Chapter 14

On the uselessness of glottochronology for the subgrouping of Tibeto-Burman

James A. Matisoff

This chapter is perhaps an exercise in necrohippomachy, or the flogging of a dead horse.[1] Few linguists nowadays rely exclusively on the traditional sort of lexicostatistics — i.e. the use of a 100- or 200-item list of 'core vocabulary' — in order to establish a genetic relationship among languages not already recognized as related. The method has seemed somewhat more useful for the subgrouping of a well-established language family. It is my contention, however, that glottochronology is quite useless when applied to the subgrouping of Tibeto-Burman (TB), for many reasons. We will return to these TB-specific problems in §1 below, but will mention some of them here in a preliminary way:

- Our lexical data is of very uneven quality and quantity for the various branches of the family, with vast stretches of the TB heartland (especially northeast India and Burma) chronically unavailable to outside fieldworkers. Sound-laws are far from having been worked out for all branches of the family, and many 'mesolanguages' remain to be reconstructed at the subgroup level.

- Except for Tibetan and Burmese (attested since the seventh and twelfth centuries AD, respectively), only a few TB languages have written records that go back more than about 100 years,[2] so that no baseline is available to calculate TB-specific rates of lexical change over millennia.

- Chinese, the other branch of Sino-Tibetan (ST), does have a very long written history, along with a vast lexicon. Yet its non-alphabetic script, marvellous as it is, makes phonological reconstruction extremely difficult, and several competing systems are now current. Many characters found in the great Chinese dictionaries are *hapax legomena* which may never have been real words at all, but merely literary embellishments of other characters. Etymologically related words are sometimes written with totally or partially different characters, and conversely a given character has often been 'loaned' to symbolize an unrelated word of similar phonological shape.

- The TB languages have been in close contact with several other language families of East, South, and Southeast Asia, notably Chinese and Indo-Aryan.[3] This makes for great typological diversity in TB grammatical and phonologi-

cal systems. Sinospheric TB languages are typically monosyllabic and highly tonal, with a minimum of morphology (except for compounding, reduplication, and elaboration). The TB languages of the Indosphere are often toneless and/or highly suffixal; the Rai or Kiranti group of Eastern Nepal boasts complex 'head-marked' morphology, with agreement suffixes on the verb that specify the number, person, and/or case of nominal or pronominal arguments. Whether or not this inflectional apparatus is of great antiquity in TB,[4] there is no way to reconstruct cognate grammatical morphemes across subgroups of the family — in stark contrast to the precious help such morphemes provide for Indo-Europeanists.[5]

- Throughout the TB family, morphemes are monosyllabic, immensely complicating the task of cognate identification. As Dixon puts it (1997, 41), 'A cognate set between polysyllabic forms provides much better evidence than one involving monosyllables, or single-segment forms. If the verb "go" is -**gimlar**- in two languges, this is a stronger evidence of relationship than if it were -**a**-.'[6] While it is true that the homophony problem is severe in phonologically depleted languages (e.g. those of the Loloish branch of TB), this is compensated for synchronically by the pervasive strategy of compounding, which presents problems of its own for the lexicostatistical approach (see below 2.2).

- As in Indo-European, TB etyma cannot be conceived of as invariant entities, but appear in slightly different phonological shapes both within a single language and cross-linguistically. These sets of phono-semantically related forms have traditionally been called 'word families' in ST linguistics. Some of these variational patterns are pervasive across TB subgroups, and must be imputed to the proto-language; others are sporadic or idiosyncratic, to the point where one cannot always be sure whether only a single etymon is involved. This also greatly complicates the identification of cognates (more on this below, 2.3).

1. A critique of traditional glottochronology as a tool for subgrouping

1.1. Defects in the family-tree model itself, especially when applied to Southeast Asia
The classical *Stammbaum* or 'family tree' metaphor for characterizing degrees of linguistic genetic relationship has been recognized as a vast oversimplification for a century. Languages rarely split off cleanly from their relatives (Icelandic being a notable exception). A much more appropriate image for what one finds in linguistic areas like Southeast Asia (SEA) might be the *thicket,* an im-

penetrable maze of intertwined branches.[7] Instead of clearcut migrations of population groups, one finds slow *percolations* or *filtrations* of small groups of people.

1.2. Distinguishing loanwords from true cognates; analytical expertise

Two languages may have a very high percentage of shared vocabulary and still be genetically unrelated (e.g. Chinese and Japanese, Korean, Vietnamese, Thai, Hmong-Mien), or at least much less closely related than the percentage would seem to warrant (e.g. English and French, Chinese and Bai). This is especially problematic when we are dealing with languages without a long written history, spoken in a complex linguistic area with borrowing in all directions, and where the borrowings may be from a related language.

No matter how elegant the computations and calculations to which data are subjected, if these data are inaccurate or uncertain to begin with, the inferences one draws from them can be no better. Garbage in, garbage out. It is often not easy to tell from simple inspection whether a given pair of forms are cognate. In many cases it takes detailed, specialized knowledge simply to be able to venture an educated guess. Even professional linguists are not necessarily equipped to decided subtle questions of cognacy involving the very language family to which their native language belongs, unless they have received specific training to that end. Often forms which look very much alike are not true cognates at all. Is Latin **dies** cognate to English **day**? How about Latin **habeo** and English **have**? Conversely, true cognates which are perfectly regular reflexes of rigorously statable 'rules of sound-change' may have no superficial phonological resemblance to each other at all. One may cite thrilling examples like Latin **duo**/Armenian **erku** 'two'; Latin **oculus**/Mod. Gk. **mati** 'eye' < (*op-)ma-ti-on < *ok^w-mn̥-ti-on; German **was**/Russian **čto** 'what'; Written Tibetan **brgyad**/Lahu **hí** 'eight'; Lahu **ɔ̂**/Burmese **lê** 'four' (see below 2.3: 'Word families and regularity of correspondence').

1.3. Inappropriateness of the list for many linguistic areas

It has been repeatedly observed that the standard Swadesh lists are culturally and grammatically inappropriate for many linguistic areas of the world, full of over- and under-differentiations.[8] I have attempted to mitigate these problems somewhat by compiling a 200-word list more appropriate for Southeast Asia, though this is comparable to applying a bandaid to a gangrenous foot.[9] A pair of distinguished Thai linguists have recently encountered problems with several items on my list in their preliminary fieldwork on a dozen Mon-Khmer languages of Laos.[10]

1.4. Artificiality of limiting the list to only 100 or 200 items

A major objection to the standard lists, besides their inappropriateness, is the paucity of the items they contain. Surely the more words we have to go on the better. An excellent list of some 3000 items, organized by semantic field, with both English and Chinese glosses, has been compiled by the Academia Sinica in Taipei, and is in standard use by Taiwanese fieldworkers.[11] Similar lists have been used in mainland China, though they are marred by the presence of dozens of historically useless items of modern vocabulary (*school, post-office, bank,* etc.), many of them obligatorily included because of their political correctness (*cadre, work-unit, party, factory,* etc.). Needless to say, words like these in the minority languages of China are always very recent loans from Chinese.

1.5. 'Core vocabulary' and the rate of lexical replacement

Idiosyncratic morphological features (e.g. parallel exceptional forms in inflectional paradigms) have long been appreciated as especially valuable indicators of genetic relationship. Unfortunately, in languages with minimal morphologies, like most of those in the Southeast Asian linguistic area, this criterion is of little use, and one is forced to rely mostly on lexical resemblances in 'core vocabulary'. Although on the whole it seems true that 'core vocabulary' is more resistant to change than what we might call 'peripheral' vocabulary,[12] this is only a matter of degree; it is easy to find striking examples of lexical replacement in any semantic/conceptual realm. Numerals may be borrowed wholesale. Kinship terms may change their referents or disappear owing to taboo, euphemism, teknonymy, or social change. Animal names may fall out of use and be replaced because of pernicious homophony (e.g. the similarity between **gallus** 'rooster' and **gattus** 'cat' in Gallo-Romance led to French **coq**), or hunters' taboos (cf. euphemisms for *bear* like 'honey-eater' or 'the brown one' in Indo-European, or the constantly replaced animal names in the Aslian branch of Mon-Khmer (Malaysia). Body-part terms are not exempt (the Thai word **cəmùuk** 'nose' is from Khmer (Lao **daŋ** is a reflex of the original Tai etymon). Even words of abstract grammatical function, basic relational particles like *and, or, not,* may be replaced by foreign borrowings (cf. the Lahu suspensive particle **lɛ** < Shan **lɛʔ**).

Conversely, the tendency for non-core vocabulary to be particularly prone to replacement is far from absolute. Under certain sociolinguistic conditions, non-core vocabulary can have a surprisingly long half-life.[13] In the speech of descendants of shifting bilinguals, it is only a handful of cute or culturally interesting words that are likely to be preserved from the 'deep substratum' of the original language.[14]

Even if we could agree on such a list, why must we assume with the glottochronologists that languages replace their basic vocabulary at a universally constant rate? Furthermore, it has been persuasively argued that the rate of linguistic change of all kinds is highly sensitive to extra-linguistic events, with long eras of relative stasis giving way to periods of rapid change prompted by military, political, or demographic upheavals at irregular intervals.[15]

1.6. Constancy of rate of change

Owing partly to the accidents of cultural history and partly to mysterious differences in their internal 'genius' or 'essence', synchronically observable languages differ considerably in their receptivity to borrowings from foreign languages, whether these foreign languages are dead, learned languages like Latin or Sanskrit, geographically contiguous contact languages, or nowadays even culturally important languages spoken on the opposite side of the globe.[16] Japanese borrows many more words from English than *vice versa*. We can conceive of Japanese borrowing the English word *kiss* (Jse. *kissu*) much more easily than our borrowing the Japanese word *seppun* 'kiss', or even a more culturally indigenous word like *o-jigi* 'a bow'.

The vicissitudes of world cultural history constantly reshuffle the patterns of linguistic dominance and submissiveness with respect to outside incursions on a language's vocabulary.

1.6.1. Dixon's 'punctuated equilibrium' model

For hundreds of years after the Norman conquest, French was accepted as a culturally superior language to English — a view in which both Englishmen and Frenchmen concurred — and the massive flow of borrowed vocabulary was almost totally unidirectional, from French to English. In the nineteenth century, with the growing political and economic power of the Anglophone world and the decline of France's role on the international stage, the tide began to turn. Words like *revolver, bifteck, redingote* (< *riding coat*) crept into the hallowed French lexicon.[17] As we approach the new millennium the contamination has reached truly alarming proportions, as the older generation decries the vile new *franglais* of French youth. The glottochronologist would have us believe that such cultural currents 'even themselves out' over a long enough period of time, but surely this is an unprovable article of faith.[18]

English and German, two quite closely related languages, seem until very recently to have differed greatly in their willingness to tolerate the replacement of inherited lexical material, or the accretion of new foreign words: the Accre-

tion Tolerance Quotient (ATQ) of English has been much higher than that of German.[19] This has manifested itself in several ways:

1. *Coinage of learned vocabulary.* While English rushes to Latin and Greek to make up technical neologisms (*television, tonogenesis, leucorrhoeia, rhino-glottophilia*), German much prefers to use native material *(Fernseher, Tonent-stehung, Weissfluss, Nasenstimmbändergegensatzliebe).*

2. *Borrowings of core vocabulary.* English words like *dog, pig, skin, animal* are innovations with respect to Common Germanic, either borrowings from some (often obscure) contiguous Germanic source (*dog, pig, skin*), or in the case of *animal* from a 'higher culture' language. German, on the other hand, tends to preserve the most widespread Common Germanic etymon (*Hund, Schwein, Haut, Tier*).

3. *Semantic shifting in core vocabulary.* It seems impressionistically as if English items of core vocabulary are more likely to undergo semantic change than are their German equivalents; this indeed is a smart strategy for a language with a high ATQ to adopt. While welcoming innovations on the one hand, English is reluctant to throw away the older word entirely.[20] Typically it is retained, either as a less frequent or stylistically 'marked' synonym of the newer word, or else with some change or increment of meaning, so that the two words, old and new, continue to coexist, each safe within its semantic domain. We have not given up the word *hat* just because we have borrowed *sombrero* from Spanish. English cognates to the German words just cited are still very much alive (*hound, swine, hide, deer*).

In sum, it is not at all obvious that languages undergo lexical replacement at the same rate (as viewed across languages) or at a constant rate (as viewed across time), either in core or peripheral vocabulary.

1.7. Semantically shifted cognates

It is a fundamental dogma of lexicostatistics that one must not look around too hard for etymological cognates in compiling the basic vocabulary lists in the languages to be examined. What is wanted is 'the usual word in the language at the present time'.[21] This of course leads to the loss of vital information that might make the lexicostatistician's judgments more subtle and refined.

Let us try a 'thought experiment' involving a hypothetical English–German bilingual, well-educated but not a professional linguist.[22] If asked to list as many words as possible in the two languages that are related (we assume he has a good layman's notion of what 'relationship' means in this context), he would be able to come up with dozens of correct answers in short order.

a. The easiest relatives to recognize would be those where the pronunciation, spelling, and meaning of the two forms are identical or nearly so:
    finger/Finger, fish/Fisch, nest/Nest, hand/Hand, hammer/Hammer, house/Haus, mouse/Maus, louse/Laus, shoe/Schuh . . .
b. Only minimally more difficult to identify would be pairs where the meanings of the two forms are still identical, but the pronunciation and spelling are 'a little more different':
    out/aus, foot/Fuss, water/Wasser, day/Tag, eye/Auge, sheep/Schaf, bride/Braut, sheath/Scheide, heath/Heide, thumb/Daumen . . .
c. Sometimes, however, even if there exists a cognate in the other language, and *even if the meanings of the two forms are still more or less the same,* our bilingual would have a lot of trouble coming up with a correct answer. Given the English word *tail,* he might not be able to cite the cognate *Zagel,* since this only survives in the Siebenbürgisch dialects of East Carpathia, and has been generally replaced by *Schwanz.* Similarly, asked to find English relatives of *Zeichen* 'sign, symbol, mark', *heissen* 'be named, be called', *gern* 'willingly', *Schaden* 'damage', or *schwarz* 'black', he would not be very likely to think of *token, hight, yearn, scathe,* or *swart.* Although the semantic changes here have not been particularly drastic, the relationships are hard to figure out on other grounds: the obsolescent, archaic, or dialectal nature of some of the forms, the fact that the cognate morphemes are usually hidden in combination with other morphemes (*unscathed, scathing, swarthy*), or the considerable surface phonological and orthographic divergence of the forms (*token/ Zeichen*).

