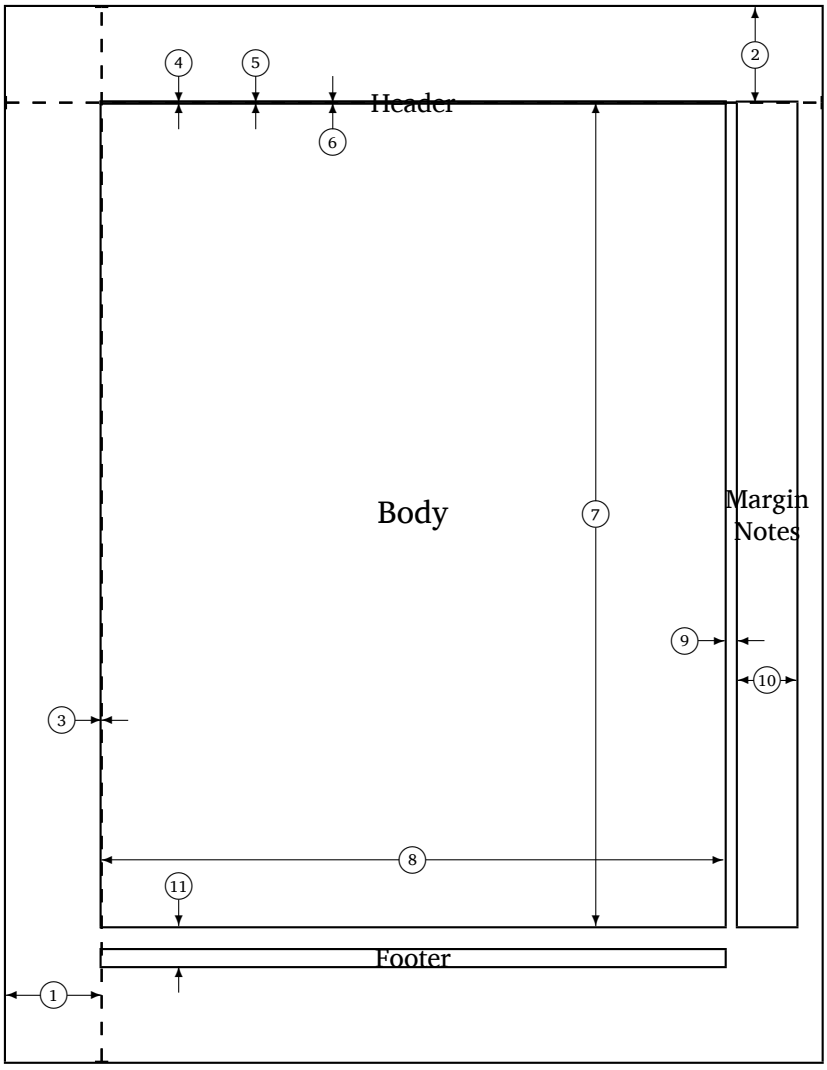

1	one inch + \hoffset	2	one inch + \voffset
3	\oddsidemargin = 0pt	4	\topmargin = 0pt
5	\headheight = 0pt	6	\headsep = 0pt
7	\textheight = 620pt	8	\textwidth = 469pt
9	\marginparsep = 10pt	10	\marginparwidth = 44pt
11	\footskip = 30pt		\marginparpush = 7pt (not shown)
	\hoffset = 0pt		\voffset = 0pt
	\paperwidth = 614pt		\paperheight = 794pt

*Acknowledgements

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, of the Language, Cognition, and Social Behavior division of NSF; and Dr. Guinevere Greist, Dr. Helen Agüera and Dr. Martha Bohachevsky-Chomiak 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, and BCS-0712570.
* 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, and PM-50072-07.

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 200x, the STEDT grants have been administered by the Linguistics Department, under the able stewardship of Paula Floro.

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.

It is a pleasure to single out several ``Stedtniki" whose contributions to this project and the present volume have been particularly outstanding, and all of whose computorial expertise infinitely outstrips my own:

- John B. (``J.B.") Lowe, the only researcher who has been continuously working at STEDT since its inception in 1987, designed our initial computer environment and has been fine-tuning it ever since, creating original database software adapted to the highly specialized needs of the project and breaking new conceptual ground in the use of the computer for etymological research.[1]

[1] J.B.'s work at STEDT has already spun off into several other etymological projects on which he has consulted here and abroad: M. Mazaudon and Boyd Michailovsky's *Reconstruction Engine* (Paris) for testing putative cognate sets in Himalayan languages; L. M. Hyman's *Comparative Bantu On-line Dictionary* (CBOLD, Berkeley); Sjors van Driem and K.B. Kepping's *Tangut Dictionary Project* (Leiden), and Sharon Inkelas' *Turkish*

- Randy J. LaPolla, now teaching at the City University of Hongkong, has also been affiliated with STEDT since the beginning. Until receiving his doctorate in 1990, he played a vital part in our activities, including the preparation of STEDT Monographs and the processing of fieldworkers' questionnaires. His superb knowledge of Chinese has been a prime asset to the project.
- Zev J. Handel (``Z as in *zebra*, V as in *violin*", as he explains over the telephone), is a specialist in Chinese historical phonology, now teaching at the University of Washington. He was active at STEDT in the 1990's, and had a major role in the formatting of our prototype ``fascicle" on the *Reproductive System* for our projected
- Bodyparts volume, adding bells and whistles like the program to insert notes at various points in the etymologies, and transforming my hand-scrawled semantic diagrams into elegant computer graphics. I am especially grateful to him for producing the concise comparison of three of the most influential systems for reconstructing Old Chinese that appears as an Appendix to this *Handbook*.

When I went off on sabbatical to Taiwan during 1995-96, I left the day-to-day running of STEDT in the capable hands of J.B. and Zev. One day I e-mailed them from Taipei, referring to them as the ``duumvirate" . Back came an aggrieved message from J.B., protesting that they really would rather be called the ``smart-virate" . No argument there.

- Kenneth VanBik is a native speaker of Lai Chin and a graduate of Rangoon University. Possessing an intimate knowledge of languages from two branches of Tibeto-Burman, he was able to identify a number of new Burmese/Chin cognates that are thus reconstructible at the PTB level. His etymologies are included in this volume, marked ``KVB".
- Richard S. Cook, currently producing a mammoth dissertation on the Eastern Han

Electronic Living Lexicon.

``Grammaticon" *Shuō Wén Jiě Zì* , has been the chief architect of the formatting of this Handbook during 2002-3. It was his idea to transfer the whole MS from Microsoft Word 5.1a to Adobe FrameMaker™, an arduous process that has paid off in the end, as the attractive appearance of the book testifies. Richard wrote Appendix B (in consultation with Zev Handel), and extracted the etymologies from the electronic Dictionary of Lahu files to supplement the Index of Proto-Forms . He wrote the computer programs to format the Index of Proto-Forms and to generate and format the indexes of *Proto-Glosses*, *Proto-Root-Syllables*, *Proper Names*, and *Chinese Character*s. He produced the kerned version of the STEDT PostScript font family, as well as the font for the rare Chinese characters found in this book.

- David Mortensen, a linguistics graduate student specializing in Hmong-Mien, has contributed equally to the production of this *Handbook*. An accomplished computorial troubleshooter, he did much formatting work, and has carried out such vital tasks as assuring the integrity of the *Handbook's* innumerable internal cross-references.

During this period J. B. Lowe devised a pioneering program called "The Tagger's Assistant", that enabled me to etymologize tens of thousands of syllables in our database by labelling them with numerical "tags" that could then be used to assemble them into cognate sets. (That is, each syllable deemed to be a reflex of a particular etymon would be tagged with the same number.) With an eye to the eventual publication of our results, J. B. also solved such essential formatting problems as how to insert footnotes at any point in a printed etymological text, whether on a semantic diagram, an etymon as a whole, or a particular supporting form.[2]

So I decided to let the thesauric side of STEDT slide for awhile, and to switch the emphasis

[2]See Section 2.9 of the *Introduction*, below.

of the project to *phonologically* presented etyma (the "D" or "dictionary" part of "STEDT"), an effort which culminated in the publication of the *Handbook of Proto-Tibeto-Burman* (2003).

It gives me special pleasure to thank Professor Zev J. Handel of the University of Washington for his work in producing the Chinese comparanda in this work.[3] Zev had originally contributed such comments to the preliminary version of the manuscript some ten years ago, evaluating my suggested Proto-Tibeto-Burman/Old Chinese comparisons in terms of the competing reconstructive systems of leading Sinologists, past and present. These updated comments, presented in a neutral, non-judgmental tone, constitute a precious guide through the minefield of Chinese historical phonology!

The structure and essential layout of the Dictionary-Thesaurus have been set for some time -- indeed, several intermediate iterations of the work have been produced, starting with the 1995 draft of the Body Parts volume produced in RTF by J. B. The "fascicle producing software has now been re-written three times, each iteration corresponding to a major revision in the underlying technology supporting the database. Dominic Yu did the initial programming of the latest and final "publication system", a set of PERL programs which extracted material from the STEDT database and created LaTex files from which the publication-ready documents were produced. These have been substanially revised in the last two years by J. B., Daniel, and Chundra.

On the back end, his efforts involved porting the entire database to a web-accessible engine using MySQL, accomplished in conjunction with David R. Mortensen and J. B. Lowe, and simultaneously converting our in-house legacy STEDT Font encoding to Unicode. The final print volume is typeset in XeLaTeX using Charis SIL as the main font.

The STEDT project has been sponsored from the beginning by the National Endowment

[3]See the *Introduction*, Section 2.8.

for the Humanities and the National Science Foundation. To both agencies I express again my enduring gratitude.[4]

The STEDT logo was designed by Nadja R. Matisoff.

Finally I would like to thank my wife Susan for her constant support, and for having taught me so much about the reproductive system over the past 46 years.

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:

Martine Mazaudon and Boyd M. Michailovsky (1987-89, 1990-91) Centre National de la Recherche Scientifique (Paris), Himalayan languages ; 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

[4]See *Grant Support*, p. i above.

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. 63 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
- Jocelyn Ahlers
- Shelley Axmaker
- Stephen P. Baron
- Leela Bilmes (Goldstein)
- Michael Brodhead
- Daniel Bruhn
- Chundra Cathcart
- Jeff Chan
- Patrick Chew
- Melissa Chin
- Isara Choosri

- Richard Cook
- Jeff Dale
- Amy Dolcourt
- Julia Elliot
- Jonathan P. Evans
- Allegra Giovine
- Cynthia Gould
- Daniel Granville
- Joshua Guenter
- Kira Hall
- Zev J. Handel
- Takumi Ikeda
- Annie Jaisser
- Matthew Juge
- Daniel Jurafsky
- David Kamholz
- Nina Keefer
- Jean Kim
- Kyung-Ah Kim
- Heidi Kong
- Aimée Lahaussois (Bartosik)

- Randy J. LaPolla
- Jennifer Leehey
- Anita Liang
- Liberty Lidz
- John B. Lowe
- Jean McAneny
- Pamela Morgan
- David Mortensen
- Karin Myrhe
- Ju Namkung
- Toshio Ohori
- Weera Ostapirat
- Jeong-Woon Park
- Jason Patent
- Chris Redfearn
- S. Ruffin
- Keith Sanders
- Marina Shawver
- Elizabeth Shriberg
- Helen Singmaster
- Tanya Smith

- Gabriela Solomon
- Silvia Sotomayor
- Jackson Tianshin Sun
- Laurel Sutton
- Prashanta Tripura
- Nancy Urban
- Kenneth VanBik
- Blong Xiong
- Dominic Yu
- Liansheng Zhang

*Preface

My involvement in Tibeto-Burman (TB) and Sino-Tibetan (ST) comparative reconstruction dates from my first fieldwork on Jingpho, Burmese, and Lahu in the 1960's, and especially from my intense contact with Paul K. Benedict when I was teaching at Columbia University (1966-69). The manuscript version of Benedict's Sino-Tibetan: a Conspectus (STC) had been lying around unpublished since its composition around 1940; it was exciting for me to contribute to its eventual publication in 1972. With its nearly 700 TB cognate sets, and over 300 TB/Chinese comparisons, the Conspectus ushered in the current renaissance of TB and ST comparative linguistics. Its rigor and precision, as well as the breadth of its vision, have made it the indispensable point of departure for subsequent work in the field. While there is certainly room for tinkering with a few details of Benedict's reconstructive scheme for Proto-Tibeto-Burman (PTB), the major features of the system itself remain basically unassailable. The real progress that has been made in the past 30 years lies elsewhere. An avalanche of new data from recent fieldwork has strengthened the support for previously reconstructed etyma and has permitted the reconstruction of hundreds of new roots at all taxonomic levels of TB, though many more undoubtedly remain to be discovered. The harnessing of the computer for etymological research has speeded up the identification of new cognates and provided a powerful tool for testing the validity of proposed reconstructions. A better understanding of the variational processes at work in TB and ST word-families has enabled us to decide more accurately whether sets of forms that bear partial phonosemantic resemblances to each other are really variants of the same etymon or etymologically independent. On the Chinese side, the successors to Karlgren have made profound changes in the reconstructive scheme for Old Chinese, and it is no exaggeration to say that the field of historical Sinology is now going through a period of ferment. Still, almost all of STC's suggested Chinese comparanda for PTB etyma

have gone unchallenged.[1]

Despite its brilliance, 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 STC 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 STC, but at the cost of considerable labor.) Towards the end of his life Benedict does seem to have felt the need to embark on such a systematic project, although it never actually got off the ground.

