Scholarship on Trans-Himalayan (Tibeto-Burman) languages of South East Asia

Nathan W. Hill

September 14, 2021

1 Introduction

The spread of the Trans-Himalayan family\(^1\) naturally paid no attention to 21st century political boundaries.\(^2\) The family includes languages with a geographic range from Balti Tibetan in Pakistan to Hokkien Chinese in Indonesia, with the foothills of the Himalayas and the South East Asian highlands as its center of gravity. Van Driem (2001) and Thurgood (2017a) provide helpful introductions to the family overall.\(^3\) Here we restrict the focus to South East Asia, in more specific terms treating those Trans-Himalayan branches that include languages spoken in today’s Myanmar and Thailand and excluding Chinese. I include discussion of subbranches entirely contained within the boundaries of the Peoples Republic of China (Rgyalrongic, Qiangic, Ersuic, and Naish), but omit treatment of primary branches entirely confined to regions under the control of China, India, Nepal or Bhutan (Bai, Tujia, Kiranti, etc.); these criteria yield Burmo-Qiangic, Kuki-Chin, Karen, Sal, Mruic, and Nungish as the branches for discussion.\(^4\)

The farther South and East a language is spoken, the more it exhibits the typical South East Asian typological profile of simple syllable structure, lack of inflection, and concatenating auxiliary verbs. Karenic, as the most southern of the Trans-Himalayan subgroups, reflects the vanguard of this transition, whereas the Rgyalrongic languages of Sichuan exhibit the opposite extreme. The frequency of the South East Asian typology in the Trans-Himalayan family is what led Meillet to despair that “la restitution d’une « langue commune » dont le chinois, le tibétain, etc., par

\(^1\)This family is also called Indo-Chinese, Tibeto-Burman, or Sino-Tibetan, (see van Driem 2014).
\(^2\)I would like to thank Guillaume Jacques, Mathias Jenny, Lai Yunfan, Alexis Michaud, and David Solnit for helpful pointers while I wrote this piece.
\(^3\)Other useful reference works include the 言語学大辞典 Gengogaku Daijiten (1988-2001), which includes over forty entries on Trans-Himalayan languages, and 云南特殊语言研究 Yunnan têshâ yîyán yânyû (2004), which focuses on languages spoken in Yunnan.
\(^4\)Bodic is also very marginally reflected because of the handful of Tibetan speaking villages of northern Burma (Suzuki 2012). But here is hardly the place for a survey of Tibetan linguistics; for part of such a survey see Hill and Gawne (2017).
example, seraient des formes postérieures, se heurte à des obstacles quasi invincibles” (1954, 26-27).⁵ Such pessimism is not entirely warranted. On the one hand, historical linguistics is still profitably undertaken even in such innovative branches as Naic (Jacques and Michaud 2011) and Karenic (Haudricourt 1946; Haudricourt 1975). On the other hand, Kuki-Chin, Sal, Mruic, and Nungish all have inflectional morphology of the kind that has facilitated progress in the reconstruction of Indo-European. As data on more languages become available, it is increasingly clear that the typological profile of the Trans-Himalayan proto-language is close to that of the Rgyalrongic languages, with complex syllable structure and ornate inflection; the more typically South East Asian languages have lost these features more or less independently. While it is inappropriate to speculate too precisely about prehistoric migrations on the basis of language distributions today, without the corroborating evidence of genetics or archaeology (pace LaPolla 2012), the broad pattern–languages with complex syllable structure and abundant inflectional morphology spoken in more mountainous terrain contrasting with languages of more simple structure spoken in flatter and more southern regions–points to an Urheimat inside of what is now China. The Lolo-Burmese seem to be a relative newcomer to South East Asia, with Nungish, Sal, Kuki-Chin, and Karen having spread earlier.

2 Research trends

The practical need of communication, particularly in commercial and diplomatic circumstances is typically what drove the study of foreign languages in the pre-modern period. The first records of Trans-Himalayan languages of South East Asia, arose in the diplomatic entanglements of successive Chinese empires. This earliest evidence of a Trans-Himalayan language of South East Asia is three songs that a Bailang delegation presented to the Chinese court (58-75 CE). The songs so delighted the emperor that he had the songs recorded for posterity. Many centuries later the diplomatic requirements of the Ming and Qing dynasty led to the compilation of the Huáyí yìyǔ (華夷詞語) vocabularies. In many cases these reflect an older stage of a particular language (in particular Rgyalrong and Tosu) than we would otherwise know.

The advent of European colonialism brought increased information on languages of South East Asia. Investigators were generally either colonial officials or missionaries. Later, the explicitly atheist Marxist ideology of the Peoples Republic of China also required the classification of and to some extent documentation of that state’s subject peoples. The major ethnolinguistic surveys of the 1950s in the PRC were only published after the turbulence of the Cultural Revolution subsided. The publications, when they did emerge, provided invaluable comparative word lists of Trans-Himalayan, namely Huáng 1992 and Sūn 1991.⁶ Today most language documentation available to academic linguists is produced by academic linguists, but missionaries also continue to produce useful resources. Most linguists engaged in the documentation of the Trans-Himalayan languages have a functional-typological rather than a historical orientation. As such, the collection and publication of lexicographical resources and text collections receives much less attention than grammar. Although language communities themselves produce a certain amount of text in their own languages, this work tends to be undervalued, underused, and not even collected.

