

The Sino-Tibetan Etymological Dictionary and Thesaurus

The
Sino-Tibetan Etymological
Dictionary
and
Thesaurus
–DRAFT–

James A. Matisoff

2014

The Sino-Tibetan Etymological Dictionary and Thesaurus Project
BERKELEY, CALIFORNIA
©2014 The Regents of the University of California
First Edition
PRINTED IN THE UNITED STATES OF AMERICA
Library of Congress Control Number: 2014911220
Library of Congress Cataloging-in-Publication Data available upon request

Compiled on 2014/11/19 at 10:59:00

Contents

| | | | |
|--|-----------|--|------------|
| Preface | v | Notational conventions and pan-allofamic formulas | x |
| Introduction | vi | Parallel but surprising examples | xii |
| History of the STEDT Project | vi | Notational conventions with respect to variation | xii |
| Original Plan of STEDT | vi | Similarities and differences between Tibeto-Burman and Indo-European patterns of variation | xii |
| The ST/TB Family in its Areal Context | vi | More on the concept of word family: blood relatives vs. adoptive ones | xii |
| Typological diversity of TB: Indosphere and Sinosphere | vi | Etymological accuracy and rectification of possible errors | xii |
| Scope and subgrouping of the TB family | vi | Desiderata for the future | xiii |
| Evolving notions of ST/TB subgrouping | vii | Acknowledgments | xiv |
| Teleo- and meso-reconstruction | vii | NSF and NEH Support | xiv |
| The problem of nomenclature | viii | Administrators | xiv |
| Special Challenges for ST/Eytmologization | viii | Contributors | xiv |
| Phonological System of Proto-Tibeto-Burman | ix | STEDTniks | xiv |
| Organization of the Etymologies | ix | Visiting Scholars | xv |
| The “tagging” process | ix | Complete List of STEDTniks | xv |
| Supporting forms and their attribution | ix | Conventions | xvi |
| Glosses of the supporting forms | ix | | |
| Notes | ix | References | 1 |
| Chinese comparanda | x | | |
| Treatment of loanwords | x | List of Languages | 3 |
| The Representation of Semantic Associations | x | | |
| Regularity and variation | x | | |

頭懸梁，錐刺股
Tóu xuán liáng, zhūi cì gǔ.
Hanging our hair from roof-beams,
pricking our thighs with awls.*

*This expression refers to the study practices of two celebrated young scholars: *Sūn Jìng* 孫敬 of the Han Dynasty would tie his hair to a roof-beam to avoid dozing off, while *Sū Qín* 蘇秦 of the Warring States period would keep himself awake by pricking his thigh with an awl. Both later had distinguished careers. At STEDT we have been inspired by their example.

Preface

THIS IS THE PREFACE

Introduction

History of the STEDT Project

The great Sino-Tibetan language family, comprising both Chinese and Tibeto-Burman (TB),¹ is comparable in time-depth and internal diversity to Indo-European, and equally important in the context of world civilization. The overwhelming cultural and numerical predominance of Chinese is counterbalanced by the sheer number of languages (some 250–300) in the rest of the family.

After the existence of the Tibeto-Burman family was posited in the mid-19th century, British scholars, missionaries, and colonial administrators in India and Burma began to study some of the dozens of little-known “tribal” languages of the region that seemed to be genetically related to the two major literary languages, Tibetan and Burmese. This early work was collected in the monumental *Linguistic Survey of India* (?), three sections of which (Vol. III, Parts 1, 2, 3) are devoted to wordlists and brief texts from TB languages.

Further significant progress in TB studies had to wait until the late 1930’s, when the eccentric amateur comparativist Robert Shafer headed a Depression-era project called “Sino-Tibetan Linguistics”, sponsored by the eminent anthropologist A.L. Kroeber of U.C. Berkeley.² With admirable thoroughness, the project staff assembled all the lexical material then available on TB languages, enabling Shafer to venture a detailed subgrouping of the family at different taxonomic levels, called (from higher to lower) divisions, sections, branches, units, languages, and dialects. This work was finally published piecemeal in a two-volume, five-part opus called *Introduction to Sino-Tibetan* (?).

Shafer’s junior collaborator Paul K. Benedict based his own work on the same body of material as Shafer, but achieved much more usable results. In an unpublished manuscript entitled *Sino-Tibetan: a Conspectus* (ca. 1942–43; henceforth ?), Benedict adopted a more modest approach to supergrouping and subgrouping than Shafer, stressing that many TB languages had so far resisted precise classification. While Shafer had included Tai in Sino-Tibetan, Benedict (?) banished it from the family altogether, relating Tai instead to Austronesian.³ Shafer’s pioneering work, valuable as it was, suffered from his mistrust of phonemics, with a consequent proliferation of pseudo-precise and arcane phonetic symbols. Benedict’s structural insight, his flair for isolating that which is crucial from masses of data, enabled him to formulate sound correspondences with greater precision, and to distinguish between regular and exceptional phonological developments.

The publication of a revised and heavily annotated version of ? in 1972, with J. Matisoff as contributing editor, laid the foundations for modern Sino-Tibetan historical/comparative linguistics. In this recension, nearly 700 Proto-Tibeto-Burman (PTB) roots were reconstructed (494 of them in numbered cognate sets, with about 200 more scattered throughout the text and footnotes), as well as some 325 comparisons of PTB roots with Old Chinese etyma, largely as reconstructed in ? (*Grammata Serica Recensa*, henceforth ?). While Benedict focused principally on five key, phonologically conservative TB languages (Tibetan, Burmese, Lushai [=Mizo], Kachin [=Jingpho], Garo), he also used data from more than 100 others, judiciously making allowances for inadequacies of transcription where necessary.⁴ How well has ? stood the test of time? The short answer is: remarkably well. The work has been reviewed about 15 times, almost always in a highly favorable tone,⁵ and has been translated into Chinese.⁶ In fact nearly all 700 of the TB cognate sets in ? have been shown to be perfectly valid, though many of the reconstructions have had to be changed slightly in the light of new data, and in a couple of cases etyma which had been reconstructed separately have been shown to be variant forms (“allofams”) of the same word-family.⁷

Despite its brilliance, however, the *Conspectus* is notoriously difficult to use, largely due to its complex apparatus of footnotes, which often (especially in the Chinese section) occupy more of the page than the text itself. These notes include Benedict’s original ones from the 1940’s, as well as those he and I added before publication in 1972. Some 200 valid etymologies are squirreled away in these convoluted notes, but they also contain a number of errors, unsubstantiated speculations, and over-complications.

Benedict himself realized the limitations of the data he had to work with, and never intended ? to be more than an overview or “conspectus” of its vast subject. Neither did he structure it as a practical handbook which systematically tabulated the sound correspondences among the major languages of the family at all canonical points of the syllable. (Such information is certainly extractable from the terse but labyrinthine pages of ?, but at the cost of considerable labor.) Towards the end of his life, which ended tragically in a car accident on July 21, 1997, Benedict does seem to have felt the need to embark on such a systematic project, although it never actually got off the ground.

After the publication of the *Conspectus*, further progress in intra-TB and TB/Chinese comparison seemed to depend on multiplying the number of reliably reconstructed etyma, as well as systematizing and refining the methodological underpinnings of the reconstructions. In the mid-1970’s, when I was attempting to apply the principles of glottochronology in order to subgroup the TB family, the very first item of “basic vocabulary” that I looked at happened to be ‘belly / stomach’. Much to my initial dismay, I quickly found that it was futile to use a simple wordlist to try to subgroup a family as complex and ramified as TB. In fact it was impossible even to deal in isolation with a single point in semantic space; etyma with the meaning ‘belly’ or ‘stomach’ spilled over into concepts like ‘cave / hole’, ‘swelling’, ‘calf of leg’, ‘liver’, ‘guts’, etc. I became preoccupied with notions of semantic variability, semantic fields, and the field of bodypart nomenclature in particular. At the same time I could not help noticing the morphophonemic variations displayed by almost every etymon previously or newly reconstructed. Instead of guiltily sweeping these variational phenomena under the rug, I began to revel in them. In *Variational Semantics in Tibeto-Burman* (?), henceforth ?) I set out to establish an explicit methodology for handling phonosemantic variation in word families, introducing the notion of “allofams” and a notation for diagramming patterns of semantic association (“metastatic flowcharts”). In those pre-computer days, I naturally had to assemble my data by hand, copying out bodypart words from dictionaries and sorting them into synonym sets on filecards, then grouping them into putative cognate sets. The older sources used by Shafer and Benedict were supplemented by an ever-increasing volume of new material in the 1970’s and 1980’s, much of it from post-Cultural Revolution China, but also from India, Nepal, and Thailand. It eventually became apparent that the job of digesting these massive amounts of new and old data would be vastly facilitated by the use of computers. The hitch was my own ignorance of computer technology beyond the level of simple word-processing. Fortunately I somehow got the idea of applying to federal granting agencies for a longterm project to create a computerized etymological dictionary of Tibeto-Burman / Sino-Tibetan based on semantic principles, i.e. an etymological thesaurus.⁸ In 1987, the Sino-Tibetan Etymological Dictionary and Thesaurus Project (STEDT) got under way, funded jointly by the National Science Foundation and the National Endowment for the Humanities. Thanks to the efforts of a succession of computer-savvy graduate students (see the *Acknowledgments*), a massive lexical database of forms from over 250 TB languages and dialects has been created, mostly of bodypart terminology at first, but then extending to all the other areas of the lexicon. It has been a race between the vertiginous progress of computer technology (when we started in the Pleistocene, 1987, we were using Mac Pluses!) and our ever-expanding needs for disk capacity, memory, and operating speed. The hard-won experience gained at the STEDT project has inspired similar lexical database projects in the U.S. and abroad.

Original Plan of STEDT

It was originally planned to publish the *Sino-Tibetan Etymological Dictionary and Thesaurus* as a series of printed volumes, each containing full details on all the etymologies in a given semantic area, starting with bodyparts and then proceeding to animal names, natural objects, verbs of motion, and all the rest of the lexicon:

| |
|---|
| Volume I: <i>Body Parts</i> |
| Volume II: <i>Animals</i> |
| Volume III: <i>Natural Objects, Plants, Foods</i> |
| Volume IV: <i>Kinship Terms, Ethnonyms, Social Roles</i> |
| Volume V: <i>Culture, Artifacts, Religion</i> |
| Volume VI: <i>Verbs of Motion, of Manipulation, and of Production</i> |
| Volume VII: <i>Adjectival Verbs</i> |
| Volume VIII: <i>Abstract Nouns and Verbs, Psychological Verbs, Verbs of Utterance</i> |
| Volume IX: <i>Shape, Size, Color, Measure, Number, Time, Space</i> |

^[1]Some scholars, especially in China, consider Sino-Tibetan to include the Tai-Kadai (TK) and Hmong-Mien (HM) (= Miáo-Yáo) language families as well. While there is definitely a striking typological similarity among Chinese, TK, and HM, this is undoubtedly due to prolonged ancient contact rather than genetic relationship.

^[2]For a readable and humorous account of this project, see ?.

^[3]To this putative megalolinguistic grouping, later to include Hmong-Mien and Japanese as well as Tai-Kadai and Austronesian, Benedict gave the name “Austro-T(h)ai” (see ??).

^[4]? followed Benedict’ s example in their choice of key TB languages, basing their Sino-Tibetan reconstructions on Written Tibetan, Written Burmese, Lushai, Jingpho, and Chinese, all of which are treated as if they belonged on the same taxonomic level. See the discussion in ?, Ch. 3.

^[5]A notable exception is the intertemper review by ?, which bitterly criticizes the fact that the notes added in 1972 sometimes modify points made in the original text (ca. 1942). See the defense of ? against Miller’s attack by ?.

^[6]See ?.

^[7]E.g. ***dyam** ✕ ***tyam** [? #226] ‘full; fill’ and ***dyam** [? #227] ‘straight’ ; see ?.

^[8]The shining example of an etymological thesaurus in the field of Indo-European is Carl Darling Buck’s *A Dictionary of Selected Synonyms in the Principal Indo-European Languages* (?).

| |
|------------------------------------|
| Volume X: <i>Grammatical words</i> |
|------------------------------------|

The sheer amount of the etymologizable data soon made it clear that this was unrealistic, and that each projected volume of STEDT would have to be split up into smaller units or “fascicles”, e.g. in the case of *Body Parts* into ten subdivisions, as follows:

- Body (general)*
- Head and Face*
- Mouth and Throat*
- Torso*
- Limbs, Joints, and Body Measures*
- Diffuse Organs*
- Internal Organs*
- Secretions and Somatophonics⁹*
- Reproductive System*

Every subpart and sub-subpart of the lexicon expanded and bloomed into a major project.

I decided to start with the reproductive system, not only because of its prurient interest but also because it seemed like the point of departure for all things. Accordingly a printed manuscript of some 480 pages was produced in 1997–98, called *Sino-Tibetan Etymological Dictionary and Thesaurus, Volume I: Bodyparts, Fascicle 1: The Reproductive System*, containing 286 pages of forms assembled into 174 cognate sets, divided into nine chapters: (1) Egg, (2) Birth, (3) Navel, (4) Breast, (5) Vagina, (6) Womb, (7) Penis, (8) Copulate, (9) Body Fluids. As part of the front matter, I put together a 60 page essay on the initial consonants and consonant clusters of Proto-Tibeto-Burman.

As it turned out, perhaps fortunately, that introductory essay soon took on a teratoid life of its own, and became an example of what one might call in Proto-Tibeto-Burman

| | | | | | |
|----------------|----------------|----------------|----------------|----------------|----------------|
| | | | | | |
| *kʷəy | lətak | rəmay | gəya:p | way | |
| dog | ACC | tail | wag | COP/NOM | |

or “the tail wagging the dog”.¹⁰ Was I not responsible for dealing with the whole proto-syllable, not just the initial consonants? I delayed publication of the “Reproductive Fascicle” until I could get the whole job done.¹¹ Eventually, the “introductory essay”, then entitled *System and Philosophy of Tibeto-Burman Reconstruction*, grew to become *Handbook of Proto-Tibeto-Burman* (?), henceforth ?).

The phonological approach of ? is complementary to the main thrust of the STEDT project, which is semantically organized.

The ST/TB Family in its Areal Context

Typological diversity of TB: Indosphere and Sinosphere

The TB family, which extends over a huge geographic range, is characterized by great typological diversity, comprising languages that range from the highly tonal, monosyllabic, analytic type with practically no affixational morphology (e.g. Loloish), to marginally tonal or atonal languages with complex systems of verbal agreement morphology (e.g. the Kiranti group of E. Nepal). While most TB languages are verb-final, the Karenic and Baic branches are SVO, like Chinese.

This diversity is partly to be explained in terms of areal influence from Chinese on the one hand, and Indo-Aryan languages on the other. It is convenient to refer to the Chinese and Indian spheres of cultural influence as the “Sinosphere” and the “Indosphere”.¹² Some languages and cultures are firmly in one or the other: e.g. the Munda and Khasi branches of Austroasiatic, the TB languages of Nepal, and much of the Northeast Indian branch of TB (notably Meithei = Manipuri) are Indospheric; while the Hmong-Mien family, the Kam-Sui branch of Kadai, the Loloish branch of TB, and Vietnamese (Mon-Khmer) are Sinospheric. Others (e.g. Thai and Tibetan) have been influenced by both Chinese and Indian culture at different historical periods. Still other linguistic communities are so remote geographically that they have escaped significant influence from either cultural tradition (e.g. the Aslian branch of Mon-Khmer in Malaya, or the Nicobarese branch of Mon-Khmer in the Nicobar Islands of the Indian Ocean).

Elements of Indian culture, especially ideas of kingship, religions (Hinduism/Brahminism, Buddhism), and *devanāgarī* writing systems, began to penetrate both insular and peninsular Southeast Asia about 2000 years ago. Indic writing systems were adopted first by Austronesians (Javanese and Cham) and Austroasiatics (Khmer and Mon), then by Tai (Siamese and Lao) and Tibeto-Burmans (Pyu, Burmese, and Karen). The learned components of the vocabularies of Khmer, Mon, Burmese, and Thai/Lao consist of words of Pali/Sanskrit origin. Indian influence also spread north to the Himalayan region. Tibetan has used *devanāgarī* writing since A.D. 600, but has preferred to calque new religious and technical vocabulary from native morphemes rather than borrowing Indic ones.