So far this is quite in conformity with glottochronological theory. But consider the next class of cases:

d. Crucially important are the cognates where one of the languages has indeed undergone a considerable semantic shift. I contend that this shift in meaning would not necessarily make it difficult for our bilingual to come up with the correct cognate right away. Asked to find the German cognates to *flesh, fowl, bone,* and *hound,* would he not quickly supply *Fleisch* 'meat', *Vogel* 'bird', *Bein* 'leg', and *Hund* 'dog'? Granted, there will be cases where the semantic divergence is so great that the cognate might escape notice even if the two forms were phonologically or orthographically quite similar: *deer/Tier* 'animal', *tiding(s)/Zeitung* 'newspaper', *knave/Knabe* 'boy'. In extreme cases, the pronunciations and/or the meanings have diverged so much that only the specialist in Germanic linguistics could be expected to recognize the relationship: G. *Schmuck* 'ornament'/E. *smug*; G. *nehmen* 'take'/E. *numb* < OE *numen,*

past participle of *niman* 'take' ('taken, seized'), and *nimble* 'quick-witted, quick to *grasp* things, quick on the up*take*'; G. *schräg* 'oblique, slanted', *schränken* 'lay crosswise'/E. *shrug, shrink*.

The point is this: why should we throw away any information at all that bears on the past history of relationship of languages? To ignore a good cognate on the grounds of semantic change is obfuscatory rather than clarificatory. The English words *girl*, *tree*, *cloud* have, as far as I know, no German cognates; conversely, German *Stimme* 'voice' has no English relatives. Surely cases like these should be 'scored' differently from *fowl/Vogel* or *hound/Hund*! Instead of the 'all-or-nothing' approach, why not use some more sensitive scoring system that can capture the whole continuum of phono-semantic relationships among the lexicons of our languages? It is true that this gets very complicated.[23] Furthermore it is subjective: we do not have principled ways of measuring degrees of semantic divergence (or phonological divergence either, for that matter). But at least it would begin to do justice to the complexity of the problem, and has the virtue of focusing our attention on whole sets of phono-semantically related forms, rather than artificially isolating a single pair of words to consider each time.

1.8. Semantic latitude and areal semantics

It is even more of an art to decide how much semantic divergence may be tolerated among reflexes of the same etymon. Roots may indeed undergo spectacular semantic changes through time, and the glottochronological dogma against accepting semantically shifted cognates in determining degrees of genetic relationship goes much too far. However, the bigger the semantic leap the better the phonological correspondence must be between the putative cognates. Otherwise the phonological and semantic arguments are like two drunks supporting each other.

Crucially, it should not automatically be assumed that semantic associations attested in one linguistic area are universally valid. Among the supposed cognates offered by Sagart (1993) to demonstrate a genetic link between Chinese and Austronesian is Proto-Austronesian (PAN) **pusuq* 'heart; central leaf' and Old Chinese **swia* (re-reconstructed **s-j-waʔ*) 'marrow', since marrow is supposedly 'the heart of a bone'. Yet, aside from the dubious phonological correspondence, there is no evidence at all that marrow has ever been conceived in a 'heartlike' way by East Asian peoples.[24] Similarly, after admitting that '. . . the abundance of comparisons of the type *water/sap* over the type of *water/water* seriously diminishes the credibility of any hypothesis of genetic relationship', Vovin (1990, 1) attempts to prove the Altaic affiliations of Japanese by such

comparisons as Proto-Japanese *momo 'peach' to Proto-Manchu-Tungus
*ñaṅ-ta 'nut', (perhaps because such an association exists in N. Caucasian lan-
guages). Sometimes a semantically dubious etymology is presented as if the
meaning association were obvious, even though it may never have been clearly
attested in any language family. As support for his Austro-Japanese theory,
Benedict (1990, 193) compares Indonesian ikan 'fish' (< PAN *Sikan) to Japa-
nese ika 'squid' (< PJse *yika), since 'squid, like fish, have long been a staple
food source for the Japanese'.

The notion of 'areal semantics' is just as valid as that of 'areal phonology'.
However, once a semantic association has already been established on inde-
pendent grounds within a linguistic area, similar associations found elsewhere
may well have confirmatory force. Just as BRAIN <—> MARROW is unmistakably
attested both in Tibeto-Burman (TB) and Indo-European (IE), so have I hypoth-
esized that two supposedly distinct but homophonous PTB roots *dyam 'full'
and *dyam 'straight; flat' are really one and the same, offering as additional
evidence the phonological similarity and intercontamination between two se-
mantically similar IE roots represented by Latin plānus 'flat' and plēnus 'full'
(Matisoff 1988b).[25]

2. Morphological and morphophonemic problems specific to Tibeto-Burman

These TB-specific complications include phonological slightness of morphemes
(monosyllabicity), unpredictable compound-formations, and sporadic morpho-
phonemic variations in etyma (resulting in 'word-families').

2.1. Monosyllabicity and homophony

The strictly monosyllabic and phonologically depleted Sinospheric TB languages
are rife with homophony. Consider the five unrelated Lahu morphemes pro-
nounced ha (all under the mid-tone, unmarked in the transcription) in the third
column of Figure 14.1.

In actual speech there is little chance of confusing these morphemes, both
because their meanings are so different and because they almost always occur
in tight combination with other elements.[26] But for the etymologist the problem
can be very serious indeed. Most dictionaries of TB languages simply inter-
alphabetize all collocations containing a given phonological syllable regardless
of its morphemic identity. While my Dictionary of Lahu (Matisoff 1988a) made
an earnest attempt to avoid this 'pernicious interalphabetization',[27] there were
many cases where I found it very difficult to determine the proper head-entry

	Proto-Tibeto-Burman	Proto-Lolo-Burmese	Lahu monosyllables	Lahu dissyllabic collocations
hundred	*b-r-gya	*ʔra¹	ha	tê ha
moon	*s-gla	*s-la³	ha	ha-pa
tongue	*s-lya	*s-l(y)a¹	ha	ha-tē̄
spirit	*s-hla	*sla³	ha	ɔ̀-ha
winnow	*g-yaː(p)	*ʔ-ya¹	ha	ha ve

Figure 14.1. *Lahu homophonous monosyllables.*

for a given collocation. (This problem is largely irrelevant to Chinese, despite the massive homophony in phonologically depleted dialects like Mandarin, since the logographic nature of the script almost always makes it clear what morpheme is involved.)

2.2. *Compounding and phonological bulk*

A cognate morpheme is often hidden in a non-initial syllable of a di- or tri-syllabic compound.[28] For some time I did not realize that Lahu preserved a reflex of PTB *g-zik 'leopard' (cf. Written Tibetan (WT) gźig, Written Burmese (WB) sac), until I found it in the compound mɔ̀ʔ-yɨ̀ʔ 'cloudy leopard' (*Felis nebulosa*), where the first syllable mɔ̀ʔ- means 'monkey' (i.e. literally 'monkey-leopard').

Cross-linguistically a given concept may be expressed by compounds with identical semantic components, though with different etyma filling the semantic slots. Words for *tears* in TB languages are usually expressed by compounds meaning EYE + WATER. While there is only one widespread root for EYE in TB (*myak ≈ *mik), there are eight or nine separate etyma for WATER that appear in these compounds.

Even closely related languages within the same subgroup — nay, even dialects of the same language! — are likely to make different selections from the proto-treasury of roots (the *Urwortschatz*) in compound formation. Consider Figure 14.2, containing the words for HEAD (certainly an item of core vocabulary) in Written Tibetan and a few Loloish (Bisu, Akha, Lisu, Lahu) and Burmish (Written Burmese, Atsi, Lashi, Maru) languages.

	Prefix	*bu ⋈ *wu 'head'	*du 'head'	*s-koŋ 'hollow'	*lum ⋈ *lim 'round'
Written Tibetan	d-	bu			
Bisu	ʔaŋ-		tù		
Akha		ʔù-	dù		
Lisu		wu¹-	du³		
Lahu Nâʔ (Black Lahu)		ó-		qō	
Lahu Nyi (Red Lahu)	ʔa-		tù-	kù	
Lahu Shi (Yellow Lh.)	ʔa-			kù	
Written Burmese		û-		khôŋ	
Atsi (Zaiwa)		u-			lum
Lashi		ʔú-			lèm
Maru (Langsu)	ʔáu-				làm

Figure 14.2. *Compounds for HEAD in WT and some Lolo-Burmese languages and dialects.*[29]

Each of these closely related Lolo-Burmese languages is a law unto itself, making idiosyncratic selections both from the stock of roots and the stock of prefixes.[30] Taken together, this 'compound family' demonstrates the futility of trying to correlate particular morphemes in compound formation with degrees of genetic relationship. The clinching case is provided by the three (mutually intelligible) Lahu dialects, which each went a different route in making up a compound for HEAD.[31]

2.3. Word families and regularity of correspondence
Since every natural language is rife with irregularities, and since every modern language is 'a proto-language with respect to the future',[32] it is unreasonable to

	Written Burmese	Lahu	Akha	Jingpho	Written Tibetan	Proto-TB
'wind'	le	mû-hɔ		būŋ-lī	rdzi	*g-ləy
'boat'	hle	hɔ-lò?-qō		lī		*m-ləy
'heavy'	lê	hɔ̂		lī	lči	*s-ləy-t
'grand-child'	Insc.mliy WB mrê	hɔ̄-ɛ̄	ǿ-pà	məlī 'young man'		*b-ləy
'four'	lê	ɔ̂	ǿ	məlī	bži	*b-ləy
'bow/ sling'	lê	hɔ̂-ma	ca-ǿ	ləlī	gžu	*d/s-ləy

Figure 14.3. *Non-obvious but parallel correspondences: reflexes of prefixed laterals plus *ǝy.*

expect that all etymologically related forms in daughter languages will exhibit perfect regularity of phonological correspondence. (In fact if forms from not particularly closely related languages are *too* similar, it should arouse one's suspicions that perhaps borrowing or pure chance is involved.) Still, if we are to do historical reconstruction at all, we must never abandon the *ideal* of regularity. While (as mentioned above §1.2) there are stunning examples of perfectly cognate forms that have little or no surface phonetic similarity, the stranger the correspondences, the more independent evidence is required to back them up. Such an inspiring set of forms is given in Figure 14.3, with six parallel examples illustrating the regular Lahu and Akha reflexes of PTB etyma with prefixed laterals plus the rhyme *ǝy.

Still it is all too easy to abuse notational devices and *ad hoc* explanations to make just about any correspondence achieve a specious air of regularity.[33] It is not enough to set up 'tables of correspondences' without presenting all the data *that either confirm or disconfirm* the fillers of the cells in the table. The trick is to steer a middle course between etymological promiscuity and a stodgy insensitivity to the mechanisms of linguistic variation.

2.3.1. Variational patterns in TB word-families

The term 'word family' (i.e. a group of phonosemantically similar but not identical forms that can be traced back to a single etymon) was first used with re-
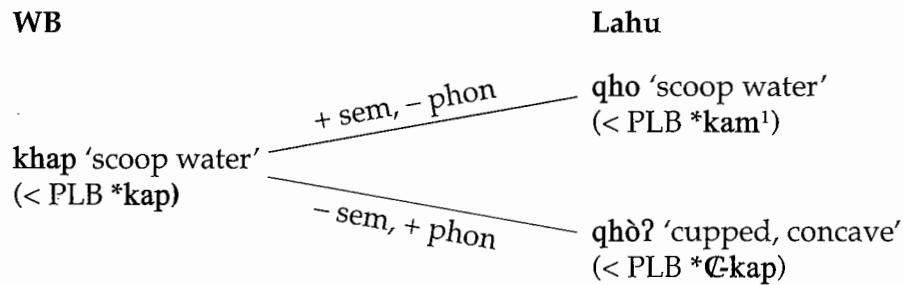
WB   **Lahu**

khap 'scoop water'
(< PLB *kap)

+ sem, – phon ——— qho 'scoop water'
                       (< PLB *kam[1])

– sem, + phon ——— qhò? 'cupped, concave'
                       (< PLB *C-kap)

Figure 14.4. *Semantic and phonological isofamy and heterofamy.*

spect to Chinese by Karlgren (1933). A book-length attempt to make this concept more precise with respect to TB was Matisoff 1978, where various terms were introduced to facilitate discussion, including: *allofam* (a particular member of a word-family); *allofamy* (the relationship in which members of a word-family stand to each other); *isofams* (forms which may be traced back to the same proto-allofam); *heterofams* (forms which descend from different proto-allofams). The relationship of allofamy is symbolized by '≍', i.e. 'A ≍ B' means that 'A is an allofam of B'. Allofams may coexist synchronically within a single language, either by common genetic descent, or as doublets where one is a borrowing from a related language; cf. Eng. *shirt* ≍ *skirt* (the latter from Scandinavian). Allofams may also of course be identified cross-linguistically, e.g. when two languages each inherit a different variant of a proto-word-family.