⁵In English, “the reconstruction of a proto-language of which Chinese, Tibetan, etc. for example, would be the descendants, faces almost nearly insurmountable obstacles”.

⁶These sources have been conveniently digitized by the STEDT project, based at the University of California at Berkeley (1987-2015).
General works on historical linguistics often present research on the Trans-Himalayan family as in keeping with the standards and methods of the discipline at large (e.g. Abondolo 1998, 8 Campbell and Poser 2008, 114). Such authors paint an overly rosy picture. In particular, the influential research tradition that Benedict (1972) inaugurated and Matisoff (2003) led, does not adhere to the comparative method in an orthodox manner (Chang 1973; Fellner and Hill 2019; Miller 1974; Sagart 2006). The contributions of Tatsuo Nishida, unfortunately rather neglected in the West, are generally more reliable; I make a point of mentioning his work in the relevant places below. An element of Nishida’s judiciousness is that he avoids proposing reconstructions, but focuses on establishing the facts of language change. Reconstructions of Trans-Himalayan subgroups typically associate proto-forms with the prominent correspondence patterns found in the group, but do not formulate sound changes and their relative chronology with sufficient precision to trace the reconstructions down to the attested forms. As such, one must approach the findings of such research with caution. If two languages share a sound change which occurred in their respective histories subsequent to a change that they do not share, then the shared sound change must be due to contact. There are a number of cases like this among the Trans-Himalayan languages of South East Asia. Chinese, Karen, Burmish, and Loloish all underwent quite similar tonal splits conditioned by initial manner (as did Thai and Vietnamese); the conditioning in each case are different. In the Burmish language the split only occurred in stop final syllables (Hill 2019, 55-56); the ‘Loloish tonal split’ appears to have occurred independently with the same conditioning, but the locus classicus for its description is a work that I find uniquely impenetrable (Matisoff 1972). In Karen the tonal split effected open and closed syllables equally, and in addition to a low register associated with original voiced onsets and a high register associated with original voiceless onsets, there is a mid register associated with implosive onsets (Kato 2018).

VanBik takes the change *s- > t- as diagnostic of Kuki-Chin, but it is a wider areal feature and also attested in Tangkhul (Mortensen 2003) Bodo-Garo (Burling 1959), and Bangru (Bodt and Lieberherr 2015). To give a third example, standard Burmese has changed r- to y- [j-] and shares this sound change with several Burmish languages (WBur. ran, SBur. jin, Lacid jaj³¹-‘chest’, but However, the Burmese dialect of Arakanese preserves r- distinct from y- (Okell 1995, 2), so the parallel change of Standard Burmese and Lashi occurred separately in the two languages.

In terms of historical morphology, causative formations and person agreement are the most widely discussed issue. It is traditional to see a devoicing *s- as responsible for alternations between voiced intransitive and voiceless transitives or alternations between unaspirated intransitives and aspirated transitives. Jacques has shown that this explanation is untenable. Rgyalrongic languages have both a s- prefix causative and a nasal prefix anticausative and the Burmese unaspirated versus aspirated pattern is cognate with the anticausative in Burmese (Bur. kya ‘fall, drop’ and Japhug ŋgra ‘fall’, versus khya ‘bring down, lower’ Japhug kra ‘bring down’ (Jacques 2014, 250, also see Jacques 2020).

As we saw, early researchers like Meillet saw an absence of agreement as characteristic of the Trans-Himalayan family overall. As more languages with agreement have come to be described, two logical possibilities arise to explain the typological diversity of languages in the family. LaPolla argues that isolating languages continue the typological pattern of the Trans-Himalayan proto-language, and that agreement emerged via pronoun incorporation, partly independently in different branches, but largely through a single common innovation that in his proposal constitute an enormous and geographically dispersed ‘Rung’ branch (LaPolla 2013). Most other researchers, most vocally DeLancey, believe that the proto-language had verb agreement that was independently lost in various branches (DeLancey 2015). This inky debate has generated more heat than
light, but slowly the detailed work of subbranch level morphological reconstruction is making progress.

Comparative syntax is only possible once phonology and morphology are advanced to a certain level and even in Indo-European, the world's best studied language family, the methodology of comparative syntax remains insecure and controversial (Watkins 1976). The publication of texts in Trans-Himalayan languages has not been a research priority and as such in most languages very few continuous texts are available for study, if any at all. The prospects of syntactic study in such circumstances are dismal. Trans-Himalayan languages are overwhelmingly verb final, the only exception in South East Asia are the Karenic languages. Their exceptional status is understood to have arisen through contact, but there is unlikely to ever be sufficient evidence to trace this transition. An area of great promise in Trans-Himalayan historical syntax is the variation between verb stems in Kuki-Chin. Broadly speaking each Kuki-Chin language has one verb stem predominantly used in finite contexts and another that favors non-finite contexts, but the specific distribution of the two forms varies quite subtly from one language to another. A close study across the Kuki-Chin family could teach a lot about the syntax of Proto-Kuki-Chin and about syntactic change in general. In addition, Scott Delancey in forthcoming work proposes that lexical doublets in Jinghpaw are explainable as cognates to different Kuki-Chin verb forms.