What is now China south of the Yangtze did not have a considerable Han Chinese population until the beginning of the current era (??). In early times the scattered Chinese communities of the region must have been on a numerical and cultural par with the coteritorial non-Chinese populations, with borrowing of material culture and vocabulary proceeding in all directions (???). As late as the end of the first millennium A.D., non-Chinese states flourished on the periphery of the Middle Kingdom (Nanchao and Bai in Yunnan, Xixia (Tangut) in the Gansu/Qinghai/Tibet border regions, Lolo (Yi) chieftaincies in Sichuan. The Mongol Yuan dynasty finally consolidated Chinese power south of the Yangtze in the 13th century. Tibet also fell under Mongol influence then, but did not come under complete Chinese control until the 18th century.

Whatever their genetic affiliations, the languages of the East and SE Asian area have undergone massive convergence in all areas of their structure phonological, grammatical, and semantic.¹³ Hundreds of words have crossed over genetic boundaries in the course of millennia of intense language contact, so that it is often exceedingly difficult to distinguish ancient loans from genuine cognates.

Scope and subgrouping of the TB family

The exact number of TB languages is impossible to determine, not only because of the elusiveness of the distinction between “languages” and “dialects”, and the fact that a number of languages remain to be discovered and/or described, but especially because of the profusion and confusion of different names for the same language.¹⁴ At the present state of our knowledge we can estimate that the Tibeto-Burman family contains approximately 440 languages, which may be broken down into population categories as indicated in Figure 1.

| | | |
|--|---------------------------|----------------------------|
| | <i>Number of Speakers</i> | <i>Number of Languages</i> |
| | more than 1,000,000 | 12 |
| | 500,000 – 999,999 | 13 |
| | 250,000 – 499,999 | 15 |
| | 100,000 – 249,999 | 41 |
| | 50,000 – 99,999 | 37 |
| | 25,000 – 49,999 | 52 |
| | 10,000 – 24,999 | 82 |
| | fewer than 10,000 | 185 |

 Figure 1: TB languages by number of speakers¹⁵

According to Ethnologue estimates, there are twelve Tibeto-Burman languages with over a million speakers: Burmese, Nuosu (Yi/Lolo), Amdo Tibetan, Meit(h)ei (= Manipuri), S’gaw Karen, Khams Tibetan, Bodo, Eastern Tamang, Central Tibetan, Eastern Pwo Karen, Garo, and Rakhine. Though care has obviously been taken to obtain the most reliable data, such figures can often be quite misleading, since ethnic identity and linguistic competence do not always coincide: of the 8 million people who identify themselves as Tujia, only about 70,000 can actively use the Tujia language, the rest having shifted to Chinese (?). Designations like *Yi* and *Karen* actually apply to large numbers of related but often not mutually intelligible “dialects”, which might as well be called separate languages. At the bottom of the scale are some 185 languages with fewer than 10,000 speakers, many of which are highly endangered,¹⁶ to the point where children are no

^[9]By ‘somatophonics’ I mean sneezes, belches, farts, and the like.

^[10]The presence of the accusative particle ***lətak** is motivated by the semantic anomalousness of this phrase, which has also caused the fronting of the object ***kʷəy** ‘dog’ to initial position.

^[11]It was later published as ?, henceforth ?.

^[12]See ? (“On megalocomparison.”)

^[13]An excellent recent study of such phenomena is ?.

^[14]See ?, and STEDT Monograph #2 (?).

^[15]Based on ?.

^[16]See ?.

longer learning to speak them.

Although much of the geographical area covered by TB languages¹⁷ has been chronically inaccessible to fieldwork by scholars from outside, there has been an explosion of data in the past 25 years, especially from China¹⁸ and Nepal, but also increasingly from India and Burma.

Evolving notions of ST/TB subgrouping

Millennia of language contact due to migrations, intermarriage, and warfare have presented us what I have described as ‘an interlocking network of fuzzy-edged clots of languages, emitting waves of mutual influence from their various nuclear ganglia. A mess, in other words.’¹⁹ While subgrouping such a recalcitrant family is difficult, there is certainly no need to go so far as van Driem by denying that TB subgroups exist at all, or by claiming that even if they do exist, there are so many of them that there is no point in talking about them!²⁰

In the published version of ? (1972), P.K. Benedict wisely refrained from offering a pseudo-precise family-tree model of the higher-order taxonomic relationships in TB, presenting instead a schematic chart where Kachin (= Jingpho) was conceived as the center of geographical and linguistic diversity in the family. See Figure 2.

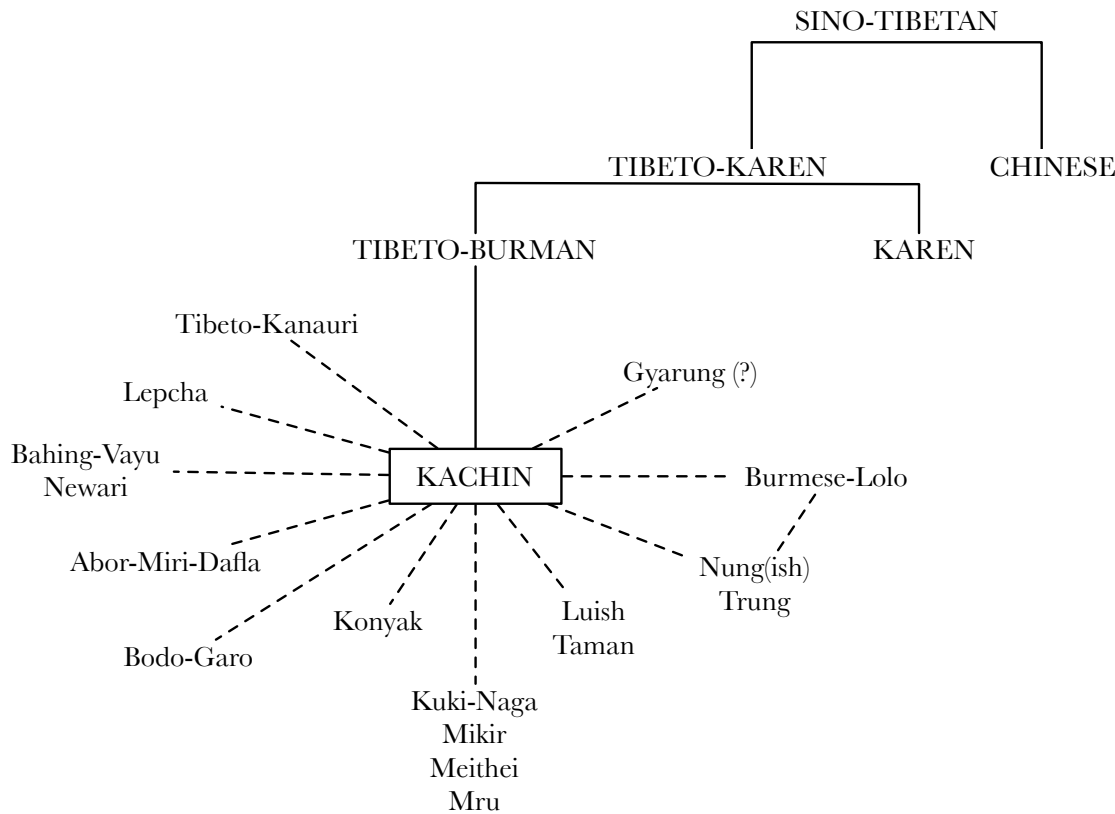


Figure 2: Schematic chart of Sino-Tibetan languages²¹

A simpler scheme was used heuristically at STEDT until very recently (see ?;5 and ?;xxix), as shown in Figure 3. It differed from ? in several respects:

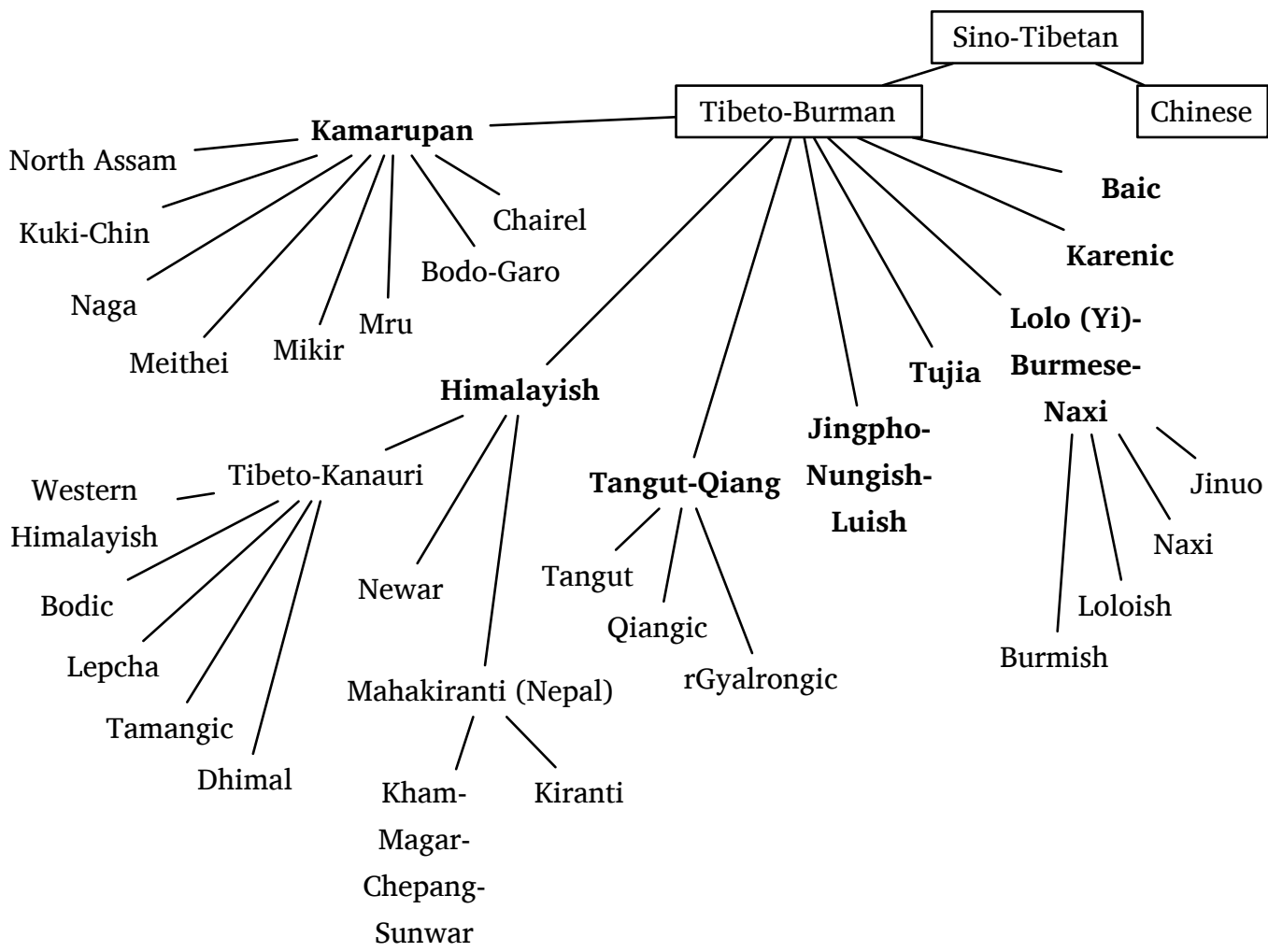


Figure 3: Previous STEDT family tree of ST languages

- Karenic is no longer regarded as having a special status, but is now considered to be a subgroup of TB proper.
- Baic, hardly mentioned (under the name ‘Minchia’) in ?, but later hypothesized by Benedict to belong with Chinese in the ‘Sinitic’ branch of Sino-Tibetan, is now also treated as just another subgroup of TB, though one under particularly heavy Chinese contact influence. Both Karenic and Baic have SVO word order, unlike the rest of the TB family.
- The highly ramified Kuki-Chin and Naga groups were provisionally amalgamated with Bodo-Garo (= Barish) and Abor-Miri-Dafla (= Mirish) into a supergroup called by the purely geographical name of *Kamarupan*, from the old Sanskrit name for Assam.
- The important Tangut-Qiang languages (deemed to include rGyalrong [= Gyarung = Jiarong] and the extinct Xixia [= Tangut]) were hardly known to Western scholars at the time ? was written (ca. 1942–3) or published (1972). It seems doubtful that a special relationship exists between Qiangic and Jingpho, or between Qiangic and Lolo-Burmese, as some Chinese scholars maintain.²²
- The Nungish and Luish languages were grouped with Jingpho (= Kachin). Jingpho was also recognized to have a special contact relationship with the Northern Naga (= Konyak) group.²³
- The somewhat idiosyncratic Mikir (Karbi), Meithei (= Manipuri), and Mru languages were included under *Kamarupan*.
- The Himalayish (= Himalayan) group is considered to include Bodic (i.e. Tibetanoid) languages, as well as Kanauri-Manchad, Tamang-Gurung-Thakali, Kiranti (= Rai), Lepcha, and Newar. It remains to be seen whether this is anything more than a geographic grouping.
- The relatively well-studied Lolo-Burmese (LB) group (= ?’s ‘Burmese-Lolo’) is deemed to include the somewhat aberrant Jinuo language of Xishuangbanna, Yunnan.²⁴ The Naxi/Moso language is quite close to LB, but stands somewhat outside the core of the family.²⁵
- The rather mysterious Tujia language of Hunan and Hubei (not mentioned in ?) has so far not been assigned to a subgroup.

¹⁷The approximate distribution of TB languages by primary country is as follows: India 133, China 133, Nepal 65, Burma 58, Bhutan 24, Bangladesh 18, Thailand 4, Laos 4, Vietnam 3 (?).

¹⁸Among the most valuable of these new sources are Sun Hongkai, Xu Jufang *et al.* (?), containing 1004 synonym sets in 52 languages and dialects; and Dai Qingxia and Huang Bufan (?), with 1822 synonym sets in 50 languages and dialects.

¹⁹?;2.

²⁰See his review (?) of ?, *The Sino-Tibetan Languages*.

²¹Reproduced from ?, p. 6; ?, p. 3; ?, p. 4.

²²A supergroup called ‘Rung’ was proposed by ?, into which he placed, among others, some Qiangic languages, Nungish, and Lepcha. This grouping was based partly on shared ‘proto-morphosyntax’, and partly on nomenclature, including the *-rong* of *rGyalrong*, the Nungish language *Rawang*, and the Lepcha autonym *Rong*.

²³The *Linguistic Survey of India* (?) recognized a ‘Bodo-Naga-Kachin’ group, an idea revived by ?, whose ‘Sal’ supergroup comprises Bodo-Garo (Barish), Northern Naga (Konyak), and Jingpho (= Kachin). Burling’s name for this grouping is derived from the etymon *sal ‘sun’ (ult. < PTB *tsyar ‘sunshine’), one of a number of roots which is attested chiefly in these languages. See ?;393–4 and below.

²⁴Chinese scholars have further divided the Loloish languages of China into six nuclei, although no attempt is made in this volume to distinguish them. In ? I examined Loloish tonal developments and the fate of the PLB rhyme *-a in terms of this six-way grouping, with inconclusive results.

²⁵I have grouped Naxi with Lolo-Burmese proper in a supergroup called ‘Burmo-Naxi-Lolo’ ?. On the basis of some shared tonal developments, I have also entertained the idea of a special relationship between Lolo-Burmese and Jingpho, to which I assigned the jocular designation *Jiburish* (< Ji-(ngpho) + -bur(mish) + (Lolo)ish). See ?.

