driem
Universität Bern

1 Tamangic and Tamangish

As Martine Mazaudon stressed in September 2010 at the 16th Himalayan Languages Symposium in London, and as she has indeed done on several occasions previously, a terminological distinction ought to be observed between Tamangic and Tamangish in keeping with Robert Shafer’s convention of using the suffix -ish for a subordinate group and -ic for a superordinate group within the language family tree. Certainly, Mazaudon is right in insisting that the group which she has previously designated either by the abbreviation TGTM or by the prolix label ‘Tamang–Gurung–Thakali–Manangba’ could aptly be designated Tamangish or, in French, tamangois or tamangais. However, this terminological distinction, which Shafer originally proposed for English and French, does not work as well for, say, German or Dutch.

Yet there are a number of languages which appear to be related to Tamangish but would not be Tamangish stricto sensu. Two such languages are Kaike and Ghale. The term Tamangic is therefore reserved to designate the hypothesis that Tamangish proper and certain Tamang-like languages such as Kaike and Ghale together form a coherent subgroup. This use of the term Tamangic is in keeping with Shafer’s terminological convention whereby three of his six larger groups ending in -ic are heterogeneous and impressionistic, i.e. Baric, Burmic, Bodic. This hypothetical usage also corresponds to the way that the term Tamangic is used in my handbook (van Driem 2001). Not all of Shafer’s taxa ending in -ic were nebulous, however. Shafer’s Daic turns out to represent a separate linguistic phylum altogether, distinct from Tibeto-Burman, as Shafer himself believed (1955: 97–98), whereas Shafer’s Sinitic and Karenic are well-defined subgroups within the language family. Shafer’s nomenclature reflected his rebellion against the Sino-Tibetan paradigm.

The Tibeto–Burman phylum was identified by Klaproth in his 1823 polyphyletic view of Asian linguistic stocks. His language family comprised Tibetan, Burmese and Chinese and languages demonstrably related to these three, and explicitly excluded Thai, Mon, Vietnamese, Japanese, Korean, Mongolic and Turkic. The competing theory Sino-Tibetan, originally ‘Indo-Chinese’, by contrast promulgated a phylogenetic model with a primary bifurcation between Sinitic and all non-Sinitic languages. Shafer was constrained to work within the Sino-Tibetan paradigm when he joined the Sino-Tibetan Philology project at Berkeley, for which Alfred Kroeber had raised funding through the Works Progress Administration instituted by Franklin Roosevelt in 1935 as an economic relief programme in the wake of the depression. Yet Shafer assailed this paradigm from the start. In the form of six main divisions, he challenged the tenet that the phylum was fundamentally divided between Sinitic and non-Sinitic languages. In fact, Shafer’s claim of a particular genetic propinquity between Sinitic and Bodic echoed the intimations of Klaproth and scholars...
who worked outside the Indo-Chinese paradigm. In 1938, Shafer also attempted in vain to oust Daic from Sino-Tibetan (Shafer 1955: 97–99).

Although Benedict (1942, 1972) later accepted Shafer’s insight about Daic, Benedict’s model in other respects represented an artefact of the Indo-Chinese lineage of thinking. The literature in our field since the early 19th century can only be properly understood with an awareness that some scholars operated within Klaproth’s Tibeto-Burman paradigm, whilst others operated within John Leyden’s Indo-Chinese paradigm.1 It has taken two centuries for the once more expansive Indo-Chinese to be whittled down to its present size and be renamed Sino-Tibetan. Yet Sino-Tibetanists have still not adduced evidence in support of their hypothesis that all non-Sinitic languages constitute a single ‘Tibeto-Burman’ taxon.

In 1856, Wilhelm Schott warned that those who continued to use the name *indo-chinesisch*, now ‘Sino-Tibetan’, would implicitly continue to endorse the particular phylogenetic model which the term designated. Today’s Sino-Tibetanists demonstrate the prescience of Schott’s admonition. In 2004, I proposed the term Trans-Himalayan as a neutral geographical name for the family (cf. van Driem 2007: 226), by analogy to Austroasiatic, Indo-European and Afroasiatic. The latter name, coined in 1914, in fact replaced ‘Hamito-Semitic’ for a similar reason: Like the Sino-Tibetanists’ truncated ‘Tibeto-Burman’ (i.e. excluding Sinitic), Hamitic was shown not to be a valid subgroup.