There is a tendency in descriptive work on Trans-Himalayan languages, particularly common in discussions of clause chaining, to speculate about particular grammaticalization pathways that give rise to converbial constructions, rather than to carefully illustrate the precise meaning of one construction versus another. My hunch is that many grammatical patterns wait to be discovered in the clause chaining of Trans-Himalayan languages; Tibetan has a switch-reference system that was only clearly explained in 2019 (Beer 2019), despite this being the best studied member of the family other than Chinese.

Looking to the future, research on Trans-Himalayan languages is constrained by the absolute number of investigators; the achievements seen in better studied families will only be achieved through methodological innovations that increase the productivity of an individual researcher. Fortunately, these innovations are nearly at hand. Automatic transcription of under resourced languages has the potential to speed up the collection and processing of fieldwork data into textual corpora (Adams et al. 2018; Do et al. 2014). Off-the-shelf natural language processing (NLP) tools speed up the glossing and translation of texts and the compilation of dictionaries. Automatic cognate detection can speed up the articulation of sound laws and the reconstruction of proto-forms (Bodt and List 2019; List and Hill 2017). To fulfill the promise of these new techniques, it is only necessary to incorporate more advanced technological training into education of the next generation of linguists.

3 Subgroups

In very few instances are the Stammämne of Trans-Himalayan languages rigorously justified by the identification of shared isoglosses. Nonetheless, a genealogically organized account provides a convenient way of structuring this survey of research on a large number of languages, and will orient the reader within the current state of discussions on subgrouping, despite the tentativeness of current understandings.
4 Burmo-Qiangic

The view that Lolo-Burmese and Qiangic languages are closely related has circulated for some time (Dempsey 1995, 13, Bradley 1997, Peiros 1998, Jacques and Michaud 2011). The subgroups within Burmo-Qiangic are Naic, Ersuic, Qiangic (including Rgyalrongic), Loloish, and Burmish. The grouping together of Loloish and Burmish as ‘Lolo-Burmese’ is universally accepted. Tangut (Rgyalrongic, txg), Burmese (Burmish, mya), and Yi (Loloish, iii) are the Burmo-Qiangic languages of medieval attestation, respectively attested from 1036, 1113, and 1485. Naxi (Naic, nxq) is written from the 18th century. Features characteristic of the Burmo-Qiangic languages include a contrast of velarized and plain vowels, a complex system of directional prefixes that double as past tense markers, and inverse agreement marking in the verb. The more conservative languages show all of these features, but in many subbranches only relics or indirect evidence is preserved.

In the 1960s and 1970s the easy accessibility of Thailand to outsiders, compared to Burma and China, made Loloish the focus of both documentation and reconstruction by Western scholars, but now the Qiangic side of Burmo-Qiangic is where one finds the vanguard of Neogrammatically progress in Trans-Himalayan reconstruction. Jacques (2014), consolidating earlier forays, devotes a monograph to Tangut-Rgyalrong comparative phonology and grammar. The clarity and comprehensiveness of his presentation has in turn enabled others to make quick discoveries. In particular, Gong (2020) finds that the mysterious Tangut ‘grades’ (等 děng) correspond to the velarized versus non-velarized syllabic contrast met in Rgyalrongic languages and Sims (2020) shows that the tonal contrasts of northern Qiang dialects correspond to the Tangut tonal contrasts.
4.1 Naic

Naic consists of Namuyi (nmy), Xumi (sxg), and ‘Naish’, with the latter made up of Naxi (nxq), Na (nru), and Laze (Jacques and Michaud 2011). For Namuyi, there is a deposit at the Endangered Languages Archive (number 0217), and three recent treatments of its grammar (Li 2017; Pavlík 2017; Yǐn 2016). Very little work has been done on the Xumi language. There is a short grammar by Sun (2014) and a series of articles by Chirkova (see Chirkova 2017). The most well-known language of the family is Naxi, famed for the representational writing system used in the liturgical writings of Dongba priests. Michaud et al. (2017) provide an up-to-date survey of research on Naxi, including its written forms. Within Naish, Yongning Na has received most attention. Lidz (2010) provides an overall grammar. Michaud (2017) is specifically devoted to the very complicated tone system. The least documented language of this group is Laze (see Michaud and Jacques 2012).

Jacques and Michaud (2011) provide a preliminary reconstruction of Proto-Naish. Li has made a few further forays into Naish reconstruction (Li 2018; Li 2020). He is paradoxically critical of his predecessors for letting non-Naic languages inform their reconstructions, but he himself uses Tibetan comparisons (not always correctly) to inform his own reconstructions.⁷

4.2 Ersuic

The Ersuic subbranch of Burmo-Qiangic consists of only three languages, Ersu (ers), Lizu, and Tosu.⁸ Tosu is recorded among the Huíyì yìyǔ from the Qianlong (1735-1796) period (Nishida 1973b). Ersu was first recorded by Baber (1882). Lizu appears to have first been studied by Sun Hongkai, with partial publication of his data in Nishida and Sun (1990). Yu (2012) surveys the existing work on the family and provides a preliminary reconstruction for this subfamily, based on

⁷Li (2020) suggests that the C₁ of his own previous reconstruction is probably a nasal because it corresponds to ‘pre-initial’ nasals in Written Tibetan. I have to admit that this ends up being a bit odd; it we allow ourselves to symbolize this as ‘N’, we end up with a contrast in the proto-language between *Nŋg and *Ng-. I would also point out that one of the pre-initials that he suggests comparison to in Written Tibetan is, in my own analysis and that of other investigators a voiced velar fricative, and not prenasalization, but this is a point of controversy in Tibetan historical phonology.