Most recent subgrouping schema

My most up-to-date reconsideration of TB subgrouping has been motivated by very recent research on the Luish languages (???), as well as on Jingpho dialects (?). It now seems clear that Jingpho’s closest relatives are the formerly obscure languages of the Luish²⁶ group, with the Northern Naga group also fairly close, and the Bodo-Garo group more distantly related. On the other hand, Nungish (once believed to be especially close to Jingpho) now seems to constitute an independent subgroup, so that the many lexical similarities it shows with Jingpho are due to contact rather than to any particularly close genetic relationship.²⁷ These matters were explored in detail in ?, which offers (p. 39) a “Jingpho-Asakian Stammbaum”, here reproduced as Figure 4.

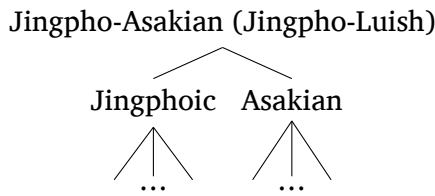


Figure 4: Jingpho-Asakian Stammbaum

Extending this position of the family tree to include Northern Naga and Bodo-Garo, we may re-interpret Burling’s *Sal* grouping, as displayed in Figure 5 (cf. ?;40).

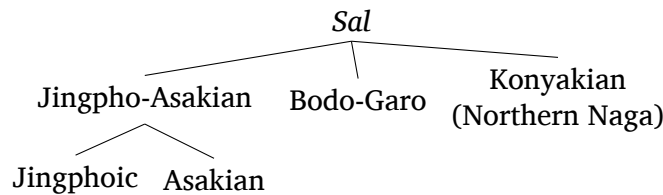


Figure 5: Elaboration of the *Sal* hypothesis

A further change in the previous STEDT schema is my recent abandonment of the cover-term “Kamarupan” for sociolinguistic reasons.²⁸ For now we are substituting “NE Indian Areal group”.

We may now present the latest version of our subgrouping scheme, as shown in Figure 6.

Finally, we are left with the major question of where Chinese fits into the picture. The standard view has always been that the primary split in the ST family is between Chinese (= Sinitic) on the one hand, and TB on the other. This is now being questioned by certain scholars who claim that there are no clear innovations shared by TB as a whole with respect to Chinese, or that there is an especially close relationship between Chinese and a particular branch of TB.²⁹ If we adopt such a point of view, that would make Sinitic just another branch of a superfamily we might call “Sino-Tibeto-Burman” (STB). It seems prudent, however, to treat this as an open question for the time being.

Teleo- and meso-reconstruction

While in the abstract it might seem super-virtuous to refrain from venturing to reconstruct at the PTB (or *a fortiori*, the PST) level until satisfactory reconstructions are available at every subgroup level, in fact this is neither possible nor necessary. The data are still uneven in the various branches of the family, ranging from the overwhelmingly copious to the tantalizingly sparse, from the phonologically accurate to the poorly transcribed. As mentioned above, as far back as the early 1940’s³⁰ Benedict made brilliant use of “teleoreconstructive” methods by relying mainly on 5 criterial, phonologically conservative languages: Written Tibetan, Written Burmese, Jingpho, Lushai (Mizo), and Bodo. Teleoreconstruction sometimes involves peeking outside a subgroup in order to come up with a meso-reconstruction; i.e. making educated guesses.³¹ Thus final stops may sometimes be reconstructed for Proto-Karenic by looking elsewhere in TB. W.T. ? used similar techniques in his reconstruction of some aspects of his Proto-Northern Naga.³² As in all scientific inquiry, the main thing is to form precise and correctable hypotheses, i.e. hypotheses that can be tested.

The mesolanguages on which we have relied include the following:

- Proto-Bodo (?)
- Proto-Ersuic (?)
- Proto-Karenic (???????)
- Proto-Kiranti (?)
- Proto-Kuki-Chin (?)
- Proto-Northern Chin (?)
- Proto-Lolo-Burmese (?????????)
- Proto-Tamangic (??)
- Proto-Tangkulic (?)
- Proto-Tani (?)

Nevertheless large gaps remain in our knowledge. We have nothing approaching well-worked out reconstructions for such key subgroups as Qiangic, rGyalrongic, Baic, and Nungish. Still unclear is the exact genetic position of many transitional languages like Chepeng, Kham, Lepcha, Newar (all lumped currently with “Himalayish”), or Meithei, Karbi, Mru (close to the Kuki-Chin branch), or Naxi/Moso (close to Lolo-Burmese), or the mysterious Tujia of Hunan/Hubei. As mentioned above, the position of the crucially important Jingpho language has been undergoing reevaluation, with current opinion returning to the notion of a special relationship with Northern Naga and Bodo-Garo, and especially with the hitherto obscure Luish (= Asakian) group of Manipur, Arakan, and Bangladesh. It remains to be seen whether the large “Kamarupan” (languages of the NE Indian area) and the “Himalayish” groups are anything more than purely geographic divisions of the family, and if so what the internal relationships among their many parts might be.

Whenever possible (i.e. when the data are sufficient and the sound-laws are well enough known) we can reconstruct a given etymon at several subgroup levels. A few examples from our database:

²⁶This group needs to be renamed (I suggest Asakian), since “Luish” is a pejorative exonym. See below, “The problem of nomenclature”.

²⁷If anything, Nungish now seems to be closer to LB than to Jingpho.

²⁸See below, “The problem of nomenclature”.

²⁹Cf. van Driem’s “Sino-Bodic” grouping (?), and my critique of it (?).

³⁰I.e., the manuscript version of ?, finally published in 1972.

³¹This method must of course be applied with due caution, and I feel that Benedict applied it too loosely with respect to the vexed question of the existence of a reconstructible tonal system at the PTB level. See e.g. ? (“Tibeto-Burman tones, with a note on teleo-reconstruction”). Neither ? nor the STEDT database makes any attempt to reconstruct tones at the PTB level, although tones can indeed be reliably reconstructed for certain subgroups (e.g., Lolo-Burmese, Tamangic, Karenic, Kuki-Chin).

³²I would in fact recommend such a teleological approach to Sinologists as well, since TB evidence will be crucial in evaluating the competing reconstructions of Old Chinese that have been proposed in the past few decades.

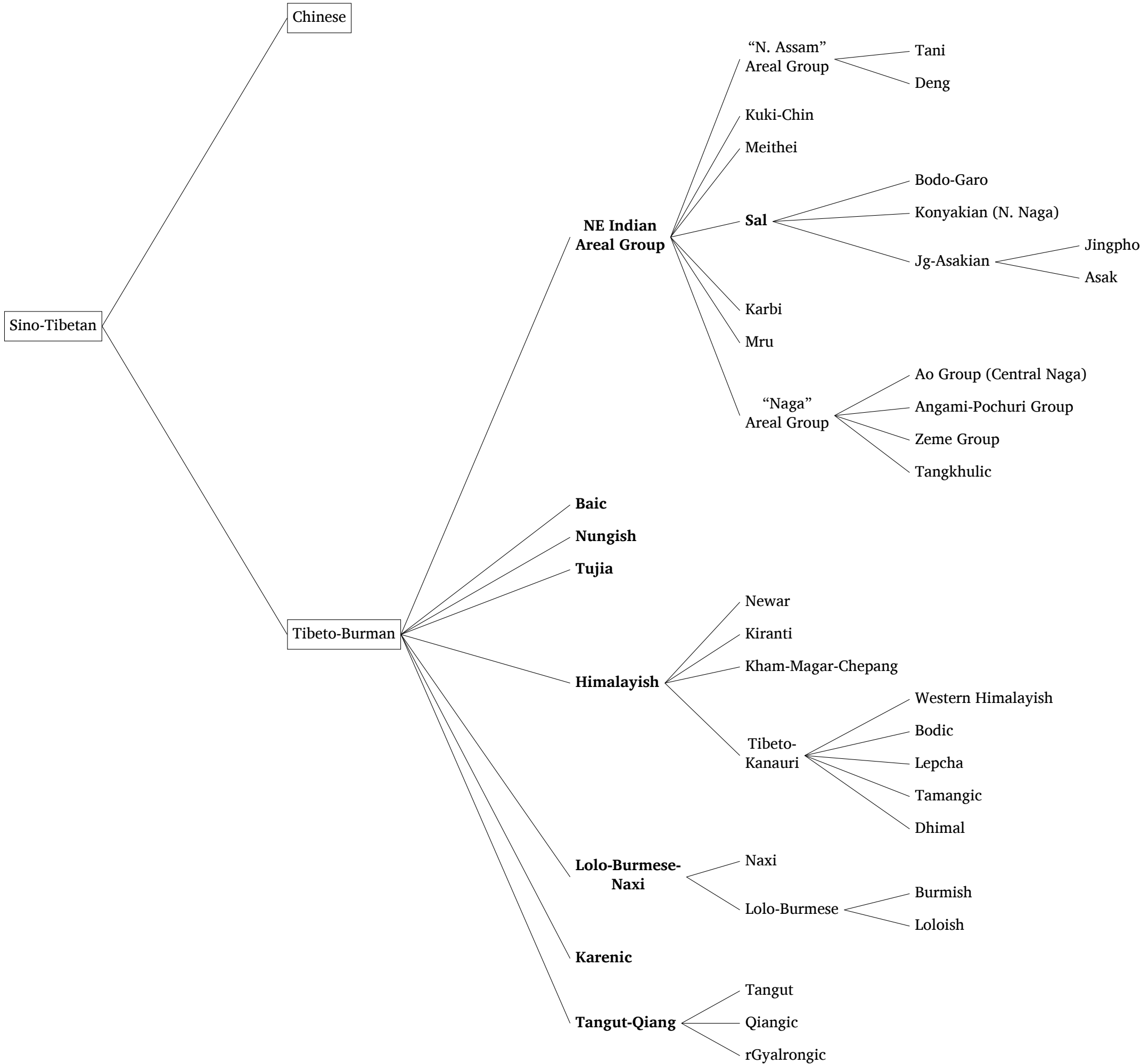


Figure 6: Updated subgrouping model of ST/TB relationships

| | | | |
|-----------|---------------|--|-------------------------------------|
| PIG | | | |
| #1006: | PTB | *p ^w ak | |
| | PKC | *wok | PKC = Proto-Kuki-Chin |
| | PNN | *wak | PNN = Proto-Northern Naga |
| | PTk | *hwok | PTk = Proto-Tangkhulic |
| | TGTM | ^b ɖwa | TGTM = Tamang/Gurung/Thakali/Manang |
| | PLB | *wak ^L | PLB = Proto-Lolo-Burmese |
| | PKar | *thoʔ ^D | PKar = Proto-Karenic |
| | PBai | *te | PBai = Proto-Baic |
| FOUR | | | |
| #2409: | PTB | *b-ləy | |
| | Proto-Tani | *pri | |
| | PKC | *lii | |
| | PNN | *bə lay | |
| | PCN | *phV-ləj | PCN = Proto-Central Naga |
| | PTk | *ti | |
| | TGTM | *bli | |
| | PLB | *b/ʔ-ləy ² | |
| | PKar | *lwi-t | |
| STONE | | | |
| #1269: | PTB | *r-luŋ ≈ *k-luk | |
| | PTani | *luŋ | |
| | PKC | *luŋ | |
| | PCN | *luŋ | |
| | PTk | *luŋ | |
| | PLB | *k-lok ≈ *k-loŋ | |
| | PKar | *loŋ ^B | |
| TREE/WOOD | | | |
| #2658: | PTB | *siŋ ≈ *sik | |
| | PTani | *suŋ | |
| | PKC | *thiŋ | |
| | PCN | *siŋ | |
| | PTk | *tʰiŋ | |
| EIGHT | | | |
| #2259: | PTB | *b-r-gyat ≈ *(b-)g-ryat | |
| | PKC | *riat | |
| | PTk | *ʃet | |
| | PLB | *ʔ-rit ^L | |
| | PKar | *khrət ^D /grət ^D | |
| ASHES | | | |
| #3514: | PTB | *hot | |
| | PKC | *wut | |
| | PTk | *hwot | |
| | Proto-Asakian | *k-but | |

When a meso-reconstruction can clearly be subsumed under a PTB etymology, we include it (along with its entire suite of supporting forms) under the PTB etymon in question. While it is always gratifying to be able to do this, in may cases it is not yet possible, so that we include the meso-reconstruction as a stand-alone item. These roots of limited distribution are actually highly important for a finer subgrouping of the TB family, since they might eventually permit the establishment of “isogloss bundles” distinguishing one subgroup from another.

The problem of nomenclature

Tibeto-Burman languages are notorious for the multiplicity of names by which they are referred to. These may include the name they use for themselves (autonym), as opposed to the name(s) other groups use for them (exonyms). Languages are frequently referred to by the principal town in which they are spoken (loconyms). Some exonyms are now felt to be pejorative, and have been abandoned, thus acquiring the status of ‘paleonyms’ for which ‘neonyms’ have been substituted.³³ A certain Angamoid Naga group call themselves and their language *Memi* (autonym), and their chief village they call *Sopvoma*; but other groups use *Mao* for this village or its people (exonym), and either *Mao* or *Sopvoma* (exonymic loconym) for their language. There is an older term *Imemai* (probably an autonymic paleonym) which refers to the same

language and people.

Some names are used in both a broader and a narrower sense, both for a specific language and for a group of languages that share a close contact relationship. The Maru, Atsi, and Lashi³⁴ (who speak Burmish languages) consider themselves to be ‘Kachin’ in the broad sense, and in this the Jingpho themselves seem to agree, even though the Jingpho language belongs to a different TB subgroup.

In recent years cultural sensitivities have forced the abandonment of many language names that had been well established in the academic literature. The important Central Chin language that used to be called *Lushai* (a name which is said to mean “long-headed”) should now properly be called *Mizo*. A Karenic group that used to be known by the Burmese exonym *Taungthu* (lit. “mountain folk”) now prefers their autonym *Pa-o*. The Southern Loloish people formerly known by the Tai exonym *Phunoi* (lit. “little people”) should now be called by their autonym *Coong*. The language group that has just been shown to be particularly close to Jingpho has traditionally been called Luish, although this has been shown now to be a pejorative exonym based on the Meithei word for “slave”. It is now recommended to change this group’s name to *Asakian* (?). Speakers of several TB languages of Nepal now object to the Indianized versions of their names with the Indo-Aryan -i suffix (e.g. *Newari*, *Magari*, *Sunwari*), and prefer to omit the suffix, even though this can lead to ambiguity between the names of the people and their languages (*Newar*, *Magar*, *Sunwar*). The autonomy of the most widely spoken TB language of Manipur is Meithei, but for political reasons the preferred name is Manipuri (with the IA suffix -i!). There is an Eastern Tani language of Assam usually call Mising, but many leading members of that speech community fear the mockery and mispronunciation of English speakers, who might consider their “Mising” language to be somehow *missing* something—so they now prefer the palatalized variant *Mishing*. The psychological dimensions of these issues are often as fascinating as they are paradoxical. Chinese linguists now feel that the term *Lolo(ish)*, widely used outside of China, is offensive, and insist that the proper respectful term is *Yi*, written with the character 彝 ‘type of sacrificial wine vessel’. Yet this is only a recent substitution for the homophonous character 夷 ‘barbarian; savage group on the fringes of the Chinese empire’.

The nomenclatural confusion is particularly acute at the subgroup level, where multiple names for the same TB subgroup have proliferated. Some of this nomenclatural variation goes back to differences between Benedict and his former collaborator and supervisor Robert Shafer, e.g. Shafer’s *Barish* and *Mirish* are the same as Benedict’s *Bodo-Garo* and *Abor-Miri-Dafla*, respectively. An important group of at least a dozen TB languages spoken in East Nepal is known either as *Kiranti* or *Rai*.³⁵ An extreme example of proliferation is furnished by the well-established and non-controversial group I call Lolo-Burmese, which has also been referred to as Burmese-Lolo, Yi-Burmese, Burmese-Yi, Burmese-Yipho, Yipho-Burmese, Yi-Myanmar, Myanmar-Yipho—and even Myanmar-Ngwi!