Many well-attested patterns of allofamy must be recognized for TB, though not all of them are of equal antiquity. Here we can do no more than list the most important ones:[34]

a. *Alternations between final homorganic stops and nasals*[35]
   E.g.: Lahu šē 'sow broadcast' (< PLB *san[2]) ≍ šê? 'pour; spill' (< PLB *sat)
   A slightly more complicated example is furnished by two Lahu forms, one of which is a perfect phonological fit with a Burmese form, but divergent semantically (i.e. phonologically isofamous but semantically divergent); while the other Lahu form is semantically identical but phonologically heterofamous:
   Lahu qho 'draw water' (< PLB *kam[1]) ≍ qhò? 'cupped; concave' ≍≍ WB **khap** 'scoop water' (see Fig. 14.4).

b. *Alternations between open syllables and those with suffixal final dentals* /-t -n -s/ (see Fig. 14.5).[36]

c. *Variations of medial glides.*
   E.g.: 'many' WB *myâ < PLB *mya[2] ≍ Lahu mâ < PLB *mra[2]

d. *Variation of the high vowels -i- and -u- in closed syllables, especially in the*

PST *na	PST *nan	PST *nat
WB na 'be sick, hurt'		WB nat 'spirit
Lh. nà 'be sick, hurt'	Lh. nê 'illness-producing spirit'	
WT na-ba 'be sick, hurt'		WT nad 'illness'
Ch. nâ 'ceremony to expel demons of illness' 亻難	Ch. nân 'be in difficulty, be suffering' 難	
		Lu. nat 'ache, be in pain'

Figure 14.5. *Some members of the word-family* *na* ≈ *nan* ≈ *nat.*[37]

environment of syllable-initial or -final labials.

E.g.: PTB *lum ≈ *lim 'round'; *mul ≈ *mil 'body hair'

e. *Variation between *-ya- and *-i-*

  E.g.: 'eye' WT mig (< PTB *mik) ≈ WB myak (< PTB *myak)

    'pheasant' WT sreg-pa ≈ śrag-pa (< PTB *s-ryak) ≈ WB rac (< PTB *s-rik)

f. *Variation between *-a- and *-u-.*

  E.g.: 'descend' Lahu yàʔ (< *zak) ≈ Jingpho ʔyúʔ (< *s-zuk)

g. *Variation between initial sibilants and affricates.*

  Among innumerable examples one could cite are the three Lahu co-allofams:
  cà- 'prefix to male names' (< PLB *dza¹) ≈ yâ 'son; child' (< PLB za²) ≈ šā
  'sibling's child' (< PLB *sa²).

h. *'Lenition' of labial stops.*[38]

  'pig' WT phag ≈ WB wak

Finally, two variational phenomena that are too pervasive and complicated to
begin to go into here:[39]

i. *Variation in voicing and aspiration in syllable-initial position.*

  This variation is due especially to the influence of prefixes.

j. *Tonal alternations in word families.*

* * *

Subtle problems constantly arise in attempting to distinguish between allofams
of the same word-family and unrelated but fortuitously similar etyma, as with
these phonosemantically similar but unrelated pairs of Lahu words:

    mu 'high' (< PLB *mraŋ³) vs mû 'sky' (< PLB *məw²)

    phu 'silver, money' (< PLB *plu¹) vs phû 'price' (< PLB *pəw²)

What partially similar forms reflect genuine patterns of allofamy and what are merely unrelated lookalikes? How much latitude can we permit in our positing of an allofamic relationship between forms? There is no mechanical answer, and there is no substitute for long experience. What is certain is that language families are very different in their ranges of allofamic variation. While voicing and aspiration discrepancies are par for the course in TB/ST word-families, Indo-Europeanists will tolerate much less of this sort of thing before declaring forms to be unrelated. Conversely, vowel alternations or ablaut are normal in IE word-families, are largely explicable in terms of consonantal influence and stress, and frequently have recognizable morphological functions, while they are much more sporadic and much less explicable in TB/ST.[40]

3. Subgrouping schemes for Tibeto-Burman[41]

In his seminal work, *Sino-Tibetan: a Conspectus* (1972; henceforth *STC*), Benedict wisely refrained from constructing a TB family tree of the conventional type, presenting instead a schematic chart where Kachin (= Jingpho) was conceived as the centre of geographical and linguistic diversity in the family, and many individual languages are not assigned to any larger group at all (e.g. Gyarung, Meithei, Mikir, Mru, Newari, Lepcha). Karen is banished from the core of TB altogether (see Fig. 14.6).

The subgrouping scheme now being used heuristically at Berkeley's *Sino-Tibetan Etymological Dictionary and Thesaurus* project (*STEDT*) is an improvement in certain respects, though its apparent neatness conceals many problems that have been temporarily swept under the rug (see Fig. 14.7).

The seven major subgroups postulated here are quite heterogeneous in terms of their internal complexity and the number of languages each contains.[42] Furthermore, as admitted above, the current state of our knowledge about the various subgroups is very different. This being said, let us take a quick look at them one by one.

- **Baic** (consisting of dialects of a single language, Bai, spoken in NW Yunnan), was hardly mentioned in *STC*, under the name 'Minchia'. It was later hypothesized by Benedict to belong with Chinese in the 'Sinitic' branch of Sino-Tibetan. I feel this is a mistake, and Baic should be treated as just another subgroup of TB, though one under particularly heavy Chinese contact influence.[43] Baic has SVO word-order, like Chinese, but so does Karenic.
- **Karenic** contains about a dozen fairly well differentiated languages, now spoken mostly in the border regions between Burma and Thailand. The Karens
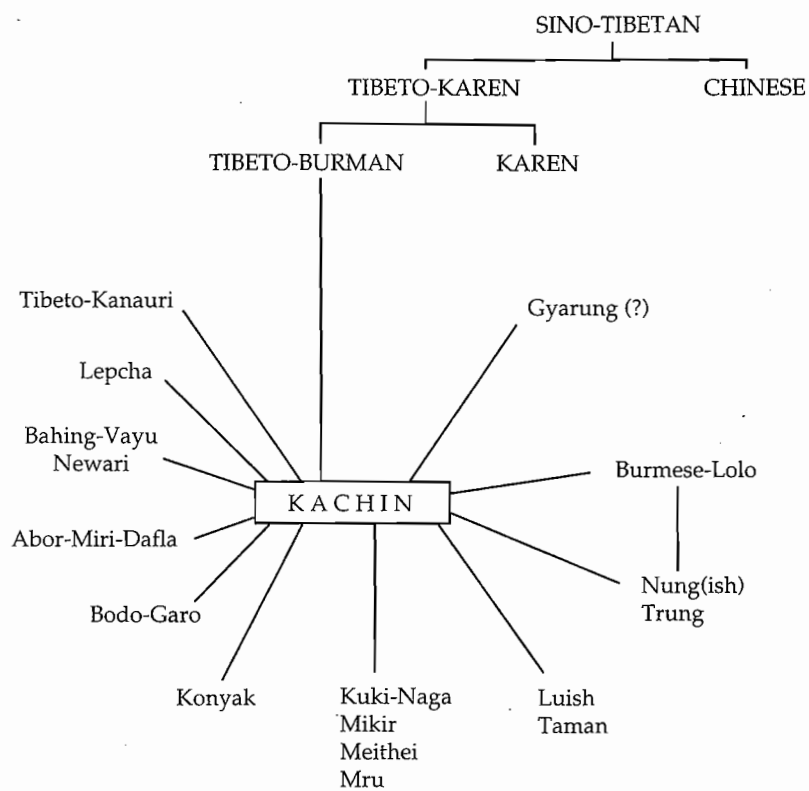
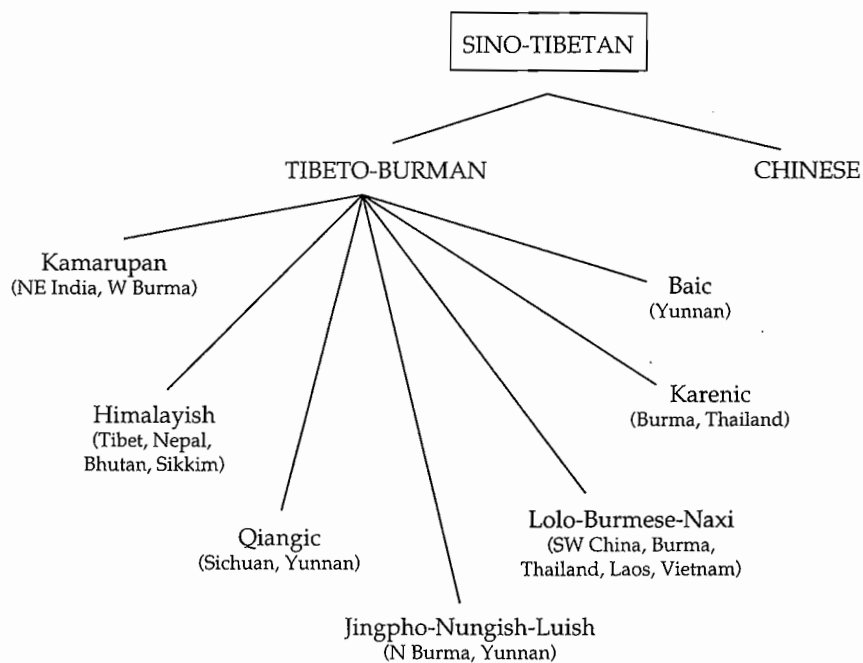Figure 14.6. *Schematic grouping of Sino-Tibetan languages (Conspectus).*

Figure 14.7. *The STEDT working model of TB subgroups.*

were among the earliest TB peoples to penetrate into peninsular SE Asia, in the late first millennium AD, and were closely associated with the Mons, with whom they are said to have lived in a slave–master relationship. This heavy contact with Mon (as well as with Tai languages) undoubtedly explains the atypical Karenic SVO word-order,[44] which had led Benedict to banish Karen from 'TB proper', postulating a primary split between Chinese and 'Tibeto-Karen'. It now seems clear that Karenic is just another subgroup of TB. (For one thing, the tones of Proto-Karen correspond quite regularly to those of Lolo-Burmese.) The reconstruction of Proto-Karenic is reasonably well advanced, with fundamental contributions by Haudricourt (1942–45; 1975), Jones (1961), and Burling (1969).

- **Lolo-Burmese**[45] is by far the best-studied branch of TB from a comparative-historical point of view, and we now have a very good picture of Proto-Lolo-Burmese (PLB). These languages (especially the northern and central Loloish branches) are highly tonal, with few initial clusters and often no syllable-final consonants. The Burmish languages (especially Written Burmese, attested since AD 1111) are more conservative, preserving precious clues to the phonological shape of PLB. The Naxi/Moso language is close to the Loloish nucleus, and is of special interest because of its complex, hieroglyphic-like writing system, now virtually unknown except to a few scholars. The Yi (Lolo) language itself, spoken by about six million people, has a syllabic writing system of considerable antiquity, which has recently been simplified and standardized (cut down from 800-odd symbols to about 350), and is coming into increased use.[46]

- **Jingpho** (also known as Kachin) is an important language spoken in northernmost Burma and adjacent areas of China and India. Since it shows phonological and lexical similarities with several other branches of TB, Benedict considered it to be 'genetically central' in the TB family, just as it is geographically central. It preserves final nasals and stops well (though $*$-k > -ʔ in native words), along with a good number of initial clusters. Most strikingly it contains a high percentage of 'sesquisyllabic' words,[47] which preserve in their 'minor syllables' many of the prefixes set up for PTB. It has long been noticed that Jingpho exclusively shares certain items of core vocabulary like *fire* and *sun* with the Bodo-Garo (= Barish) and Northern Naga (= Konyak) groups, leading Grierson & Konow (1903–28) to set up a 'Bodo-Naga-Kachin' nucleus. This idea has subsequently been revived by Burling (1971; 1983), who has dubbed this grouping the 'SAL' languages (named for their shared word for *sun*). I myself have entertained the idea of a 'Ji-bur-ish'

supergroup consisting of Jingpho and Lolo-Burmese, on the basis of certain suggestive but exception-ridden correspondences between their tonal systems (Matisoff 1974; 1991b). All in all, however, it seems much preferable to consider Jingpho to be the main representative of a separate subgroup of TB, which probably also includes the Nungish languages (a dialectally highly diversified group of Northern Burma and Yunnan: see LaPolla 1987), and the obscure and moribund Luish group of Manipur and N. Burma.

- One of the most exciting recent developments in TB studies is the discovery of a 'new' branch of the family, hitherto virtually unknown to Western scholars: the **Qiangic** languages of Sichuan and Yunnan (see Sun Hongkai 1985; 1990), containing around a dozen languages notable for their phonological complexity and the extreme dialectal diversity shown by each language. Historical reconstruction of this subgroup is still in its infancy, undoubtedly because much of the data so far recorded seems to be overtranscribed, so that it is hard to tell which phonetic features are phonemic and which are redundant or allophonic. The Qiangic languages are characterized by initial consonant clusters comparable in complexity to those of Written Tibetan. (This is especially true of rGyalrong and Ergong, which seem to form a special nucleus within the group.[48]) Some languages in this group have developed typologically strange syllable-final consonants, owing to the fusion of second syllables in compounds, e.g. N. Qiang **tśhaz** 'seed' (cf. S. Qiang **zuə-za**). Most interestingly, tonal contrasts seem to be of demonstrably recent origin in this subgroup; some languages, like Qiang itself, have both tonal (S. Qiang) and non-tonal (N. Qiang) dialects. Morphologically, the Qiangic languages are characterized by a unique system of 'directional prefixes' on verbs, which specify the literal or figurative spatial orientation of the verbal event (up, down, in, out, over, under, back, around, etc.).

I have saved the two most complicated and problematic subgroups for last: Himalayish and Kamarupan.