The continued use of Shafer’s system of nomenclature when we speak of Tamangic commemorates his lone stand at Berkeley against the Indo-Chinese paradigm. Solid historical comparison is required to establish whether the Tamangic subgroup is more than just a hypothesis or whether Kaike and Ghale together constitute an independent subgroup. In this context, I should like to quote from my correspondence with the late David Watters, who was no doubt the most knowledgeable scholar on the topic of Kaike and Ghale, and it is to his opinion that I presently defer. The following extract comes from David’s message to me of May 3rd, 2007.

As you know, I’ve been working on a grammar of Kaike, and I’ve been finding some very interesting stuff. It turns out, for example, that Kaike is Tamang-like but not TGTM. The pronouns don’t fit and the tones don’t fit. Kaike and Ghale fall together on both counts, but in addition, Kaike shows a fairly strong Kham substrate. As we all know, the low register in Bodish (and Tamangic) comes from the devoicing of voiced initials, but the story is demonstrably different for both Kaike and Kham (plus Magar to a certain extent). The source of low register in Kaike falls in with Kham, not Tamangic. In addition 12.5% of 1,200 items of Kaike vocabulary are clear cognates with Kham.

Here follows an excerpt from David’s message of August 25th, 2008:

Yes, I’m still writing a grammar of Kaike, though I haven’t had time to give to it for several months now. Just three weeks ago I returned home to the U.S. after an absence of six years. I’ll be returning to Nepal sometime in the Spring for more Kaike data collection. It’s full of fascinating stuff.

---

1 In the British Isles, the term Tibeto-Burman was used for Klaproth’s phylum, e.g. Cust (1878), Forbes (1878), Houghton (1896), even by those who, for a brief time, incorporated Tibeto-Burman within Max Müller’s sweeping Turanian construct, such as the famous Himalayan scholar Brian Houghton Hodgson.
About six months ago I wrote a little article entitled ‘The Position of Kaike in Tibeto-Burman (How a tiny language can shed light on big questions)’. It talks mostly about linguistic evidence for migrations, etc., but isn’t very solid yet. I did it mostly for fun (I even gave a talk on it in Australia) and sent a copy to a friend who mistakenly thought I was submitting it to LTBA! He sent it off for the review process and it came back thoroughly trashed [rightly so] on the grounds that the evidence was too thin and that it was too speculative. Okay, but I already knew that!

Far from deserving a trashing, David’s thoughts on the ethnolinguistic prehistory of the Himalayas were sophisticated and complex. I never saw David’s article on the position of Kaike, but we corresponded both on linguistic subgrouping and on human genetic phylogeography. David raised the issue of the prevalence in Tibeto-Burman language communities of the Y chromosomal haplogroups defined by the mutations M122 and YAP. In the revised 2008 nomenclature of the Y Chromosome Consortium (Karafet et al. 2008), these Y chromosomal haplogroups are currently labelled O3 (M122) and DE (YAP) respectively.

David inquired specifically about the occurrence of the YAP mutation in Magaric and Kiranti language communities of the Himalayas, and I was regrettably compelled to dodge the question. As I apologised to David, I was ‘beholden unto the geneticists whose findings cannot be disclosed until they have actually appeared in print’. In fact, some of the relevant data for Himalayan populations are still under embargo due to computational work in the laboratories in Cambridgeshire holding up the publication of the findings.

Some of these genetic findings will certainly prompt a rethink of the population prehistory of the Himalayas when they are finally published, and I wish that David could have been here to share in the new vista that will open up through this research. Meanwhile, suffice it to say that today’s geneticists often end up corroborating the intimations of linguists and astute ethnographers. David was, of course, keenly aware that population genetics and linguistics present two distinct versions of prehistory and that the linguistic and biological ancestors of a language community were not always the same people. In this context, David’s insightful reflections on the population prehistory of the Himalayas merit quotation:

Kham was clearly one of the very early Tibeto-Burman migrations into Nepal. They scooted in ahead of the rapid Tibetan expansion while the area was still pretty empty and their language was still ‘Kiranti’ ([in the sense that] their language was unaffected by Bodish). Shamans retrace the old migration route by naming stop-over points along the gTsang-po in their funerary chants. I can assume that Kaike came much later, but early in the Tibetan diffusion across the plateau (it has Bodish syntax, etc., but pre-Bodish pronouns [like Kham], and Kham-like register tone). Tamangic most likely represents a later wave, probably coincident with Srong-bstan sGam-po’s Tibetan kingdom from the seventh century.