As mentioned above, I had used the name *Kamarupan* as a cover-term for the dozens of TB languages of NE India and adjacent areas (see Figure 3), but several scholars have objected to this largely on the grounds that it has distinctly Indo-Aryan connotations, which might irritate TB groups. See, e.g. R. Burling, ‘On *Kamarupan*’ (?), and the reply by Matisoff, ‘In defense of *Kamarupan*’ (?). I am now certainly willing to abandon this term, although I cannot think of any alternative more concise terms than “TB languages of NE India and adjacent areas”. For now, we have chosen the acronym “NEIA” for these languages, and this term appears in body of this work.³⁶

Special Challenges for ST/Eymtologization

The relative lack of anciently attested languages—only Chinese, Tibetan, Tangut (Xixia), Burmese, Meithei, Newar, and a few scantily-recorded others—is largely compensated for by the sheer number of modern TB languages (250–300). Special difficulties are presented by the monosyllabic nature of the ST/TB morpheme, a problem shared by other mono- or sesqui-syllabic language families of mainland Southeast Asia (Hmong-Mien, Tai-Kadai, Mon-Khmer).³⁷ Furthermore, in many TB subgroups these monosyllables are severely eroded phonologically, often losing all the syllable-final consonants.

There are in fact great difference amount the TB subgroups with respect to their presentation of proto-entities. WT is especially valuable for its presentation of prefixes; both WT and Kuki-Chin faithfully preserve nasal, stopped, and liquid finals. On the other hand, groups like Loloish and Asakian have retained only indirect traces of earlier prefixes, and have largely or completely undergone erosion of their post-vocalic consonants. rGyalrongic has developed syllable-initial consonant combinations which often surpass those of WT in complexity, but many of the will undoubtedly prove to be secondary. It is therefore a boon to find languages which preserve and erode differently: Hmong-Mien specialists are fortunate that Hmong preserves *initials better, while Mien preserves the pHM *finals better. Recent advances in research in

³³The terminology for the various types of TB language names was developed in ?: ‘The languages and dialects of Tibeto-Burman: an alphabetic/genetic listing, with some prefatory remarks on ethnonymic and glossonymic complications.’ In John McCoy and Timothy Light, eds., *Contributions to Sino-Tibetan Studies*, pp. 1–75. This article was later ? expanded into a STEDT Monograph, with the assistance of J.B. Lowe and S.P. Baron.

³⁴Referred to as Langsu, Zaiwa, and Leqi in Chinese sources.
³⁵Neither of these names is considered to be pejorative. According to K.P. Malla (p.c. 2008), “*Kirāt* is a loose label in Old Indo-Aryan for the cave-dweller, attested in late Vedic texts as well as in the *Mahābhārata*.” Rai is “a Nepali word, linked to IA *raaya* ‘lord’, given to the Khambu chiefs by the Gorkhali rulers in the late 18th century.”
³⁶This would be similar the the “SHWING” group of Austronesian, i.e. the “South Halmahera West New Guinea Group”.
³⁷As opposed, e.g., to Austronesian, with its juicy disyllabic roots. For the term *sesquisyllabic*, see below paragraph 4.

Proto-Asakian³⁸ have been facilitated by the fact that Kadu preserves *finals better, while Sak preserves *initials better. Such differential preservation often makes for extreme differences in surface form between perfect cognates, e.g. WT **brgyad** / Lahu **hí** ‘eight’.

Phonological erosion makes it tough to distinguish between phonologically similar etyma in the same semantic field:

LOUSE ***s(y)ar** and ***s-r(y)ik**. What do you do with a form like **ḡeʔ**?

MIND/BREATH ***s(y)am** and ***sak**

MOUTH ***mu:r** and ***muk**

FACE ***s-may** and ***s-myal**

As an extreme example, we have reconstructed no fewer than 11 etyma meaning NECK/THROAT with initial velars or velar clusters:³⁹ #389 ***s-gwa-n** ≈ ***r-gwa-n** NECK/NAPE (probably related to #495 ***kwa** THROAT/NECK); #481 ***s-key-k** ≈ ***m-key-k** NECK; #486 ***l-kwak** THROAT/NECK; #488 ***k-raw** ≈ ***k-law** THROAT/PALATE; #489 ***k/s-rawk/ŋ** ≈ **k-rwak/ŋ** THROAT; #491 ***grey-k** THROAT; #493 ***kak** THROAT; #494 ***ka** THROAT/NECK; #3361 ***ḡwaŋ** NECK/THROAT; #5651 ***ku** NECK/THROAT.

Some of these putative etyma may ultimately be combinable (e.g. #389 and #495, #481 and #491).⁴⁰ It is necessary to constantly rectify one’s reconstructions in the light of new data (see below, “Etymological accuracy and rectification of possible errors”), and equally imperative to try to establish ‘sound laws’ by finding parallel examples (see below, “Parallel but surprising examples”).

Phonological System of Proto-Tibeto-Burman

Most reconstructions conform to the PTB syllable canon posited in ʔ:11–13, shown in Figure 7:

(P₂) (P₁) C_i (G) V (:)
(C_f) (s)

Figure 7: Original PTB syllable canon

The initial consonant C_i could be preceded by up to two prefixes, with the inner prefix (P¹) usually assumed to be historically prior to the outer one (P²).⁴¹ Since one must also assume that some sort of schwa-like vowel must have intervened between the prefix and the root-initial (especially in doubly-prefixed forms like WT **brgyad** ‘eight’), this canon actually crosses the boundary between monosyllable and sesquisyllable.⁴² While it is sometimes possible assign a clear meaning to a prefixal element⁴³ This is not generally the case.⁴⁴ Should it be objected that the term “prefix” should be reserved for fully meaningful elements, we could use the neutral term “formative” instead.) In fact, the semantic and morphophonemic behavior of the prefixes is one of the most interesting aspects of TB morphology.⁴⁵ New prefixes are being created all the time, via the reduction (a.k.a. prefixization) of destressed initial constituents in compounds (e.g. PLB ***bəw²-rwak** ‘ant’ with the first syllable meaning ‘bug’ > WB **pərwak**). Morphophonemically the prefixes can interact with the root-initial in a variety of ways, including what I have called “prefixal preemption”, when a prefix drives out a weak initial like a nasal, liquid, or semivowel.

The C_i may be followed by one of four glides */-w-, -y-, -r-, -l-/, or even by certain sequences of two of them (especially -yw-).

The PTB vowel system in open syllables is displayed in Figure 8 (see ʔ:159–160). Low frequency open-syllable rhymes are in parentheses or square brackets:

| | | | | | |
|------|-------|---|------|-------|------|
| | (-i) | | | (-u) | |
| | -əy | | | -əw | |
| [-e] | | | | | [-o] |
| | -ey | | | -ow | |
| | (-ew) | | | (-oy) | |
| | -ay | | -aw | | |
| | -a:y | | -a:w | | |
| | | a | | | |

Figure 8: PTB open vowels

The only monophthong of high frequency is *-a. Although *-i and *-u (especially *-u) are reconstructible, in many languages they have merged with *-ey and *-ow, respectively. The core of the system is *-a plus a set of falling diphthongs.

Very recently I have banished the weakly attested mid-vowels (-e, -o) from PTB reconstructions altogether, both in open and closed syllables, substituting for them a set of rising and falling diphthongs */-wa(-), *-ya(-), *-aw(-), *-ay(-)/.⁴⁶ These formerly reconstructed mid vowels appear in square brackets in Figure 8.

The reinterpretation of some of the formerly reconstructed mid vowels as falling diphthongs necessitates a slight revision in the syllable canon, to permit sequences like */-wk-, -yk-, -wŋ-, -yŋ/. We may now revise the PTB syllable canon as shown in Figure 9:

(P₂) (P₁) C_i (G₁) (G₂) V (:)
(w/y) (C_f) (s)

Figure 9: Revised PTB syllable canon

Vowel length is reconstructed for PTB, with most of the evidence coming from the Kuki-Chin branch. This is an area where variation is rife, and a length contrast is only shakily reconstructible for the family as a whole.⁴⁷

The set of final consonants includes nasals, stops, liquids, and -s: */-**m** -**n** -**ŋ** -**p** -**t** -**k** -**r** -**l** -**s**/. The rarest of these is *-s, reconstructible in only a few etyma. In both open and closed syllables, a suffixal *-s appears in languages like WT (e.g., WT **zas** ‘food’ < ***dzya** ‘eat’; **sbugs** ‘hole’ ≈ **sbug** ‘pierce’).

No attempt is made to reconstruct tones beyond the subgroup level, since it is far from proven that a single system of tonal contrasts can be set up for PTB.

It should be noted that many TB languages have much simpler syllable canons, e.g., Lahu (Central Loloish), where native syllables consist maximally of an initial consonant, a vowel, and a tone:

 T
(C_i) V

Organization of the Etymologies

The “tagging” process

Etymologization of the thousands of forms in the STEDT database is a laborious and time-intensive process. In principle we are responsible for etymologizing every syllable of every form in the database, by subsuming each syllable under a particular etymon or word-family. This is done by “tagging” the syllables with a numeral representing its provenience, whether it be a root already established in the prior literature or one newly reconstructed by STEDT. The computer can then assemble all similarly tagged forms to create a cognate set. Under favorable circumstances every syllable of a compound can indeed receive a proper tag, e.g. Lahu **ḡû-tu-ṣī** ‘navel’ can be shown to be formed from three well-established roots **ḡû** < #137 ***p’am** ‘belly, stomach, waist’, **tu** < #520 ***ḍu** ‘navel, umbilical cord’, and **ṣī** < #1019 ***sey** ‘fruit, rose, round object’. But in reality, of course, a great many of our millions of syllables remain to be tagged, and many will never be etymologized. Such is the case with the five-syllable Lahu word **á-lā-mi-ṣí-jə** ‘rainbow’, none of which can yet be assigned to an etymon.⁴⁸

Besides our numerical tags a number of alphabetical symbols are also used: *p* for ‘prefix’, *s* for ‘suffix’, *b* for ‘borrowing’,⁴⁹ and *m* for ‘morpheme’. The latter tag is assigned to an etymologically obscure constituent of a compound the other members of which have been assigned an etymology.

³⁸e.g. ʔ, ʔ, ʔ.

³⁹This is perhaps further confirmation of the frequently noted tendency of words for these parts of the body to have ‘guttural’ initials.

⁴⁰This classic problem of distinguishing co-allofams from reflexes of separate roots has bedeviled Tibeto-Burmanists from the beginning. Benedict (ʔ) hesitated before assigning ***g-yak** and ***s-rak** ASHAMED to two separate roots (#452 and #431); similarly with ***m-da** and ***b-la** ARROW (pp. 111–112 and #449). I now believe that two separate roots I set up in *GSTC* (ʔ) actually reflect the same etymon, so I would now combine CATTLE/DOMESTIC ANIMAL ***dzay** (#129) with ELEPHANT/CATTLE ***tsay** (#143). In 1988 I tried to show that two homophonous roots set up in Benedict (ʔ) (***dyam** STRAIGHT and ***dyam** FULL/FILL) are actually one and the same root (ʔ). For many similar examples, see ʔ.

⁴¹To take a synchronic example, the etymon ***m-lyak** ‘lick’ must be reconstructed with a nasal prefix that is found in several branches of the family (e.g. Chungli Ao **mánák** ~ **məzək**, Jg. **mətáʔ**, Akha **myəʔ**). The reflex in Tangkhul Naga, however, is **khəməlek**, with a secondary velar prefix that now occurs before all TN verbs.

⁴²This term, introduced in ʔ, describes morphemes that are “one syllable and a half” in length.

⁴³The pioneer in attempting to do so at the PTB level was ʔ.

⁴⁴The stronghold of sesquisyllabicity in Southeast Asia is the Mon-Khmer family, but specialists in that family have despaired of attaching any particular meanings to the “minor syllables” with schwa-like vocalism.

⁴⁵See the extended discussion in ʔ:87–156.

⁴⁶See ʔ. Notations like *-wa(-) mean “*wa in both open and closed syllables”.

⁴⁷See ʔ:237–45.

⁴⁸Words for RAINBOW tend to be multisyllabic and unanalyzable in many TB languages, undoubtedly because of the mythic significance of rainbows. Not a single root for RAINBOW can be reconstructed at the PTB level.

⁴⁹For more details about the treatment of borrowings, see below, “Treatment of loanwords”.

Supporting forms and their attribution

Each record in our database is ascribed to a particular source. Many of these forms, especially from well-documented languages like Written Tibetan, Written Burmese, or Jingpho, are cited from more than one source. Even when the form is cited identically in multiple sources, we include each citation.⁵⁰ When a given form is cited somewhat differently in the sources, our practice of “following copy” is especially useful, particularly when one or more of the sources might not be totally accurate phonemically, or where subphonemic phonetic detail is provided in one source but not another.⁵¹ Great care has also been taken to attribute etymologies to their original source, if they were not newly created at STEDT.

The STEDT database contains forms from sources of many different kinds, including:

- printed books, monographs, articles, especially dictionaries and grammars of individual languages
- synonym lists (i.e. groups of forms from different languages with the same meaning, but with no re-constructions provided), e.g. ʔ; ʔ (ZMYYC); ʔ (TBL)
- semantically based questionnaires solicited by STEDT from fieldworkers working on particular languages
- monographs and treatises of an etymological nature, including works which provide reconstructions at the subgroup level

The supporting forms for an etymology are cited subgroup by subgroup, in a fixed order. As explained above, meso-reconstructions are subsumed (along with their own supporting forms) under general PTB etymologies whenever possible.

For some key languages, special points need to be made:

Written Tibetan

In Jäschke 1881/1958 the letter known as “a-chung” is written with a subscript circle in initial position, e.g. **bu** ‘insect’. For this we have substituted the symbol **ḥ**, e.g. **ḥbu** ‘insect’.

Jäschke symbolizes aspiration by a backward apostrophe, for which we substitute **-h-**, e.g. **kha** ‘bitter’, instead of **k’a**.

Jäschke writes the palatal series of stops with the cumbersome symbols “č, č’, j”, for which we substitute “**c**, **ch**, **j**”.

Written Burmese

Already in the Rhyming Dictionary of Written Burmese (1941/1977), Benedict’s original “**ts**, **tsh**, **dz**” were changed to “**c**, **ch**, **j**”. The unaspirated member of this series can also occur syllable finally, so that, e.g. Benedict’s tshats ‘joint’ has been reinterpreted as **chac**.

The low back nasal rhyme is sometimes written as “**-oŋ**” (as in the RDWB) and sometimes as “**-auŋ**” (as in HPTB).

The three WB tones in open syllables have been described and symbolized in many different ways, but our preferred notation is: no tonemark for the modal tone (e.g. **la** ‘come’), a circumflex for the heavy tone (e.g. **lâ** ‘yes/no question particle’), and an apostrophe for the creaky tone (e.g. **la’** ‘moon’).

Jingpho

Tones are not indicated in the classic source, Hanson (1906/1954). Dai Qingxia et al. (1983) use the system devised by Chao Yuen-ren for Chinese dialects, i.e. a system of super- or sub-script numerals. Other authors (e.g. Matisoff 1974, 1991) use diacritical marks. Thus, ‘house’ may be written as **nta** (Hanson), **n⁵⁵ta**⁵¹ (Dai et al.), or **ñ-tâ** (Matisoff).

Chinese

Old (= “Archaic”) Chinese forms cited from Karlgren’s *Grammata Serica Recensa* (1957) are cited exactly as he wrote them. OC forms cited from Coblin (1986) are according to Li Fang-kuei, while those cited from Schuessler (2007) or Baxter/Sagart (2013) are in the latest versions of their own systems.

Glosses of the supporting forms

In almost all cases, the gloss given in each particular source is preserved, unless it is so awkward or misleading as to require emendation.⁵² Even if the glosses in consecutive records are identical, the gloss is repeated for each individual record, instead of using a symbol like the “ditto-mark”.

If a gloss is too long to fit onto a single line, it is ‘wrapped’ so that the additional lines are indented under the first one.

Many glosses are ambiguous due to the peculiarities of homophony or homography in the English lexicon, e.g. *bark*, *bear*, *boil*, *calf*, *drop*, *fly*, *quiver*, *right*, *scale*, *sow*, *tear*, *wind*, etc.⁵³ Some sources take pains to disambiguate such glosses by providing additional information, e.g. ‘bark (of dog)’ vs. ‘bark (of tree)’; ‘wind (n.)’ vs. ‘wind (v.)’, etc. Many other sources, however, are not so helpful, and it has been up to us to try to figure out which alternative is correct, on the basis of comparative evidence.

