

1 Introduction

The great Sino-Tibetan language family, comprising Chinese on the one hand and Tibeto-Burman (TB) on the other,¹ is comparable in time-depth and internal diversity to Indo-European, and equally important in the context of world civilization. The overwhelming cultural and numerical predominance of Chinese is counterbalanced by the sheer number of languages (some 250-300) in the 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 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.² With admirable thoroughness, the project staff assembled all the lexical material then available on TB languages, enabling Shafer to venture a detailed subgrouping of the family at different taxonomic levels, called (from higher to lower) divisions, sections, branches, units, languages, and dialects. This work was finally published piecemeal in a two-volume, five-part opus called *Introduction to Sino-Tibetan* (1966-67; 1974).

Shafer’s junior collaborator Paul K. Benedict based his own work on the same body of

¹Many 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).

²For a readable and humorous account of this project, see Benedict 1975b (LTBA 2.1:81-92).

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 In an unpublished manuscript entitled *Sino-Tibetan: a Conspectus* (ca. 1942-43; henceforth *STC*), Benedict adopted a more modest approach to supergrouping and subgrouping than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.³ Shafer's pioneering work, valuable as it was, suffered from his mistrust of phonemics, with a consequent proliferation of pseudo-precise and arcane phonetic symbols. Benedict's structural insight and his flair for isolating that which is crucial from masses of data enabled him to formulate sound correspondences with greater precision, and to distinguish between regular and exceptional phonological developments.

The publication of a revised and heavily annotated version of *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.⁴ 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

³To 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).

⁴In 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).

is: remarkably well. The work has been reviewed about 15 times, almost always in a highly favorable tone,⁵ and has been translated into Chinese.⁶ In fact nearly all 700 of the TB cognate sets in 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.⁷

1.1 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.⁸ 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,⁹ there has been a recent explosion of new data, especially from China¹⁰ and Nepal.

⁵A 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).

⁶See Le Saiyue and Luo Meizhen 1984.

⁷E.g. *dyam *tyam [STC #226] ‘full; fill’ and *dyam [STC #227] ‘straight’; see JAM 1988a.

⁸See JAM 1986a, and STEDT Monograph II (JAM 1996a).

⁹Very 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.

¹⁰Among 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.

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.¹¹ 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-Dafra (=Mirish) into a supergroup called by the purely geographical name of Kamarupan, from the old Sanskrit name for Assam.¹²
- The important Qiangic languages (deemed to include rGyalrong [=Gyarung=Jiarong] and the extinct Xixia [=Tangut]) were hardly known to non-Western scholars 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 many Chinese scholars maintain.

¹¹The 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).

¹²Issue has been taken with this term by Burling (1999), but see the reply by JAM (1999c).

- The Nungish and Luish languages are grouped with Jingpho (=Kachin).¹³ 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.¹⁴¹⁵

1.2 Typological diversity of TB: Indosphere and Sinosphere

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

This diversity is partly to be explained in terms of areal influence from Chinese on the one hand, and Indo-Aryan languages on the other. It is convenient to refer to the Chinese and Indian spheres of cultural influence as the “Sinosphere” and the “Indosphere”.¹⁶ 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

¹³The 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.

¹⁴As 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.”

¹⁵Various 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).

¹⁶See JAM 1990a (“On megalocomparison.”)

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 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.¹⁷ 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-reconstructionthe 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,¹⁸ 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).¹⁹ It remains to be seen whether the 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

¹⁷An excellent recent study of such phenomena is Thomason and Kaufman 1988.

¹⁸See, 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).

¹⁹Cf. 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).

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.²⁰ 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.²¹ 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. 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.²² It is for this reason that TB evidence will prove to be so crucial in evaluating the multitude of competing reconstructions of Old Chinese.

²⁰See the discussion in JAM (VSTB) 1978a:3-12.

²¹This 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”).

²²Many of the features of W. French’s excellent reconstruction of Proto-N.-Naga (1983) were motivated by extra-Naga evidence.