It looks now like Kaike absorbed an earlier Kham population in their current homeland. They call themselves Magars (which may or may not mean anything), and their mythology talks about one of their (semi-divine)
founders intermarrying with a Kham woman. Furthermore, Kham speakers refer to the Tarakot region as an early Kham settlement area.

My surmise would be that the absolute dating of the Tamangic wave would have been considerably earlier than David conjectured. It is also conceivable that the Kham shamanic traditions represent younger mythologies. However, there may very well be something to the relative chronology outlined by David, which envisages the anteriority of the Magaric and Kiranti waves of settlement succeeded by a Tamangic wave, in turn succeeded by a Bodish wave.

Moreover, David and his son Stephen Watters put me and Aashish Jha in touch with the right people and so enabled us to collect DNA samples from the Kusunda in 2006. The Kusunda genome campaign was conducted by informed consent with the logistical assistance and active participation in the field of the Ādivāśi Janjāti Uthān Rāṣṭrīya Pratiṣṭhān (National Foundation for the Development of Indigenous Nationalities, NFDIN) under the Śrāṇīya Vikāś Mantrālaya (Ministry of Local Development) of the Government of Nepal. The success of this campaign owes a fundamental debt to David and also to Stephen.

Finally, in the context of Tamangic, I should like to add an observation on Chantyal. In addition to the texts which appeared in Michael Noonan's grammar (Noonan 1999), Michael had envisaged publishing an additional corpus of analysed Chantyal texts with the same publisher. In the end, this project could not be realised due, in part, to a lack of interest on the part of the publisher for bringing out a bound volume containing an analysed text corpus due to marketability considerations. Fortunately, this valuable material has been made available through Himalayan Linguistics (Noonan 2005). It is a tribute to the foresight of Michael Noonan and Carol Genetti that such internet-based publication has enabled our field to liberate itself from a number of the irrelevant economic constraints of commercial publishing.

2 Limbu and Tibetan

I first visited Richard Keith Sprigg at his home, a beautiful stone cottage overlooking Kalimpong to the west and the Himalayas to the north, in 1984. When I arrived at the top of the hill, I heard the sound of bagpipes playing ‘Scotland the Brave’, and the strains of this melody amidst the tree ferns created a surreal impression. As I advanced towards the house, I could hear that the sound of the bagpipes emanated from the portion of the garden to the left of the cottage. As I strayed from the path and followed the music across the lawn, I discerned the figure of Professor Sprigg standing with his bagpipes at the end of the garden on the edge of the ridge, facing away from me, with his gaze directed toward the Himalayas. After he ceased playing the bagpipes, he continued to look to the north in silence. After some time, he turned around and, spotting me in his garden, sauntered towards me and greeted me, saying ‘van Driem, I presume’.

This was my first visit to the Sprigg home. In those days, it was necessary to obtain a special permit in Darjeeling in order to visit Kalimpong. A permit could be granted with a validity of either 24 or 48 hours. I was still working on my grammar of Limbu at the time, and Sprigg told me of his own sojourns at Myanglung and Tehrathum, which he had visited with his wife Ray Sprigg, née Williams, in the 1950s. The picture which he painted of Tehrathum was more idyllic and bucolic than Tehrathum had become in the 1980s. Since then, however. Tehrathum has been
connected to the rest of Nepal by tarmac road, and this once lovely mountain bazar has now become a bustling roadside town. Today even the Tehrathum that I once knew is an idyll of the past.

From our very first discussions, it at once became apparent that Sprigg had done far more work on Limbu than he had ever published. Needless to say, one of the topics that we touched upon was Limbu verbal morphology. Sprigg immediately pointed out that something remarkable happens with the Limbu open-stem verbs such as *camaʔ* ‘to eat’. I hastened to say that I had observed the alternation in the stem vowel, and that the switch from */a/* to */ɔ/* in forms of *camaʔ* with a third person patient, e.g. *cɔ* ‘he ate it’, was clearly connected with the third person patient *<-u>*, which we would otherwise expect, whereas the tense-motivated apophony in the stem involving the switch from */a/* to */eˑ/* was clearly connected with the preterite morpheme *<-ɛ>*, which we could observe elsewhere. ‘So you’ve figured that out, have you? Very good. Now how about Tibetan?’.