- The languages lumped together under the rubric of **Himalayish** do not constitute an orderly genetic group, but rather an aggregate of linguistic nuclei spoken in the same general geographical area (Tibet, Nepal, Bhutan, Sikkim, Himachal Pradesh). These include, first and foremost, Tibetan and its close relatives (the so-called Bodic languages). Written Tibetan, attested since the early seventh century AD, is our most ancient source of data in the TB family; it faithfully preserves all the prefixes set up for PTB, as well as the medials */ -r- -l- -y- / (but not * -w-), and the full array of final consonants imputed to the proto-language, */ -p -t -k -m -n -ŋ -r -l -s/. Other well-established

nuclei in this area include the Tamang-Gurung-Thakali group of W. Central Nepal;[49] and the Rai or Kiranti group of E. Nepal (including what used to be known as Bahing-Vayu),[50] characterized by complex agreement morphology on the verb. Other important TB languages of Nepal (Kham, Magar, Chepang, Sunwar) have yet to be assigned with certainty to any larger grouping. Newar, spoken in the Kathmandu valley, is of considerable cultural importance, with a highly Sanskritized literature that goes back many centuries; the Newars have historically been peripatetic traders and merchants, and their language has been influenced more by Indo-Aryan than other languages of Nepal. The conservative Lepcha language of Sikkim, as well as the several languages of Bhutan (including the national language Dzongkha), are receiving renewed attention, as are the westernmost TB languages of Himachal Pradesh (e.g. Pattani). Himalayish has become one of the 'growth points' in TB studies, with yearly Himalayish Symposia being held in various parts of the world, most recently in Santa Barbara, Pune, and Kathmandu.[51]

- The most diverse and unruly branch of the TB family comprises the languages of NE India and Western Burma, to which I have assigned the purely geographical label of **Kamarupan**, from an old Sanskrit term for part of NE India, Kāmarūpa. This area, which includes the Chin Hills of Burma and the modern districts of India called Assam, Meghalaya, Mizoram, Manipur, Tripura, Nagaland, and Arunachal Pradesh, is the center of diversification of the whole TB family. Nagaland alone, with an area of only 6350 square miles, is home to some 90 TB languages and dialects.[52] See the complex ethnolinguistic map in Figure 14.8.

The Kamarupan complex of languages includes those that belong to three traditional groups, Kuki-Chin-Naga, Bodo-Garo (called 'Barish' by Shafer), and Abor-Miri-Dafla (Shafer's 'Mirish'), plus a number of languages that seem to fall outside any of these, including Mikir, Meitei (the chief language of Manipur, also known as Manipuri, with a literary tradition dating at least from the eighteenth century), and Mru (spoken in the Chittagong Hills of Bangladesh and the Arakan region of Burma).[53] The numerous closely-related Kuki-Chin languages of Western Burma and Mizoram are quite conservative phonologically, preserving final stops and nasals, and often even final liquids and/or traces of final *-s in the shape of -ʔ. Some of them are toneless, while others have relatively simple tone systems compared to phonologically more depleted languages.[54] The verbs of many Chin languages have two morphophonemic variants with a complex distribution, roughly corresponding to their appearance in main vs subordinate clauses, but also dependent on a number of other factors. The relatively few
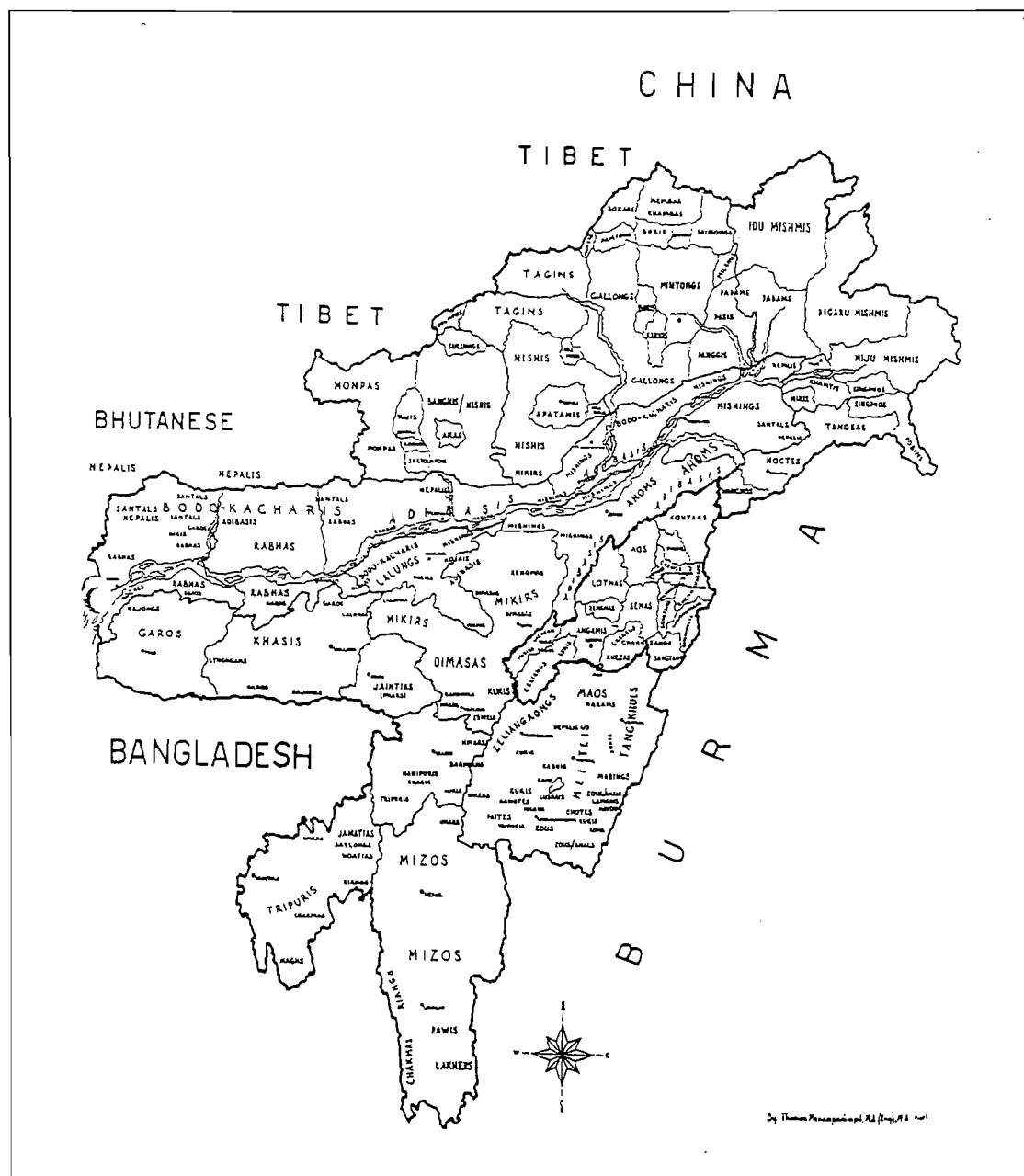
Figure 14.8. *Linguistic diversification in Kamarupa.*

Bodo-Garo languages, spoken in Meghalaya and Tripura, are mostly toneless, also with good preservation of final consonants, sometimes via final echo-vowels (e.g. Garo do?o 'bird'). The Naga languages (spoken in Nagaland and SE Arunachal Pradesh: see Fig. 14.9) are a large group, which can be further subdivided into five smaller nuclei on the basis of various phonological and lexical
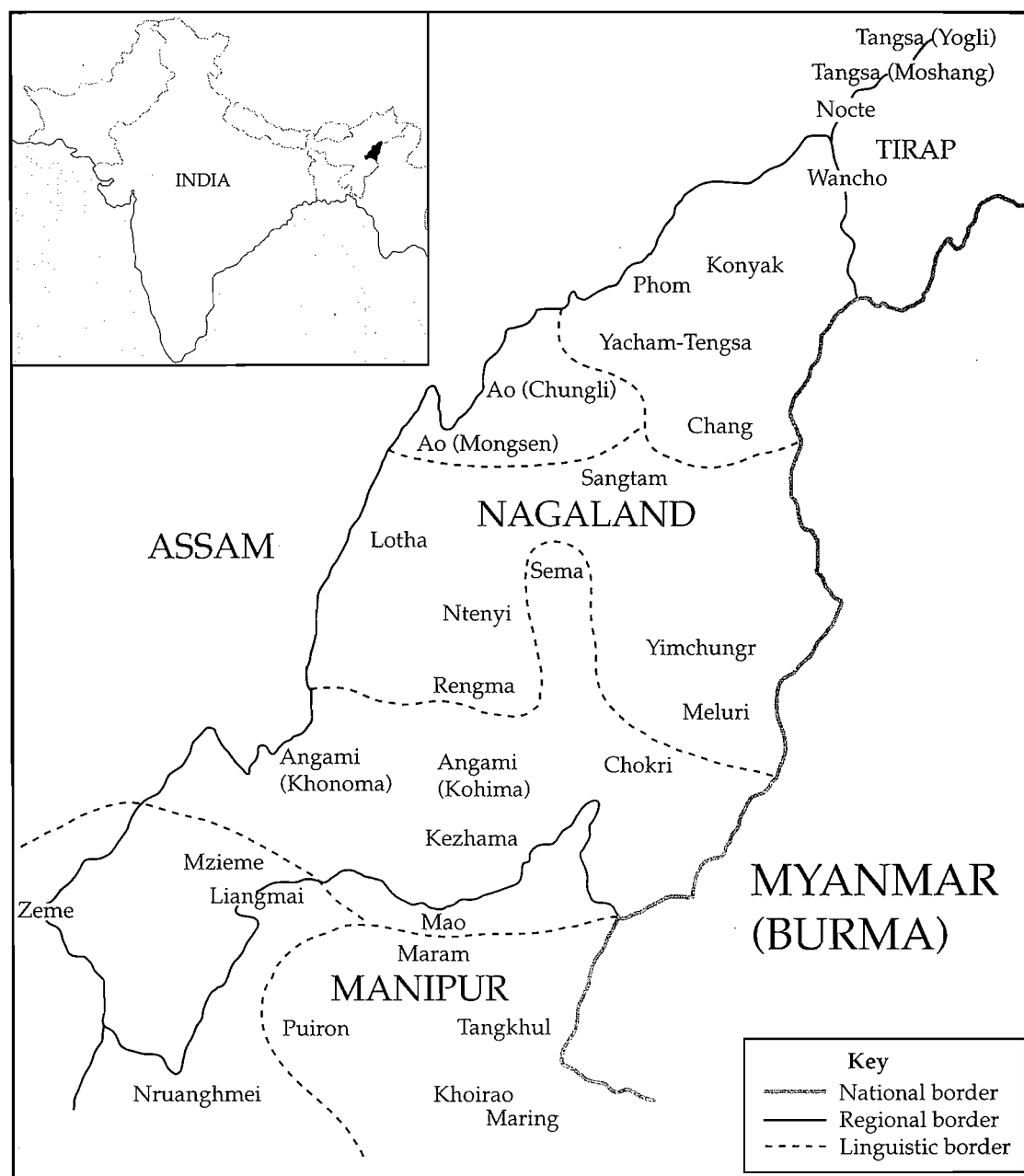
Figure 14.9. *Location of the principal Naga languages.*

criteria (see Fig. 14.10). Some (especially the Angamoid group) are fully tonal, others only marginally so (Lothoid), while some lack tones altogether (e.g. the Luhupa group). As usual, there seems to be an inverse relationship between the degree of consonantal preservation and the proliferation of tones. The northern Naga languages (those with certain lexical resemblances to Jingpho: see above) have been particularly well studied (Marrison 1967; French 1983). The
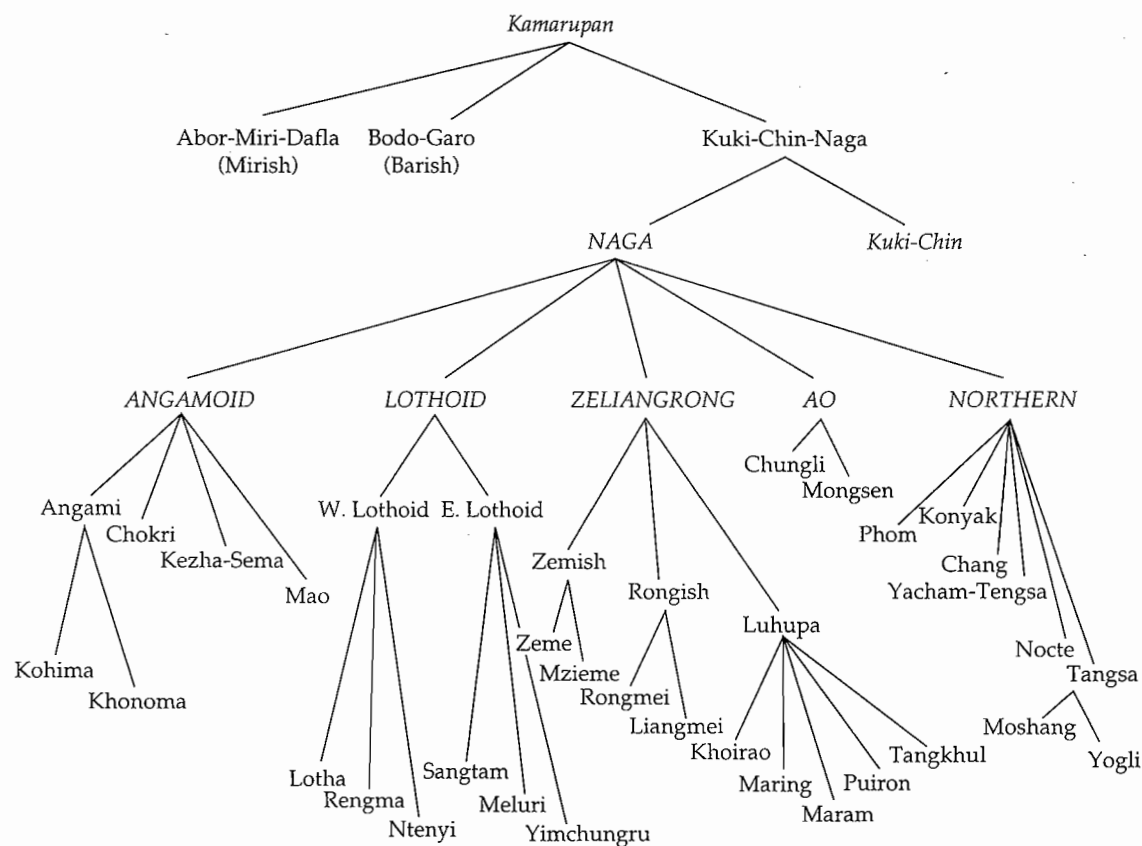
Figure 14.10. *The subgrouping of the Naga languages.*[55]
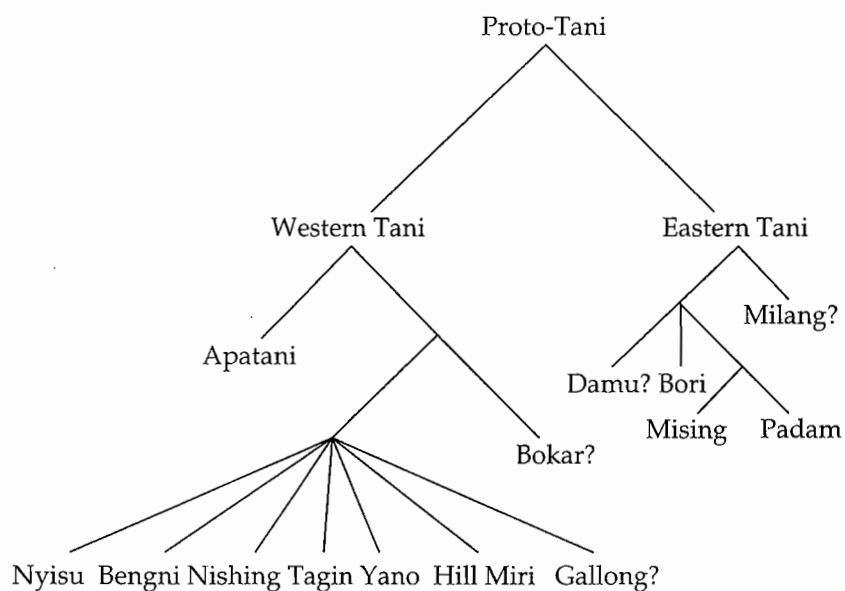
Figure 14.11. *The Tani group of Kamarupan languages.*