⁸This Lizu language is not to be confused with the similarly named Lisu of the Loloish subbranch.
Ersuic and Lizu, as he had insufficient Tosu data available. Unfortunately, most of the changes, e.g.
the brightening of *-a- to *-i- are not very diagnostic.

Since Yu’s study considerably more data on these three languages have become available. For
Ersu, there is a deposit at the Endangered Archives Project hosts deposits for Ersu (MPI655457)
and a PhD dissertation Zhang (2013). Katia Chirkova has published new data on Tosu (2014),
including a grammar (Han et al. 2019).

4.3 Qiangic

Qiangic has the four daughters Prinmi (pmi, pmj), Muya (mvm), Rma (eng, qxs), and Rgyalrongic.
It is convenient to treat here work on those Qiangic languages other that Rgyalrongic, since the
body of scholarship on the latter is rather large. The documentation of Qiangic languages is not vast
but is increasing. For Prinmi, we have grammars by Daudey (2014) and Ding (2014). The Munya
languages is documented in a few articles of Ikeda (2002; 2006; 2008) and two PhD theses (Bai
2019; Gao 2015) The Rma language, spoken in Sichuan, consists broadly of two sets of dialects,
a Southern set and a Northern set, LaPolla and Huang (2003, 16-17) surveys early work on Rma,
and themselves offer a grammar of northern Rma as spoken in Ronghong Village.

4.4 Rgyalrongic

Apart from the dead language Tangut, the Rgyalrongic languages are confined to Sichuan. Under
pressure from the expanding Tibetan empire the ancestors of the Tanguts left the Rgyalrongic
homeland heading North East; they founded a polity in 984 in what is today Ningxia. On the basis
of lexical isoglosses and the distribution of morphological features such as case marking and direc-
tional prefixes on the verb, Lai et al. (Forthcoming) divide the Rgyalrongic branch into a western
and eastern subbranch, with Tangut on the western branch. Galambos (2015) provides a conve-
nient entry point to Tangut studies for the anglophone reader. The remaining West Rgyalrongic
languages are less well studied. For a survey of Horpa (ero) see Jacques et al. 2017. The Geshiza
variety of Horpa was more recently the subject of a Phd (Honkasalo 2019). On Khroskyub (jiq),
Lai has published several articles and a PhD dissertation (Lai 2015; Lai 2016; Lai 2017). The
Figure 4: The Stammbaum of Rgyalrongic.\(^a\)

Rgyalrongic

West Rgyalrongic

Tangut

Horpa

Khroskyabs

Stau

Geshiza

Cogtse

East Rgyalrongic

Situ

Japhug

Tshobdun

Zbu

Kyomkyo

Bragbar, etc.

\(^a\) Based on Lai et al. (Forthcoming), incorporating suggestions of Guillaume Jacques.

documentation of East Rgyalrongic languages begins in the 18th century with a Huáyí yiyă vocabulary of 800 words (Nishida and Sun 1990). Jacques (2004, 9-13) gives a history of Rgyalrong studies. In the intervening 16 years since his survey, the field his been very active. Jacques’ many contributions have made Japhug Rgyalrong (jya) among the best described Trans-Himalayan languages (Jacques 2017a). Sun has a number of papers on Tshobdun (jya) (Sun 2017), and has also contributed to the study of other Rgyalrong varities. For Situ ((jya)) there is a grammar of the Kyom-kyo variety (Prins 2016), and a collection of text in Cogtse (Lin 2016). For an overview of the East Rgyalrongic languages see Jacques (2017b).

4.5 Burmish

There are around seven Burmish languages, spoken in the hills of North Burma and on the other side of the border in China, namely Ngochang (also called Achang, acn), Zaiwa (Atsi, atb), Pela (Bela, bxd), Lacid (Leqi, Lashi, lsi), Lhao Vo (Langsu, Maru, mhx) and Hpun (Phon, hpo), and of course Burmese itself. The Burmese wandered out of the ancestral homeland and onto the planes around the 10th century, to the linguistic disadvantage of the Pyu. The Burmish group of languages itself splits into two, which Nishi calls Burmic and Maruic, following (1999, 70). The Burmic subbranch is characterized by the merger of inherited plain and preglottalized stops as aspirates, as seen in the merger of *kruk ‘six’ (Bur. khrrok, Longchuan Ngochang ʐoʔ⁵⁵ versus Zaiwa kjuʔ⁵⁵ Pela kjaʔ⁵⁵) and *ʔkruk ‘frighten’ (Bur. khrrok, Longchuan Ngochang ʐoʔ⁵⁵ versus Zaiwa kjuʔ⁵⁵, Pela kjaʔ⁵⁵).