Notes

Notes may appear at any point in an etymology. The loosely-hierarchical taxonomy of notes used is as follows:

- “**chapter notes**”, which introduce new thesaurus (semantic) sections; note that this moniker is not completely accurate: chapter notes may introduce volumes and fascicles as well.
- “**etymon notes**”, which present points about an etymon as a whole. Details on their composition is given further below.
- “**subgroup notes**” pertain to a particular subgroup.
- “**source notes**”, copied from the original source document, appear as footnotes with record notes, start with the indicator *[Source Note]*.
- “**record notes**” are associated with a specific supporting form. They appear as alphabetcally numbered footnotes at the end of the tabular cognate set display (and the numbered restarts with each set).
- “**internal notes**” are reminders made by STEDTniks in the course of their work and do not appear in the printed work.

Notes at the level of a semantic category or etymon are shown “inline”, that is, as paragraphs in the running text. “internal notes”“source notes” “record notes”, and so on.) However, only source notes are shown in this work,

Many etyma have **Chinese comparanda** and these are rendered as a separate section following their associated etymon (see details below).

Etymon notes provide such information as:

- richness of attestation:** This ranges from a minimum of two perfectly cognate forms, each from a different branch of the family (e.g. #2625 ***d-rum** ‘pine for’), to etyma with support from dozens of languages (e.g. #230 ***s-hyway-t** ‘blood’).
- distribution of the etymon among the subgroups:** e.g., “***dum** ‘kiss/suck’ is firmly established in Tani, N. Naga, and Barish, with convincing but so far isolated cognates in Chepeng (Himalayish) and Nesu (Loloish).”
- distribution of a phonological feature:** e.g. “In ***m-prwaŋ** ‘mouth’, the ***m-** prefix is only directly attested in Qiangic, Tani, and S. Chin.”
- meso-roots** to be subsumed under the PTB reconstruction.
- allofams:** variant forms that follow well-established patterns (see below, “Notational conventions and pan-allofamic formulas” & following).
- irregularities:** apparently related forms that vary in a non-structured way, perhaps due to borrowing from a related language, or to a now lost morphological element. Such forms are included in some etymologies in hopes that the irregularities will be explained in the future.

⁵⁰Paper is not a problem electronically, but in printed publications (e.g. ʔ) the form only appears once, with multiple attributions separated by commas.

⁵¹Cf. the multiple transcriptions of the Written Burmese form for BREAST/MILK under “***s-naw**: **no**’ (ʔ:281.39); **nui**’ (ʔ:327; ʔ:419); **núi**’ (ʔ:48); **nui**’ (ʔ; ʔ; ʔ); **nui**. (ʔ); **nuiw**’ (ʔ). Records marked with the abbreviation “JAM-Ety” refer to my own etymological notes compiled in the pre-STEDT era, derived especially from older, classic sources. These specific sources can easily be tracked down from the *Bibliography*.

⁵²Our practice of “following copy” does not extend to obvious errors, e.g. misspellings in the gloss. It is always best to bear in mind Ralph Waldo Emerson’s apothegm “A foolish consistency is the hobgoblin of little minds...”

⁵³This problem is of course not unique to English. Cf. the French verbs *voler* ‘fly; steal’ and the corresponding nouns *le vol* ‘flight; theft’.

Chinese comparanda

After the evidence for a TB etymon is presented, one or more Chinese comparanda are often suggested in the interests of pushing the reconstruction further back to the PST stage.⁵⁴ However, no attempt is made to choose among the often contradictory reconstruction schemes for Old Chinese, and I usually have not tried to set up PST forms, as ? sporadically attempted to do.

In addition to these sources, OC reconstructions are also cited from ??, ? (?), and the new Baxter/Sagart system ?.⁵⁵

Many of the comparanda have been discussed by Zev Handel; his contributions are indication by the inclusion of his initials “ZJH” at the end of the section. Stephen Baron recently returned to the project and has reviewed and improved a number of the comparanda. His initials (“SPB”) appear along with items he has revised. A comparandum cited without attribution is the responsibility of JAM.

Treatment of loanwords

As mentioned above, obvious loanwords from non-TB languages are tagged as such in our database, with the symbol “b” for “borrowing”. When the source of the borrowing is clear, this information is provided, e.g. “bChinese”, “bIndic”, “bMon”.

If an etymon of Indic origin has been borrowed into a large number of TB languages, as is frequently the case in the Himalayan and Northeast Indian areas, we sometimes include the Indic root among our etymologies, marked as “IA” (Indo-Aryan). A couple of examples:

- IA ***ghoDā** HORSE > Apatani **go-ra**, Mizro **sa-kor**, Lotha Naga **ku¹ra¹**, Tibetan (Jirel) **gorā**, Khaling **ghora**, Bodo **goray**, Chang **kuri**, etc.
- IA ***battakh** DUCK > Chinbon **papak**, Mongsen Ao **phàtàk**, Chaudangsi **bə̀tək**, Phom **bat⁵⁵æk⁵⁵**, Jingpho **khai³³pjek⁵⁵**, Raji **bə̀tək**, etc.

It sometimes happens that there is an accidental phonological resemblance between a non-TB etymon and a genuine TB root with a similar meaning. In the rare cases where it is hard to decide between borrowing and chance resemblance, we tag the lexical item as “b?”.

The Representation of Semantic Associations

As opposed to the organization of ?, where the etymologies are presented according to their phonological shape, the etymologies in this *Etymological Dictionary and Thesaurus* are arranged by their semantic content.⁵⁶

Every gloss that appears anywhere in our database has been assigned to one or more semantic categories, according to a system we have developed known as SEMCAT.

Many of the points in semantic space in this Dictionary/Thesaurus, as well as each of the chapters of ?, are provided with semantic diagrams, called “metastatic flowcharts” in STEDT parlance. These diagrams were first introduced in ?, and have been used subsequently in many of my articles.⁵⁷

They are intended to represent the paths of semantic association undergone by etyma, as established by comparative/historical and/or internal synchronic evidence. An association between two points (X,Y) in semantic space may be established either synchronically or diachronically, either on the basis of a single language or comparatively. I rely on three basic types of evidence:⁵⁸

- Synchronic intra-lingual vagueness.* A given daughter language has a single form that means X or Y according to context, e.g. Karbi **ar̥tho** means either ‘blood vessel’ or ‘tendon’ or ‘muscle’ or ‘nerve’. In many Chin languages reflexes of ***m-luŋ** can mean either HEART or LIVER.
- Inter-lingual semantic shift of phonological cognates*, i.e. reflexes of the same etymon mean X in Lg. A but Y in Lg. B, e.g.:

PTB ***r-kliŋ** ‘marrow/brain’ > Karbi **arkleŋ** ‘marrow’, Dimasa **buthluŋ** ‘brain’;

PTB ***s-p^wik** ‘bowels/stomach’ > Karbi **phək** ‘bowels’, Lahu **ႤႬႬ-qō** ‘stomach’.
- Association via compounding.* Three points (A,B,C) in semantic space are related, such that in some language a compound of two morphemes, A + B, has the meaning C. In other words, an etymon appears as a constituent in compounds, such that part of the meaning of the compound derives from it, e.g.:

FOOT + EYE → ANKLE (Lahu **khi-m̥ḥ[?]-ṣi** < **khi** ‘foot’ + **m̥ḥ[?]-ṣi** ‘eye’); similarly Indonesian **mata-kaki** ‘ankle’ (< **mata** ‘eye’ + **kaki** ‘foot’), which establishes the association EYE ↔ ANKLE⁵⁹

Certain graphological conventions are observed in the metastatic flowcharts:

- Points in semantic space between which an association has been established are connected by solid lines. If a point is a bodypart, it is labeled in capital letters. An association between two points that are both bodyparts is an ‘intra-field association’, e.g.:



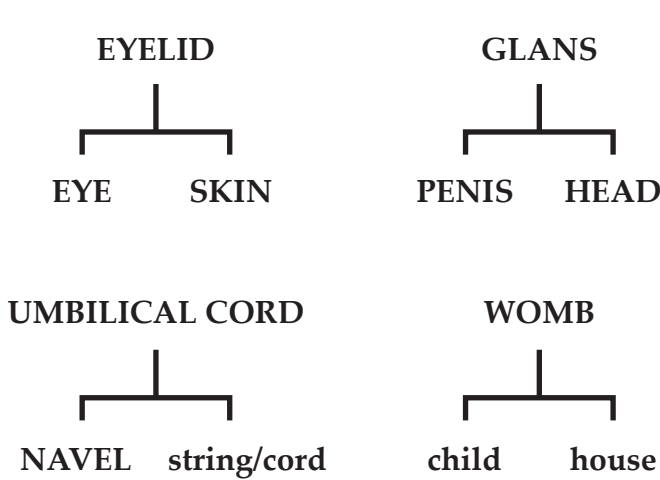
- If the association crosses into another semantic field (e.g., if it is between a bodypart and a non-bodypart), the non-bodypart point is labeled with small letters, e.g.:



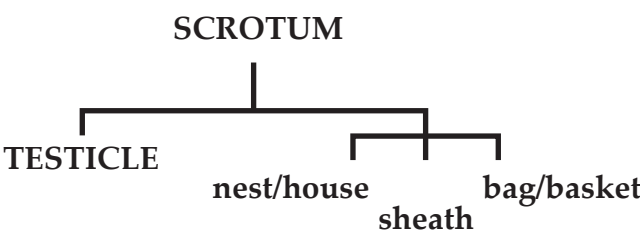
- Antonymic associations (cases where the etymon has acquired opposite meanings) are diagrammed by a curved *yin-yang* type of line, e.g.:



- Compounds are diagrammed by a pitchfork-like symbol, with the two constituents appearing at the points of the fork, and the overall meaning of the compound indicated at the tip of the handle, e.g.:



The same convention with respect to capital vs. small letters applies to compounds. In cases where several different combinations of morphemes are attested in compounds with the same meaning, graphic constraints sometimes require geometric reorientations of the pitchfork, e.g.



The category of “reproductive bodyparts” is construed broadly to include related verbs (e.g., KISS, SUCK, LOVE, SQUIRT).

More complex flowcharts have been constructed for the associations of SKIN, based on the analysis of etyma with that meaning. These are presented in Figures 10 and 11.⁶⁰

In all language families, semantic associations and variation are much less structured than phonological variation. Sometimes the same unexpected semantic association is found in widely different language families, e.g. BRAIN and MARROW, exemplified both in IE and TB.⁶¹ Other associations seem to be confined to a particular family, e.g. YEAST and SALT, so far discovered only in TB.⁶²

The vagaries of semantic history occasionally result in strange allofamic bedfellows, e.g. Eng. **black** and **blank**, from the same PIE root ***bhel-** SHINE/FLASH/BURN/SHINING WHITE AND VARIOUS BRIGHT COLORS, or Eng. **science** and **shit** < ***skei-** CUT/SPLIT.

Given this unpredictability, how much semantic latitude is prudent in our etymologies, and under what circumstances can semantic leaps be plausibly justified? On the wild side we have megalocomparative fantasies with implausible semantic associations offered which are totally unsupported by reasonable phonological correspondences or areal semantic tendencies, as in Jan Braun’s *Sumerian and Tibeto-Burman* (?), in which he claims genetic connections between pairs of words like the following:

| <i>Sumerian</i> | <i>Written Tibetan</i> |
|--------------------------------|---|
| gim ‘axe’ | ḡgem-pa ‘destroy’ |
| alad ‘protective demon’ | lad-mo ‘imitation’ |
| ṣinig ‘tamarisk’ | snyigs-pa ‘degenerated; grown worse’ |
| peṣ ‘three’ | gsum ‘three’ |

Deciding how much semantic latitude to allow among putative cognates is definitely an art rather than a science.⁶³ Here as elsewhere a middle-of-the-road approach is necessary, neither overly conservative nor too wildly speculative.

Regularity and variation

Every spoken language is rife with variation on both the phonological and semantic planes, some of it rule-governed and some of it sporadic. Thus, despite the qualms of certain extreme neogrammarians, we must assume that this is true for proto-languages as well. It must be admitted that a lot of guesswork is involved in etymologizing material from hundreds of languages and dialects at once, without having established the “sound laws” in advance. The problems are especially acute when comparing phonologically depleted languages with those having richer syllable canons. When there is a partial phonological similarity between distinct etyma with the same meaning, it is not easy to decide by simple inspection to which etymon we should assign a phonologically slight form in a daughter language.

There is a dialectical relationship between synchronic data and sound laws. The “laws” are derived by inference from the data in the first place, but once proto-forms are reconstructed, they can be used to guide us in our hunt for cognates in languages not yet examined (even if they have undergone semantic change). Almost every TB/ST etymology so far proposed presents problems and complications—irregularities—in some language or other, which is par for the course even in the much better known Indo-European family. Part of our task is to indicate where the exceptions, problems, and irregularities lie, in the hope that they can ultimately be explained. The concept of “regularity” itself is by no means simple, nor does it mean the same thing to different scholars.⁶⁴

Those who lack what I have called “Proto-Sprachgefühl”⁶⁵ can produce abstract, formulaic reconstructions bristling with strange symbols but devoid of any phonetic or typological plausibility.⁶⁶ Given sufficient semantic latitude and proto-forms that are complex enough, one can formulate “sound laws” in such a way that they appear completely regular and exceptionless. At an extreme level we find “megalocomparative” proposals of genetic relationship that turn received notions upside down (e.g. Sino-Mayan, Sino-Caucasian, Sino-Austronesian, Japanese-Dravidian), and which can lead the unwary down fruitless paths, obscuring the differences among cognates, borrowings, and chance resemblances.⁶⁷ Various tricks of analysis that I have lumped under the rubric of “proto-form stuffing” can help the Nostraticist or Sino-Mayanist convince himself that his fantastical comparisons are “perfectly regular”. Paradigmatically, one can multiply the number of proto-phonemes. If you reconstruct 35 proto-vowels, any anomalous vowel correspondence can be regarded as “regularly reflecting” a separate proto-vowel. Syntagmatically, if you reconstruct etyma like ***mrgsla**, and the monstrous proto-cluster ***mrgsl-** occurs only in a single etymon, any set of reflexes in the daughter languages can be said to be “regular”.⁶⁸

The time-depth of PST is perhaps 6000 years B.P., about at the limits of the comparative method. We can hardly afford to insist on “perfect regularity” of correspondence among our putative cognates. But instead of resorting to “proto-form stuffing” to try to explain away problems, what is needed is an explicit theory of variational phenomena. TB and ST etyma, like those of other language families, are not independent isolated entities, but stand in complex phonosemantic relationships with each other. It has been recognized for a long time that words in Chinese and TB languages participate in morphophonemic groups of partially resemblant forms that have been called “word families”.⁶⁹ In ? I developed the notion of the *allofam*, or individual member of a word-family, and advocated the formulation of “allofamic reconstructions” that accommodated all the well-attested variants deemed to descend from the same proto-word-family. The symbol ≋ was introduced to symbolize an allofamic relationship between variant forms, i.e., “A ≋ B” means that “A and B are synchronic allofams of each other”, while “*A ≋ *B” means that there is a word-family relationship between A and B at the proto-level.⁷⁰

Needless to say, extreme care must be used in claiming that different forms are variants of the same etymon. Allofamic theory must be applied in a controlled and constrained way.⁷¹ Not everything may be said to vary with everything else! It is sometimes quite difficult to decide whether partially resemblant forms represent separate etyma or whether they are merely allofams of the same word-family. Not only must each proto-allofam fit our canonic template (above, “Phonological System of PTB”), but the type of variation posited must be abundantly replicated in other examples. We do not attempt to conceal such uncertainties, but frequently entertain the possibility that etyma set up as independent might actually be co-allofams, or *vice versa*.