Abor-Miri-Dafla group (Arunachal Pradesh) contains some of the most lexically aberrant languages from the standpoint of TB in general, with particularly strange numerals (see Matisoff 1997b). A recent dissertation by J.T. Sun (1993) has greatly clarified the interrelationships of a subset of these languages to which he has given the name 'Tani' (see Fig. 14.11).

4. Old-fashioned alternatives to lexicostatistics

What Benedict said back in 1940,[56] 'supergroups within Tibeto-Burman cannot safely be set up at the present level of investigation' (1972, 11), remains substantially true today. This should not be unduly discouraging, however, since the same can be said of Indo-European after nearly 200 years of scholarship. Historical linguists have traditionally relied heavily on the concept of *shared innovations* in their subgrouping efforts; but, as the proponents of the 'wave theory' of linguistic change pointed out long ago, there is an indefinitely large number of 'innovations' in the history of any language family, and the selection of different innovations can yield contradictory results. See O. Schrader's classic chart of overlapping phonological and morphological innovations among the established subgroups of IE, reproduced here as Figure 14.12.[57]

Judicious selection of particular innovations can in fact give results which are at variance with common sense. It is obvious to the expert that the Central Loloish languages
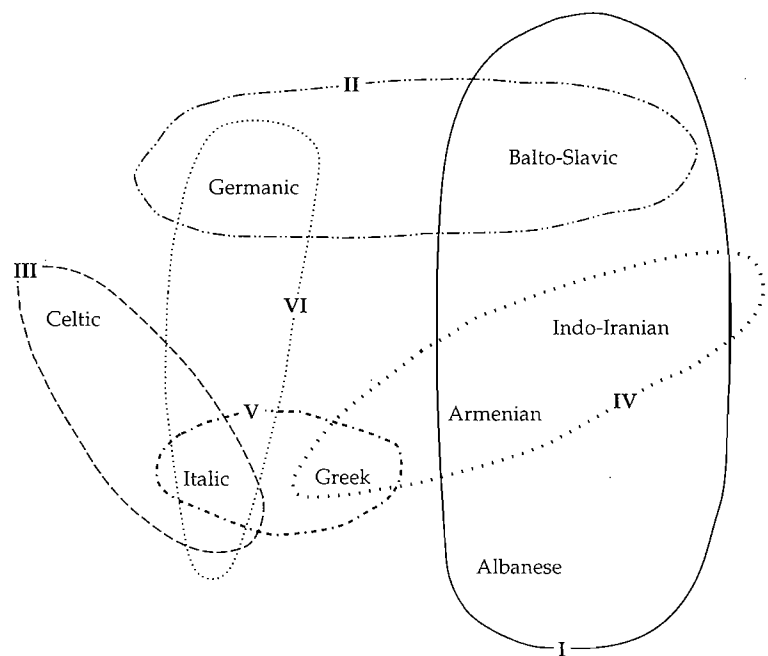
Figure 14.12. *Some overlapping features of special resemblance among the Indo-European languages: I) Sibilants for velars in certain forms; II) Case-endings with [m] for [bh]; III) Passive-voice endings with [r]; IV) Prefix ['e] in past tenses; V) Feminine nouns with masculine suffixes; VI) Perfect tense used as general past tense. (After Bloomfield 1933, 316.)*
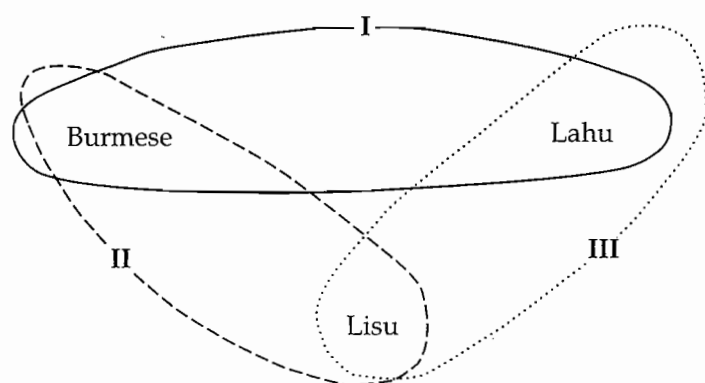
Figure 14.13. *Overlapping shared innovations in Lolo-Burmese: I) Devoicing of Proto-Lolo-Burmese *voiced obstruents; II) Fronting of *a before velar stop; III) Tonal split in checked syllables. (Reproduced from Matisoff 1978, 11.)*

Lahu and Lisu, with some 80 per cent of shared basic vocabulary, are more closely related to each other than either is to Modern Burmese. Yet it is not hard to find phonological isoglosses that pair up the three languages in three different ways, as if they were all on a genetic par (see Fig. 14.13).

It could be argued that these innovations are not of equal importance. But the point is that there is no mechanical way to assign relative weight to conflicting patterns of innovative rule-sharing. The analyst must make such decisions on the basis of hard-earned intuitions as to what is critical and what is of lesser importance. One is reminded of the Dutch epigram about generative grammar,[58] *Het beoefening van de generatieve grammatik is hetzelfde als zelfverstopte Paaseieren terugfinden* ('The operation of generative grammar is the same as finding Easter eggs one has hidden himself').

One could say the same about lexicostatistics. If we already know that certain languages are closely related, since we have discovered that they display regular sound-correspondences (and perhaps also striking morphological similarities), and we are sure we can tell genuine cognates from mere lookalikes or borrowed forms, then lexicostatistics can confirm what we already knew — and maybe even help us decide relative degrees of closeness of relationship. If, on the other hand, we are dealing with languages that are distantly related at best, or whose genetic affiliations are in doubt, we are confronted with the 'Garbage in, garbage out' problem.

Still, there is no reason to despair. Even if glottochronology is not the 'magic bullet' that it was once supposed to be, historical linguists are not defenceless. We can limp along as we always have, using a multiplicity of criteria — geographical, surface phonological, morphosyntactic, lexical — always realizing that language is an aspect of human behaviour that is highly sensitive to outside interference by a host of extralinguistic factors.

5. Perils and caveats

There is a constant temptation among linguists to be the first to 'discover' a new subgroup within a recognized family (cf. Thurgood 1984; van Driem 1997), or a farflung relationship between language families on opposite sides of the globe.

The easiest proposals to dismiss as chimerical are those which depend entirely on surface similarity among forms from modern languages, without bothering to attempt reconstructions of proto-forms in the languages to be compared. In this category belongs Greenberg 1987, an attempt to group all the languages of the Western Hemisphere into three families: 'Amerind', Eskimo-Aleut, and Na-Dene.[59] As an exercise in megalocomparison, I had no difficulty in coming up with about 50 good-looking 'cognates' between Amerind and Proto-Sino-Tibetan or Proto-Tibeto-Burman.[60] This exercise took about three hours by the clock, but did not fill me with any exhilaration, only a vague depression that this sort of thing was so easy.

Even more dangerous are serious megalocomparative efforts that are clothed in the trappings of the traditional comparative method, that use reconstructed forms, and that purport to show 'regular correspondences' among 'cognates'.[61] Despite the occasional brilliance of such endeavours,[62] what they all have in common is tortured sound-correspondences, disregard of counter-examples to 'sound laws', and unconstrained semantic latitude.

We all have to take a deep breath, and admit that the comparative method has intrinsic, ineluctable limitations. The remote linguistic past is a dark tunnel, and the torch of the comparative method can only illuminate it so far. It goes without saying that lexicostatistics — which is at best but a feeble adjunct to the comparative method — cannot push the light back any farther.

Notes

1. As the inventor of this hideous neologism, I suggest that it should be pronounced with antepenultimate stress, viz. [nɛkrɔʷhɪpɔ́məkɪy].
2. The most interesting exception is Xixia (also known as Tangut), once the language of a powerful empire in the Western China/Tibetan borderlands, extinct since the Mongol conquests of the thirteenth century. Its extremely complicated logographic writing system has been largely deciphered through analysis of bilingual Chinese and Tibetan texts (many of which were unearthed in the famous Tun Huang caves), though much uncertainty remains as to its phonological reconstruction, with competing systems offered by Russian, Japanese, and Chinese scholars. It is now thought that Xixia belongs to the newly proposed *Qiangic* subgroup of TB.
3. I call these areas of influence the *Sinosphere* and the *Indosphere*. An extreme case is the Bai language of NW Yunnan, some dialects of which are said to have as much as 75 per cent of their lexicons consisting of Chinese loanwords.

4. I personally believe this agreement morphology (traditionally called 'verb pronominalization' by Tibeto-Burmanists) to be quite secondary with respect to PTB, though a number of scholars (especially those whose experience is entirely with Himalayish languages) disagree.

5. Several *prefixes* are reconstructible for PTB, though they are mostly quite vague in meaning, and have left only very indirect traces in Sinospheric subgroups like Lolo-Burmese. Many prefixes reconstructible at the subgroup level are of demonstrably secondary origin, deriving from the reduction of initial syllables in compounds.

6. This is certainly true in general, but occasionally even single-segment forms are reconstructible. I have reconstructed a solid PTB numeral *a 'one' on the testimony of Qiang (Sichuan) and Hruso of Arunachal Pradesh (two obscure languages that could not have been in contact), a root which has yet to be uncovered elsewhere in TB (Matisoff 1997a, 23).

7. Those familiar with children's books might also be reminded of the tangled interconnected horns of a Dr Seuss animal. Dixon 1997 is a recent study emphasizing that *convergence* (diffusional phenomena, contact influence) is just as important as divergence in the history of languages.

8. See the critique in Matisoff 1978, 133–40.

9. This 'CALMSEA' list (Culturally and Linguistically Meaningful for Southeast Asia) appears in Matisoff 1978, 284–96, and is reproduced here as Appendix 14.1.

10. Personal communication, Theraphan L. Thongkum, May 1999.

11. I used this list to good advantage in March 1996 in Kunming, while working on Prinmi (Pumi), a TB language of the Qiangic group.

12. Thus in Tibeto-Burman (TB) languages one would never expect to find cognates for the names of most insects or other lower animals (e.g. spiders and snails). In Lahu, e.g., such words show extreme dialectal variation and tend to be long polysyllabic compounds, pointing to fresh creation by each generation of children (who use such creatures for playthings).

13. G. Diffloth observes that core vocabulary is likely to be replaced by taboo in the Nicobarese and Aslian subgroups of Mon-Khmer, while unusual words seem to have greater survival value (p.c., 1985). In his defense of the 'Austric hypothesis' (below §2), Diffloth (1993) relies especially on arcane lexical items like *scruff* and *smegma*.

14. One thinks of the survival of Yiddish words and expressions like *shmate* 'rag; worthless object'; *nudnik* 'pest', *shlep* 'drag laboriously', *hakn a tshainik* 'nag or prate noisily' (lit. 'bang on a teakettle') in the speech of otherwise monolingual third-generation American Jews.

15. Dixon (1997) refers to this as the 'punctuated equilibrium' model of linguistic change.

16. I have called this differential receptivity a language's 'RTQ' or *replacement tolerance quotient* (Matisoff 1978, 96–9), but have recently modified this concept to 'ATQ' or *accretion tolerance quotient* (Matisoff 1998b), since a borrowing does not necessarily displace a previously existent item from the receptive language.

17. For some strange reason, English deverbal nouns in -*ing* have received a particularly warm welcome into French: *le meeting, le building, le smoking* ('smoking jacket').