When I failed to grasp what Sprigg meant, he kindly helped me along, pointing out that the preterite and imperative stem *zos* of the Tibetan *za* ‘to eat’ appeared precisely to correspond to the phenomenon that we observe in Limbu. He challenged me to write an article about this correspondence. I pointed out that he should write the article himself, since he had observed the correspondence, not I. Then I hastened to ask if he could think of any other instances in which Tibetan had preserved a possible reflex of the third person patient agreement suffix *<-u>*, well reflected in Kiranti languages such as Limbu. Sprigg said that he could think of no other instance, but that it would be a challenge to see whether old Tibetan texts, such as those found in Dūnhuáng, might contain other attestations of possible vestiges of the same phenomenon. Moreover, it might be going out on a limb, he mused, to write about a morphological correspondence of this nature between Tibetan and Limbu in light of the queer hostility that some scholars of Tibeto-Burman at the time harboured with respect to historical morphology.

When I pointed out that just one case of apophony in a single Tibetan verb was not much to go on, Sprigg stressed that this Tibetan verb precisely represented the environment in which such a vestige of a cognate verbal desinence would have been likely to have been preserved. Sprigg went on to say that he was willing to wager that if I would not write about the probable historical source of this Tibetan apophonic pattern, it would be just a matter of time before someone else did.

Lo and behold, just a few months ago a brilliant paper by Guillaume Jacques was published, entitled ‘A possible trace of verbal agreement in Tibetan’. When a friend of mine in Bhutan sent me Jacques’s well-written exposition of the argument, obviously I could not help but recall Keith Sprigg’s wager. In this way, Jacques’s lovely piece on the possible vestige of verbal agreement in Tibetan evoked the memory of the warm hospitality which Keith Sprigg and his wife Ray extended to me in their cosy cottage up on the hill above Kalimpong in the 1980s and 1990s. The relationship observed by Sprigg and later by Jacques also highlights the importance of considering obscure but crucial tidbits of evidence.

### 3 The Black Mountains and Gwendolyn Hyslop’s acuity of vision

Some argue that the New World was not really discovered by a Genoese seaman sailing for the Spanish crown in 1492 because the ancestors of native American peoples had colonised the Americas via the Bering passage many millennia before. Likewise, when another Italian mariner, sailing
under an English flag, discovered Newfoundland in 1497, he was oblivious to the fact that Norse-
men had set ashore there several centuries before him. In this sense, a language too can never be
discovered, since any language is presumably already known to its speakers. For the rest of us, how-
ever, it would be fair to say that the New World was discovered several times.

With this qualification in place, the discovery of two new languages in Bhutan in 1990,
Gongduk and Black Mountain Mönpa, filled me with sheer delight. Of course, these languages
were only new in the sense of having been previously unknown to linguistics and, for that matter,
to the Royal Government of Bhutan, for whom I conducted the first Linguistic Survey of Bhutan
from 1989 to 1991. Whilst Gongduk and Black Mountain Mönpa were genuine discoveries, the
existence of a third unknown language, Lhokpu, was not entirely new to linguistics.

Bābu Kṛṣṇakānta Bos, alias ‘Kishen Kant Bose’, the famous Bengali spy sent by the British
to the Drû desi in 1815 to settle frontier disputes along the Bhutanese duars and foothills, men-
tioned the existence of a tribe by the Nepali name of ‘Doya’ (1825: 13), but he made no mention
of their language. Charles John Morris became the first to report that the ‘Daoyas’ had their own
Tibeto-Burman language (1935: 210). Later, Olschak mentioned the existence of ‘an archaic lan-
guage in the south’ (1979: 25), although she was unable to locate with complete accuracy where it
might be spoken. Other than these three references, nothing was known about this language.

The discovery of two new languages and the first linguistic material on this third unknown
language did not at the time, however, make a big media splash, as did the recent ‘discovery’ of Koro
in the neighbouring Indian state of Arunachal Pradesh. I had no formidable publicity machine
behind me, nor was there as much funding available for endangered language research in 1990 as
there is today, which might have justified an ostentatious presentation of the discoveries. Notably,
the subsequent discovery of a previously unknown Bodish language is eastern Tibet by Nicolas
Tournadre drew even less attention. As Post and Blench (2011) have pointed out, Koro was actu-
ally known to be a distinct linguistic group as early as Hesselmeyer (1868), and Koro data have
been presented in the literature under various names such as ‘Miri Aka’ and ‘Angka Miri’. Post and
Blench have also presented evidence for the identification of a new subgroup comprising Koro and
Milang, which they have christened Siangic.