I previously expressed reservations about the validity of Maruic, because Nishi characterized it by the shared retention of preglottalized initials (Hill 2019, 51-52), but the validity of the subgroup is confirmed by what we can call the ‘chicken-mouse’ split, whereby for ‘chicken’ and ‘mouse’ Burmic reflects respectively *grak (Bur. krak, Xiandao Ngochang ʐoʔ⁵⁵) and *grok (OBur.
Hill Trans-Himalayan languages of SE Asia

kro₁k, Longchuan Ngochang k₆'oʔ⁵⁵ whereas Maruic instead reflects *rak (Lhao Vo γ₀ʔ³¹, Pela γ₀ʔ³¹pha⁵⁵) and *rok (Lhao Vo γ₀ʔ³¹hoʔ³¹, Pela γ₀ʔ³¹haoʔ³¹).

Dai Qingxia and his collaborators have contributed massively to the documentation of Burmish languages including book length grammars of Ngochang (Dài and Cuī 1985), Pela (Dài et al. 2007), Lacid (Dài and Li 2007), and Lhao Vo (Dài 2005), Chashan (Dài et al. 2010). The best documented Burmish language, other than Burmese itself is Zaiwa, with three book-length grammars (Xú and Xú 1984, Lustig 2010, Zhū and Lèpáizǎozā 2013).


The oldest document in Burmese is the Myazedi inscription of 1113 CE (Nishida 1955; Nishida 1956; Yabu 2006). Essentially all documents in Old Burmese are stone inscriptions recording land grants to Buddhist establishments (Frasch 2018). The family that consists of languages descending from Old Burmese, known as Burmese dialects, includes ‘Standard Burmese’ of Rangoon (Yangon) and Mandalay, Tavoyan (Dawei) and the closely related Palaw, Yaw, Merguese (Beik), Intha, Danu, Arakanese (Rakhine) and the closely related Marma and Taung’yo (Bernot 1965, Jones 1972, Okell 1995, Naksuk 2012). I call this the ‘Mranmaic’ language family. Compared to its older cousin Tangut, serious study of the history of Burmese from a linguist perspective is languishing. Hill (2019, 46-83) provides a general orientation, to Burmese and its place within the Burmish family.

There have been a handful of forays into the reconstruction of proto-Burmish, none of them wholly successful. Neither Burling 1967 nor Mann 1998 employ Written Burmese or Old Burmese data in their correspondence sets. Nishi 1999 brings together a number of cognate sets, assembled from Huáng 1992. A disadvantage of Nishi’s approach is that he only offers cognate sets where there is a cognate in Burmese. In addition, he throws out the full words. Dempsey offers two insightful contributions on Burmish language history (2001; 2003), but did not integrate his findings into an overall reconstruction.

4.6 Loloish

There are many Loloish languages, some among them are quite well studied. The majority, however, are only known from comparative word lists used in more or less naïve lexicostatistical studies (Lama 2012; Satterthwaite-Phillips 2011). There was a great deal of work on this branch during the 1970s; work continued subsequently, but at a slower pace. Figure 6 presents the Loloish Stammbaum according to Bradley (2007), but the “classification of the Loloish languages must be re-established in the future by considering more data sets and also grammatical features” (Gerner 2013, 7). The general disjuncture between linguistic reality and the official Chinese classification of its subject peoples reaches its apogee in the treatment of speakers of Loloish languages. Katso

---

⁹An easy solution for reconstructing these words at the proto-Burmish level is to posit *gərak ‘chicken’ and *gərok ‘mouse’.

¹⁰In addition, Yabu (1982) and Wannemacher (1994) and Wannemacher (1998) have made smaller contributions.
(kaf) speakers are absurdly classified as Mongolian. Speakers of Hani (hni), Lahu (lhu), Lisu (lis), and Jino (jiu, jiy) are recognized as having their own respective nationalities. The remaining speakers of Loloish languages are mostly grouped together as ‘Yi’, although this term is particularly associated with Nosu (iii), the official language of the Yi nationality and the only Loloish language to enjoy limited government support (Gerner 2016). A number of the languages called Yi in China, used, and to some extent still use, a family of logographic scripts, predominantly for religious texts (Iwasa 2018; Wasilewska 2014).

If we concentrate on those languages for which a reasonable level of documentation has been achieved, for Northern Loloish we have only Nosu (Gerner 2013). The Central Loloish branch contains the best studied languages including Lahu, for which there are a number of grammars (Cháng 1986; Li 2014; Matisoff 1973) and dictionaries (Matisoff 1988; Matisoff 2006). There are dictionaries of both Northern (Bradley 1994) and Southern (Bradley 2006) dialects of Lisu. The Southern Lisu use the Fraser script, invented by the eponymous missionary. Turning to Southern Loloish, Lewis and Bai prepared comprehensive dictionaries for both Hani (Lewis and Bai 1996) and Akha (ahk) (Lewis 1968), as well as text collection again for both Hani (Lewis and Bai 2002) and Akha (Lewis 2003). The 12 volumes of the series ‘Regional Culture Investigation of International Hani/Aka’ published in 2011, provides information on Hani and Akha as spoken inside China at a county by county level. Tatsuo Nishida has done important work in the Central and Southern Loloish, including field work on Akha (1966), Lisu (1967; 1968), Lahu (1969), and particularly Bisu (bzi), a language that he was the first to document (1966; 1973). Not many resources are available for the Southeastern Loloish languages, but Pelkey (2011) provides some useful lexical lists for Phula.