18. Benedict (1975) has persuasively claimed that in prehistoric times Chinese may well have borrowed extensively from Tai, since the Tai were probably then more numerous than Han Chinese in the area south of the Yangtze. In more recent times, of course, the borrowing has gone overwhelmingly in the opposite direction.

19. It must be said, however, that in the post-World War II period, the ATQ of German with respect to English has drastically changed, so that the speech of the younger generation is peppered with as much English as that of their French counterparts.

20. The same is true of Japanese, another language with an enormously high ATQ, which has by now absorbed virtually the entire Chinese and English lexicons while very seldom abandoning the original native word.

21. This immediately relies on two questionable assumptions: (a) that it is always possible to decide on one particular 'usual' word in the target language to translate each concept on the core vocabulary list; and (b) that the language in which this list is written itself (e.g. English) has a single clear and unambiguous word for each of the core concepts.
22. This discussion is somewhat abridged from Matisoff 1978, 100–104.
23. It must also be admitted that this kind of detailed historical information is unavailable for most TB languages.
24. What MARROW *is* related to conceptually (both within and without the SEA'n linguistic area) is BRAIN.
25. A desirable, but perhaps unattainable, goal would be to create a data base of universal vs areal vs language-specific semantic associations, a mapping of semantic space that I have wistfully called the 'human semome project'.
26. The syllable ha 'hundred' is a classifier, and must always be preceded by a numeral, as in tê ha 'one hundred'; the -pa in 'moon' is a meaningless suffix, ubiquitous in TB (cf. WT zla-ba 'moon'); the -tē in 'tongue' may once have had an independent meaning, but now occurs nowhere else in the language; the prefix ɔ̀- in 'spirit' (< PTB *ʔaŋ-) occurs as a bulk-provider before hundreds of Lahu roots; the particle ve in 'winnow' is a nominalizer that appears in the citation form of verbs (much like English *to*), serving incidentally to distinguish verbs or adjectives from any homophonous nouns.
27. I first used this phrase in my review of Denise Bernot's otherwise excellent *Dictionnaire Birman-Français* (Matisoff 1987).
28. For this reason I made an effort in my Lahu dictionary to list most non-initial syllables of compounds as head-entries (lemmata).
29. This chart is adapted from Nishida (1967, 68) and Matisoff (1978, 64). See the extended discussion in the section 'Compounding and genetic relationship' in the latter work, pp. 58–72.
30. The four roots in these forms do not begin to exhaust the morphemes that appear in compounds for HEAD in the TB family as a whole.
31. An analogy would be cases where speakers of English differ widely in the compounds they use for a given concept, e.g. the device from which piped running water emerges: *water-tap, water-faucet, water-spigot, water-cock, water-spout*, etc.
32. This epigram is due to Mary R. Haas.
33. Cf. the discussion of 'pseudo-micritizing devices' to make megalocomparisons look plausible (Matisoff 1990a, 116–17). These include the liberal use of slashes, brackets, and parentheses, and in extreme cases can lead to 'split cognates', i.e. supposedly cognate forms in daughter languages that have no reflex of any phoneme of a polysyllabic proto-form in common. Cf. the Proto-Austro-Tai word for 'rabbit' **m/l̩oto[k/a], which is said to yield Proto-Tai *thoʔ but Proto-Hmong-Mien *ʔ[l̩]ɔ(t) (Benedict 1975, 359–60).
34. For more details see Matisoff 1978, 21–58.
35. This type of alternation is equally characteristic of Chinese. See the list of Cantonese examples in Bauer & Benedict 1997, 92–4.
36. This alternational pattern was already recognized by Wolfenden (1929; 1936; 1937).
37. The cognacy of the Chinese form for 'ceremony to expel demons' has been questioned, since it does not show the expected shift of PST *-a > Chinese -o. See Matisoff 1978, n. 140.
38. Matisoff 1998b is entirely devoted to this phenomenon.
39. See Matisoff 1978, 47–54 *et seq.*
40. A partial exception to this generalization is the ablaut relationships in the principal parts of WT verbs.
41. For reasons of space, we forgo discussion of earlier subgroupings of TB, the most influential of which were Grierson & Konow 1903–28 and Shafer 1966/67.
42. This in itself is not a serious objection, since the same can be said of Indo-European: branches like

Albanian, Armenian, Greek, Tocharian, and Hittite contain dialects of a single language, while subgroups like Indo-Aryan, Germanic, Balto-Slavic, Italic, and Celtic are highly ramified.

43. I have found several etyma which Bai and Lolo-Burmese have in common, but which are not shared by Chinese, including CRAZY (PLB *ru²) and EGG (PLB *ʔu³).

44. All branches of TB except for Baic and Karenic have SOV order.

45. Also known as Burmese-Lolo, Burmese-Yi, and Burmese-Yipho. The term 'Lolo' is now felt to be pejorative in China, where 'Yi' is currently preferred. For a list of references to works on the historical phonology of this subgroup, see Matisoff 1991a.

46. An interesting feature of this syllabary is that the tone is 'built in', so that each combination of initial consonant and vowel is written with totally different symbols according to the syllable's tone.

47. This term was introduced in Matisoff 1973 to characterize words that are 'a syllable and a half' in length, consisting of a 'minor syllable' with schwa vocalism followed by a 'major syllable' with a full vowel.

48. rGyalrong (also spelled rGyarong, Gyarong, Jiarong, etc.) was considered of uncertain affiliation in STC (see Fig. 14.6). It now seems likely that the extinct Xixia language (above, note 2) also belonged to the Qiangic group, with its most plausible modern descendant now known as Tosu.

49. See Mazaudon 1973; 1976. Other languages in this group include Manang, Chantyal, and Narphu.

50. See Michailovsky 1988.

51. Quite properly the Indo-Aryan languages of the region (e.g. Nepali, Hindi, Maithili) are also well represented at these symposia, since their influence on the coterritorial TB languages is pervasive.

52. Kamarupa has also been populated by speakers of Mon-Khmer (Khasi), Munda (Santali), Tai (Khamti and Ahom, the latter now extinct, but whose name survives in the word *Assam*), and Indo-Aryan (Assamese, Bengali).

53. According to the Swiss linguist Lorenz Löffler (p.c.), there is a Mru dialect with secondarily developed SVO word-order, almost unique in TB except for Bai and Karenic (see above). Wheatley (1985) cites certain Northern Loloish languages that also are beginning to postpose certain nominal arguments to the verb, under heavy Chinese influence. This illustrates the danger of using syntactic arguments uncritically for subgrouping purposes.

54. Weidert 1987 is a sophisticated and data-packed treatment of the tonology of Kamarupan languages, though it is marred by its disorganized presentation and overly formalistic approach.

55. Since the sound correspondences among these languages have not yet been worked out, this classification is based on a variety of surface-phonological criteria, including the presence of prenasalized obstruents, the degree of preservation of syllable-final consonants, the complexity vs simplicity of tonal systems, etc.

56. In the original MS of *Sino-Tibetan: a Conspectus*, finally published with extensive annotations in 1972.

57. This was quoted in Bloomfield 1933, 316 (where I first saw it), and repeated in Matisoff 1978, 5.

58. Cited to me in Leiden in the mid-1980s.

59. For a more-or-less serious review of this book, see Matisoff 1990a. At least Greenberg's method of 'mass comparison', which had worked well for him in Africa where the languages were better known, does not limit itself to using a list of only 100 or 200 items.

60. See Matisoff 1990b, reproduced here as Appendix 14.2.

61. With respect to the affiliations of Chinese, cf. such recent proposals as Sagart's 'Sino-Austronesian', Robert Jones' 'Sino-Mayan', and Sergei Starostin's 'Sino-Caucasian'. Greenberg's inclination would be to accept all these proposals, thereby proving that Austronesian, Mayan, and Caucasian were also all related to each other.

62. I am thinking especially of Benedict 1990.

References

Bauer, R.S. & P.K. Benedict, 1997. *Modern Cantonese Phonology*. (Studies and Monographs 102.) Berlin: Mouton de Gruyter.

Benedict, P.K., 1972. *Sino-Tibetan: a Conspectus*, contributing editor, J.A. Matisoff. Cambridge: Cambridge University Press.

Benedict, P.K., 1975. *Austro-Thai Language and Culture, with a Glossary of Roots*. New Haven (CT): Human Relations Area Files Press.

Benedict, P.K., 1990. *Japanese/Austro-Tai*. (Linguistica Extranea. Studia 20.) Ann Arbor (MI): Karoma Press.

Bloomfield, L., 1933. *Language*. New York (NY): Holt, Rinehart & Winston.

Burling, R., 1969. Proto-Karen: a reanalysis. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics* 1, 1–116.

Burling, R., 1971. The historical place of Jinghpaw within Tibeto-Burman. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics* 2, 1–54.

Burling, R., 1983. The SAL languages. *Linguistics of the Tibeto-Burman Area* 7(2), 1–32.

Diffloth, G., 1993. The Lexical Evidence for Austric, So Far. Unpublished paper presented at CAMAC (Conference on Austronesian-Mainland Asian Connections), Honolulu. 14 pp.

Dixon, R.M.W., 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

French, W.T., 1983. Northern Naga: a Tibeto-Burman Mesolanguage. Unpublished Ph.D. dissertation, City University of New York.

Greenberg, J.H., 1987. *Language in the Americas*. Stanford (CA): Stanford University Press.

Grierson, G.A. & S. Konow (eds.), 1903–28. *Linguistic Survey of India*, vol. III, parts 1–3: *Tibeto-Burman Family*. Delhi, Varanasi & Patna: Motilal Banarsidass. (Reprinted 1967.)

Haudricourt, A.-G., 1942–45. Restitution du karen commun. *Bulletin de la Société de Linguistique de Paris* 42(1), 103–11.

Haudricourt, A.-G., 1975. Le système de tons du karen commun. *Bulletin de la Société de Linguistique de Paris* 70, 339–43.

Jones, R.B., 1961. *Karen Linguistic Studies: Description, Comparison, and Texts*. Berkeley & Los Angeles (CA): University of California Press.

Karlgren, B., 1933. Word families in Chinese. *Bulletin of the Museum of Far Eastern Antiquities, Stockholm* 5, 5–10.

LaPolla, R., 1987. Dulong and Proto-Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 10(1), 1–43.

Marrison, G.E., 1967. The Classification of the Naga Languages of Northeast India. 2 vols. Unpublished D.Phil. thesis, University of London.

Matisoff, J.A., 1972. *The Loloish Tonal Split Revisited*. (Research Monograph 7, Center for South and Southeast Asia Studies.) Berkeley (CA): University of California.

Matisoff, J.A., 1973. Tonogenesis in Southeast Asia, in *Consonant Types and Tone*, ed. L.M. Hyman. Los Angeles (CA): University of California, 71–95.

Matisoff, J.A., 1974. The tones of Jinghpaw and Lolo-Burmese: common origin vs independent development. *Acta Linguistica Hafniensia* 15(2), 153–212.

Matisoff, J.A., 1978. *Variational Semantics in Tibeto-Burman*. Philadelphia (PA): Institute for the Study of Human Issues.

Matisoff, J.A., 1987. Review of Denise Bernot, *Dictionnaire Français-Birman*, fascicules 1–8. *Bulletin of the School of Oriental and African Studies* 50(1), 191–5.

Matisoff, J.A., 1988a. *The Dictionary of Lahu*. London, Berkeley & Los Angeles (CA): University of California Press.

Matisoff, J.A., 1988b. Universal semantics and allofamic identification — two Sino-Tibetan case-studies: STRAIGHT/FLAT/FULL and PROPERTY/LIVESTOCK/TALENT, in *Languages and History in East Asia*, ed. A. Sato. Tokyo: Shokado, 2–14.

Matisoff, J.A., 1990a. On megalocomparison. *Language* 66(1), 106–20.

Matisoff, J.A., 1990b. (under pseudonym *Ffositam A. Semaj*) Proto-Sino-Tibetan and Amerind lookalikes. Unpublished manuscript. 5 pp.

Matisoff, J.A., 1991a. Sino-Tibetan linguistics: present state and future propects. *Annual Review of Anthropology* 20, 469–504.

Matisoff, J.A., 1991b. Jiburish revisited: tonal splits and heterogenesis in Burmo-Naxi-Lolo checked syllables. *Acta Orientalia* (Copenhagen) 52, 91–114.

Matisoff, J.A., 1994. Regularity and variation in Sino-Tibetan, in *Current Issues in Sino-Tibetan Linguistics*, eds. H. Kitamura, T. Nishida & Y. Nagano. Osaka: National Museum of Ethnology, 36–58.

Matisoff, J.A., 1997a. *Sino-Tibetan Numeral Systems: Prefixes, Protoforms, and Problems.* (Pacific Linguistics B-114.) Canberra: Australian National University.

Matisoff, J.A., 1997b. Issues in the subgrouping of Tibeto-Burman in the post-Benedict era. Unpubl. paper presented at the 30th International Conference on Sino-Tibetan Languages and Linguistics, Beijing.

Matisoff, J.A., 1998a. Dàyáng Pumi phonology and adumbrations of comparative Qiangic. *Mon-Khmer Studies* 27, 171–213.

Matisoff, J.A., 1998b. Genetic vs areal linguistics in Southeast Asia: prosodic diffusibility in Southeast Asian languages. Unpublished paper presented at *Workshop on Areal Diffusion and the Genetic model of Linguistic Relationship*. Canberra: Research Centre for Linguistic Typology, Australian National University.

Matisoff, J.A., 1998c. An extrusional approach to *p-/w- variation in Sino-Tibetan. Unpubl. paper presented at the 31st International Conference on Sino-Tibetan Languages and Linguistics, University of Lund.

Mazaudon, M., 1973. *Phonologie Tamang.* Paris: Société d'Études Linguistiques et Anthropologiques de France.