The present work may be viewed largely as an updating, clarification, and expansion of STC. It aims to build on the valid etymologies already proposed, but also to present new ones that conform to established sound correspondences. When necessary, previously proposed etymologies are modified in order to accommodate new data. In this Handbook, I have organized the discussion according to the inventory of proto-entities at the various points of the syllable: initial consonants; medial glides; prefixes; simple and diphthongal vocalic nuclei; closed syllable rhymes (with final nasals, stops, liquids, and -s); and suffixes.[2] Wherever possible, the regular reflexes in major languages of these

[1]The over 300 TB/OC comparisons made in STC are conveniently indexed in the excellent review by Chou Fa-kao (1972).

[2]Similarly organized examples of the Handbook genre in Southeast Asia include Li Fang-Kuei's *A Handbook of Comparative Tai* (1977) and Wang Fushi's *Miáoyǔ fangyán shēngyùnmǔ bǐ jiào* (Comparison of the

syllabic elements are displayed in tabular form. The best etymologies illustrating each sound-correspondence are presented, and exceptional or problematic cases are discussed, with alternative analyses suggested.

That is the ``systematic" part. The ``philosophical" aspects of this book are more elusive, but implicit throughout. First of all, I have striven for clarity and simplicity of presentation, for ``user-friendliness". Being understandable rather than obscurantist poses certain risks, in that one's opinions are clear and therefore falsifiable in the light of new data, but it has the advantage of encouraging feedback from others.[3] Secondly, I operate under a theoretical framework according to which the proto-lexicon is not conceived of in terms of monolithic, phonosemantically invariant etyma, but rather as a collection of word families that may each exhibit some internal variation on both the phonological and semantic planes, but according to certain reasonable principles. Distinguishing between such valid variational phenomena and wild speculative leaps is not always easy.

* * *

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

Initials and Rhymes of the Miao Dialects; 1979).

[3]The difficulty of STC can be used as an excuse for not studying it thoroughly. It would be tragic if its fundamental insights were to be forgotten.

/ 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 (1978) 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.[4] 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 rapidly extending

[4]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* (1949).

to 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.

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. 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 bodyparts into ten subdivisions including head, limbs, internal organs, diffuse organs, reproductive system, etc., each to be published separately. 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

*k^{w}əy	lətak	rəmay	gəya:p	way
dog	ACC	tail	wag	COP/NOM

or ``the tail wagging the dog".[5] 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. The ``introductory essay", then entitled System and Philosophy of Tibeto-Burman Reconstruction, eventually grew to its present length of some 600 pages. It gradually dawned on me that it would be preferable to publish it as a stand-alone book, indeed a Handbook. This decision has much to recommend it. In its present form, the phonological approach of this Handbook is complementary to the main thrust of the STEDT project, which is semantically organized. Both prongs of attack are certainly necessary. Henceforth each set of etymologies in the various semantic areas of the lexicon can be put up on the worldwide web as soon as they are deemed ready to go, rather than waiting until they can appear in print form. Many trees will be spared as reams of paper are saved. As each series of etymologies is released, it will be possible to solicit comments and criticisms from colleagues all over the world, and it will be simplicity itself to incorporate any addenda or corrigenda. It is extremely wasteful of space to print out computer records from a database -- who wants to see the gloss `egg' printed out hundreds of times? Since STEDT has had a policy of ``following copy", the same form from a given language (especially well documented ones like Written Burmese or Written Tibetan) is likely to appear several times in slightly different transcriptions used in the various sources. Instead of trying to ``normalize" these, or indeed to delete totally identical records from different sources, we can just include them all, thereby saving much drudgery, since space will not be an issue. Perhaps the greatest advantage of having this Handbook appear before the semantically organized etymologies are promulgated is that it can serve as a standard or ``template" against which each newly proposed etymon can be tested. Let us say, e.g., that a hypothetical new PTB root *b-zer-s has been recon-

[5]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 *k^{w}əy `dog' to initial position.

structed with the meaning `tonsil'. The supporting forms for this etymology can then be compared for consistency with other data that motivate the reconstructions of the same proto-elements, i. e. other etyma with prefixal *b- (§4. 4. 3), with initial *z- (§3. 3), with the liquid-final rhyme *-er (§9. 2. 3), and with suffixal *-s (§11. 4). Before long the Handbook itself can be put up on the web, so that these new etymologies may be plugged directly into it. Much obviously remains to be done. The data are still uneven in the various branches of the family, ranging from the overwhelmingly copious to the tantalizingly sparse. Most strikingly perhaps, this Handbook makes no attempt to reconstruct tones at the PTB level, although this can already be done at the level of certain individual subgroups (e.g. Lolo-Burmese, Tamangic, Karenic). Some reconstructions are given at the subgroup level, when they are available, and a number of roots are marked as being confined to certain subgroups (e.g. Himalayan, Kiranti, Kamarupan, Lolo-Burmese, Karenic). It is precisely these roots of limited distribution, or ``cognate isoglosses'', that will prove to be important for a finer subgrouping of the TB family. However, new data frequently forces us to revise our judgments of etyma distribution: many roots considered to be confined to a single subgroup in STC must now be set up for TB as a whole. These are usually noted in the text. As emphasized in the Conclusion (Ch. XIII), the approach of this Handbook is definitely conservative, in that speculative etymologies are almost always avoided, or at any rate suitably hedged. Variational phenomena are handled with care; phonosemantically non-identical roots are not claimed to be co-allofams unless the morphophonemic relationship between them is paralleled in other word families. Semantic leaps are kept to a minimum, and detailed justification is provided when the meanings of putative cognates diverge significantly. Many solid Chinese comparanda to TB etyma are offered, but no attempt is made to choose among the often contradictory reconstruc-

tive schemes for Old Chinese;[6] for now I just use the classic reconstructions of Karlgren (with some modifications[7]), a policy which STC also followed.[8] I usually have not tried to set up PST forms, as STC sporadically tries to do. I just give the best comparanda. That is why this is basically a Tibeto-Burman handbook, even though its system and methodology apply to all of Sino-Tibetan (hence the subtitle). The primary organization of this Handbook is by rhyme, since this is the most stable part of the syllable.[9] In sharp contrast to Indo-European, the manner of initial consonants (voicing and aspiration) in TB/ST is highly variable, due to the pervasive phenomenon of prefixation (see Ch. IV). Chinese comparanda (I usually avoid the term ``cognate'') are given mostly under the proto-rhyme of their TB counterparts. Most correspondence charts of reflexes also appear under the rhymes. Still there is a certain unavoidable repetitiveness, in that the same root might be discussed in different contexts, e.g. with respect to its initial, its rhyme, and/or its variational pattern. The Indexes will facilitate finding all references to a given etymon. A few words about nomenclatural and transcriptional matters:[10]

- Names for TB languages have undergone frequent changes, as exonyms are replaced by autonyms, and as names felt to be pejorative become politically incorrect.[11] However, certain older language names have been retained, just because they are more widely used in the literature: thus I use ``Lushai'' instead of the now-preferred self-designation ``Mizo''.

[6]See ``A Concise Introduction to Old Chinese Phonology'' by Zev Handel (below, Appendix A), which treats the major differences in the reconstructive systems of Karlgren, Li Fang-Kuei, and W.H. Baxter.

[7]One minor change is that we write the velar nasal as ``ŋ'' instead of ``ng''.

[8]Despite of the fact that Karlgren's system has been superseded and simplified in some respects by subsequent scholars, GSR remains the best-known, most copious, and most convenient reference for OC. I conventionally do not precede OC reconstructions with an asterisk. Asterisks do, however, appear before the OC forms cited in Appendix A.

[9]Hence the great utility of rhyming dictionaries for TB languages; Benedict put several such to good use during the compilation of the Conspectus.

[10]For more details about the transcriptional systems used for key languages, see Citational and Transcriptional Conventions, below.

[11]For a discussion of the issues surrounding the proliferation of language names in TB, see JAM 1986a.

- Subgroup names can be particularly confusing. Occasionally I use equivalent names for the same subgroup, e.g. ``Himalayish" or ``Himalayan", ``Bodo-Garo" or ``Barish", ``Kuki-Naga" or ``Kuki-Chin-Naga". My use of ``Kamarupan" as a geographical cover term for the subgroups of Northeast India (including Abor-Miri-Dafla, Bodo-Garo, and Kuki-Chin-Naga) remains controversial, although it is certainly useful.[12]
- Tones are not marked for every language that has them, especially not for those where no good tonological description is available. Tones are consistently marked for Lolo-Burmese languages and for Jingpho, as well as for the tonal languages cited in Sun et al., 1991 (ZMYYC) and Dai et al., 1992 (TBL); but they are only sporadically provided for such languages as Lushai and Lai Chin.

Great care has been taken to ascribe etymologies to their original source. Any TB etymology or part thereof not specifically ascribed to a prior source is original with me, as far as I know. In any case, the responsibility for the TB reconstructions is mine alone.

It is hoped that this Handbook will prove useful to specialists and general linguists alike, and that it will help to demystify the most important understudied language family in the world.

[12]See JAM 1999c (``In defense of `Kamarupan' ").

*Introduction

Apropos Sino-Tibetan and the STEDT project

This Front Matter includes material from a number of previous publications and oral presentations: HPTB, TBRS, VSTB, ICSTLL 46.

The Sino-Tibetan (ST) language family, comprising Chinese on the one hand, and the hundreds of Tibeto-Burman (TB) languages on the other, is one of the largest in the world, with well over a billion and a half speakers.[1] Yet the field of ST linguistics is only about 70 years old, and many TB languages remain virtually unstudied. The *Sino-Tibetan Etymological Dictionary and Thesaurus* project (STEDT) was begun in August 1987, with the goal of reconstructing the lexicon of Proto-Sino-Tibetan and Proto-Tibeto-Burman from both the phonological and the semantic point of view.

The great Sino-Tibetan language family, comprising Chinese on the one hand and Tibeto-Burman (TB) on the other,[2] 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 TB branch.

After the existence of this vast and ramified family of languages was posited in the mid-19th century, British scholars 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

[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. See Benedict 1975a (*Austro-Thai Language and Culture, with a glossary of roots*).

[2]Many scholars, especially in China, interpret "Sino-Tibetan" to include the Tai and Hmong-Mien families as well, though a consensus is developing that these latter two families, while possibly related to each other, have only an ancient contact relationship with Chinese (Benedict 1975a; JAM 1991a:486-90).

genetically related to the two major literary languages, Tibetan and Burmese. This early work was collected in the monumental Linguistic Survey of India (Grierson and Konow 1903-28), 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.[3] 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 (1966-67; 1974).

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 STC), 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 (1942) banished it from the family altogether, relating Tai instead to Austronesian.[4] 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.

[3]For a readable and humorous account of this project, see Benedict 1975b (LTBA 2.1:81-92).

[4]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 Benedict 1975a, 1990).

The publication of a revised and heavily annotated version of STC 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 (491 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 by Karlgren (1957). While Benedict focussed 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.[5] the moment of writing (September, 1997) marks the 30th anniversary of the publication of STC in 1972. The recent tragic death of Benedict in a car accident (July 21, 1997) makes this a particularly appropriate time to take stock. How well has STC 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,[6] and has been translated into Chinese.[7]In fact nearly all 700 of the TB cognate sets in STC 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.[8]

[5]In a recently published work, Peiros and Starostin (1996) follow 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 Handel (1998, Ch. 3).

[6]A notable exception is the intemperate review by Miller (1974), 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 STC against Miller's attack by JAM (1975a).

[7]See Le Saiyue and Luo Meizhen 1984.

[8]E.g. *dyam & *tyam [STC #226] 'full; fill' and *dyam [STC #227] 'straight' ; see JAM 1988a.