The only book-length treatment of Loloish reconstruction is Bradley 1979. Bradley compares Lahu, Lisu, Bisu, Phunoi, Akha and Mpi (mpz). His work very thoroughly discusses the history of research on these languages and previous sources and previous efforts at reconstruction. The reconstruction per se is however not very successful, out of date at the time of its publication (Thur-

5 Kuki-Chin

Kuki-Chin languages are spoken throughout the hilly terrain of Myanmar (the Chin hills), North East India, and Bangladesh (the Chittagon hills). The verbal systems of these languages typically have person agreement. Their verbal systems are also noteworthy for the use of two verbal stems, the second usually having a more complex final and typically thought to originate from nominalization. The patterns of use of these two stems varies considerably from language to language and should be investigated more.¹¹ Peterson (2017b) discusses subgrouping extensively. He divides the branch into the Maraic, Northwestern, Central, and Peripheral branches.

Lorrain 1951 is the *locus classicus* for Mara (mrh), which unfortunately does not mark tones. Löffler (2002) provides a discussion of earlier work on Mara as well as insightful comments on Mara in its comparative context. Northwestern (formerly called Old Kuki) centers on the Indian state of Manipur; it is a poorly studied group of languages. A synthesis of work on this group is a major desideratum of Kuki-Chin studies. In contrast, the Central branch has received most attention, with Mizo (lus) as probably the best studied of the Kuki-Chin languages. Lewin (1874) provides an early textbook. Lorrain (1940) offers a full Mizo dictionary; this work has been extremely influential in Trans-Himalayan studies, used by Benedict (1972) and Peiros and Starostin

¹¹It merits mentioning that the recent fad in anthropological circles to discuss ‘Zomia’ at least etymologically is referring to this area. Their autonym is Zo in many languages.
Chhangte has written both a preliminary grammar (1986) and a more complete study of syntax (2001). For Hakha Lai (cnh) Peterson 2017a and references therein. Reichle (1981) offers a grammar of Bawm (bgr). According to Peterson the Peripheral branch itself includes the three subbranches Northeastern, Southern, and Khomic. Of these three, the Northeastern branch is better studied and has played more of a role in reconstruction. Henderson’s (1965) treatment of Tedim (ctd), despite its unambitious title, is one of the more extensive treatments of any Kuki-Chin language. Sizang (csy) was the subject of some early pedagogical grammars (Naylor 1925; Rundall 1891), two articles in the 20th century (Stern 1963; Stern 1985), and is now being studied earnestly by Davis (2017). The change r- > g- is particularly characteristic of the Northeast branch (Solnit 1979); Peterson (2017b) discusses this change, convincingly showing that it is relatively recent and spread through contact.

In general the southern languages have received less attention than the Northeastern languages. (Houghton 1892) offers an early, and extremely insightful Asho (csh) vocabulary with comparative notes, which would reward careful study in light of more recent developments. There are full grammars of Daai (dao) So-Hartmann (2009) and Hyow (Zakaria 2018). The Language and Social Development Organization, a missionary organization affiliated with the Summer Institute of Linguistics, conducted a dialect survey of the Southern Chin languages from 2005 to 2014. The data, recently published, is an invaluable resource for future research (LSDO 2019a; LSDO 2019b).

David Peterson is researcher most active on the Khomic group. He has written on Khumi (cnk) extensively (Peterson 2019 and citations therein). Peterson is also documenting Rengmitca (2014), a highly endangered and phonologically archaic language.

There was a long hiatus in Kuki-Chin reconstruction following the pioneering efforts of Ohno (1965). Efforts in the reconstruction of Proto-Kuki-Chin accelerated in the 21st century (Button 2012; Hill 2014; Khoi 2001; VanBik 2009). In general the more northern languages present rimes in something like their ancestral form but radically simplify the onset whereas the southern languages have more conservative onsets and innovate more in the rimes. The existing reconstructions rely mostly on Northern languages, because more data has been available on them.

6 Karen

The most southern members of the Trans-Himalayan family are known for their SVO syntax and their simple syllable structure; most language have only open syllables. According to Manson “it would appear from the literature that there are between 20-30 distinct Karen languages” and sixteen of them “have been reasonably documented” Manson (2017, 150). Reasonably well documented varieties include Sgaw (ksw) (Binney 1883), Pwo (Kato 2017; Purser and Aung 1920), Kayah Li (Red Karen, eky, kyu, kvy, kxf) (Brown 1900; Solnit 2017), and Bwe (bwe) (Henderson and Allott 1997). The Pa-o language, because it retains nasal finals, is of great importance for reconstruction, but is unfortunately relatively understudied. Some Pwo data is included by Jones (1961) and two Pwo dialects are compared by Nishida (1967a). David Solnit is nearing completion on a new subgrouping of Karen varieties, in deference to which I omit a tree diagramm here.