Notational conventions and pan-allofamic formulas

The best attested patterns of variation in ST/TB are all exemplified in the etymologies of Fascicle 1.9 (*The Reproductive System*, previously published as ?). They include the following:

- Voicing vs. voicelessness of the initial consonant:⁷²

| | | |
|-------------------------------|-------|----------------------|
| * gwap ≋ * kwap | (11a) | HATCH/INCUBATE/COVER |
| * prat ≋ * brat | (75) | BREAK/WEAN |
| * tuŋ ≋ * duŋ | (44a) | NAVEL |

⁶⁰As a desideratum for the future, one might envision three-dimensional semantic diagrams like those used to model molecules in organic chemistry!

⁶¹See ?, pp. 182–3, 203.

⁶²See ? (?:34) and ?. Yeast is traditionally used for brewing liquor rather than for baking bread in East and SE Asia. Both salt and yeast are efficacious substances that have dramatic effects on the taste of food or drink; their lack renders the food or drink insipid. Cf. Proto-Kuki-Naga ***m-tsyi** ‘salt’ and Jingpho **mətsi**, Lahu **di** ‘yeast’ (Lahu voiced initials come from *prenasalized ones).

⁶³As in many other respects, Benedict was also a pioneer in the field of SE Asian areal semantics. Cf. his brilliant article “Semantic differentiation in Indo-Chinese” (?).

⁶⁴See ? (“Following the marrow”) and ? (“Regularity and variation”).

⁶⁵See ?.

⁶⁶Recent examples of this genre include ?; ???; ?.

⁶⁷See ? (“On megalocomparison”). Megalocomparison has the apparent advantage of non-falsifiability, since, as Haudricourt has observed, one can never prove that any two languages are not related. But non-falsifiable hypotheses are not scientific. When presented with alternative non-falsifiable proposals it is impossible to choose among them.

⁶⁸This is actually the proto-form offered in ?:25 for an etymon meaning ‘spirit, ghost, shadow’ (reconstructed as ***m-hla** in ? #475). As I have observed (?;22), “It is always possible and sometimes necessary to invent an *ad hoc* explanation for an anomalous case. It is even true that some such *ad hoc* ‘solutions’ are more plausible than others. The only harm is in deluding oneself that an explanation which covers only a single case establishes a ‘regularity’.”

⁶⁹See the pioneering study of Karlgren (?), “Word families in Chinese”.

⁷⁰This symbol ≋, a combination of > ‘goes to’ and < ‘comes from’, is meant to suggest that neither variant is necessarily deemed to have temporal priority, but that both must be set up to account for attested forms. Other related symbols are A ≋ B “A is not an allofam of B,” and A ≋? B “is A an allofam of B?”

⁷¹See the extended discussion in Ch. XII of ? (pp. 491–534), “Allofamic variation in rhymes”.

⁷²Nothing is more common in TB word families than variation of voicing in initial consonants, largely due to the pervasive influence of prefixes on the manner of the initial. This is in sharp contrast to the situation in Indo-European or Austronesian, where such variation in manner is quite rare, and is usually not tolerated in reconstructions.

⁵⁴Most of these comparanda have already been included in ?.

⁵⁵See “A Concise Introduction to Old Chinese Phonology” by Zev Handel (?;543–574), which treats the major differences in the reconstructive systems of Karlgren (?), ??, and William ?.

⁵⁶This is also the organizational plan of ?, which deals entirely with a single semantic area, the Reproductive System. (This was originally supposed to constitute Fascicle Nine of Vol I (Body-parts). Most of the examples in this section are drawn from ?, since it is there that the semantic associations of the etyma have been investigated most thoroughly.)

⁵⁷See, e.g., ? (“Stars, moon, and spirits”), ? (“God and the ST copula”), ? (“Arm, hand, wing”), ? (“Property, livestock, talent”), ? (“Grammatization in Lahu”), ? (“Mother of all morphemes”), ? (“Buttock and heel”), ? (“Three TB word-families”), and ? (“Areal semantics”).

⁵⁸See the discussion in ?;194–200.

⁵⁹The same formation is found in many other TB languages, e.g.:

| | <i>FOOT</i> | <i>EYE</i> | <i>ANKLE</i> |
|-----------------|-----------------|----------------------------|---------------------------------|
| Lalung | ia-thong | mi | ia-thong-mi |
| Limbu | lāŋ | mik | lāŋ-mik |
| Lushai (Mizo) | ke | mit | ke-mit |
| Meithei | khu | mit | khu-mit |
| Tangkbul | phei | mik-ra | phei-mik-ra |
| Written Burmese | khre | myak-ci¹ | khre-myak-ci¹ |

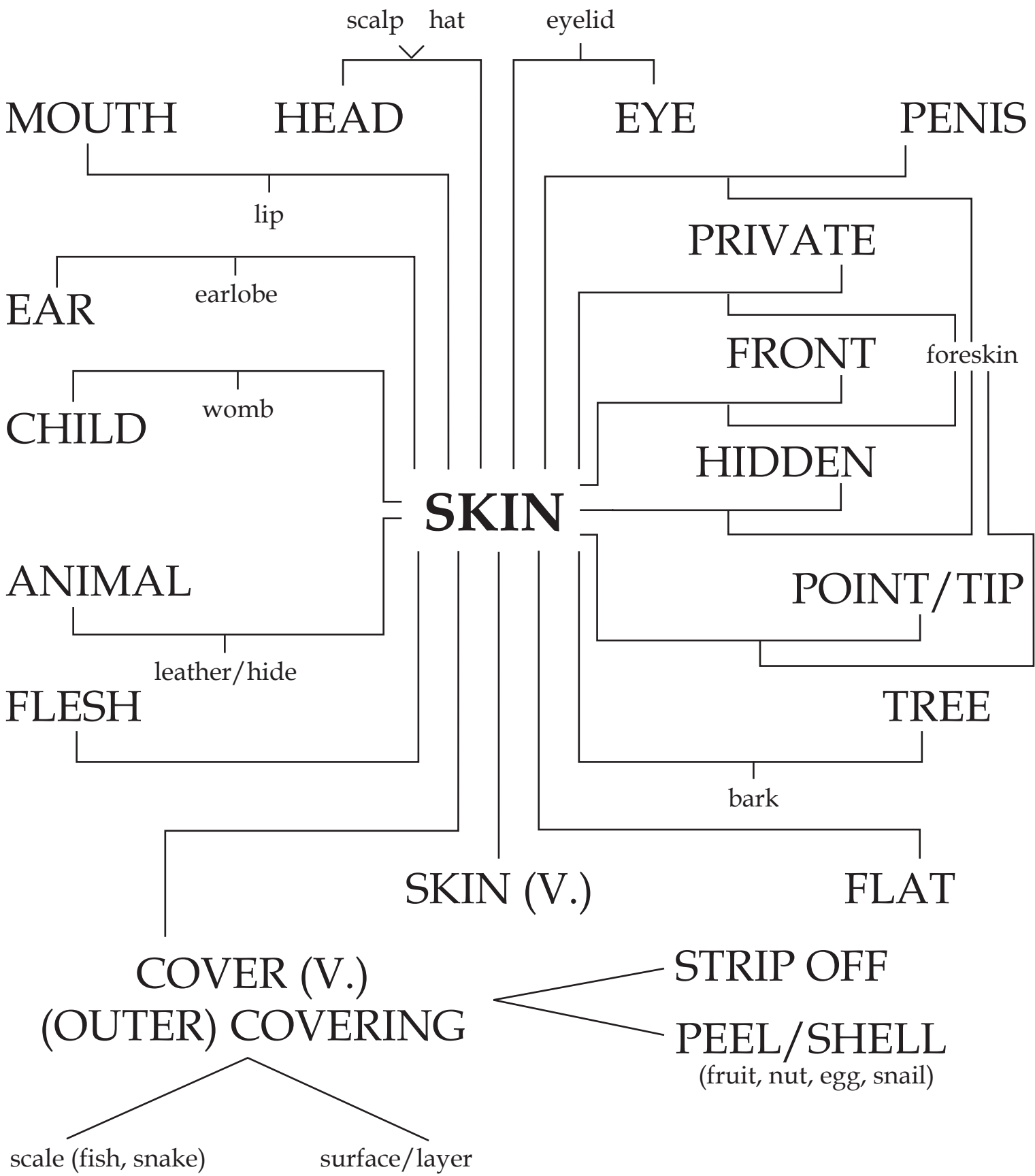


Figure 10: SKIN flowchart #1

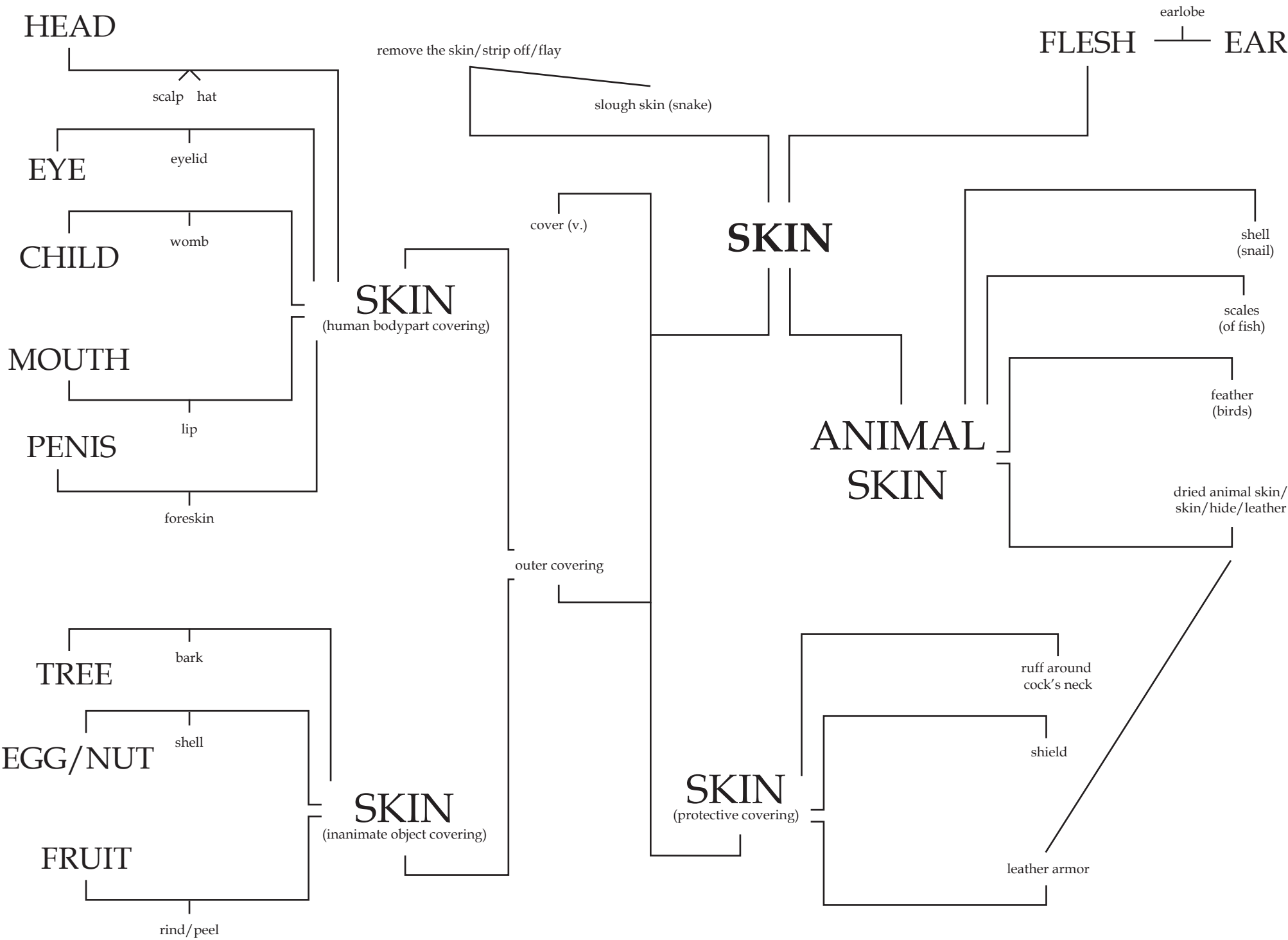


Figure 11: SKIN flowchart #2

2. Variation between fricative and affricate:

 - ***(t)sum** (45) NAVEL
 - ***(t)sip** × ***(t)sup** (107) NEST/WOMB/SCROTUM⁷³

3. Presence vs. absence of medial -y-:

 - ***b(y)at** (81) VAGINA
 - ***l(y)ap** (151) COPULATE

A special case of this pattern is the alternation between dental and palatal fricatives and affricates:

 - ***s(y)awk** (61) DRINK/SUCK/SMOKE
 - ***dz(y)əw** (56) BREAST/MILK
 - ***ts(y)u:ŋ** (44b) NAVEL

4. Variation between labial stop and labial semivowel:
- ***pu** × ***wu** (1a, 1b) EGG
 - ***pam** × ***wam** (98a, 98b) WOMB/PLACENTA/NEST

5. Variation between different prefixes:

 - ***r-ga** × ***N-ga** × ***d-ga** × ***s-ga** (141) COPULATE/LOVE/WANT
 - ***n-tow** × ***s-tow** (3) EGG
 - ***m-ŋal** × ***l-ŋal** (100) WOMB/PLACENTA

6. Variation between -u- and -i- in closed syllables:

 - ***dul** × ***dil** (2b) EGG/TESTICLE
 - ***m-dzup** × ***m-dzip** (55) SUCK/SUCKLE/MILK/KISS
 - ***tsyur** × ***tsyir** (66) MILK/SQUEEZE/WRING

7. Variation between medial -ya- and -i-:

⁷³This etymon also illustrates #6, below.

1. Inherited Germanic material
- (a) ***wod-ōr-** [suffixed o-grade] > PGmc ***watar** > OE *wætar* > WATER

(b) ***wēd-o-** [suffixed lengthened grade] > PGmc ***wēd-** > OE *wæt, wēt* > WET

(c) ***wod-** [o-grade] > PGmc ***wat-skan** > OE *wæscan, wacsan* > WASH

(d) ***we-n-d-** [nasalized form] > PGmc ***wintruz** ‘wet season’ > OE *winter* > WINTER

(e) ***ud-ro, *ud-r-** [suffixed zero-grade] > PGmc ***otraz** > OE *otor* > OTTER
2. Borrowings from other Indo-European languages
- (a) ***ud-ōr-** [suffixed zero-grade] > Greek *hudōr* ‘water’ > HYDRO- (including CLEPSYDRA, DROPSY)

(b) ***u-n-d-ā-** [suffixed nasalized zero-grade] > Latin *unda* ‘wave’ > UNDULATE, INUNDATE, ABOUND, REDUNDANT, SURROUND

(c) ***ud-skio-** [suffixed zero-grade] > Scot. and Ir. Gaelic *uisge* ‘water’ > *uisquebaugh* ‘eau de vie’ > WHISKEY

(d) ***wod-ā-** [suffixed o-grade] > Russ. *voda* ‘water’, with *-ka* ‘diminutive’ > VODKA

Figure 16: English reflexes of the PIE etymon ***wed-** ≈ ***wod-** ≈ ***ud-** WATER (adapted from AHD 73)

Treating a loanword as native

I was at first delighted when I ran across the Jingpho form **wéʔ-wū** ‘screw’, since its first syllable looked like an excellent match with Lahu **ṣ-vèʔ** ‘id.’, for which I then had no etymology. Could this be a precious example of the rare PTB rhyme ***-ekʔ**? But the screw is hardly an artifact of any great antiquity, and it would be *prima facie* implausible that a root with such a meaning would have existed in PTB. The truth quickly became apparent. The modern Burmese form for ‘screw’, **wéʔ-ʔu** (WB **wak-ʔu**), the obvious source from which both Jingpho and Lahu borrowed these words, means literally “pig-intestine”. The semantic association is the corkscrew-like appearance of a pig’s small intestine. This etymology is also interesting from the viewpoint of distinguishing native vs. borrowed co-allofams. The usual, native words for ‘pig’ in Jingpho and Lahu are **wàʔ** and **vàʔ**, respectively; but the doublets borrowed from Burmese have front vowels, as in spoken Burmese. Unless a native speaker of Jingpho knows Burmese, s/he is unlikely to realize that the first syllable of **wéʔ-wū** means ‘pig’, especially since this syllable is in the high-stopped tone, while ‘pig’ is low-stopped. The native Lahu speaker is even less likely to recognize the source of **ṣ-vèʔ**, since the morpheme for ‘intestine’ has been completely dropped from the original Burmese compound, rather like the way our word *camera* (< Lat. ‘room; chamber; vaulted enclosure’) is a shortening of the old compound *camera obscura* (“dark chamber”).⁸¹

Combining reflexes of unrelated roots

When two forms bearing a semantic resemblance in a phonologically depleted language differ only in tone, it is tempting to try to relate them. I once entertained the possibility that such pairs of Lahu forms as **phu** ‘silver, money’ / **phû** ‘price, cost’ and **mu** ‘high, tall’ / **mû** ‘sky’ were co-allofams, though they can easily be shown to descend from quite separate etyma: **phu** < PTB ***plu** (? p. 89) / **phû** < PTB ***pəw** (? #41); **mu** < PTB ***mraŋ** (? p. 43) / **mû** < PTB ***r-məw** (? #488).⁸²

Failure to recognize that separately reconstructed etyma are really co-allofams

An opposite type of error is to overlook the etymological identity between sets of forms, assigning them to separate etyma when they are really co-allofams. Thus ʔ sets up two independent PTB roots, both with the shape ***dyam**, one meaning ‘full; fill’ (? #226) and the other glossed as ‘straight’ (? #227). Yet it can be shown that the latter root also means ‘flat’, and that all reflexes of #226 and #227 may be subsumed under a single etymon, with the underlying idea being “perfection in a certain dimension”.⁸³

Similarly, I was slow to recognize that two roots I had set up separately, PLB ***dzay**² ‘cattle; domestic animal’ (? #129) and NE Indian ***tsa:y** ‘elephant; cattle’ (#143) are really one and the same.⁸⁴

Double-dipping

This embarrassing situation occurs when an author inadvertently assigns the same form in a daughter language to two different etyma, perhaps within the pages of the same book, but more likely in separate articles. At different times I have compared Chinese **chún** 唇 ‘lip’ (OC **ḍiʷən**) to both PTB ***dya**l and ***m-ts(y)ul**, finally deciding in favor of the latter.⁸⁵ It is of course perfectly legitimate to change one’s mind, as long as one explains why. The best course is to present the alternative etymologies together, inviting the reader to choose between them.