Original plan of STEDT

As originally conceived, STEDT was to produce a series of large print volumes, each devoted to the exhaustive presentation of the reconstructed roots in a specific semantic area, covering the entire lexicon, approximately as follows:

Volume I: *Body Parts*

Volume II: *Animals*

Volume III: *Natural Objects, Plants, Foods*

Volume IV: *Kinship Terms, Ethnonyms, Social Roles*

Volume V: *Culture, Artifacts, Religion*

Volume VI: *Verbs of Motion, of Manipulation, and of Production*

Volume VII: *Adjectival Verbs*

Volume VIII: *Abstract Nouns and Verbs, Psychological Verbs, Verbs of Utterance*

Volume IX: *Shape, Size, Color, Measure, Number, Time, Space*

Volume X: *Grammatical words*

Each volume was in turn to be divided into a number of smaller units called “fascicles”. Thus Vol. I *Body Parts* was to comprise the following nine fascicles:

1. *Body (general)*
2. *Head and Face*
3. *Mouth and Throat*
4. *Torso*
5. *Limbs, Joints, and Body Measures*
6. *Diffuse Organs*
7. *Internal Organs*

8. *Secretions and Somatophonics*[9]

9. *Reproductive System*

Every subpart and sub-subpart of the lexicon expanded and bloomed into a major project. Concentrated on NBP's --- TBRS was fascicle 9 of Vol. I.

- The end/culmination of the Sino-Tibetan Etymological Dictionary and Thesaurus project (STEDT), 1987-2014.
- Lexical database of nearly 500,000 forms from ca. 300 TB languages.
- Thousands of reconstructed roots, both at Proto-Tibeto-Burman and subgroup levels, along with Chinese comparanda.
- A monograph series. HPTB, TBRS, ELL.
- Final product: a printed volume of several hundred pages (only a few copies for libraries!), to be made generally available electronically.

Phonological vs. semantic criteria

Buck and Roget

Comparison with HPTB: In HPTB the etyma were discussed, sorted and analyzed according to their phonological shapes, regardless of their meanings.

Comparison with VSTB and TBRS: phonologically disparate etyma assembled according to their meanings. This is also the strategy of our Root Canal extractions: HAND/ARM/WING; SKIN; ZODIAC.

This is also the strategy of this Final Product

[9]By “somatophonics” I mean sneezes, belches, farts, and the like.

Methodology

The "tagging" process. The Lahu trisyllable for NAVEL.

Terminology

[[INSERT JIM'S PROSE]]

Subgrouping

[[INSERT JIM'S PROSE]]

Subgroup names

Tibeto-Burman is an extremely complex language family, with great internal typological diversity, comparable to that of modern Indo-European. This diversity is due largely to millennia of language contact, especially with the prestigious cultures of India and China,[10] but also with the other great language families of Southeast Asia (Austroasiatic, Tai-Kadai, Hmong-Mien), as well as with other TB groups. We are thus faced with 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."[11] 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

[10]I have called the Indian and Chinese areas of linguistic and cultural influence the *Indosphere* and the *Sinosphere*. See Matisoff 1973.

[11]Matisoff 1978 (*VSTB*), p. 2.

they do exist, there are so many of them that there is no point in talking about them![12]

In the published version of *STC* (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 Fig. 1.

Figure 1. Schematic Chart of Sino-Tibetan Languages[13]

A simpler scheme represents the heuristic model now used at STEDT. See Fig. 2.

Figure 2. Simplified STEDT Family Tree of ST Languages

This diagram differs from *STC* in several respects:[14]

- 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 *STC*, 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 have provisionally been 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

[12]See his review (2003) of G. Thurgood & R.J. LaPolla, eds. (2003), *The Sino-Tibetan Languages*.
[13]Reproduced from *STC*, p. 6; *VSTB*, p. 3; *HPTB*, p. 4.
[14]See *HPTB*, pp. 5-6.

at the time *STC* 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.[15]

- The Nungish and Luish languages are grouped with Jingpho (=Kachin). Jingpho is also recognized to have a special contact relationship with the Northern Naga (=Konyak) group.[16]
- The somewhat idiosyncratic Mikir, Meithei (=Manipuri), and Mru languages are 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.
- The relatively well-studied Lolo-Burmese group (= *STC*'s "Burmese-Lolo") is deemed to include the aberrant Jinuo language of Xishuangbanna, Yunnan.[17] The Naxi/Moso language is quite close to LB, but stands somewhat outside the core of the family.[18]
- The mysterious Tujia language of Hunan and Hubei (not mentioned in *STC*) has so far not been assigned to a subgroup.

[15]A supergroup called "Rung" was proposed by Thurgood (1984), into which he placed, among others, some Qiangic languages, Nungish, and Lepcha. This grouping was based partly on shared "protomorphosyntax", and partly on nomenclature, including the *-rong* of *rGyalrong*, the Nungish language *Rawang*, and the Lepcha autonym *Rong*.

[16]The *Linguistic Survey of India* (Grierson and Konow, 1903-28) recognized a "Bodo-Naga-Kachin" group, an idea revived by Burling (1983), 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 *HPTB*:393-4.

[17]Chinese scholars have further divided the Loloish languages of China into six nuclei, although no attempt is made in this volume to distinguish them. In a recent talk (Matisoff 2007b) I examined Loloish tonal developments and the fate of the PLB rhyme ***-a** in terms of this six-way grouping, with inconclusive results.

[18]I have grouped Naxi with Lolo-Burmese proper in a supergroup called "Burmo-Naxi-Lolo" (Matisoff 1991c). 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 Matisoff 1974, 1991c.

Still, a schema like Fig. 2 hardly does justice to the complexity of the problem of subgrouping the TB languages. In particular, the "Kamarupan" and "Himalayish" groupings are based more on geographical convenience than on strong constellations of similar characteristics.[19] More detailed subgroupings are certainly possible, as in STEDT Monograph #2,[20] which makes distinctions like the following:

Kamarupan

- Abor-Miri-Dafla (=Mirish)[21]
- Kuki-Chin
- Naga
  - · Konyak (=Northern Naga)
  - · Angamoid
  - · Central
  - · Eastern
  - · Southern
  - · Southwestern
- Meithei
- Mikir
- Mru
- Bodo-Garo (=Barish)
- Chairel

Himalayish

- Western (Bunan, Kanauri, Manchad/Pattani)

[19]Several scholars have objected to the term Kamarupan, largely on the grounds that it has distinctly Indo-Aryan connotations, which might irritate TB groups. See, e.g. R. Burling, "On *Kamarupan*" (1999; *LTBA* 22.2:169-71), and the reply by Matisoff, "In defense of *Kamarupan*" (1999; *LTBA* 22.2:173-82). The only alternative term suggested so far to refer to these geographically contiguous languages collectively is the verbose "TB languages of Northeast India and adjacent areas".

[20]J. Namkung, ed. (1996), *Phonological Inventories of Tibeto-Burman Languages*, pp. 455-457.

[21]A well-defined subgroup of AMD has been dubbed *Tani* by J. Sun (1993).

- Bodic (Tibetanoid)
- Lepcha
- Tamangic (incl. Chantyal, Gurung, Tamang, Thakali, Manang, Narphu)
- Dhimalish
- Newar
- Central Nepal Group (Kham, Magar, Chepang, Sunwar)
- Kiranti (=Rai), including Bahing and Hayu

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.[22] At the present state of our knowledge we can estimate that the Tibeto-Burman family contains approximately 250 languages, which may be broken down into population categories as indicated in Table 1: there are 9 TB languages with over a million speakers (Burmese, Tibetan, Bai, Yi (=Lolo), Karen, Meithei, Tujia, Hani, Jingpho), and altogether about 50 with more than 100,000 speakers; at the other end of the scale are some 125 languages with less than 10,000 speakers, many of which are now endangered (JAM 1991b). Though much of the geographical area covered by TB languages has been chronically inaccessible to fieldwork by scholars from outside,[23] there has been a recent explosion of new data, especially from

[22]See JAM 1986a, and STEDT Monograph II (JAM 1996a).

[23]Very approximately, the distribution of TB languages by country is as follows: India 107, Burma 75, Nepal 69, China 50, Thailand 16, Bangladesh 16, Bhutan 9, Laos 8, Vietnam 8, Pakistan 1.

China[24] and Nepal.

As far as subgrouping this unruly conglomerate of languages goes, Benedict wisely refrained from constructing a family tree of the conventional type, presenting instead a schematic chart where Kachin (=Jingpho) was conceived as the center of geographical and linguistic diversity in the family. See Figure 1:Schematic Stammbaum of Sino-Tibetan Languages [STC, p. 6] The genetic schema now being used heuristically at the STEDT project differs from this in several respects.[25] See Figure 2:Provisional STEDT Family Tree

- 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 STC , 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-Naga group has provisionally been 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.[26]
- The important Qiangic languages (deemed to include rGyalrong [=Gyarung=Jiarong] and the extinct Xixia [=Tangut]) were hardly known to non-Western scholars at the

[24]Among the most valuable of these new sources are Sun Hongkai, Xu Jufang et al. (ZMYYC; 1991), containing 1004 synonym sets in 52 languages and dialects; and Dai Qingxia and Huang Bufan (TBL; 1992), with 1822 synonym sets in 50 languages and dialects.

[25]The STEDT project' s working hypotheses regarding the subgrouping of individual languages may be found in the indices to STEDT Monograph III (J. Namkung, ed. 1996:455-7).

[26]Issue has been taken with this term by Burling (1999), but see the reply by JAM (1999c).

time STC 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 many Chinese scholars maintain.

- The Nungish and Luish languages are grouped with Jingpho (=Kachin).[27] Jingpho is also recognized to have a special contact relationship with the Northern Naga (=Konyak) group.
- The somewhat idiosyncratic Mikir, Meithei (=Manipuri), and Mru languages are included under Kamarupan.
- The Himalayish (=Himalayan) group is considered to include Bodic (i.e. Tibetanoid) languages, as well as Kanauri-Manchad, Kiranti (=Rai), Lepcha, and Newar. [28][29]

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 Chi-

[27]The obscure Luish group, also known as Kadu-Andro-Sengmai, includes a few languages spoken by groups that were once exiled to a remote corner of NE India by the Rajah of Manipur. See Grierson 1921.

[28]As part of a recent trend to purge TB language names of Indo-Aryan suffixes, specialists in Himalayish languages are no longer using the name "Newari" for this language, but rather "Nepal Bhasha" or simply "Newar" . Similarly, the language known formally as Magari is now preferably referred to as "Magar."

[29]Various other subgroupings have been proposed, e.g. "Rungic" (Thurgood 1984) and "Sino-Bodic" (van Driem 1997). See a critique of the latter by JAM (2000b).

nese and Indian spheres of cultural influence as the "Sinosphere" and the "Indosphere".[30] 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 Kamarupan 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 (Ramsey 1987, Norman 1988). In early times the scattered Chinese communities of the region must have been on a numerical and cultural par with the coterritorial non-Chinese populations, with borrowing of material culture and vocabulary proceeding in all directions (Benedict 1975; Mei and Norman 1976; Sagart 1990). As late as the end of the first millennium A.D., non-Chinese states

[30]See JAM 1990a (``On megalocomparison.")

flourished on the periphery of the Middle Kingdom (Nanchao and Bai in Yunnan, Xixia 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.[31] 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.

Teleo- and meso-reconstructionsthe current state of comparative/historical TB research is quite uneven. While some branches of the family are relatively well studied, to the point where "mesolanguages" have been reconstructed at the subgroup level,[32] large gaps remain we have nothing approaching well-worked out reconstructions for such key subgroups as Qiangic, Baic, Luish, and Nungish. Still unclear is the exact genetic position of many transitional languages like Chepang, Kham, Lepcha, Newar (all lumped currently with "Himalayish"), or Meithei, Mikir, Mru (close to the Kuki-Chin-Naga branch), or Naxi/Moso and Jinuo (close to Lolo-Burmese), or the mysterious Tujia of Hunan/Hubei. The position of the crucially important Jingpho language is undergoing reevaluation, with current opinion returning to the notion of a special relationship with the Bodo-Garo-Konyak group (Burling 1971, Weidert 1987).[33] It remains to be seen whether the

[31]An excellent recent study of such phenomena is Thomason and Kaufman 1988.

[32]See, e.g. Proto-Karen (Haudricourt 1942-5, 1975; Jones 1961; Burling 1969; Solnit, in prep.); Proto-Bodo (Burling 1959); Proto-Lolo-Burmese (Burling 1968, JAM 1969, 1972a; Bradley 1978); Proto-Tamang-Gurung-Thakali (Mazaudon 1978); Proto-Kiranti (Michailovsky 1991); Proto-N.-Naga (W. French 1983); Proto-Tani [Mirish] (J.T. Sun 1993).

[33]Cf. the volume of Grierson and Konow (1903-28) called "Bodo-Naga-Kachin." Elsewhere (JAM 1974, 1991c) I have discussed the pros and cons of lumping Jingpho and Lolo-Burmese together into a supergroup facetiously called "Jiburish" (Jingpho-burmish-Loloish).

large "Kamarupan" (NE India) and "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.