Mazaudon, M., 1976. Consonantal mutation and tonal split in the Tamang subfamily of Tibeto-Burman. *Kailash* (Kathmandu) 6(3), 157–79.

Michailovsky, B., 1988. *La Langue Hayu: phonologie, morphologie, syntaxe.* Paris: Éditions du Centre National de la Recherche Scientifique.

Nishida T., 1967. A comparative study of the Bisu, Akha, and Burmese languages, part II. [in Japanese] *Tōnan Azia Kenkyū* (Kyoto) 4(3).

Sagart, L., 1993. Chinese and Austronesian: evidence for a genetic relationship. *Journal of Chinese Linguistics* 21, 1–62.

Shafer, R., 1966/67. *Introduction to Sino-Tibetan*, parts I & II. Wiesbaden: Otto Harrassowitz.

Sun Hongkai, 1985. Peoples and languages of the Six River Valleys and their genetic classification. [in Chinese] *Minzu Xuebao* 3, 99–274.

Sun Hongkai, 1990. Languages of the 'ethnic corridor' in Western Sichuan. *Linguistics of the Tibeto-Burman Area* 13(1), 1–31.

Sun, J.T., 1993. A Historical-comparative Study of the Tani (Mirish) Branch of Tibeto-Burman. Unpublished Ph.D. dissertation, University of California, Berkeley.

Thurgood, G., 1984. The Rung languages: a major new TB subgroup. *Proceedings of the 10th Meeting of the Berkeley Linguistic Society*, 338–49.

van Driem, G., 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies* 60(3), 455–88.

Vovin, A., 1990. Is Japanese related to Austronesian? Unpublished paper presented at Conference on Austronesian-Mainland Asian Connections, University of Hawaii. 10 pp.

Weidert, A., 1987. *Tibeto-Burman Tonology.* Amsterdam & Philadelphia (PA): Benjamins.

Wheatley, J.K., 1985. The decline of verb-final syntax in the Yi (Lolo) languages of southwestern China, in *Linguistics of the Sino-Tibetan Area*, eds. G. Thurgood, J.A. Matisoff & D. Bradley. Canberra: Pacific Linguistics C-87, 401–20.

Wolfenden, S.N., 1929. *Outlines of Tibeto-Burman Linguistic Morphology.* London: Royal Asiatic Society.

Wolfenden, S.N., 1936. On certain alternations between dental finals in Tibetan and Chinese. *Journal of the Royal Asiatic Society* 401–16.

Wolfenden, S.N., 1937. Concerning the variation of final consonants in the word families of Tibetan, Kachin, and Chinese. *Journal of the Royal Asiatic Society* 625–55.

Appendix 14.1. *The CALMSEA 200-word list. * in the English word column = words which also appear on the Swadesh 200-word list; †= corresponds to Swadesh's 'feather'.* STC = *Benedict's* Sino-Tibetan: a Conspectus *(1972); TSR =* The Loloish Tonal Split Revisited *(Matisoff 1972); JAM = James Matisoff; PLB = Proto-Lolo-Burmese. Numbers in brackets refer to page numbers in the respective volumes. (Adapted from Matisoff 1978, 284–96.)*

A. Body parts

1.	belly (exterior)*	STC *pu·k (358)
2.	blood*	STC *s-hwiy (222)
3.	bone*	STC *rus (6)
4.	ear/hear*	STC *g-na (453)
5.	egg*	STC *twiy (168)
6.	eye*	STC *mik ⪰ *myak (402)
7.	fat/grease*	STC *ryak (204); *tsil (pp. 16, 168, 173); *sa·w (272)
8.	foot*	STC *kriy (38)
9.	guts*	STC *pik (35)
10.	hair (head)*	STC *s-kra (115); *ney (292); *tsam (73)
11.	hair (body)*†	STC *mul (2); *tsam (73)
12.	hand/arm*	STC *lak (86)
13.	head*	STC *m-gaw ⪰ *s-gaw (490); *d-bu (p. 117)
14.	heart*	STC *s-niŋ (367)
15.	horn*	STC *kruw (37); *ruŋ (85)
16.	liver*	STC *m-sin (234)
17.	mouth*	STC *ak (106); *m-ka (468)
18.	neck*	STC *ke (251); *liŋ (96); *tuk (392)
19.	nose*	STC *s-na ~*s-na·r (101)
20.	skin/bark*	STC *s-graw (121); *kok (342)
21.	spit*	STC *m-ts(y)il (231); *m-tuk ⪰ *s-tu·k ⪰ *s-du·k (pp. 58, 75, 126, 132, 146); *twiy (168)
22.	tail*	STC *r-may (282)
23.	tongue*	STC *m-lay ⪰ *s-lay (281)
24.	tooth*	STC *s-wa (437)
25.	wing*	JAM *s-toŋ; *(s/k-)l-w-ak; *(s/p-)lyap
26.	nail/claw*	STC *m-(t)sin (74)

In Swadesh, but not in our list:

	back	STC *s-ga·l (p. 18); *s-nuŋ (354)
	knee	STC *du (p. 21); *(m-)ku·k (pp. 120, 159, 182); *put (7)
	leg	STC *(r-)kaŋ (70, 142)

The following are not in Swadesh, but are on our list:

27.	finger/toe	STC *m-yuŋ (355)
28.	palm	STC *pwa or *b-wa (418)
29.	penis	STC *li ⪰ *m-ley (262)
30.	vagina *or* breast/milk	STC *nuw (419) 'breast'; *dzo·p (69) 'suck'; TSR *b(y)et (5) 'vagina' (TSR 5)
31.	brain	STC *nuk (483)
32.	navel	STC *la·y; *s-tay (299)
33.	shit	STC *kliy (125); *n(y)ik (235); *s-baŋ (p. 21); *r-kyak ⪰ *s-kyak (pp. 26, 146); *e·k (26, 146)

34.	piss·	*STC* *ts(y)i ⩲ *ziy (77)
35.	sweat	*PLB* *ʔkrwiy₂
36.	snot	*STC* *s-nap (102)
37.	vomit	*PLB* *Npat (*TSR* 38); *STC* *on (343)
38.	marrow	*STC* *kliŋ (126)
39.	breath/life	*STC* *sak (485); *TSR* *C-sak (123)

B. Pronouns/kinship terms/nouns referring to humans

40.	person/human being*	*STC* *r-mi(y) (pp. 107, 119, 158)
41.	thou*	*STC* *na(ŋ) (407)
42.	I*	*STC* *ŋa (406) ⩲ *ŋay (285)
43.	child/son*	*STC* *za ⩲ *tsa (59)
44.	grandchild (nephew)	*STC* *b-liy (448)
45.	son-in-law	*STC* *ma·k (324); *TSR* *z-ma·k (153)
46.	name*	*STC* *r-miŋ (83)

C. Foodstuffs

47.	peas, beans	*STC* *be (253); *TSR* *s-nok (140)
48.	poison [antifood]	*STC* *duk ⩲ *tuk (472)
49.	mushroom/fungus	*STC* *g-muw (455) ⩲ (*JAM* *s-muw)
50.	liquor	*STC* *yu(w) (94)
51.	plantain/banana	*STC* *s-ŋak (477)
51a.	medicine/juice	*STC* *tsiy (65)
51b.	rice (in fields)	*STC* *b-ras (pp. 17, 123)
51c.	rice (cooked)	*PLB* *haŋ²

D. Animal names or animal products

52.	meat/animal*	*STC* *sya (181)
53.	bird*	*STC* *bya (177); *TSR* *s-ŋak (141)
54.	dog*	*STC* *kwiy (159)
55.	fish*	*STC* *ŋya (189)
56.	louse*	*STC* *s-rik (439); *sar ⩲ śar (pp. 15, 53, 84, 147, 172, 189)
57.	snake*	*STC* *b-ru·l (447)
58.	frog	*STC* *s-bal (15, 21, 107)
59.	insect, bug, vermin	*STC* *buw (27)
60.	bee	*PLB* *bya₂; *STC* *kwa·y (157); *was (17); *taŋ (494)
61.	dove	*STC* *m-kruw (118); *kuy 'pigeon' (495)
62.	monkey	*STC* *mruk (pp. 43, 112); *woy (314); *PLB* *m-yuk
63.	pig	*STC* *pwak (43); *PLB* *wak (*TSR* 168)
64.	fowl	*STC* *k-rak (pp. 88, 107, 187–9)
65.	otter	*STC* *s-ram (438)
66.	horse	*STC* *s-raŋ ⩲ *m-raŋ (145)
67.	ant	*STC* *rwak (199)
68.	bear	*STC* *d-wam (461)
68a.	leech	*STC* *r-pat
68b.	water leech	*STC* *m-li·t
69.	rat/rodent	*TSR* *k-r-wak (188); *STC* *bwiy 'bamboo rat' (173); *STC* *b-yuw 'rat, rabbit' (93)

E. Natural objects or phenomena; the inanimate landscape; vegetable and mineral kingdoms

70.	ashes*	STC *pla (137)
71.	cloud*	PLB *C-tim$_1$
72.	earth*	STC *r-ka (97); *mliy (152)
73.	fire*	STC *bar ≍ *par (220); *mey (290)
74.	flower*	STC *ba·r (1)
75.	fruit*	STC *sey (57)
76.	grass*	STC *mrak (149)
77.	leaf*	STC *la (486); *lap (321); *pak (40)
78.	moon*	STC *s-la ≍ *g-la (144)
79.	mountain*	PLB *kaŋ$_1$
80.	rain*	STC *r-wa (443)
81.	river/valley*	STC *klu·ŋ (127); *kor (349); PLB *laŋ^1
82.	road*	STC *lam (87)
83.	root*	STC *bul ≍ *pul (pp. 166, 173); *r-sa (442)
84.	salt*	STC *g-ryum (245); *tsa (214)
85.	sky*	STC *muw (488)
86.	smoke*	STC *kuw (256)
87.	star*	STC *s-kar (49)
88.	stick*	PLB *da$_1$ (cf. Lahu á-tà)
89.	stone*	STC *brak 'rock' (134); *r-luŋ (88); PLB *k-lok ≍ *k-loŋ (TSR 190)
90.	sun/day*	STC *nam (48); *niy (81); *tsyar (187)
91.	tree/wood*	STC *siŋ (233); PLB *sik (TSR 118)
92.	water*	STC *twiy (168)
93.	wind*	STC *g-liy (454)
94.	branch	STC *ka·k (327); *ku·ŋ (359)
95.	silver	STC *d-ŋul (pp. 15, 173)
96.	bamboo	STC *g-pa (44)
97.	shade/shadow	STC *g-rip ≍ *s-rip (p. 113)
98.	joint	STC *tsik (64)
99.	thorn (prick)	STC *tsow (276)
100.	night*	STC *ya (417)
101.	iron	STC *si·r (372); *syam (228)
102.	field	
102a.	swidden	PLB *hya$_1$
102b.	irrigated paddy field	PLB *ʔdan$_3$; *C-mi$_1$

F. Artefacts and social organization

103.	arrow	STC *m-da (pp. 96, 111–12, 118); *b-la (449)
104.	needle	STC *kap (52); PLB *k-rap (TSR 191)
105.	house	STC *kim ≍ *kyim ≍ *kyum (53)
106.	bow	STC *d-liy (463)
107.	boat	STC *m-liy (474)
108.	mortar	STC *tsum (75)
109.	village	STC *r-wa ≍ *g-wa (444); TSR *kak (22)

G. Spatial/directional

110.	left side*	STC *bay (47)
111.	right side*	STC *g-ya ≍ *g-ra (98)
112.	far* (v.)	STC *dzya·l (229); *wiy₂ (PLB)
113.	near* (v.)	STC *ney (291)
114.	year*	STC *niŋ (368); PLB *C-kok (TSR 34)

H. Numerals and quantifiers

115.	twenty/score	STC *(m)-kul (397)
116.	one*	STC *it (pp. 94, 162); *kat (p. 94); *t(y)ik ≍ *(g-)tyik (pp. 84, 94, 169, 189)
117.	seven	STC *s-nis (5)
118.	ten	STC *gip (16)
119.	hundred	STC *r-gya (164)
120.	two*	STC *g-nis (4)
121.	three*	STC *g-sum (409)
122.	four*	STC *b-liy (410)
123.	five*	STC *l-ŋa ≍ *b-ŋa (78)
124.	six	STC *d-ruk (411); PLB *C-krok (TSR 35)
125.	eight	STC *b-r-gyat (163)
126.	nine	STC *d-kuw (13)
127.	(be) many (v.)	STC *mra (148)

I. Verbs of utterance, body position or function

128.	be born	STC *braŋ (135); *kruŋ (382)
129.	sleep/lie down*	STC *ip (114); *mwiy (196); *n(y)it (236)
130.	weep	STC *krap (116); *ŋuw (79)
131.	laugh*	STC *m-nwi(y) (191); *rya-t (202)
132.	die*	STC *siy (232)
133.	awaken	STC *m-sow (295); PLB *nuw₂ ≍ *s-nuw₂
134.	cough	STC *su(w) (423)
135.	stand*	STC *g-ryap (246)
136.	sit*	STC *tu·ŋ (361)

J. Verbs of motion

137.	fall* (from a height)	STC *kla (123)
138.	climb, ascend	STC *l-tak (pp. 52, 110, 123); *syar 'rise' (p. 28, n. 90)
139.	descend	STC *yu(w) (101, n. 90); *ʔ-yuk; *zak (TSR 121)
140.	fly	STC *pur ≍ *pir (398); *pyam (p. 29, n. 93; p. 51, n. 171); *pyaw 'fly, swim float' (176)
141.	hide	STC *p(w)ak (46); PLB *wak ≍ *ʔ-wak (TSR 178)
142.	run/flee	STC *ploŋ (140)
143.	emerge	STC *twak (17); *pro (248)