---

12 James Matisoff claims to be in possession of a copy of Lorrain’s dictionary to which Siamkima Khawlhring added tone marks, but has not made it available to the scholarly public.

13 Justin Watkins has unpublished materials on Burmese Khumi.
Comparative work on the branch begins with a brilliant essay of A. G. Haudricourt (1946). He produced a reconstruction of Proto-Karen relying only on data from Pwo and Sgaw. The crux of his insight was to realize that the complex correspondence patterns for the onset manner and tones that he saw between the two languages reflected the aftermath of a tonal split according to manner class, just like also happened in Thai, Vietnamese, and Middle Chinese. Discussions with Gordon Luce, and the addition of data from Red Karen and White Karen allowed him to correct a few mistakes, to yield the reconstruction of onsets that is still held to today (Haudricourt 1953). In his third intervention, he relies on subsequent work of Luce (1959), Jones (1961), and Burling to add a fourth tone. Kato (2018) provides a synthesis of the reconstructed tonal system of Haudricourt’s first and third papers, helpfully adding data in the original Sgaw and Pwo scripts. Although a number of researchers have been active in Karen reconstruction since, very little further progress appears to have been made (Luangthongkum 2019).

He reconstructs three inherited manner types (the same as proto-Burmish incidentally) by making use of a tonal split, the same as seen in Chinese and Tai. He reconstructs a series of voiceless nasals, which are later confirmed by fieldwork of Luce. Haudricourt proposes *ŋ- > *ɲ- as an isogloss for the Karenic branch.

7 Sal

Burling (1983) proposes to call a group of languages ‘Sal’ on basis of a shared lexical item for ‘sun’, that is otherwise not seen in Trans-Himalayan languages. This family consists of three sub-branches, Bodo-Garo, Konyak, and Jingpho-Luish (Post and Burling 2017, 224-227). Of these three, only the third has members in South East Asia, so is our focus here. The Bodo-Garo family, spoken mostly in Northeast India, but also parts of Bangladesh, is one of the better studied Trans-Himalayan families. Joseph and Burling (2006) and Jacquesson and Breugel (2017) discuss subgrouping and provided reconstructions. More recent publications include a complete grammar of Rabha (Joseph 2007) and a survey article on Garo (Burling 2017). French (1983) provides a
reconstruction of the Konyak branch, but much descriptive work has subsequently taken place to merit a fresh look at this reconstruction.

The third Sal subbranch is, Jingpho-Luish, which consists of only seven languages, five on the Luish side and two on the Jingpho side. Progress in both subfamilies is now steady, thanks mostly to the intrepid efforts of two young Japanese scholars. Jingpho (kac) is the major language of Kachin state in Myanmar and is also spoken in India and China. As such, it was extensively researched already in the British Colonial period, and has received a lot of attention from Chinese linguists as well. It is one of the five languages used in Benedict’s (1972) influential study. The apparently archaic iambic syllables of the language, along with its preservation of final -r and -j, does make it a useful language for comparative linguistics. Kurabe provides a summary article (2017), a grammar (Kurabe 2016), and a discussion of historical phonology (Kurabe 2015). The major obstacle to the use of Jingpho for comparative linguistics has been its isolated status, but better study of the Luish languages has improved this situation. On the Luish side, Huziwara has come out with a grammar of Cak (ckh) (2008), a Cak dictionary (2016), and a reconstruction of the Luish subbranch (2012). The relationship between Jingpho and Luish will doubtless be an exciting area of research in the coming years.

8 Nungish

The Nungish family consists of the three languages Trung (duu), Rawang (raw), and Anong (nun), spoken on both sides of the Sino-Burmese border. For Trung there are a few survey articles (LaPolla 2017; Nishida 1987) a short grammar (Sun 1982) and a longer grammar (Perlin 2019). Perlin has also made available a deposit on Trung at the Endangered Languages Archive (Id: 0235). For Rawang there is a pedagogical grammar (Barnard 1934) and a text collection (LaPolla and Poa 2001). Thurgood (2017b) provides an overview of Anong, including a discussion of previous studies. Straub compiled an extensive bibliography of work on the Nungish languages (Straub 2020). To date there have been no efforts to reconstruct Proto-Nungish.

9 Mruic

The languages Mru (mro) and Anu-Hkongso (anl) together make up the small Mruic branch of Trans-Himalayan. These languages are spoken in the highlands of Burma and across the border in Bangladesh. Löffler (1966) elaborates correspondences between Mru, Mizo, Burmese, and reconstructed stages of Chinese.¹⁴ His work provides an excellent starting point for further historical consideration of Mru, but naturally must be updated, in particular with reference to newer Chinese reconstructions. Quite rare for Trans-Himalayan languages Mru distinguishes final -r, -l, and -j, although it may not be inherited. In a few interesting cases, Mru vocabulary compares suspiciously well with Tibetan (Mru pak, Tib. phag ‘pig’, Mru kim, Tib. khyim ‘house’, Mru tom, Tib dom ‘bear’); these similarities are indicative of Trans-Himalayan archaisms. The numerals are intriguing and thus merit repetition here: chum ‘three’, tali ‘four’, tanga ‘five’, taruk ‘six’, ranit ‘seven’, riat ‘eight’, taku ‘nine’. My proposal is that ta- is etymological only in ‘six’ and ‘nine’ and the ta- of ‘six’ spread through contamination to ‘four’ and ‘five’. One may be tempted to see the ch- of ‘three’ as somehow preserving a *k- prefix (compare Tib. gsum), but the word ching ‘tree’ (compare Tib. śin) instead suggests a regular change *s- > ch-.