Although it remains true that "supergroups within TB cannot safely be set up at the present level of investigation" (STC, p. 11), the same can be said of Indo-European (IE) after nearly 200 years of scholarly investigation. Thus while it is obvious that the closely related Baltic and Slavic languages constitute a valid IE supergroup, "Balto-Slavic" (just as, e.g. the Loloish and Burmish languages clearly group together as "Lolo-Burmese"), higher order IE lumpings (e.g. "Italo-Celtic", "Italo-Germanic", "Italo-Greek") remain highly controversial, since patterns of shared innovations, or overlapping features of special resemblance, may be found between virtually any two major subgroups of the family.[34]Meso-level reconstruction per se is not one of the goals of the STEDT project; nor does the project's reconstruction of PTB depend strictly on the direct comparison of meso-level reconstructions. However, such reconstructions are used when available in reconstructing roots at the Proto-Tibeto-Burman level. We therefore treat meso-level proto-forms as lexical data records, just like attested forms in individual languages. I follow Benedict in caring little for a chimerical methodological purity in this respect, and generally endorse his philosophy of "teleoreconstruction", by which salient characteristics of the proto-language may be deduced by inspection of attested forms in well-chosen languages from different subgroups, thereby "leap-frogging" the need for step-wise reconstruction.[35] This in fact has been the only practical methodology for reconstructing TB given the uneven state of our present knowledge. It goes without saying that one's teleo-hypotheses are subject to constant revision in the light of new data at the level of individual languages or subgroups.

[34]See the discussion in JAM (VSTB) 1978a:3-12.

[35]This 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. Benedict 1973 ("Tibeto-Burman tones, with a note on teleo-reconstruction").

As in all scientific inquiry, the process of formulating falsifiable hypotheses lies at the heart of the reconstructive enterprise. I feel that it is perfectly justifiable to "take a peek" outside a given subgroup in order to help one choose between alternative reconstructions that might be equally plausible on the basis of intra-group evidence alone.[36] It is for this reason that TB evidence will prove to be so crucial in evaluating the multitude of competing reconstructions of Old Chinese.

[[INSERT JAM PROSE]]

Language names

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.[37] 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,

[36]Many of the features of W. French's excellent reconstruction of Proto-N.-Naga (1983) were motivated by extra-Naga evidence.

[37]The terminology for the various types of TB language names was developed in Matisoff 1986a: “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 (1996) expanded into a STEDT Monograph, with the assistance of J.B. Lowe and S. P. Baron.

and Lashi[38] (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*. 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 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'.

[[LUISH > ASAKIAN?]]

Naturally enough, what is true for the names of individual languages is also true for the names of subgroups. Some of this nomenclatural variation goes back to differences between Benedict and his former collaborator and supervisor Robert Shafer,[39] e.g. Shafer's

[38]Referred to as Langsu, Zaiwa, and Leqi in Chinese sources.

[39]Shafer and Benedict collaborated on the Depression-era *Sino-Tibetan Linguistics* project at Berkeley (1939-40), which aimed to assemble all data then available on TB languages. The direct fruits of this project were Shafer's *Introduction to Sino-Tibetan* (1967-73), 5 vols. (Wiesbaden: Otto Harrassowitz) and the MS of Benedict's *STC*. Benedict produced (1975) an entertaining account of this seminal project in LTBA 2.1:81-92: “Where it all began: memories of Robert Shafer and the *Sino-Tibetan Linguistics* project, Berkeley

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*.[40] 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, etc.—and even Myanmar-Ngwi!

Bearing all these complicating factors in mind, an attempt has been made in this volume to use maximally clear and consistent designations for the TB languages and subgroups.

Meso-Reconstructions

Stepwise vs. teleo-reconstruction Both techniques are indispensable. Benedict (1972) 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. Thus final stops may sometimes be reconstructed for Proto-Karenic by looking elsewhere in TB. W.T. French (1983) used similar techniques in his reconstruction of some aspects of his Proto-Northern Naga. The main thing is to form precise and correctible hypotheses, i.e. hypotheses that can be tested.

In favorable circumstances (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:

(1939-40)."

[40]According to K. P. Malla (p.c. 2008), "*Kirãt* 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."

PIG &

#1006: PTB *p?ak PIG &

PKC *wok PIG PKC Proto-Kuki-Chin

PNN *wak PIG PNN Proto-Northern-Naga

TGTM *??wa PIG TGTM Tamang/Gurung/Thakali/Manang

PLB *wak? PIG PLB Proto-Lolo-Burmese

PKar *th???/ th&#65533;u? PIG PKar Proto-Karenic

PBai *te PIG PBai Proto-Baic

PTk *hwok PIG PTk Proto-Tangkhulic

FOUR #2409: PTB *b-l?y FOUR Proto-Tani *pri FOUR PKC *lii FOUR PNN *b? l?y FOUR TGTM *?bli FOUR PLB *b/?-l?y? FOUR PKar *lwi-t FOUR

STONE #1269: PTB *r-lu? ? k-luk STONE PTani *l?? STONE PKC *lu? STONE / ROCK PLB *k-lok ? k-lo? STONE PTk *lu? STONE

TREE/WOOD #2658: PTB *si? ? sik TREE, WOOD PTani *s?? WOOD / TREE PKC *thi? TREE / WOOD PTk *t?i? WOOD

EIGHT #2259: PTB *b-r-gyat ? (b-)g-ryat EIGHT PKC *riat EIGHT PLB *?-rit? EIGHT PTk *??t EIGHT

ASHES #3514: PTB *hot ASHES PKC *wut ASHES / DUST Proto-Asakian *k-but ASHES PTk *hwot ASHES

[[SHOULD THIS GO HERE?]]

Language contact: loanwords vs. cognates

We include reconstructions of a number of Indo-Aryan roots, if they've been borrowed into a large number of TB languages (esp. in the Himalayas and NE India). Even when the donor language is not related to the receptor TB language, there may well be an accidental phonological resemblance between the form in the donor language and a similar genuine TB root. To facilitate recognition of loanwords, we have adopted the practice of including e.g. Nepali roots (identified as such) in our list of etyma if they have been borrowed into several TB languages.

Special challenges for Sino-Tibeto-Burman etymologization

- Relative lack of anciently attested languages: only Chinese, Tibetan, Tangut (Xixia), Burmese, Meithei, Newar, and a few others.
- Monosyllabicity 'as opposed, e.g., to Austronesian, with juicy disyllabic roots (trisyllables if one follows Benedict's 'Austro-Tai' (1975).
- Unevenness of our knowledge with respect to the various subgroups of Tibeto-Burman

Subgrouping issues:

- Phonologically eroded forms in many subgroups

This presents problems common to reconstruction in other mono- or sesqui-syllabic languages of mainland Southeast Asia (Hmong-Mien, Tai-Kradai, Mon-Khmer).

There are great differences among the subgroups with respect to degree of preservation of proto-entities. Contrast, e.g. Written Tibetan and Kuki-Chin on the one hand, to Loloish, Asakian, and Qiangic on the other. So it's good to find languages which erode differently

(e.g., Hmong preserves initials better, while Mien preserves finals better; Kadu preserves finals better while Sak preserves initials better). This differential preservation makes it exciting when you can show that WT brgyad and Lahu h&#65533; are perfect cognates for EIGHT.

Phonological erosion makes it tough to distinguish between phonologically similar etyma in the same semantic field: LOUSE *sar and *s-r(y)ik. What do you do with a form like ?? ? MIND/BREATH *sem and *sak MOUTH *mu?r and *muk FACE *s-ma?y and *s-mel As an extreme example, we have reconstructed no fewer than 11 etyma meaning NECK/THROAT with initial velars or velar clusters:[41] 1 #389 *s-gwa-n ? *r-gwa-n NECK/NAPE (probably related to #495 *kwa THROAT/NECK); #481 *s-ke-k ? *m-ke-k NECK; #486 *l-kok THROAT/NECK; #488 *k-ro THROAT; #489 *k/s-rok ? *k-ro? THROAT; #491 *gre-k THROAT; #493 *kak THROAT; #494 *ka THROAT/NECK; #3361 *go? NECK/THROAT; #5651 *ku NECK/THROAT. Some of these putative etyma may ultimately be combinable (e.g. #389 and #495, #481 and #491).[42] It is necessary to constantly rectify one's reconstructions in the light of new data (see below XYZ), and equally imperative to try to establish 'sound laws' by finding parallel examples (see below XYZ).

Sources and Source Abbreviations

[[INSERT JAM TEXT]]

[41]This is perhaps further confirmation of the frequently noted tendency of words for these parts of the body to have 'guttural' initials.

[42]This classic problem of distinguishing co-allofams from reflexes of separate roots has bedeviled Tibeto-Burmanists from the beginning. Benedict (1972) 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 (JAM 1985) actually reflect the same etymon, so I would now combine CATTLE/DOMESTIC ANIMAL *dzay (#129) with ELEPHANT/CATTLE *tsa?y (#143). In 1988 I tried to show that two homophonous roots set up in Benedict 1972 (*dyam STRAIGHT and *dyam FULL/FILL are actually one and the same root. For many similar examples, see JAM 2013.

Source abbreviations

[["true, in this volume, but not generally; even when identical]]

Each supporting form is ascribed to a particular source. Many forms are cited in more than one source in our database. If the form is not identical in different sources, we include them all. This is especially useful in cases where one or more of the sources might not be totally accurate phonemically, or where subphonemic phonetic detail is provided. When the forms in different sources are identical, the form only appears once, but there are multiple source abbreviations, separated by commas. Forms from well-studied languages (e.g. Written Tibetan, Written Burmese, Jingpho) are likely to appear in several sources used by 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 reconstructions provided), e.g. Luce 1986 (PPPB); Sun Hongkai et al. 1991 (ZMYYC); Dai Qingxia, Huang Bufan et al. 1992 (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, e.g.:

    Proto-Bodo: Burling 1959

    Proto-Karenic: Haudricourt 1942-45/1975, Jones 1961, Burling 1969, Benedict 1972 (*STC*), Solnit, in prep.

    Proto-Kiranti: Michailovsky 1991

Proto-Kuki-Chin: VanBik 2003

Proto-Lolo-Burmese: Burling 1968, Matisoff 1969/1972, Bradley 1979

Proto-Northern-Naga: French 1983

Proto-Tamangic: Mazaudon 1978

Proto-Tani: Sun Tianshin 1993

The abbreviations used in these source attributions are in general quite transparent,[43] e.g.:

CK-YiQ	Chen Kang, "Yi Questionnaire"
JZ-Zaiwa	Xu Xijian, *Outline Grammar (Jiǎnzhì) of Zaiwa*
AW-TBT	A. Weidert, *Tibeto-Burman Tonology*
GHL-PPB	G. H. Luce, *Phases of Pre-Pagán Burma*
JAM:MLBM	J. A. Matisoff, "Mpi and Lolo-Burmese microlinguistics"
EJAH:BKD	E. J. A. Henderson, *Bwo Karen Dictionary*

The abbreviation "JAM-Ety" refers 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*.

Supporting forms in the individual languages

[[margin notes: "paper is not a problem electronically"; maryama fix for absolutely identical ones]]

The forms which support the reconstructions are cited according to the notation of the particular source. Although this policy of "following copy" often leads to redundancy (see 2.7 below), since one and the same form in a given language may be transcribed in

[43]For a complete list of the source abbreviations that appear in this volume, see the *Appendix*.

a variety of different ways,[44] it seems preferable to a policy of "normalization", which might have the effect of losing some phonetic detail that is captured in one source but not in another.

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. 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.

Organization of the etymologies

[[INSERT JAM PROSE]]

Supporting forms in the individual languages

The forms which support the reconstructions are cited according to the notation of the particular source. Although this policy of "following copy" often leads to redundancy (see 2.7 below), since one and the same form in a given language may be transcribed in

[44]Cf. the multiple transcriptions of the Written Burmese form for BREAST/MILK under ***s-nəw**, #53 below: **no**[1] (ZMYYC:281.39); **nuí** (AW-TBT:327; *STC*:419); **núi** (WSC-SH:48); **nui'** (JAM-Ety; GEM-CNL; PKB-WBRD); **nui.** (GEM-CNL); **nuiw'** (GHL-PPB). For these source abbreviations, see the *Appendix*. Similarly, cf. the many slightly different forms from the Bodic and Tamangic groups that reflect the etymon ***tsaŋ** NEST/WOMB/PLACENTA (#103 below).

a variety of different ways,[45] it seems preferable to a policy of "normalization", which might have the effect of losing some phonetic detail that is captured in one source but not in another.

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. 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.

Source abbreviations

Each supporting form is ascribed to a particular source. Many forms are cited in more than one source in our database. If the form is not identical in different sources, we include them all. This is especially useful in cases where one or more of the sources might not be totally accurate phonemically, or where subphonemic phonetic detail is provided. When the forms in different sources are identical, the form only appears once, but there are multiple source abbreviations, separated by commas. Forms from well-studied languages (e.g. Written Tibetan, Written Burmese, Jingpho) are likely to appear in several sources used by STEDT.