Yet, in sequel to the survey, the Royal Government of Bhutan in 1999 formally accorded
a high priority to describing the ‘three gems’ of the country’s ethnolinguistic heritage in order to
preserve these three endangered native languages of the kingdom and to document them for pos-
terity. The generosity of the Royal Government of Bhutan enabled me to conduct three additional
longer missions to the Black Mountain Mönpa, Gongduk and Lhokpu language communities
under auspices of the Dzongkha Development Commission in order to describe and document the
grammars of the three gems in the years 2000, 2001 and 2002.

I had enthusiastically processed many of my field notes, but by May 2004 the work came
to a sudden impasse. A new operating system, called OS X, rendered impossible the printing of
any document containing tailor-made linguistic fonts. The new incompatibility adversely affected
phonetic fonts, rendered impossible the printing of documents and disabled all documents writ-
ten in the font called ‘Dzongkha’, the development of which had been commissioned by the Royal
Government of Bhutan specifically for the Apple platform. The new ‘fourth generation’ machines
supported the older operating system OS 9 inadequately as an emulation program which invariably
froze or crashed with any document containing one of these tailor-made fonts. A good number of
experts around the world were experimentally able to ascertain that the encoding problems were
not trivial or amenable to being remedied by cleverly designed macros or similar solutions.
Other data-oriented linguists have also faced the irreparable loss of data resulting from the planned obsolescence of software. The entire comparative Tibetan dialects dictionary project in Bern presently operates only on older machines which still support older fonts and older software, and the Bern Linguistics Institute was this year fortunately able to allocate funding to have the database re-engineered. Based on the similar experiences of quite a number of colleagues, Gérard Diffloth has long expressed skepticism regarding the longevity of digital information, and in retrospect his preference for data in handwritten notebooks now seems prudent. Due to the volume of linguistic material which I had processed in a variety of fonts, some projects were stalled for years. Recovering the completed work ultimately necessitated considerable investment in funding and man-hours. The dedicated software specialist Atanu Majindar came to Europe and spent months in 2008 salvaging the work that had been done on the three gems.

Meanwhile, the three publications which contained a modicum of Black Mountain Mönpa data (van Driem 1993, 1994, 2001) also tentatively classified the language as an archaic East Bodish language. However, my field notes and the electronic manuscript of the grammar, which had become frozen in time, boldly stated the hypothesis that Black Mountain constituted a totally distinct subgroup within Tibeto-Burman, poorly disguised by a plethora of East Bodish loans. The data had already compelled me to establish that both Lhokpu and Gongduk constituted distinct subgroups in their own right within the Tibeto-Burman linguistic phylum. Yet I was in no hurry to stake the same claim for Black Mountain Mönpa in print, since I somehow felt that I could and would come out with the Black Mountain grammar at any moment. Until then, I had chosen not to depict Black Mountain as a separate leaf in the agnostic Fallen Leaves model of the linguistic phylum, a diagram of which has appeared in several publications.

![Diagram 1: The agnostic Fallen Leaves model of the Tibeto-Burman linguistic phylum challenges linguists to discover the structure of the family tree by working up from the firmer ground of lower-level subgroups to the higher levels of superordinate subgroups. The conventional use of the terms ‘higher’ and ‘lower’, of course, inverts the metaphorical tree so that its roots are up in the air.](image-url)
Then at the 16th Himalayan Languages Symposium in London in September 2010, Gwendolyn Hyslop discussed the published Black Mountain Mönpa data and began, ever so gingerly and diplomatically, to question the contention, which had already been enshrined in print whilst the electronic manuscript of the grammar languished, that the language was a member of the East Bodish subgroup. After her talk, I had to publicly proclaim mea culpa and confess that her insights, on the basis of just the few data available to her, were entirely correct. The view which she had advanced in ever so delicate a fashion was entirely borne out by the language data which she had not yet seen. In tribute to Gwen’s acuity of vision, I reproduce here an updated Fallen Leaves diagram with Black Mountain Mönpa depicted as a distinct subgroup. It now appears that all three Bhutanese gems represent independent Tibeto-Burman subgroups in their own right.

**References**


Hyslop, Gwendolyn. 2010. ‘On the placement of East Bodish in Tibeto-Burman’, lecture given at
the 16th Himalayan Languages Symposium, Saturday, 4 September 2010, School of Oriental and African Studies, London.