¹⁴David Peterson has done extensive fieldwork on Mru that remains as yet unpublished.
10 Trümmersprachen

There are at least two ancient Trans-Himalayan languages of fragmentary attestation (Trümmer-sprachen) that fall within the scope of our survey. The older, and more fragmentary of the two is Bailang, know only from the three ‘Songs of Bailang’ (白狼歌). These are poems transliterated with Chinese characters and translated into Chinese during the Han dynasty (specifically 58-75 CE). In 1979, making extensive use of previous research, W. S. Coblin (1979) published a study of these songs. The author of these lines revisited the songs in light of ensuing progress in Chinese reconstruction (Hill 2017) and referring to the intervening more small scale studies (Mā and Dāi 1982, Zhèngzhāng 1993, Beckwith 2008). Whereas Coblin and Beckwith propose that the language is Loloish, I do not think that this conclusion is obvious. There are many etyma of obvious Trans-Himalayan provenance, but they are etyma also attested in other branches, e.g. 蘇 *sa ‘meat’ (Tib. ša, Bur. sāḥ, Mizo. sà ‘meat), 螗 *ruai ‘rain’ (Bur. rwā, Mizo. ṛāḥ), 鼙 *bi ‘give’ (OBur. piyḥ, Mizo pè). The only candidate lexical innovation that Sagart et al. (2019) say is indicative of Burmo-Qiangic that is found in the Bailang songs is 寬 *mus ‘heaven’, which compares with OBur. muivḥ ‘sky’ and Japhug tɯ-ɯ ‘rain, sky, weather’, but this word occurs in other branches as well (Tib. dmu ‘sky god’, Rawang dvmo ‘celestial spirit’).

As these examples show Bailang is phonologically quite innovative, and lest one credit the filter of Chinese transcription entirely, note that the Chinese were perfectly able a few centuries later to transliterate Tibetan clusters where they head them. For example, the Sino-Tibetan treaty inscription of 821-822 has Tib. stag ‘tiger’ (in a name) as 悉諾 [sir-*dak] (Preiswerk 2014, 51).

The Pyu (pyx) were the urban civilization that proceeded and was absorbed by the Burmese. Pyu literary activity covers the 6th to the 13th century CE. Griffiths et al. (2017) provides a complete inventory of Pyu inscriptions and a detailed study of a previously unpublished Sanskrit-Pyu bilingual inscription around the base of a headless Buddha statue. Miyake discusses the rimes of Pyu (2018) and Pyu grammar (2019). The team responsible for this renaissance of Pyu scholarship envisions a number further studies on aspects of Pyu linguistics and epigraphy. Pyu has quite conservative phonology, quite atypically for Trans-Himalayan languages of South East Asia, compare täk ‘one’, kni ‘two’, plä ‘four’, pā.ṅa ‘five’, tko ‘nine’, and tdü ‘water’, which look quite Tibetan in their tolerance for cluster onsets, compare Tibetan gcig ‘one’, gnis ‘two’, bźi ‘four’, lṅa ‘five’, dgu ‘nine’, and chu ‘water’.

Having admitted bafflement with the wider relations of either of our Trümmersprachen, it is interesting to note that they both share the same form of the first person pronoun, namely Bailang *ke ‘we, us’ and Pyu giy. In contrast, the vast majority of Trans-Himalayan languages have velar nasal initial first person pronouns. Ken VanBik (this volume) points to a first person pronoun like kay as indicative of Kuk-Chin, but before we attach our Trümmersprachen to that family, note similar looking forms like Olekha kō ‘I’ and Puxi Qiang qa ‘me’ (Jacques 2007).

References
Hill 

**Trans-Himalayan languages of SE Asia**

Figure 8: The Stammbaum of Sinitic

Old Chinese

Min

Middle Chinese

Bai, Jin, Xiang, Gan, Wu, Hakka, Yue, Mandarin


Dü, Qingxiä, Ying Jiâng, and Zhî’ên Kông. 2007. Boîlôyû yûnjî. 民族出版社 Minzú chūbānshè.

Dü, Qingxiä and Jié Li. 2007. Lèqîyû yûnjî. 中央民族大学出版社 Zhōngyâng mínzú dàxué chūbānshè.


Hill Trans-Himalayan languages of SE Asia


Hill

Trans-Himalayan languages of SE Asia


– 1956. Myazedi hibun ni okeru chūku Biruma-go no kenkyū (Studies in the ancient Burmese language through the Myazedi inscriptions) (2). In *Palaeologica* V.1, 17–32.


Hill Trans-Himalayan languages of SE Asia