[45]Cf. the multiple transcriptions of the Written Burmese form for BREAST/MILK under ***s-nəw**, #53 below: **no**[1] (ZMYYC:281.39); **nuí** (AW-TBT:327; *STC*:419); **núi** (WSC-SH:48); **nui'** (JAM-Ety; GEM-CNL; PKB-WBRD); **nui.** (GEM-CNL); **nuiw'** (GHL-PPB). For these source abbreviations, see the *Appendix*. Similarly, cf. the many slightly different forms from the Bodic and Tamangic groups that reflect the etymon ***tsaŋ** NEST/WOMB/PLACENTA (#103 below).

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 reconstructions provided), e.g. Luce 1986 (PPPB); Sun Hongkai et al. 1991 (ZMYYC); Dai Qingxia, Huang Bufan et al. 1992 (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, e.g.:

  Proto-Bodo: Burling 1959

  Proto-Karenic: Haudricourt 1942-45/1975, Jones 1961, Burling 1969, Benedict 1972 (*STC*), Solnit, in prep.

  Proto-Kiranti: Michailovsky 1991

  Proto-Kuki-Chin: VanBik 2003

  Proto-Lolo-Burmese: Burling 1968, Matisoff 1969/1972, Bradley 1979

  Proto-Northern-Naga: French 1983

  Proto-Tamangic: Mazaudon 1978

  Proto-Tani: Sun Tianshin 1993

The abbreviations used in these source attributions are in general quite transparent,[46] e.g.:

[46]For a complete list of the source abbreviations that appear in this volume, see the *Appendix*.

CK-YiQ	Chen Kang, “Yi Questionnaire”
JZ-Zaiwa	Xu Xijian, *Outline Grammar (Jiănzhì) of Zaiwa*
AW-TBT	A. Weidert, *Tibeto-Burman Tonology*
GHL-PPB	G. H. Luce, *Phases of Pre-Pagán Burma*
JAM:MLBM	J. A. Matisoff, “Mpi and Lolo-Burmese microlinguistics”
EJAH:BKD	E. J. A. Henderson, *Bwo Karen Dictionary*

The abbreviation “JAM-Ety” refers 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*.

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 Proto-Sino-Tibetan stage. For all of these comparanda Zev J. Handel has kindly provided comparisons of the Old Chinese reconstructions cited in Karlgren's (1957) system with those of Li Fang-kuei (1971, 1976, 1980) and William Baxter (1992),[47] evaluating the plausibility of the putative TB/Chinese comparison according to each of these systems.[48] Handel's invaluable contributions are marked with his initials “ZJH”. Comparisons between TB and OC etyma that are not explicitly ascribed to a particular scholar are original with me, as far as I know.

[47]Handel also contributed a detailed comparison of these systems in his *A Concise Introduction to Old Chinese Phonology*, which appeared as Appendix A to *HPTB*, pp. 543-74.

[48]Handel also frequently refers to several other reconstructive systems for OC that are to be found in the literature, e.g. those of W. South Coblin (1986), Axel Schuessler (1987), Laurent Sagart (1999), Gong Hwang-cherng (1990, 1994, 1995, 1997, 2000), and Pan Wuyun (2000).

Notes

Footnotes may appear at virtually any point in the text. They may refer to an entire chapter, to a semantic diagram, to an etymon as a whole, to a specific supporting form, or to a Chinese comparandum.

Semantic diagrams

Semantic flowcharts

[[INSERT JAM TEXT]] Each chapter begins with a semantic diagram. These diagrams, called "metastatic flowcharts" in STEDT parlance, were first introduced in Matisoff 1978a (*VSTB*), and have been used subsequently in many of my articles.[49] 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.[50] I rely on three basic types of evidence:[51]

1. *Synchronic intra-lingual vagueness.* A given daughter language has a single form that means X or Y according to context, e.g. Mikir **artho** means either 'blood vessel' or 'tendon' or 'muscle' or 'nerve'. In many Chin languages reflexes of ***m-luŋ** can mean either HEART or LIVER.

[49]See, e.g., Matisoff 1980 ("Stars, moon, and spirits"); 1985a ("God and the ST copula"); 1985b ("Arm, hand, wing"); 1988b ("Property, livestock, talent"); 1991a ("Grammatization in Lahu"); 1991b ("Mother of all morphemes"); 1994b ("Buttock and dull"); 2000b ("Three TB word-families"); 2004 ("Areal semantics").

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

[51]See the discussion in *VSTB*: 194-200.

2. *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' > Mikir **arkleŋ** 'marrow', Dimasa **buthluŋ** 'brain';

   PTB ***s-p^{w}ik** 'bowels/stomach' > Mikir **phek** 'bowels', Lahu **ɔ̀-*ɬɨ́*-qō** 'stomach'.

3. *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 **khɨ-mêʔ-šī** < **khɨ** 'foot' + **mêʔ-šī** 'eye'); similarly Indonesian **mata-kaki** 'ankle' (< **mata** 'eye' + **kaki** 'foot'), which establishes the association EYE ⟷ ANKLE[52]

Certain conventions are observed in the metastatic flowcharts of this volume:

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

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

[52]The same formation is found in many other TB languages, e.g.:

	FOOT	*EYE*	*ANKLE*
Lalung	**ia-thong**	**mi**	**ia-thong-mi**
Limbu	**lāŋ**	**mik**	**lāŋ-mik**
Lushai	**ke**	**mit**	**ke-mit**
Meithei	**khu**	**mit**	**khu-mit**
Tangkhul	**phei**	**mik-ra**	**phei-mik-ra**
Written Burmese	**khre**	**myak-ci'**	**khre-myak-ci'**

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

4. 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). This volume also includes some non-bodypart terms which frequently appear in compounds with etyma referring to the reproductive system. See especially Ch. IX, “Body fluids”.

Deciding how much semantic latitude to allow among putative cognates is definitely an art rather than a science. Here as elsewhere a middle-of-the-road approach is necessary, neither overly conservative nor too wildly speculative. As a positive example of a promising new etymology involving a semantic leap, we may offer ***m-t(s)i** 'salt / yeast' [*HPTB* 3.3.1]. Although forms in the daughter languages sometimes mean 'salt' and sometimes 'yeast', the phonological correspondences between both semantic groups of forms are good. On the other hand, the semantic association between 'salt' and 'yeast' has yet to be

attested in other language families, even though it has great initial plausibility. Both are efficacious substances that have dramatic effects on the taste of food or drink; their lack renders the food or drink insipid.[53]

In all language families, semantic associations/variation is 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.[54] Other associations seem to be confined to a particular family, e.g. YEAST and SALT, so far discovered only in TB.[55]

The vagaries of semantic history occasionally result in strange allofamic bedfellows, e.g. Eng. black and blank, from the same PIE root *bhel- (see above), or Eng. science and shit < *skei- CUT/SPLIT (AHD 58-9).

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 megalo-comparative 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 (2001), in which he claims genetic connections between pairs of words like the following: Sumerian Written Tibetan gim 'axe' h?gem-pa 'destroy' alad 'protective demon' lad-mo 'imitation' ?inig 'tamarisk' snyigs-pa 'degenerated; grown worse' pe? 'three' gsum 'three'

Benedict 1939 was a pioneering article in the field of Southeast Asian areal semantics. At STEDT we have tried to make our semantic speculations as reasonable as possible. Each word in our lexical database is accompanied by a symbol which indicates the se-

[53]Yeast is used for brewing liquor rather than for baking bread in East and SE Asia.

[54]See JAM 1987 (VSTB), pp. 182-3, 203.

[55]See 2003a (HPTB:34), and JAM 2003b. 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 m?ts, Lahu d 'yeast' (Lahu voiced initials come from *prenasalized ones).

mantic field(s) to which it belongs. Within each point in semantic space, we attempt to trace the patterns of association for which there is evidence, by means of diagrams called 'metastatic flowcharts' in our jargon.3 FIGURES 7A and 7B are two attempts to indicate the attested semantic associations of SKIN in TB/ST.

[[INSERT SKIN GRAPHICS]]

Phonological system of Proto-Tibeto-Burman

The PTB syllable canon: HPTB, p. 12 / TBRS p.xxvi TBRS:xxvi ❖The reconstructions all conform...can be set up for TB.❖ An ongoing process -- am now rethinking the mid-vowels -e(-) and -o(-)

❖Importance of having a template or syllable canon for etyma. (See JAM 2003:12)

[T] (P?) (P?) Ci (G) V (?) (Cf) (s)

Clarification of STEDT❖s reconstructions, especially with respect to prefixes. The Index of HPTB lists the naked roots, stripped of prefixes and suffixes. One must look in the text to find the details. There has been an inconsistency in noting the nasal prefix: *m- or *N- or *n-. This will be regularized during the ❖rectification❖ process (below VII).

❖Not all morphological elements in our formulas are claimed to be of equal antiquity. The only claim is that the item has been attested someplace in the family at some time. The rationale for setting up such an element (e.g. a prefix) is generally explained in an ❖etymon note❖ at the beginning of an etymology, E.g., in *m-p(r)(w)a? MOUTH: the *m- is only directly attested in Qiangic, Tani, and Southern Chin.

Reconstructed PTB etyma

After the semantic chart which begins each chapter, the reconstructed PTB roots of the chapter are presented one after the other, roughly in the order of the strength of their attestation. After preliminary remarks about the distribution of the etymon, the "supporting forms" for the reconstruction are listed, subgroup by subgroup.

The reconstructions all conform to the syllable canon posited for the proto-language,[56]

(P^2) (P^1) C_i (G) V (:) (C_f) (s),

where the initial consonant (C_i) may be preceded by up to two prefixes (with the inner prefix P^1 assumed to be historically prior to the outer one (P^2); the C_i may optionally be followed by one of four glides (G), */-y-, -r-, -w-, -l-/, and the vowel, which may be long (:), may be followed by a final consonant (C_f); if the syllable does contain a C_f, it may also (although quite rarely) end with suffixal -s. It should be noted that many daughter TB languages have much simpler canons, e.g. Lahu, where native syllables consist maximally of an initial consonant, a vowel, and a tone:

T

(C_i) V

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.

Reconstructions at the subgroup level (i.e. "meso-reconstructions" like Proto-Lolo-Burmese (PLB), Proto-Northern-Naga (PNN), Proto-Tani) are listed as individual records along with their supporting forms.

A few notational conventions with respect to my PTB reconstructions should be mentioned:

[56]See *HPTB*, pp. 11-13.

- Variant reconstructed forms are indicated in several ways. They are usually written with the "allofam symbol" ⩙ between them, e.g.: ***glim** ⩙ ***glip** BROOD / INCUBATE; ***s-riŋ** ⩙ ***s-r(y)aŋ** LIVE / ALIVE / GREEN / RAW / GIVE BIRTH. Sometimes, however, I use an alternative notation with parentheses, e.g.: ***(t)si** COPULATE/LOVE; this is equivalent to ***si** ⩙ ***tsi**. Slashes may also be used, e.g. ***p/buk** ⩙ ***p/bik** BORN/GIVE BIRTH; this is equivalent to ***puk** ⩙ ***buk** ⩙ ***pik** ⩙ ***bik**. Finally, still another way of indicating proto-variation is by means of a "vertical reconstruction", e.g.:

  * **t** / **d** **uŋ** NAVEL. This means the same as ***tuŋ** ⩙ ***duŋ.**

- Parentheses are especially appropriate for those frequent cases where there is variation or indeterminacy between dental and palatal fricates; in fact that is one of my principal motivations for writing the palatal series as sequences of dental plus **-y-**, rather than writing them with *hačeks* or grave accents, e.g.:

  ***ts(y)uːŋ** NAVEL / CENTER (= ***tsuːŋ** ⩙ ***tšuːŋ**)

  ***s(y)ok** DRINK / SUCK / SMOKE (= ***sok** ⩙ ***šok**)

- Etyma which show variation between initial ***p-** and ***w-** are reconstructed with the morphophonemic symbol ***p^{w}-,** which is roughly equivalent to treating the stop element as a prefix (***p-w-**).[57] Thus, a reconstruction like ***p^{w}u** EGG / BIRD / ROUND OBJECT implies the existence of two sub-roots, ***pu** and ***wu,** whatever the ultimate explanation for this variation might prove to be.

- In the original version of Benedict 1972 (henceforth *STC*, ca. 1943), Benedict reconstructed two PTB high long vowels ***-iy** and ***-uw,** contrasting with the much less frequent short high vowels ***-i** and ***-u.** In the published version (1972) he modified the reconstruction of these long vowels to ***-əy** and ***-əw,** a practice which I follow

[57] For extended discussion of this issue, see Matisoff 2000a.

myself. Occasionally, however, when the evidence does not permit us to decide between a long and a short high proto-vowel, it is convenient to revert to the earlier notation, with parentheses, e.g. ***b-ni(y)** 'petticoat' (*STC* #476); ***sru(w)** 'relative' (*STC* p. 108). There are no such cases among the etyma in this volume, however.

For more discussion of variational patterns in PTB, see "Regularity and variation", section 3.1 below.