Misanalyses of compounds

A vast number of words in TB languages are di- or tri-syllabic compounds, a fact which greatly complicates the task of etymologization. Many traps lie in wait for the analyst, leading to potential errors of several kinds.

- Wrong segmentation** This can happen when a form in an inadequately transcribed source is not syllabified. The Pochury and Sangtam forms for ‘star’, transcribed as **awutsi** and **chinghi**, respectively, in the little glossaries compiled by the *Nagaland Bhasha Parishad*,⁸⁶ should be segmented as **a-wu-tsi** and **ching-hi**, and not as **a-wut-si** and **chi-nghi**, as I imprudently did in ʔ:21.
- Misunderstanding the meaning of a constituent** A special case of this problem is mistaking an affix for a root, especially likely to occur when no grammatical description exists for a language. Several Naga languages have disyllabic forms for ‘moon’ with similar final syllables, e.g. Chang **litnyu**, Konyak **linnyu**, Phom **linnyü**, Sangtam **chonu**, Liangmai **chahiu**. Yet these final elements do not constitute a new root meaning ‘moon’, as I had originally guessed; rather they represent an abstract formative, ultimately grammaticalized from a root ***n(y)u** ‘mother’, that occurs in nouns from all sorts of semantic fields (e.g. Chang **chinyu** ‘center’, **henyu** ‘ladder’, **lamnyu** ‘road’, **pinyu** ‘snake’).⁸⁷
- Choosing the wrong syllable of a compound for an etymology** This can happen when two different syllables of a compound are phonologically similar, especially if one is dealing with a poorly known language with depleted final consonants, e.g. forms like Guiqiong Ganzhi **tʃha**⁸⁸**sá**⁸⁸ and Ersu **ʂ**⁸⁹**ji**⁸⁹ ‘otter’. Which syllables are to be ascribed to PTB ***sram**?

By its very nature, comparative/historical investigation of such a complex language family as TB/ST is an open-ended enterprise. As described above, changes continue to be made in the details of the STEDT reconstructive system,⁸⁸ as well as in our conception of the subgrouping of the family.⁸⁹

Desiderata for the future

I am acutely aware of the incompleteness of this Dictionary-Thesaurus. We still have a long way to go before comparative/historical TB/ST studies are as advanced as they deserve to be. Despite the quickening pace of research, our knowledge of the various branches of this multifarious family remains highly uneven. With a few important exceptions,⁹⁰ reliable reconstructions at the subgroup level are not yet available, so that “teleo-reconstruction” has to be resorted to.⁹¹ Many more roots remain to be reconstructed at all taxonomic levels of the family.

Much remains to be done on the Chinese side as well, and we seem destined for a period of flux until the dust settles and competing reconstructions have sorted themselves out.

A large lacuna in this work is the lack of a systematic treatment of tone. This is not because the topic does not interest me, but rather because it merits a book-length treatment by itself. We are only just coming to appreciate the richness and variety of TB tone systems as more and more data become available.⁹² The big questions are still open, especially the key issue of monogenesis vs. polygenesis: Can we reconstruct a single tonal system at the PTB level? At the PST level? If so, was this original system primarily phonational or melodic? Or are tonogenesis and tonoxodus cyclical processes, with tones having arisen repeatedly and independently in the various branches of TB, so that even if there was an “original” system, it can no longer be recovered?

Despite the fact that changes continue to be made in our reconstructions (see above), STEDT continues to build upon previous research rather than repudiating it. Perpetual revolution gets to be fatiguing after

⁸¹There is a difference in detail between the two cases, however: the deleted ‘intestine’ is the head of the compound “pig-intestine”, but the deleted *obscura* is the modifier in the collocation “dark-chamber”.
⁸²See ʔ:29; such speculations were debunked in the 2nd printing (1982), p. 675.
⁸³See ʔ:4–9.
⁸⁴I have argued that a third root set up in ʔ (*GSTC* #106), ***(t)say** ≈ ***(d)zay** ‘temperament / aptitude / talent’, is also related, the common notion being ‘property (either material or intellectual)’. See ʔ:44–45; ʔ:10–13.
⁸⁵See ʔ 9.2.1, 9.22(4), 9.2.4.
⁸⁶? (*Hindi Pochury English Dictionary*), ? (*Hindi Sangtam English Dictionary*).
⁸⁷See ʔ (“Stars, moon, spirits”), p. 35; for the suffixal use of morphemes meaning ‘mother’, see ʔ (“The mother of all morphemes”).
⁸⁸Cf. the recent reinterpretation of the PTB mid-vowels ***-e(-)** and ***-o(-)** as rising and falling diphthongs (above).
⁸⁹Cf. the recent reevaluation of the genetic position of Jingpho, Luish and Nungish (above).
⁹⁰These bright spots include Proto-Karen (Haudricourt, Jones, Solnit), Proto-Northern Naga (French), Proto-Tani (J. Sun), Proto-Tamangic (Mazaudon), Proto-Kiranti (Michailovsky), Proto-Kuki-Chin (VanBik), Proto-Northern Chin (Button) Proto-Lolo-Burmese (Burling, Matisoff, Bradley). See the References.
⁹¹See ʔ.
⁹²For a rough typology of TB tone systems, see ʔ. ʔis an attempt to reduce all TB tone systems to a single proto-phonational contrast among clear, breathy, and creaky voice qualities, but is marred by an over-formalistic and disorganized presentation which renders it virtually incomprehensible. See the review by JAM (?).

a while.⁹³ Historical linguistics is a cumulative enterprise. Thanks to the foundation laid by pioneering scholars, especially Paul K. Benedict, a solid body of TB/ST etymological knowledge has been accumulated, in terms of which new etymologies must be evaluated. No longer can one get away with reconstructing whatever one pleases, no matter how typologically bizarre or ad hoc or mechanistic the reconstruction might be.

All in all, it is hard not to be optimistic about the future of TB/ST linguistics, as fieldwork opportunities increase and new generations of talented researchers enter the discipline. Eventually it seems inevitable that scholars throughout the world will share their information more and more, granting mutual access to their databases for the common good. On the other hand, too many TB languages are endangered, and may well disappear before they have been adequately recorded.

In any case, “the reconstruction of PTB is a noble enterprise, where a spirit of competitive territoriality is out of place. We should pool our knowledge and encourage each other to venture outside of our specialized niches, so that we begin to appreciate the full range of TB languages...”⁹⁴

⁹³As Leon Trotsky found to his cost in 1940.

⁹⁴ʔ:41.

Acknowledgments

NSF and NEH Support

I am deeply grateful to the National Science Foundation and the National Endowment for the Humanities for their unswerving support of the *Sino-Tibetan Etymological Dictionary and Thesaurus* (STEDT) Project since 1987, even through times of budgetary stringency. I would especially like to thank Dr. Paul G. Chapin and Dr. Joan Maling, of the Language, Cognition, and Social Behavior division of NSF; and Dr. Guinevere Greist, Dr. Helen Agüera, Dr. Martha Bohachevsky-Chomiak and Dr. Nadine Gardner of the Research Tools Division of NEH. I can only hope that the fruits of this project will repay their confidence and patience.

Grants to the *Sino-Tibetan Etymological Dictionary and Thesaurus* project from:

- * The National Science Foundation (NSF), Division of Behavioral & Cognitive Sciences, Grant Nos. BNS-86-17726, BNS-90-11918, DBS-92-09481, FD-95-11034, SBR-9808952, BCS-9904950, BCS-0345929, BCS-0712570, and BCS-1028192.
- * The National Endowment for the Humanities (NEH), Preservation and Access, Grant Nos. RT-20789-87, RT-21203-90, RT-21420-92, PA-22843-96, PA-23353-99, PA-24168-02, PA-50709-04, PM-50072-07, and PW-50122-08.

Administrators

Several Organized Research Units and academic departments of the Berkeley campus have given their moral or practical support to the STEDT project, including the Center for Southeast Asia Studies, the Center for Chinese Studies, the Department of Linguistics, the Department of South and Southeast Asian Studies, the Department of East Asian Languages and Cultures, and especially the Institute of International and Area Studies, to whose administrative staff I am deeply obliged: Karin Beros, Management Services Officer and all-around trouble-shooter, who was instrumental in solving the practical problems of getting the project started back in 1987; Jerilyn C. Foushée, who has handled our budget and helped with our grant proposals and reports since 1987; and Nell Haskell (1987–95) and Kerttu K. McCray (1995–2002), who have kept track of personnel matters. Since 2009, the STEDT grants have been administered by the Linguistics Department, under the able stewardship of Paula Floro and Belén Flores.

Contributors

I would like to thank Anthony Meadow, founder of the Bear River Associates software development firm, now in Oakland, who generously gave many hours of his time during 1986–87 in *pro bono* consultations about how to formulate the computer needs of the project in our original grant proposals to NSF and NEH.

The importance of data contributors cannot be overstated, and STEDT would like to express the deepest gratitude to the many linguists, native speakers, and other individuals who advanced the project by contributing lexical and/or etymological data to the database. Below is a list of some of these contributors in alphabetical order of surname/family name. Some contributors have no doubt been inadvertently missed, and we apologize for any omissions. (Asterisks indicate contributions that have not (yet) been imported.)

- *Name* | *Contribution(s)*
- N. Achumi | Sumi
- M. Balawan (via Karl-Heinz Grüssner) | Tiwa
- Peri Bhaskararao | Mizo, Tiddim Chin
- Balthasar Bickel | Belhare*, Puma*
- Naomi Bishop | Sherpa Helambu
- Sri Chhimed Bodh | Spiti Tibetan
- David Bradley | Ugong, Bisu
- Philip & Cecilia Brassett | Tujia
- Seino van Breugel | Atong
- Daniel Bruhn | Central Naga languages
- Robbins Burling | Garo
- Robbins Buring & U.V. Joseph | Bodo-Garo languages*, Proto-Bodo-Garo*
- Christopher Button | Northern Chin languages*, Proto-Northern Chin*
- Ross Caughley | Chepang
- CHEN Kang | Tujia
- Katia Chirkova | Uppper Xumi*, Lizu*
- Terry Cooke | Manang*
- Alec Coupe | Mongsen Ao
- DAI Qingxia | various & sundry languages (including Deng, Himalayish, Tibetan, Qiangic, rGyalrongic, Nungic, Burmish, & Loloish languages)
- George van Driem | Dumi, Limbu
- Karen Ebert | Chamling*
- Jonathan Evans | Southern Qiang languages
- Jonathan Evans, John B. Lowe, & Jackson Tianshin Sun | Mandarin Chinese
- Michael Ferlus | Phunoi
- Carol Genetti | Newar
- Karen Grunow-Hårsta | Magar*
- Thien Haokip | Thado
- He Zhiwu | Naxi*
- HUANG Chenglong | Yadu Qiang
- HUZIWARA Keisuke | Sak, Marma, Ganan*, Kadu*, Usoi Tripura*, Proto-Luish*
- François Jacquesson | Deori
- Tej Ratna Kansakar | Baram
- John King | Dhimal
- Tamara Kohn | Yakha
- Donald Kosha | Moyon
- Shree Krishan | Chaudangsi, Darma, Raji
- Randy LaPolla | Rawang, Dulong
- Paul Lewis | Akha
- Li Fanwen (via Richard Cook & Andrew West) | Tangut
- Li Shaoni | Jianchuan Bai*
- Liberty Lidz | Yongning Na
- Theraphan Luangthongkum | Karenic languages, Proto-Karen
- LUO Meizhen | Ahi
- Anton Lustig | Zaiwa*
- MA Erji | Daofu
- MA Xueliang | Sani
- Kamal Malla | Newar
- Ken Manson | Kayan
- LaRaw Maran et al. | Jingpho*
- James Matisoff | various & sundry languages
- Martine Mazaudon | Nepali*, Tamangic languages

- Boyd Michailovsky | Chepang, Kham, Kiranti languages, Proto-Kiranti, Magar, Nepali, Newar, Written Tibetan
- Alexis Michaud | Yonging Na, Laze
- Dinkapar Moral | Boro*, Garo*, Koch*, Tintekiya Koch*
- David Mortensen | Sorbung, Tankghulic languages, Proto-Tangkhulic
- NAGANO Yasuhiko et al. | rGyalrongic languages
- Vikuosa Nienu | Angami, Chokri, Lotha
- Michael Noonan et al. | Chantyal
- Jamin Pelkey | Phula languages
- Audra Phillips | Moulmein Pwo
- Heleen Plaisier | Lepcha
- Mark Post | Apatani*
- Mark Post et al. | Galo
- Krishna Prasad Rai, Anna Holzhausen, & Andreas Holzhausen (via Boyd Michailovsky) | Kulung
- Roland Rutgers | Yamphu
- Daya Ratna Shakya & David Hargreaves | Newar
- Suhnu Ram Sharma | Bunan, Byangsi, Manchati, Rongpo
- SHINTANI Tadahiko | Pyen
- Ch. Yashwanta Singh | Meithei
- Helga So-Hartmann | Southern Chin languages
- David Solnit | Eastern Kayah Li*, Pa-O Karen
- SUN Hongkai | various & sundry languages (including Qiangic, Bodic, rGyalrongic, & Loloish languages)
- Jackson Tianshin Sun | Caodeng, Mandarin Chinese, Mawo Qiang, rGBenzhen, Shili, Amdo Tibetan, Written Tibetan, “North Assam” languages, Tani languages, Proto-Tani
- Tabu Taïd | Mising*
- T. Temsunungsang | Chungli Ao, Mongsen Ao
- TIAN Desheng | Tujia*
- TOBA Sueyoshi & Allen Kom | Kom Rem
- Prashanta Tripuri & Dan Jurafsky | Kokborok
- Kenneth VanBik | Kuki-Chin languages, Prto-Kuki-Chin
- Nienu Vikuosa | Angami, Chokri, Lotha
- WANG Ersong | Haoni*
- David & Nancy Watters | Kham
- Julian Wheatley | Luquan*
- XU Xijian | Langsu, Lashi, Zaiwa*
- Shiro Yabu | Khezha
- ZHANG Fengyu | Batang Tibetan
- ZHANG Liansheng | Written Tibetan
- ZHAO Yansun (via Grace Wiersma) | Bai

STEDTniks

Out of the approximately 70 individuals (undergraduates, graduate students, post-docs) who have worked at STEDT since 1987, I am here singling out 14 for special mention because of the significance of their contributions to the project. They are cited approximately in the order in which they have received their highest degrees from U.C. Berkeley or elsewhere. It should be understood that these brief mentions in no way do justice to the brilliance of their accomplishments, or to the love I feel for them all.