K. Verbs of emotion, cognition and perception

144.	fear/frighten*	STC *grok ≍ *krok (473); *kri(y) (416); *b-ray (450)
145.	know*	STC *m-kyen (223); *syey (182)
146.	ashamed	STC *kyeŋ (162); *s-rak (431); *g-yak (452)

147. forget   *STC* *b-laʻp (335)
148. dream   *STC* *maŋ (82); *PLB* *s-mak ≍ *s-maŋ (*TSR* 144)
149. see*   *STC* *mraŋ (149)
150. smell*   *STC* *m-nam (464)

L. Stative verbs with human patients
151. thin*   *STC* *ba (25)
152. old*   *STC* *r-ga (445)
153. alive   *STC* *kruŋ (382); *s-riŋ ≍ *s-raŋ (404); *PLB* *dat (*TSR* 1)
154. ill   *STC* *na (80); *nyuŋ (194)
155. fat   *STC* *tsow (277)
156. itchy   *STC* *g-ya (451); *m-sak (465)

M. Stative verbs with non-human patients
157. full*   *STC* *bliŋ ≍ *pliŋ (142); *dyam ≍ *tyam (226)
158. long/tall*   *STC* *duŋ (20; p. 75, n. 231); *low (279); *s-riŋ (433)
159. sweet   *STC* *dz(y)im (71); *twi(y) (166)
160. cold*   *STC* *graŋ (120); *kyam (224)
161. bitter   *STC* *ka (8)
162. sour   *STC* *kri(y) (413); *s-kyur ≍ *suʻr (42)
163. red*   *STC* *kyeŋ (162); *r-ni (pp. 46, 91); *tsyak (184)
164. heavy*   *STC* *s-liy (95)
165. warm*   *STC* *lum (381)
166. round*   *STC* *z-lum (143); *wal (91)
167. ripe/well-cooked *STC* *s-min (432)
168. soft (to touch) *STC* *now (274)
169. white*   *STC* *bok (181); *ŋow (296); *plu (pp. 41, 46, 60–61, 89)
170. black*   *STC* *tyaŋ (225); *(s-)nak (pp. 88, 102, 155; *TSR* 142);
                      *sim ≍ *syim (380)
171. thick*   *STC* *r-ta-t (426); *tow (319); *tuʻk (356)
172. new*   *STC* *sar (147, 172, 189); *PLB* C-šik (*TSR* 126)
173. sharp*   *STC* *s-ryam (pp. 53, 171, 189); *tak (87; *TSR* 41)
174. lightweight   *STC* *r-gyaʻŋ (328)

N. Action verbs with human agent
175. eat*   *STC* *am (481); *dza (66)
176. drink*   *STC* *am (481); *PLB* *Ndaŋ₁ (*TSR* p. 15)
177. give*   *STC* *biy (427); *pek (pp. 101, 149)
178. tie*   *STC* *du-t ≍ *tu-t (421); *kik (484)
179. steal   *STC* *r-kuw (33)
180. lick   *STC* *m/s-lyak (211)
181. bite*   *STC* *gam (491); *hap (89); *kuʻk (388); *wa (424)
182. scratch/scrape* *STC* *hyak (230); *krak ≍ *Nkrak (*TSR* 96);
                      *kret ≍ *Nkret (*TSR* 97); *kut (*STC* 383); *pruk (*STC* 391)
183. cook/boil   *STC* *klak (124); *prut (131); *pryo (250)
184. grind   *STC* *krit (119)
185. wash*   *STC* *kruw (117); *m-s(y)il (492)
186. dig*   *STC* *klaw (269); *r-ko-t (420); *laʻy (288); *tu (258)

187.	let go; set free; loosen	*STC* *g-lwat (209)
188.	extinguish	*STC* *mit (374)
189.	blow*	*STC* *s-mut (75)
190.	buy	*STC* *b-rey (293)
191.	sew*	*STC* *d-rup (456); *TSR* *(ʔ-)grup ≋ *ʔdrup (63); *byar ≋ *pyar (*STC* 178)
192.	kill*	*STC* *g-sat (58)
193.	weave	*STC* *tak (17)
194.	rub*	*STC* *nu·l (365); *s(y)wiy (180)
195.	squeeze*	*STC* *nyap (192); *tsyur (188)
196.	shoot	*STC* *ga·p (219)
197.	kick	*PLB* *tek (*TSR* 14)
198.	sell	*STC* *par (35); *ywar (pp. 15, 51, 89)
199.	put, place	*STC* *s-ta (19)
200.	drive/hunt*	*PLB* *rak ≋ *Ngak (*TSR* 162)
200a.	burn*	*STC* *bar ≋ *par (220); *ploŋ (139); *tsow (275); *ka·ŋ (330); *duk ≋ *ʔduk (*TSR* 62); *put (*TSR* 8)
200b.	cut*	*STC* *da·n (22); *lep (351); *mrak (147); *ra-t (458); *ri·t (371); *tsywar (240); *kut (383); *tsyat (185); *tuk (387); *TSR* *Ntök ≋ *ʔtök (101); *ʔtwap ≋ *C-dwap (69)

Appendix 14.2. *Proto-Sino-Tibetan and Amerind look-alikes. The numbers next to the Amerindian forms refer to pages from Greenberg's* Language in the Americas (LIA) *(1987). Proto-Sino-Tibetan/Proto-Tibeto-Burman reconstructions are from Benedict's* Sino-Tibetan: a Conspectus (STC) *(1972) and various works of Ffositam (JAM). LTBA = The journal* Linguistics of the Tibeto-Burman Area. *Since such reconstructions as are to be found in* LIA *are only of individual subgroups, not of Proto-Amerind itself, we have sometimes simply selected those forms from individual Amerind languages that look most like ST/TM etyma. (Adapted from Matisoff 1990b.)*

above/head   PTB *l-tak 'ascend, above'
Chiquito ta, ita 'top of', Coroado takuen 'above'. Blackfoot -itoxk ('the -k may be an old locative suffix') (182, 232)

all/many   PTB *mra ≋ *mya
Lake Miwok muʔe, Mixe maj (145)

ant   PTB *g-rwak ≋ *p-rwak
Cochabamba Quechua čhaka, Huave čok, Bribri t'a-vak (183)

arm/give   PTB *bəy 'give'
Guahibo -pi 'arm', Amuesah apa 'give', Apolista apaj 'give', Toyeri mpe, upi 'arm' (184)

arrow   PTB *m-da ≋ *b-la
Opaie mãa 'arrow', Caddo baʔ 'bow' (184)

ashes   PTB *pla
Cayapo pra 'embers', Krenje pro, Toyeri palo, Uncasica bura (185)

aunt   PTB *m-na 'mother; o. sister; daughter-in-law'
Macca nana, Guahibo ena 'mother', Catuquina inai 'grandmother', Maku nō 'mother,
aunt' (186–7)

bathe/swim/wash   Mandarin fú 'float', fú-shŭi 'swim'
Karok ik-puh, Yana puu 'swim', Mixe puh 'wash', Klamath pew 'bathe, swim' (188)

bee¹   PTB *bya
Botocudo pā 'bee, honey' (190)

bee²/mosquito   Old Chinese *mi̯wən 'mosquito' (Mand. wén-zi)
Proto-Ge *meñ 'bee, honey'; Mataco apina 'mosquito', Choroti wona 'wasp',
Amahuaca wina 'bee, wasp', Proto-Tacanan *wini 'wax' (190)

belly¹   PTB *pam ⋉ *wam
Pokomchi pam (191)

belly²/guts/skin   PST *bok 'belly'
Tucano and Bara paga 'belly', Papury bok 'skin' (95)
Moseten vokko 'belly', Proto-Panoan *poko 'guts' (228)

bite   PTB *N-gwya 'chew' [JAM, LTBA 9.1]
Cotoxo, Meniens kua 'eat', Vejoz okua 'bite', Proto-Tacanan *ik(i)a 'eat, bite', Sabela
kǣi (192–3)

bitter   PTB *b-ka
Moseten bikka, Proto-Panoan *moka (193–4)

blood   PST *s-hywəy
Proto-Tupi *uwɨ, Oto-Mangue *we 'red' (86, 196)

broad/flat¹   PTB *peːr
Wiyot bel 'flat, wide', Yurok pel (199)

broad/flat²   PTB *brak ⋉ *prak
Proto-Siouan *pra 'flat, broad' (199)

burn/sun²   PTB *tuk ⋉ *duk 'burn'
Amarakaeri taʔak 'fire', Campa taka 'burn', Shukuru itoka 'burn' (200–201); also
Ayoman dug 'fire' (261)

child/copulate   Old Chinese *ńi̯ə̆g 'child'; Jingpho nèʔ 'copulate'
Crengez ani 'copulate', Guenoa ineu 'son', Mayna ni- 'child', Catio niu 'boy, son' (203)

dark¹   PTB *syim
Itene issim 'night', Xinca syma 'night, black' (209) [NB: Chiquimulilla suʔmax 'black'
is to be compared rather with PTB *s-nak 'black']

die/dream/sleep   PTB *s-maŋ ⋊ *s-mak 'dream'
Choropi imak 'sleep', Araucanian umaq 'dream', Tequistlatec maʔa 'kill' (210–11)

dig   PTB *du
Chama teo, Iranshe toʔu, Ticuna to (212)

dog/deer/animal   PST *gwa 'fox'; *s-kywal 'wild dog, jackal' [JAM]
Chilanga akʔuan 'deer', Achomawi kuan 'silver fox', Yana kuwan-na 'lynx',
Yurumanguí kwan (212)

far   PTB *wəy
Colorado bara, Atabapa pel (219); Proto-Sapiens *b-wəl

fear/shame   PTB *śrak ⋊ *gyak
Botocudo krak 'shame' (219)

finger/one   PTB *g-t(y)ik 'one'
Zamora čikičik 'one', Karok tiːk 'hand, finger' (220) [cf. also PIE *deik- 'point':
LIA, p. 62]

go   PTB *(y)ay [JAM]
Motilon ja 'walk', Azteco-Tanoan *ja 'go, carry', Tsimshian jeː, Algonquian *jaː 'go' (226)

hair   PTB *(t)sam
Kashaya sime 'hair', Mohave sama 'root', Woccon summe 'feather' (229)

hear/ear   PTB *r-na ⋊ *g-na
Uto-Aztecan naka 'ear' (232)

large   PST *ta-y ⋊ *da-y
Ramkokamekran ti 'large', Palmas tei 'long, high', Opaie ta, Puri tahe 'large', Lake
Miwok ʔadi 'large', Klamath ʔaːdi 'long' (237)

liver   PST *m-sin 'liver'
Kwakiutl mas 'bile', Musqueam məs-ən 'gall' (240)

make/do   PTB *day [JAM]
Lule ti 'make', Salinan tiː 'do', Squamish ti 'make' (242)

male/virility   PTB *suw [JAM] (cf. written Burmese sui 'animal's penis',
ʔəsûi 'virility; testicles', Lahu -šɔ̄ 'uncastrated animal')
Wintun siw-ij 'male', Chitimacha ʔasi, Huave na-šej 'man' (242)

mouth/chin/jaw   PTB *m-ka
Mbaya aka 'chin', Mocovi aka 'chin, beard' (246)

night/dark²   PTB *r-muːk 'fog(gy); dark; dull'
Chimariko hi-mok 'evening', Kekchi mux 'become cloudy', Molala muka 'night' (247)

nose   PTB *s-na
     Choroti nus, Aymara nasa (248) (cf. also perhaps PIE *na(:)s-)

pain/devil/demon   PTB *na ⋉ *nan ⋉ *nat (cf. Lahu nà 'hurt', written Burmese nat 'demon')
     Orejon ana 'devil', Yameo ana 'pain' (249)

raw   PTB *s-riŋ (cf. Jingpho kətsiŋ 'fresh, green, raw', Kanauri šöŋ 'alive')
     Pomo kahšo 'raw, alive', Tequistlatec ikʔašwi 'green' (139)

seed/fruit[1]   PTB *mruw 'seed'
     Ecuadorean muju 'fruit', Andoa imio 'seed' (253) (Yanomam mak is to be compared
     rather with Proto-Tai *hmaak 'seed; fruit; round object')

seed/fruit[2]   PTB *sey
     Proto-Ge *zɨ 'seed'; Subtiaba iši 'grain of corn, corn' (253)

sit/be/dwell   PTB *nəy (cf. written Burmese ne 'dwell'; ult. the same etymon as
                         'sun/day' [JAM])
     Papury nii 'be, live', Ticuna niʔĩ, Tucano nii (255)

sun[2]/moon   PTB *s-(g)la 'moon'
     Chimariko alla 'sun', East Pomo la: 'sun, moon' (260)

testicle   Written Tibetan *rlig-pa
     Wiyot me-lak, Chemakum łaqʔu (169)

vagina   PTB *byet [JAM]
     Guahibo petu, Wachipairi ped, Karok vi:θ, Tequistlatec la-bešu (264–5)

water[1]   Lahu í-kâʔ, á-kâʔ 'water'
     Chimariko aqa, Yuma axa, Chickasaw okaʔ, Zuni kʔa, Algonquian *akwa 'from
     water', Musqueam qaʔ (cf. also Latin aqua)

water[2]   PTB *rəy
     Rama ari 'liquid'; 'in Tarascan the suffix -ri is found in many words for liquids' (121)
     (cf. the use of Lahu -ɠ̂ɨ[< *rəy], as in pê-ɠɨ 'honey', nú-ɠɨ 'snot')

water[3]/river   PTB *t(w)əy
     Hahahay te 'rain', Ocaina tiahe 'river', Leco ndowa 'water', Gualaca to 'water'

white   PTB *bok
     Comecrudo pepok, Coeur d'Alene peq (266)

yellow   PTB *s-rwəy (written Burmese hrwe 'gold', Lahu ši 'yellow, gold)
     Oto-Mangue *(n)si 'yellow, white, bright', Uto-Aztecan *si 'yellow' (270)