Many of the PTB etyma in this volume are here reconstructed for the first time in print, and a good number of the TB/Chinese comparanda are likewise here proposed for the first time. If references are not explicitly given to *STC* and/or *HPTB* in the introductory note for an etymon it may be assumed that the reconstruction is new.[58]

Regularity and variation

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 (e.g. ***sem** and ***sak** 'mind / breath'; ***muːr** and ***muk** 'mouth'; ***s-maːy** and ***s-mel** 'face'; ***s-r(y)ik** and ***s(y)ar** 'louse'), it is not easy to decide by simple inspection to which etymon we should assign a phonologically slight form in a daughter language (e.g. **sɒ** 'mind', **mɔ** 'mouth', **hmɛ** 'face').

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 recon-

[58]References to HPTB as labelled with "**H:**" followed by a page number, e.g. **(H:165)** ***wa ≍ *wu** BIRD / FOWL means that the root is discussed chiefly on page 165 of *HPTB*.

structed, 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.[59] The concept of "regularity" itself is by no means simple, nor does it mean the same thing to different scholars.[60]

Those who lack what I have called "Proto-Sprachgefühl"[61] can produce abstract, formulaic reconstructions bristling with strange symbols but devoid of any phonetic or typological plausibility.[62] 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.[63] 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

[59]The computer can be very useful in deciding between alternative etymologies. Once "sound-laws" have been formulated, computer checking can test whether a particular reconstruction follows the laws, identifying inconsistencies in the reflexes of the same proto-element in a given language. Such a methodology has been applied to the Tamangic languages, using the "reconstruction engine" developed by J.B. Lowe at STEDT in collaboration with Martine Mazaudon and Boyd Michailovsky during their sojourns at Berkeley as visiting scholars (1987-89, 1990-91).

[60]See Matisoff 1992 ("Following the marrow") and 1994a ("Regularity and variation").

[61]See Matisoff 1982.

[62]Recent examples of this genre include Sedláček 1970; Weidert 1975, 1979, 1981, 1987; Peiros & Starostin 1996; Sagart 2007.

[63]See Matisoff 1990a ("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.

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".[64]

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".[65] In Matisoff 1978 (*VSTB*) 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.[66]

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

[64]This is actually the proto-form offered in Weidert 1981:25 for an etymon meaning 'spirit, ghost, shadow' (reconstructed as ***m-hla*** in *STC* #475). As I have observed (Matisoff 1982:22), "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'."

[65]See the pioneering study of Karlgren (1933), "Word families in Chinese".

[66]This 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.

way.[67] 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 2.2), but the type of variation posited must be abundantly replicated in other examples. This volume does not attempt to conceal such uncertainties, but frequently entertains the possibility that etyma set up as independent might actually be co-allofams, or *vice versa*.

The best attested patterns of variation in ST/TB are all exemplified in the etymologies of this volume. They include the following:

1. Voicing vs. voicelessness of the initial consonant:[68]

   ***gop ≍ *kop** (11a) HATCH/INCUBATE/COVER

   ***prat ≍ *brat** (75) BREAK/WEAN

   ***tuŋ ≍ *duŋ** (44a) NAVEL

2. Variation between fricative and affricate:

   ***(t)sum** (45) NAVEL

   ***(t)sip ≍ *(t)sup** (107) NEST/WOMB/SCROTUM[69]

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

   ***b(y)at** (81) VAGINA

   ***l(y)ap** (151) COPULATE

   A special case of (c) is the alternation between dental and palatal fricatives and affricates:

[67] See the extended discussion in Ch. XII of *HPTB* (pp. 491-534), "Allofamic variation in rhymes".

[68] Nothing 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, where such variation in manner is quite rare, and is usually not tolerated in PIE reconstructions.

[69] This etymon also illustrates (f), below.

*s(y)ok (61) DRINK/SUCK/SMOKE

*dz(y)əw (56) BREAST/MILK

*ts(y)uːŋ (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-**:

*s-riŋ ≋ *s-ryaŋ (39) LIVE/ALIVE/GREEN/RAW/GIVE BIRTH

*s-nik ≋ *s-nyak (124) PENIS/COPULATE

*b-rim ≋ *b-ryam (46) NAVEL/UMBILICAL CORD

8. Alternation between medial **-wa-** and **-u-**:

*tsyul ≋ *tsywal (105) WOMB/PLACENTA

9. Alternation between final homorganic stops and nasals:

*glim ≋ *glip (15) BROOD/INCUBATE EGGS

*s-nəwn ≋ *s-nəwt (53c) BREAST/MILK/SUCK

*tsiŋ ≋ *tsik (78) VAGINA

As some of the above examples illustrate, some roots show more than one type of varia-

tion. When a posited allofamic reconstruction (e.g. ***sir*** ≍ ***sit*** (6) EGG) does not fall into a well-attested variational category, I comment on it. Handel makes similar remarks with respect to some of my TB comparisons with OC.

Occasionally, when the phonosemantic variation among the allofams is considerable, and when each variant is amply attested, I split up the presentation of the data into subroots that are designated by the same number but with different lower case letters, e.g.: ***p-wu*** (1) EGG is split into ***wu** (1a) and ***pu** (1b); ***m/s-la(ː)y*** ≍ ***s-tay*** (40) NAVEL/CENTER/SELF is split into ***m/s-la(ː)y*** (40a) and ***s-tay*** (40b); ***m-ley*** ≍ ***m-li*** ≍ ***m-ney*** (114) PENIS is broken down into ***m-ley*** ≍ ***m-li*** (114a) and ***m-ney*** (114b).

As I put it 35 years ago, “We must steer an Aristotelian middle path between a dangerous speculativism and a stodgy insensitivity to the workings of variational phenomena in language history.”[70]

Regularity and exceptions

Exceptionlessness vs. parallel examples.1 How many parallel examples do we need to establish ﬧregularityﬧ? Are exceptions more to be tolerated in ﬧcoreﬧ than in ﬧpe-ripheralﬧ vocabulary? In very favorable circumstances we might find several parallel examples that prove the validity of a correspondence that at first glance seemed most unlikely. See Figure XYZ.

FIGURE: Lahu Reflexes of *-?y after complex laterals

Old Burmese Lahu Akha Jingpho WT Proto-TB ﬧwindﬧ le mﬧ-h? b??-l? rdzi *g-l?y ﬧboatﬧ hle h?-lﬧ?-q? l? *m-l?y ﬧheavyﬧ lﬧ h? ve l? l?i *s-l?y-t ﬧgrandchildﬧ mliy > mrﬧ h?-? ?-pﬧ m?l? ﬧyoung manﬧ *b-l?y ﬧfourﬧ lﬧ ? ? m?l? b?i *b-l?y

[70]Matisoff 1972b (“Tangkhul Naga and comparative TB”), p. 282.

�bow/sling� l� h?-ma ca-? l?l? g?u 2 *d/s-l?y

Problem: How to deal with Lahu ?, which lacks initial h-? Parallelism of variational patterns is a big plus, both within a language and across languages. See FIGURE XYZ: FIGURE: Parallel patterns of variation in Lolo-Burmese and Chinese: SCATTER/POUR

PST/PTB *sywar > OC *s�n ? *s�t and PLB *swan ? *swat (1) PTB *sywar �flow/pour/scatter� WT h?tshor-ba �escape; flow out, run over�; Lepcha t?hor �the pouring of water; Garo sol-a? �flow�; Dimasa di-sor �id.�; Jingpho ??n �flow (as tears, sweat, water poured on ground)� (2) PLB *swan ? *swat *swan?/? > WB swan �pour out, spill, shed�, sw�n �pour upon, cast by pouring liquid into a mold�; Lahu ?? �pour; sow broadcast; Akha s? �sow seeds, sj? �pour�; Mpi se? �sow broadcast� *swat > Lahu ?�? �pour, spill�; Akha sj?q; Sani s???; Bisu ??t (3) OC *s�n ? *s�t *s�n > ? [GSR 156a-b] (Mand. s�n) �disperse� *s�t > ? [AD 767] (Mand. s?, s?) �scatter; let loose�

For the ST/TB variational pattern of homorganic final stops and nasals, see below XYZ.

Variational phenomena Allofam theory: since VSTB The three symbols: A is an allofam of B A is not an allofam of B is A and allofam of B?

�Isotopes are allofams of elements.�

. Patterns of phonosemantic variation in Tibeto-Burman vs. Indo-European3

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.

�Manners of initial consonants In ST/TB variation in the manner of articulation (voiced or not, aspirated or not) of initial consonants is commonplace, largely due to the effect

of prefixes. This is not so in Indo-European or Austronesian, where a voicing discrepancy can be a fatal objection to an etymology. ◆Vocalic alternations Conversely, vowel alternations (◆ablaut◆ or ◆gradations◆) are thoroughgoing and well-structured in IE, but quite sporadic in ST/TB, either confined to particular languages (e.g. Written Tibetan), or due merely to phonetic factors, without semantic content (e.g. alternation between medial -i- ? -u-, especially in the environment of a labial Ci and/or Cf. Even less phonetically motivated is the variation between -i- ? -ya- found in several roots, notably EYE *mik ? *myak. Very occasionally a single etymon exhibits both types of alternation, e.g. BODY HAIR *mul ? *mil ? *myal. ◆Variation between homorganic final stops and nasals This is common both in Chinese and TB, perhaps due to the influence of suffixes which have since disappeared; but this phenomenon is unknown in IE. See Fig. XYZ, above (SCATTER/POUR). ◆Variation in the prefixes attested with a given root This is extremely common in TB, partly because prefixes may arise from the reduction of the first syllables of compounds (see ANT in Fig. XYZ above), and also because certain TB languages seem to prefer a particular prefix to generalize throughout large parts of the lexicon. Consonantal prefixes play a very minor role in IE etymologies. The only example I can think of is the sigma mobile, a prefixal s- that appears with a certain root in some languages but not in others (e.g. SNOW: PIE *sneig?h-, which gives, e.g. Russian sneg, but Latin nix/nivis). [BUT THIS IS THE REGULAR DEVELOPMENT OF *SN- IN LATIN.] GIVE BETTER EXAMPLES < Dartmouth banker◆s box

◆Alternation between initial labial stop and w- In as many as 20 TB etymologies, some languages have a labial stop while others have w- (even though the ◆regular◆ correspondences are between stops and stops and semivowels and semivowels); e.g. PIG WT phag/WB wak. This phenomenon, which I have called ◆extrusional◆, occurs mostly before the vowel -a-.4 ◆Tonal variations Nothing is more common than for a ST/TB etymon to show tonal variation, either within a single language, or within a single subgroup, or

across subgroups. These variations may be highly structured, even rule-governed (with clear conditioning), or they may be quite sporadic. Nevertheless tones usually do correspond very well within a subgroup, so that tonal anomalies must receive special attention. Tonally impoverished IE, on the other hand, must do without this sort of correspondence (except perhaps with respect to the pitch accents of Greek and Sanskrit, which sometimes do correspond regularly).

Here is a PAF (in vertical format) which illustrates several of the types of variation just mentioned: BELLY/STOMACH p *s- u b (?) (k) *?- i w

�Root extensions/augments in final position One variational feature shared by both TB and IE is the phenomenon of extra consonants tacked onto the end of a root, with no clear reason for it. First a couple of examples from TB:

TONGUE/LICK (w)

*m- (y) l (y) a (?)

s- (m)

(k)

FIRE/BURN/KINDLE/ROAST5

*p r

w l b a y m, n, ?

h t

A very similar example from Indo-European involves the morpheme *wer- TURN/BEND, called �the conventional basis of various Indo-European roots� (AHD: 76-77). See Figure XYZ.

FIGURE: *wer- TURN/BEND �conventional basis of various Indo-European roots� AHD

76-77

I. *wert- > -WARD (forward, toward, etc.) / Lat. vert- (convert, revert, divert, etc.)

II. *wreit- > WREATH / WRITHE / WRATH

III. *wergh- > WORRY / WRING

IV. *werg- > WRENCH / WRINKLE

V. *wreik- > WRY / WRIGGLE / WRIST / WRESTLE

VI. *wrib- > RIBALD

VII. *werb- > WARP

VIII. *werp- > WRAP / RHAPSODY

IX. *wr?mi- > WORM

V. More on the concept of word family: blood relatives vs. adoptive ones

According to the broad definition of "word family" adopted at STEDT, a family may include "blood relatives" (i.e. native words directly descended from an earlier stage of a given language), as well as words borrowed from a related language ("adoptive relatives"). It is the task of the analyst to distinguish these two types of family members. In TB, this problem is often acute, e.g. with respect to the Baic languages of Yunnan, which have been overwhelmed with different strata of loanwords from Chinese over the millennia. Indo-European also has countless examples of this mixed type of word family, as in Figure XYZ.