Randy J. LaPolla. (STEDT, 1987–90); Ph.D. Berkeley, 1990.) Has since taught at City University of Hong Kong, La Trobe University (Melbourne), and now at Nanyang Technological University (Singapore). Specialties include Nungish, Qiangic, comparative TB grammar. Edited early STEDT monographs and questionnaires.

Jackson (Tianshin) Sun. (STEDT 1990–1992; Ph.D. Berkeley, 1993.) Now teaching, conducting research, and administrating at Academia Sinica, Taipei. Specialties include Qiangic, rGyalrongic, and the NE Indian branch of Tibeto-Burman to which he assigned the now generally accepted name “Tani”.

John B. Lowe (STEDT 1987–2000; 2010–2014; Ph.D. Berkeley, 1995.) An experienced library scientist and software engineer before joining STEDT as one of the founding members of the team, John’s role has been crucial to the success of the project. He designed and implemented most of the computational apparatus used by STEDT over the years; he invented “tagging” and in 1987 presciently designed the organization and layout of the Dictionary-Thesaurus substantially as it is today.

Zev J. Handel. (STEDT 1991–1998; Ph.D. Berkeley, 1998.) Now Associate Professor of Chinese Language and Linguistics at University of Washington. Specialties include Chinese historical phonology, Sino-Xenic writing systems. Has contributed extensive notes on the Chinese comparanda cited in connection with TB etymologies in several STEDT publications.

Ju Namkung. (STEDT 1993–1997; M.A. Berkeley, 1995.) Currently a web and software developer in the Seattle area. She edited the widely used STEDT Monograph #3, Phonological Inventories of Tibeto-Burman Languages (1996). Editorial assistant (1996–1997) for the journal then published at STEDT, Linguistics of the Tibeto-Burman Area (LTBA).

Jonathan P. Evans. (STEDT 1990–2005; Ph.D. Berkeley, 1999.) Now Associate Research Fellow at Academia Sinica, Taipei. Specialties include Qiangic grammar and historical phonology (especially the development of rudimentary tonal systems), and language contact between Chinese and coterritorial TB languages.

Richard S. Cook. (STEDT 1998–2011; Ph.D. Berkeley, 2003.) Independent researcher, specializing in the history of Chinese characters and the Chinese lexicographical tradition. Played a central role in the production of the Handbook of Proto-Tibeto-Burman. Created electronic versions of several key reference works, which have become “ancillary STEDT databases”.

Kenneth VanBik. (STEDT 1997–2006; Ph.D. Berkeley, 2006.) Now teaching at San Jose State University. A native speaker of Lai Chin and Burmese, he has contributed many PTB etymologies based on his newly discovered Chin/Burmese cognates. His dissertation was a full reconstruction of Proto-Kuki-Chin, with over 1350 cognate sets.

David Mortensen. (STEDT 2003–2006; Ph.D. Berkeley, 2006.) Now teaching at University of Pittsburgh. Originally with a background in the Hmong-Mien family, he now also specializes in the Tangkhulic branch of the Naga languages, as well as in theoretical phonology. Played a major role in the production of the English-Lahu Lexicon.

Nina J. Keefer. (STEDT 2000–2008; M.A./M.J. Berkeley [Group in Asian Studies, Graduate School of Journalism].) Editorial Assistant for LTBA (2001–2004) and my chief assistant in the production of the English-Lahu Lexicon. Background in Burmese culture and religion. Currently working as a forensic mental health nurse in Canada.

Liberty A. Lidz. (STEDT 1998, 2009-2014). Ph.D. University of Texas at Austin (2010). Specialist in languages of the Na(xi) group, closely related to Lolo-Burmese, and author of a dissertation grammar on Yongning Na. She did extensive data analysis for the STEDT project, having tagged roughly 43% of the forms tagged in the database at the time of publication, as well as contributing a number of new etymologies and working closely with the P.I. to rectify existing PTB roots.

Dominic Yu. (STEDT 2004–2012; Ph.D. Berkeley, 2012.) Now employed as an International UI Software Engineer at Apple. Has specialized in Qiangic languages, especially those in the Ersuic group. Formatted and produced The Tibeto-Burman Reproductive System (2008). He continues to make key contributions to the structure of the STEDT database and website.

Daniel W. Bruhn. (STEDT 2009–2014; Ph.D. Berkeley, 2014.) Currently on the STEDT staff. Has specialized in the languages of the Central Naga group. A Jack-of-all-trades, he is contributing significantly to all

ongoing STEDT projects, including the incorporation of new sources into our database, modifications to our underlying code, improvements in our website, and formatting our publications.

Chundra A. Cathcart. (STEDT 2011–2014; Ph.D. Berkeley, expected 2014–5.) Currently on the STEDT staff. Specializes in the comparative/historical study of the Indo-Aryan family, and frequently identifies loans from IA languages into Tibeto-Burman. Is assisting the P.I. in the preparation for publication of a large corpus of Lahu texts.

Visiting Scholars

The STEDT project has been greatly enriched by the specialized expertise, unpublished data, and intellectual stimulation provided by a succession of visiting scholars, who have spent anywhere from a few weeks to more than two years at the project headquarters. Two especially deserve special mention:

Martine Mazaudon. (STEDT 1987–89, 1990–91; Doctorat 3e Cycle, Paris V, 1971; Doctorat d’Etat, Paris III, 1994.) Now a Researcher Emerita at the Centre National de la Recherche Scientifique (Paris), specializing in the Tibeto-Burman languages of the Tamangic group of Central Nepal. Spent two full years at STEDT.

Boyd Michailovsky. (STEDT 1987–89, 1990–91; Ph.D. Berkeley, 1981; Doctorat 3e Cycle, Paris III, 1981; Doctorat d’Etat, Paris III, 2004.) Now a Researcher Emeritus at the Centre National de la Recherche Scientifique (Paris), specializing in the Tibeto-Burman languages of the Kiranti (Rai) group of Eastern Nepal. Spent two full years at STEDT.

DAI Qingxia and XU Xijian (Oct.–Nov. 1989) Nationalities University (Beijing), TB languages of China; ZHANG Jichuan (Nov. 1990) Chinese Academy of Social Sciences (Beijing), Tibetan dialects; the late Rev. George Kraft (1990–99), Khams Tibetan; Nicolas Tournadre (Feb. 1991) University of Paris III, Tibetan; SUN Hongkai and LIU Guangkun (April–May, 1991) Chinese Academy of Social Sciences (Beijing), TB languages of China; YABU Shiro (April–Aug. 1994) Osaka Foreign Languages University, Burmish languages and Xixia; William H. Baxter III (May, 1995) University of Michigan, Old Chinese; Balthasar Bickel (Sept.–Oct. 1996; Feb.–Mar. 1997) University of Zürich and Johann Gutenberg University (Mainz), Kiranti languages; LIN Ying-chin (1997–98) Academia Sinica, Taipei (Xixia, Muya); David B. Solnit (1998–) STEDT, Karenic; WU Sheng-hsiung (Spring, 2002) Taiwan Normal University, Chinese phonology; IKEDA Takumi (2002–03) Kyoto University, Qiangic languages; Yashawanta Singh Manipur University (Imphal) Meithei; Maung Maung (Aaron Tun) (2011) Payap University (Chiangmai), Lahu; LI Yan (2013–14) Shaanxi Normal University (Xi’an) Historical linguistics and dialectology.

Complete List of STEDTniks

Most of all, I am indebted to the phalanges of talented students, past and present, who have been working at STEDT anywhere from five to 20 or 30 hours per week, performing a host of vital tasks such as the inputting and proofreading of hundreds of thousands of lexical records, the development of special fonts and relational database software, computer maintenance and troubleshooting, formatting articles for our journal *Linguistics of the Tibeto-Burman Area*, and editing the publications in the STEDT Monograph Series. 70 researchers have contributed to the STEDT project since 1987, mostly graduate students working as research assistants in the Berkeley Linguistics and East Asian Languages and Cultures Departments, but also including several undergraduate volunteers and non-enrolled or former students. Here they are, in an alphabetical honor roll:

- | | | |
|----------------------------|-------------------------------|------------------------|
| • Madeleine Adkins | • Zev J. Handel | • Toshio Ohori |
| • Jocelyn Ahlers | • Takumi Ikeda | • Weera Ostapirat |
| • Shelley Axmaker | • Alexander Jacobson | • Jeong-Woon Park |
| • Kenny Baclawski | • Annie Jaisser | • Jason Patent |
| • Stephen P. Baron | • Matthew Juge | • Chris Redfearn |
| • Leela Bilmes (Goldstein) | • Daniel Jurafsky | • S. Ruffin |
| • Michael Brodhead | • David Kamholz | • Keith Sanders |
| • Daniel Bruhn | • Nina Keefer | • Marina Shawver |
| • Chundra Cathcart | • Jean Kim | • Elizabeth Shriberg |
| • Jeff Chan | • Kyung-Ah Kim | • Helen Singmaster |
| • Patrick Chew | • Heidi Kong | • Tanya Smith |
| • Melissa Chin | • Aimée Lahaussais (Bartosik) | • Gabriela Solomon |
| • Isara Choosri | • Randy J. LaPolla | • Silvia Sotomayor |
| • Richard Cook | • Jennifer Leehey | • Jackson Tianshin Sun |
| • Jeff Dale | • Anita Liang | • Laurel Sutton |
| • Amy Dolcourt | • Liberty Lidz | • Prashanta Tripura |
| • Julia Elliot | • John B. Lowe | • Margaret Urban |
| • Jonathan P. Evans | • Jean McAneny | • Nancy Urban |
| • Allegra Giovine | • Pamela Morgan | • Kenneth VanBik |
| • Cynthia Gould | • David Mortensen | • Blong Xiong |
| • Daniel Granville | • Karin Myrhe | • Dominic Yu |
| • Joshua Guenter | • Ju Namkung | • Liansheng Zhang |
| • Kira Hall | | |

Conventions

Abbreviations for Languages Commonly Referred to in the text

| | |
|-----------------------------------|--------------------------------|
| Gk. Greek | PIE Proto-Indo-European |
| HM Hmong-Mien (= Miao-Yao) | PLB Proto-Lolo-Burmese |
| IA Indo-Aryan | PNC Proto-Northern Chin |
| IE Indo-European | PNN Proto-Northern Naga |
| Jg. Jingpho (= Kachin) | PST Proto-Sino-Tibetan |
| KC Kuki-Chin | PTB Proto-Tibeto-Burman |
| LB Lolo-Burmese | Skt. Sanskrit |
| Lh. Lahu | ST Sino-Tibetan |
| MC Middle Chinese | TB Tibeto-Burman |
| NN Northern Naga | TK Tai-Kadai |
| NT Northern Tai | WB Written Burmese |
| OC Old Chinese | WT Written Tibetan |

Conventions Used in Citing the Source of Individual Wordforms

The wide variety of sources included in the database make it difficult to provide a single pointer to a wordform in a source. Typically, an inline bibliographic citation includes something like the author's name, a year (with alphabetic disambiguator if needed) and a page number.

In the STEDT database, many sources do not have page numbers; and at any rate a page number is often not the most useful identifier. We therefore developed an informal system for marking provenance:

- If the source is short and has no page numbers, we have usually omitted source identifiers. A reader who is curious to examine

the wordform in its context must locate the hardcopy and find the item using the collation and indexing conventions provided.

- For etymological sources, e.g. the *Conspectus* (?), we used the etymological set number; since many forms are cited that do not belong to sets, we have indicated the page and note number where the form is to be found, e.g. 123n5.
- A number of sources utilize some sort of hierarchical (e.g. semantic) organization of the forms. In most of these cases, we used the hierarchical numbering system provided.

Miscellaneous

| | | | |
|------------------------------------|--|-----------------------------|--|
| allofam | a variant of an etymon | root canal | the process through which a STEDT etymology is made public, so that it can be commented upon |
| allofam box | a computer display where all the allofams of an etymon are grouped together within a box on the screen | semantic flowchart | SEE metastatic flowchart |
| chapter | the smallest section of a fascicle, containing all the etyma reconstructed with a given gloss, e.g. KIDNEY | semcat | a semantic category; each STEDT etymology is assigned to one or more of them |
| extra-fascicular etyma | etyma that are cited outside of their normal fascicle to illustrate a particular point | supporting forms | forms from individual languages that support an etymology |
| fascicle | a major section of a STEDT volume, comprising many chapters with closely related glosses, e.g. INTERNAL ORGANS | volume | one of the ten major divisions of STEDT, e.g. BODY-PARTS |
| metastatic flowchart | a diagram illustrating the patterns of semantic association displayed by an etymon | weakly attested root | a promising etymology, but with insufficient support |
| meso-reconstruction | reconstruction of an etymon at the subgroup level, deemed to be descended from that same etymon at a higher taxonomic level | > | goes to; becomes |
| meso-root | an etymon reconstructed at the subgroup level | < | comes from; is derivable from |
| pan-allofamic formula (PAF) | an abstract reconstruction intended to display simultaneously all the patterns of variation attested for a given etymon | * | indicates a reconstructed form (e.g. PTB * b-ləy), or acts as a synonym for “proto-” (e.g. *labials, *Kuki-Chin) |
| rectification | the process by which etymologies are subjected to reanalysis, to determine whether they should be accepted as is, accepted with modifications, or rejected | A × B | A and B are co-allofams; A and B are members of the same word-family |
| | | A ∞ B | Are A and B co-allofams? Do A and B belong to the same word family? |
| | | A ≠ B | A and B are not co-allofams; A and B do not belong to the same word family |
| | | Clf. | classifier |
| | | lit. | literally |
| | | ult. | ultimately |

Abbreviations for People

| | |
|---|-------------------------------------|
| AH André-Georges Haudricourt | KVB Ken VanBik |
| B & S (William H.) Baxter & (Laurent) Sagart | WHB William H. Baxter |
| JAM James A. Matisoff | WTF Walter T. French (cf. ?) |
| JS Jackson Sun | ZJH Zev J. Handel |

Abbreviations for Sources and Venues

| | |
|--|---|
| AD <i>Analytic Dictionary of Chinese and Sino-Japanese</i> (?) | (L)TSR <i>The Loloish Tonal Split Revisited</i> (?) |
| DL <i>The Dictionary of Lahu</i> (?) | STC <i>Sino-Tibetan: a Conspectus</i> (?) |
| GL <i>The Grammar of Lahu</i> (?) | TBL 藏缅语族语言词汇 [A <i>Tibeto-Burman lexicon</i>] (?) |
| GSR <i>Grammata Serica Recensa</i> (?) | TBRs <i>The Tibeto-Burman Reproductive System: Toward an Etymological Thesaurus</i> (?) |
| GSTC/G&C <i>God and the Sino-Tibetan Copula</i> (?) | VSTB Variational Semantics in Tibeto-Burman: The ‘Organic’ Approach to Linguistic Comparison (?) |
| HCT <i>A Handbook of Comparative Tai</i> (?) | ZMYYC 藏缅语语音和词汇 [<i>Tibeto-Burman phonology and lexicon</i>] (?) |
| HPTB Handbook of Proto-Tibeto-Burman (?) | |
| ICSTLL International Conference on Sino-Tibetan Languages and Linguistics | |
| LTBA <i>Linguistics of the Tibeto-Burman Area</i> | |

References

List of Languages