FIGURE: English reflexes of the PIE etymon *wed- ? *wod- ? *ud- WATER (adapted from AHD 73)

(a) Inherited Germanic material 1. *wod-?r- [suffixed o-grade] > PGmc *watar > OE w"tar > WATER 2. *w?d-o- [suffixed lengthened grade] > PGmc *w?d- > OE w?t,

w?t > WET 3. *wod- [o-grade] > PGmc *wat-skan > OE w�scan, wacsan > WASH 4. *we-n-d- [nasalized form] > PGmc *wintruz �wet season� > OE winter > WINTER 5. *ud-ro, *ud-r?- [suffixed zero-grade] > PGmc *otraz > OE otor > OTTER

(b) Borrowings from other Indo-European languages 6. *ud-?r- [suffixed zero-grade] > Greek hud?r �water� > HYDRO- (including CLEPSYDRA, DROPSY) 7. *u-n-d-?- [suffixed nasalized zero-grade] > Latin unda �wave� > UNDULATE, INUNDATE, ABOUND, REDUNDANT, SURROUND 8. *ud-skio- [suffixed zero-grade] > Scot. and Ir. Gaelic uisge �water� > uisquebaugh �eau de vie� > WHISKEY 9. *wod-?- [suffixed o-grade] > Russ. voda �water�, with -ka �diminutive� > VODKA

Rectification/revision of reconstructions

As of January 2014, about XYZ

We have nothing to apologize for in changing our reconstructions (look at all the revisions of Baxter/Sagart!). In our case, we�re not just rechewing the same old data in ever more implausible ways, but rather trying to accommodate all the new data that keep pouring in. In cases where the reconstructions in this Final Product differ from those in HPTB or other previous publications, the FP wins (providing the etymon has been properly rectified).

Etymological accuracy and rectification of possible errors TBRS, Section 3.2, pp. xxxviii-xl HPTB, pp. 538-40

Etymological accuracy and rectification of possible errors

There are all too many ways in which one can make etymological mistakes, and I have been guilty of all of them at one time or another.[71]A rough taxonomy of errors would

have to include the following:

- Treating a loanword as native

  I was at first delighted when I ran across the Jingpho form **wé?-wū** 'screw', since its first syllable looked like an excellent match with Lahu **ɔ̀-vɛ̀?** 'id.', for which I then had no etymology. Could this be a precious example of the rare PTB rhyme ***-ek**? 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', **wé?-?u** (WB **wak-?u**), 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 **wà?** and **và?**, 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 **wé?-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 **ɔ̀-vɛ̀?**, 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").[72]

- Combining reflexes of unrelated roots

  When two forms bearing a semantic resemblance in a phonologically depleted lan-

[71]The discussion in this section is adapted from *HPTB*, pp. 538-40.

[72]There 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".

guage 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** (*STC* p. 89) / **phû** < PTB ***pəw** (*STC* #41); **mu** < PTB ***mraŋ** (*STC* p. 43) / **mû** < PTB ***r-məw** (*STC* #488).[73]

- 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 *STC* sets up two independent PTB roots, both with the shape ***dyam**, one meaning 'full; fill' (*STC* #226) and the other glossed as 'straight' (*STC* #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".[74]

  Similarly, I was slow to recognize that two roots I had set up separately, PLB ***dzay²** 'cattle; domestic animal' (Matisoff 1985a #129) and Kamarupan ***tsa:y** 'elephant; cattle' (#143) are really one and the same.[75]

- 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 **d̂i̯wən**) to both PTB ***dyal** and ***m-ts(y)ul**, finally

[73]See Matisoff 1973b (*GL*:29); such speculations were debunked in the 2nd Printing (1982) of *GL*, p. 675.
[74]See Matisoff 1988b:4-9.
[75]I have argued that a third root set up in Matisoff 1985a (*GSTC* #106), ***(t)sa:y ≍ *(d)za:y** 'temperament / aptitude /talent', is also related, the common notion being 'property (either material or intellectual)'. See Matisoff 1985a:44-45; 1988b:10-13.

deciding in favor of the latter.[76] 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.

  1. 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*,[77] should be segmented as **a-wu-tsi** and **ching-hi**, and not as **a-wut-si** and **chi-nghi**, as I imprudently did in Matisoff 1980:21.

  2. 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 dissyllabic 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',

[76] See *HPTB* 9.2.1, 9.22(4), 9.2.4.

[77] Kumar et al., *Hindi Pochury English Dictionary* (1972); *Hindi Sangtam English Dictionary* (1973). Kohima: Linguistic Circle of Nagaland.

pinyu ‘snake’).[78]

3. 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 Ganzi **tʃhə^{55}sɑ̃^{55}** and Ersu **ʂl^{55}ji^{55}** ‘otter’. Which syllables are to be ascribed to PTB ***sram**?

Future work

interactivity Final Product is of course not final at all!

Philosophical ruminations

Both phonological and semantic reconstructions are arts as well as sciences, arts which require a feel for proto-languages (what I have called Proto-Sprachgef�hl), as well as a commitment to �steer an Aristotelian middle path betwen a dangerous speculativism and a stodgy insensitivity to the workings of variational phenomena in language history.�1

We Tibeto-Burmanists need not apologize to Indo-Europeanists, or be envious of them. Even relative neophytes can make important discoveries in our field, which is a maximal contrast with the cutthroat world of Indo-European, where you can spend your whole life defending a single �new� etymology.

I much prefer the maxim Errare humanum est to the dreadful judgment Falsum in uno, falsum in omnibus.

[78]See Matisoff 1980 (“Stars, moon, spirits”), p. 35; for the suffixal use of morphemes meaning ‘mother’, see Matisoff 1991b (“The mother of all morphemes”).

It is admirable to correct one's own mistakes, both typographical and analytical. We should indeed suffer from our errors, but also correct them at every opportunity. As STEDT goes off into the twilight, we are doing all we can to "rectify" our etymologies, accepting most of them without modification, but also revising many others, and throwing out a relatively small number of them altogether.

Although it is admirable to change one's mind for good reason, it is frivolous to change one's reconstructions just for the sake of being different.

It is admirable to evaluate the reconstructions of others critically, and to offer better alternatives if possible. But it is frivolous to make trivial changes in previously established reconstructions simply in order to call them one's own.

Cumulativity vs. perpetual revolution. Don't throw out the baby with the bath-water. It is better to build on previous etymological work, rather than attempting to tear it all down and start over. Avoid Freudian patricidalism.

Do not be discouraged as you stumble along from one half-truth to another in this marvelous enterprise we call etymologization.

Etymology is an art as much as a science. But just as in any art, there are better and worse practitioners. We owe it to the field to be responsible and to hold ourselves to the highest standards of intellectual honesty and collegiality.

As STEDT goes off into the twilight, there are hopes that it will continue in a more 21st century way. Creation of an international consortium: the interactive "Root Canal".

Conclusion

One cannot very well end a book with the word ``horse-leech", and so a few concluding philosophical remarks seem appropriate. Perhaps the best way to organize this discussion is in terms of a set of adjectives with the -ive suffix.

Cumulative

In linguistics as in other disciplines, it is a constant temptation to try to overthrow the work of one's predecessors, so that one's own research will appear to be the *fons et origo* of the truth.[79] This tendency has been especially characteristic of generative grammar, where ``theories of language" have a built-in planned obsolescence, with each new theory claiming to invalidate all previous ones.

Historical linguists are hardly exempt from this primal urge for revolutionary novelty the desire to be different just for the sake of being different. This can take many forms, some of them trivial and innocuous, like replacing a phonetic symbol in a previous reconstruction by a new but equivalent one; or changing the name of a subgroup of a language family.[80] More serious is the itch to carve out totally novel subgroupings,[81] a process rather similar to the decennial gerrymandering of congressional districts in the House of Representatives. 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

[79]I have referred to this phenomenon in Freudian terms as ``patricidal" (keynote address at the Summer Institute of the Linguistic Society of America, Ann Arbor 1973).

[80]Hence the proliferation of alternate names even for well-established subgroups like Lolo-Burmese (Burmese-Lolo, Yi-Burmese, Burmese-Yi, Burmese-Yipho, Yi-Myanmar, Myanmar-Yipho, etc.). T. Shintani has recently (2002) proposed the euphonious neologism ``Brakaloungic" for Karenic.

[81]I have called this ``neosubgroupitis" (JAM 2000b ``On Sino-Bodic"). Trying to establish higher-order combinations of TB subgroups is premature at best. Even Indo-Europeanists are still unable to do so unequivocally for their much better documented family. See the discussion in JAM 1978a (VSTB):1-12.

paths, obscuring the differences among cognates, borrowings, and chance resemblances.[82] Perpetual revolution gets to be fatiguing after a while.[83] Surely it is preferable to build on the past rather than to repudiate it. 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 adhoc or mechanistic the reconstruction might be.[84]

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.[85] The Nostraticist or Sino-Mayanist can convince himself that his fantastical comparisons are ``perfectly regular''. Given sufficient semantic latitude and proto-forms that are complex enough, one can formulate ``sound laws'' in such a way

[82]See JAM 1990a (``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.

[83]As Leon Trotsky found to his cost in 1940.

[84]Those who lack what I have called ``Proto-Sprachgefühl'' (JAM 1982a) can produce reconstructions bristling with strange symbols but devoid of any phonetic or typological plausibility; see e.g. Sedláček 1970; Weidert 1975, 1979, 1981, 1987; Peiros & Starostin 1996.

[85]See JAM 1992 (``Following the marrow'') and 1994a (``Regularity and variation'').

that they appear exceptionless. 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''.[86]

As the alternative to such ``proto-form stuffing'', one must have recourse to proto-variation, though in a controlled and constrained way. Not everything may be said to vary with everything else.[87] This Handbook places special emphasis on variational patterns in TB/ST, and attempts to classify them according to how well attested they are.[88] As I put it 30 years ago, ``We must steer an Aristotelian middle path between a dangerous speculativism and a stodgy insensitivity to the workings of variational phenomena in language history.''[89]

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'', though we must never settle for a roseate Greenbergian haze.[90]

Self-corrective

A few of our etyma are only set up provisionally, and some individual forms are assigned only tentatively to a certain etymon. It must be admitted that a lot of guesswork

[86]This is actually the proto-form offered in Weidert 1981:25 for an etymon meaning `spirit, ghost, shadow' (reconstructed as *m-hla in STC #475). As I have observed, ``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'.'' (JAM 1982a:22).

[87]This issue is the major theme of JAM 1978a (VSTB).

[88]Note the large percentage of PTB roots for which proto-variation is posited in the Index of Proto-Forms, below.

[89]JAM 1972b (``Tangkhul Naga and comparative TB''):282.

[90]See Greenberg 1987, and my review of it (JAM 1990a).

is involved in etymologizing material from hundreds of languages and dialects at once, without having established the ``soundlaws'' 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 (e.g. *sem and *sak `mind / breath'; *muXXXf2 r and *muk `mouth'; *s-maXXXf2 y and *s-mel `face'; *s-r(y)ik and *s(y)ar `louse'), it is not easy to decide by simple inspection to which etymon we should assign a phonologically slight form in a daughter language (e.g. sXXX81 `mind', mXXXbf `mouth', hmXXXe4 `face').[91]

There are all too many ways in which one can make etymological mistakes. A rough taxonomy of errors would have to include the following:Treating a loanword as nativeI was at first delighted when I ran across the Jingpho form w|eXXXd6 -wu `screw', since its first syllable looked like an excellent match with Lahu

XXXbf -v

XXXe4 XXXd6 `id.', for which I then had no etymology. Could this be a precious example of the rare PTB rhyme *-ek ?[92] 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', w|XXXe4 XXXd6 -XXXd6 u (WB wak-XXXd6 u), the obvious source from which both Jingpho and Lahu borrowed these words, means literally ``pig-intestine''. The semantic association is the squiggly corkscrew-like appearance of a pig's small intestine.[93] 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 w

aXXXd6 and v

[91]See JAM 1994a (``Regularity and variation''):54-55.

[92]See above, 8.4(1).

[93]See the photographs of a pig being butchered in a Lahu village in JAM 1978a, between pp. 168 and 169. This same semantic association is to be found with the root *riXXXf2 l, above 9.3.2(3).

aXXXd6 , 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 w|eXXXd6 -wu 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

XXXbf -v

XXXe4 XXXd6 , since the morpheme for `intestine' has been completely dropped from the original Burmese compound,[94] rather like the way our word camera (< Lat. `room; chamber; vaulted enclosure') is a shortening of the old compound camera obscura (``dark chamber'').[95]

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' / phXXXc8 u `price, cost' and mu `high, tall' / mXXXc8 u `sky' were co-allofams, though they can easily be shown to descend from quite separate etyma: phu < PTB *plu (STC p. 89) / phXXXc8 u < PTB *pXXXfa w (STC #41); mu < PTB *mraXXXb3 (STC p. 43) / mXXXc8 u < PTB *r-mXXXfa w (STC #488).[96]

[94]The Lahu cognate to WB XXXd6 u `intestine' is XXXa9
u (usually in the compound
XXXbf -XXXa9
u-tXXXc8 XXXe4 XXXd6).

[95]There 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''.

[96]See JAM 1973b (GL:29); such speculations were debunked in the 2nd Printing (1982) of GL, p. 675.

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 STC sets up two independent PTB roots, both with the shape *dyam, one meaning `full; fill' (STC #226) and the other glossed as `straight' (STC #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".[97] Similarly, I was slow to recognize that two roots I had set up separately, PLB *dzayXXXaa `cattle; domestic animal' (GSTC #129) and Kamarupan *tsaXXXf2 y `elephant; cattle' (GSTC #143) are really one and the same.[98]

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.[99] At different times I have compared Chinese XXXae B `lip' (OC XXX84 ``iwXXXfa n) to both PTB *dyal (above 9.2.1) and *m-ts(y)ul, finally deciding in favor of the latter (above 9.22(4), 9.2.4). It is 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.

[97]See JAM 1988a:4-9, and above 3.4.2(c), 7.5(6).

[98]I have argued that a third root set up in GSTC (#106), *(t)saXXXf2 y & *(d)zaXXXf2 y `temperament / aptitude / talent', is also related, the common notion being `property (either material or intellectual)'. See JAM 1985a:44-45; 1988a:10-13; and above 5.5.2(1b), 5.5.2(2).

[99]The computer can be very useful in deciding between alternative etymologies. Once ``sound-laws" have been formulated, computer checking can test whether a particular reconstruction follows the laws, identifying inconsistencies in the reflexes of the same proto-element in a given language. Such a methodology has been applied to the Tamangic languages, using the ``reconstruction engine" developed by J.B. Lowe at STEDT in collaboration with Martine Mazaudon and Boyd Michailovsky during their sojourns at Berkeley as visiting scholars (1987-89, 1990-91).

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.[100] Many traps lie in wait for the analyst, leading to potential errors of several kinds, all of which I have been guilty of at one time or another:

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,[101] should be segmented as a-wu-tsi and ching-hi, and not as a-wut-si and chi-nghi, as I imprudently did in JAM 1980:21.

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 dissyllabic forms for `moon' with similar final syllables, e.g. Chang litnyu, Konyak linnyu, Phom linnyXXXd9 u, 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').[102]

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

[100]See JAM 1978a (VSTB):58-72.

[101]Hindi Pochury English Dictionary (1972); Hindi Sangtam English Dictionary (1973). Kohima: Linguistic Circle of Nagaland.

[102]See JAM (``Stars, moon, spirits'')1980:35; for the suffixal use of morphemes meaning `mother', see JAM 1991e.

is dealing with a poorly known language with depleted final consonants, e.g. Guiqiong Ganzi tXXXc6 hXXXfa XXXb0 XXXb0 s XXX81 XXXb0 XXXb0 and Ersu XXXa8 XXXbd XXXb0 XXXb0 jiXXXb0 XXXb0 `otter'. Which syllables are to be ascribed to PTB *sram ?[103]

Semantic leaps Deciding how much semantic latitude to allow among putative cognates is definitely an art rather than a science. Here as elsewhere a middle-of-the-road approach is necessary, neither overly conservative nor too wildly speculative. As a positive example of a promising new etymology involving a semantic leap, we may offer *m-t(s)i `salt / yeast'.[104] Phonologically the reflexes correspond perfectly well. On the other hand, the semantic association between `salt' and `yeast' has yet to be attested elsewhere, even though it has great initial plausibility. Both are efficacious substances that have dramatic effects on the taste of food or drink; their lack renders the food or drink insipid.[105]

Although I feel that we are entering a new era of etymological responsibility in TB/ST studies the bar has been raised, as it were I am not suggesting that we turn our field into a ``tough neighborhood" like that of the Indo-Europeanists. In particular I hope we can avoid the ``Gotcha!" attitude,[106] whereby if a single error, real or fancied, is found in an article or book, the whole work is impugned. This attitude is encapsulated in the dreadful

[103]See above 7.1(1).
[104]Above 3.3.1.
[105]Yeast is used for brewing liquor rather than for baking bread in East and SE Asia.
[106]Non-American readers might need a word of explanation here: ``Gotcha" is an attempt to render the colloquial pronunciation of ``(I've) got you (now)!", a triumphant phrase used by someone who feels he has won an argument.

maxim *Falsum in uno, falsum in omnibus*.[107]

Historical linguists cannot afford to be too thin-skinned, as long as criticism is fair, constructive, and proportionate. As I have said in print, ``I ask nothing better than to be corrected."[108] Or again, ``We can take comfort from our mistakes. Reconstruction of a proto-lexicon is a piecemeal process. It is hardly surprising that we stumble along from one half-truth to another, as we try to trace the [phonological and] semantic interconnections among our reconstructed etyma. We should not be discouraged if we barge off down blind alleys occasionally, or if the solution to one problem raises as many questions as it answers."[109]

After all, a computerized etymological enterprise by its very nature is eminently revisable.

Desiderative

I am acutely aware of the incompleteness of this Handbook. As noted in the Preface, 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,[110] reliable reconstructions at the subgroup level are not yet available, so that ``teleo-reconstruction" has to be resorted to.[111] Many more roots remain to be

[107]``If one thing is wrong, it's all wrong." This was the approach of Miller's (1974) bitter review of STC, in which he tried to kill the Conspectus just as it was born. In my ``rejection" of his ``Conspectus inspection" I characterized his strategy as follows: (a) make some criticism of a particular point, no matter how trivial, irrelevant or obfuscatory; (b) claim that tout se tient, and that the entire work stands or falls on the point in question; (c) beat the point elaborately to death; (d) avoid any substantive comments by pleading lack of space (JAM 1975a:157).

[108]JAM 1985b (``Out on a limb"):422.

[109]JAM 1988a:13.

[110]These bright spots include Proto-Karen (Haudricourt, Jones, Solnit), PNN (French), Proto-Tani (J. Sun), Proto-Tamangic (Mazaudon), Proto-Kiranti (Michailovsky), Proto-Central Chin (VanBik), Proto-Lolo-Burmese (Burling, Matisoff, Bradley). See the References.

[111]See Benedict 1973.

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 Handbook 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.[112] 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 tonoexodus 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?

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...''[113]

[112]For a rough typology of TB tone systems, see JAM 1999a. Weidert (1987) is 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 (1994c).

[113]JAM 1982a:41.

[[VIRTUALLY IDENTICAL PROSE]]

Looking toward the future of ST/TB studies

Although I feel that we are entering a new era of etymological responsibility in TB/ST studies—the bar has been raised, as it were—I am not suggesting that we turn our field into a "tough neighborhood" like that of the Indo-Europeanists. In particular I hope we can avoid the *"Gotcha!"* attitude,[114] whereby if a single error, real or fancied, is found in an article or book, the whole work is impugned. This attitude is encapsulated in the dreadful maxim *Falsum in uno, falsum in omnibus.*[115] Historical linguists cannot afford to be too thin-skinned, as long as criticism is fair, constructive, and proportionate. As I have said in print, "I ask nothing better than to be corrected."[116] Or again, "We can take comfort from our mistakes. Reconstruction of a proto-lexicon is a piecemeal process. It is hardly surprising that we stumble along from one half-truth to another, as we try to trace the [phonological and] semantic interconnections among our reconstructed etyma. We should not be discouraged if we barge off down blind alleys occasionally, or if the solution to one problem raises as many questions as it answers."[117] After all, a computerized etymological enterprise by its very nature is eminently revisable. The reconstructive process by its very nature is provisional and open-ended. Our STEDT etymologies undergo a constant process of "rectification", and may be roughly divided into three types: (a) those to be accepted as is; (b) those to be accepted with modifications; (c) those to be rejected. As with all scientific hypotheses, our reconstructions are falsifiable in the light of new data or better analyses.

[114]Non-American readers might need a word of explanation here. "Gotcha!" is an attempt to render the colloquial pronunciation of "(I've) got you (now)!", a triumphant phrase used by someone who feels he has won an argument.
[115]"If one thing is wrong, it's all wrong."
[116]Matisoff 1985b:422 ("Out on a limb").
[117]Matisoff 1988a:13.

We still have a long way to go before comparative/historical TB 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 mentioned above, reliable reconstructions at the subgroup level are not yet available. 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 of OC have sorted themselves out.

Nevertheless, 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...."[118]

[118]Matisoff 1982:41. There is nothing more satisfying than to have inadequate data on a language of which one has no firsthand knowledge corrected by a specialist in that language. The STEDT project has recently (summer of 2007) benefited tremendously from the kindness of K.P. Malla, who edited all the Newar(i) forms in our database, identifying loanwords, putting verbs into their proper citation forms, and correcting the transcription of vowels and consonants used in our previous sources.

Books, Monographs, Monograph Series*

AHD *American Heritage Dictionary*

CISTL Kitamura, Nishida, and Nagano, eds. (1994)

CSDPN Hale (1973)

CTT Hyman, ed. (1973)

GL Matisoff (1973b/1982)

GSR Karlgren (1957)

GSTC Matisoff (1985a)

HCT Li (1977)

HPTB Matisoff (2003)

HRAF *Human Relations Area Files* (New Haven)

NHTBM Nishi, Matisoff, and Nagano, eds. (1995)

OED *Oxford English Dictionary*

OPWSTBL *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics*

PPPB Luce (1986)

STC Benedict (1972)

TBL Dai et al., eds. (1992)

TBT Weidert (1987)

TSR Matisoff (1972a)

UCPL *University of California Publications in Linguistics* (Berkeley, Los Angeles, London)

SELAF *Soci√©t√© d'Etudes Linguistiques et Anthropologiques de France* (Paris)

VSTB Matisoff (1978a)

ZMYYC Sun et al., eds. (1991)

Journals

*Here listed only by author and date. For full citations see the References, pp. **??-??**.

AM *Asia Major* (Leipzig; London; Taipei)

AO *Acta Orientalia* (Copenhagen)

BIHP *Bulletin of the Institute of History and Philology* (Taipei)

BMFEA *Bulletin of the Museum of Far Eastern Antiquities* (Stockholm)

BSLP *Bulletin de la Soci√©t√© de Linguistique de Paris* (Paris)

BSOAS *Bulletin of the School of Oriental and African Studies* (London)

GK *Gengo Kenky≈´* (Tokyo)

HJAS *Harvard Journal of Asiatic Studies* (Cambridge, MA)

IJAL *International Journal of American Linguistics* (Chicago)

JAOS *Journal of the American Oriental Society* (New Haven)

JASB *Journal of the Asiatic Society of Bengal* (Calcutta)

LTBA *Linguistics of the Tibeto-Burman Area* (Berkeley; Chico, CA; Melbourne)

MSOS *Mitteilungen des Seminars f√ºr orientalische Sprachen an der k√∂niglichen Friedrich-Wilhelms-Universit√§t zu Berlin* (Berlin)

TAK *T≈çnan Azia Kenky≈´* (Southeast Asian Studies) (Kyoto)

ZDMG *Zeitschrift der deutschen morgenl√§ndischen Gesellschaft* (Wiesbaden)

Conferences

ICSTLL International Conferences on Sino-Tibetan Languages and Linguistics

SEALS Southeast Asia Linguistics Society

Research Units

AS Academia Sinica (Taipei)

CIIL Central Institute of Indian Languages (Mysore)

EFEO Ecole Fran√ßaise d'Extr√™me Orient (Hanoi/Paris)

ILCAA Institute for the Study of Cultures of Asia and Africa (Tokyo)

POLA Project on Linguistic Analysis (Berkeley)

SIL Summer Institute of Linguistics (Dallas)

STEDT Sino-Tibetan Etymological Dictionary and Thesaurus (Berkeley)

Languages

HM Hmong-Mien (= Miao-Yao)

IA Indo-Aryan

IE Indo-European

Jg. Jingpho (= Kachin)

KC Kuki-Chin

LB Lolo-Burmese

Lh. Lahu

MC Middle Chinese

OC Old Chinese

PIE Proto-Indo-European

PLB Proto-Lolo-Burmese

PNN Proto-Northern Naga

PST Proto-Sino-Tibetan

PTB Proto-Tibeto-Burman

ST Sino-Tibetan

TB Tibeto-Burman

TK Tai-Kadai

WB Written Burmese

WT Written Tibetan

Miscellaneous

semcat

pan-allofamic formula ("PAF")

extra-fascicular etyma

mesoroots

supporting forms

"add-sourcing"

"Mariama Fix"

metastatic flowcharts

weakly-attested root

> goes to; becomes

< comes from; is derivable from

A ,™ § **B** A and B are co-allofams; A and B are members of the same word-family

Clf. classifier

JAM James A. Matisoff

lit. literally

OICC Obscure internal channels and connections,Äù (see Ch. III)

ult. ultimately

WHB William H. Baxter

ZJH Zev J. Handel