Introduction

The goal of this work is to present the sound laws relating Tibetan, Burmese, and Chinese, and to reconstruct the linguistic unity from which these three languages descend, so far as current knowledge permits. Tracing the development of etyma from their primitive origins into the living tongues of today would bring the narrative satisfaction of accompanying a hero through his struggles, but it is dishonest to present historical phonology as the trials of reconstructed forms progressing through sound changes towards an ultimate destiny in history. The end of this journey, the attested corpus of related languages, is fixed, but the original linguistic unity is the protean and mercurial product of research. Rather than presenting reconstructions picked out of the air and discussing their development, I present sound changes in reverse chronological order. I subsequently reiterate these sound changes in chronological order, so that, after seeing how the reconstructions are arrived at, one can see how it is that the reconstructed forms become the attested forms.

My starting point in this study were the cognate sets assembled by Gong (1995), with some amendments from two of my own previous papers (Hill 2012b; 2014a). While it would doubtless be of profit to consult and discuss the cognate proposals of all previous scholars (e.g. Conrady 1896, Houghton 1898, Simon 1929), this toil would in most cases replicate Gong's work, albeit in a more explicit form. In addition, because earlier generations of scholars relied on now outdated reconstructions of Chinese, among their cognate judgements there is as much chaff as grain. A thorough presentation of previous etymological proposals would needlessly expand the girth of the current work at little benefit.

The work of reconstruction here builds from the ground up, making almost no reference to previous reconstructions.[1] In Chapters 1, 2, and 3 a two-part

[1] General works on historical linguistics often present research on the Trans-Himalayan family (also called Indo-Chinese, Tibeto-Burman, or Sino-Tibetan; cf. van Driem 2012) as in keeping with the standards and methods of the discipline at large (e.g. Abondolo 1998: 8, Campbell and Posner 2008: 114). Such authors paint an overly rosy picture; the only available reconstructions, those of Benedict (1972), Peiros and Starostin (1996), and Matisoff (2003, with an expanded and updated version in press), do not predict attested forms and were not arrived at through the comparative method (Chang 1973, Miller 1974, Sagart 2006).

format – first working backwards step by step and then reiterating the changes in chronological order – is repeated respectively for Tibetan, Burmese, and Chinese. The third chapter on Chinese is disproportionately larger (and more derivative) because of the inherent complications in extracting phonetic information from the Chinese written tradition and presenting these complications to a non-sinological audience. The fourth and final chapter sums up the view of the proto-language which the preceding investigation permits. In order to maximize the ease of other investigators in their research, the provision of appendices and indices is here intentionally liberal.

Those friends and colleagues who have assisted me in myriad ways over the decade I have worked on this project are too numerous to single out individually for thanks. Instead, I limit myself to acknowledging my gratitude to the British Academy and the European Research Council for their material support, the University of California at Berkeley for providing an exceedingly pleasant environment for the sabbatical during which I finished the draft of this work, and Helen Barton, my patient and encouraging editor at Cambridge University Press.

1 Tibetan

§1. Tibetan originated as the language spoken in the Yarlung valley, the cradle of the Tibetan empire (Takeuchi 2012a: 4). Together with the troops of this empire the Old Tibetan language colonized the entire Tibetan plateau, extinguishing almost all of the languages formerly spoken across that territory (Takeuchi 2012a: 6). Evidence is available for three such languages. Most famous is Źaṅ-źuṅ, the language of a pre-existing polity in West Tibet and the sacred tongue of the Bon faith. Źaṅ-źuṅ is preserved in one bilingual cosmological text, the *Mdzod phug*, and a number of short passages in Bon texts (cf. Martin 2010). The closest living relative of Źaṅ-źuṅ is the Darma language of Uttarakhand state in India (Martin 2010: 17–21, 2013). Aside from Źaṅ-źuṅ, samples of two Trans-Himalayan languages are preserved among the collection of documents found at Dunhuang. F. W. Thomas, who first published the manuscripts containing these two languages, confusingly dubs them 'Źaṅ-źuṅ' (Thomas 2011) and 'Nam' (Thomas 1948); there is no evidence to accept these identifications (Martin 2010: 10, 2013).[1]

During the empire's initial expansion, writing was introduced *c.*650 CE to facilitate administration. In 648 CE the Chinese were asked to send paper and ink to Tibet (Laufer 1918, Pelliot 1961: 6, Lee 1981: 13). Two years later, in 650 CE, dated entries in the *Old Tibetan Annals* begin (cf. Dotson 2009: 83; Takeuchi 2012a: 3), indicating that systematic government record keeping began in this year.

§2. A gap of about one century separates the invention of the Tibetan alphabet and the earliest securely datable extant Tibetan documents. The monument generally recognized as bearing the oldest sample of Tibetan writing (post-763) is a stele inscription which now stands in the former village of Źol, in front of the Potala palace in Lhasa.[2]

[1] For more recent research on these two languages see respectively Takeuchi and Nishida (2009) and Ikeda (2012).

[2] This stele's original place of erection was Ṅan-lam Sri; it was moved by the fifth Dalai Lama, Ṅag-dbaṅ blo-bzaṅ rgya-mtsho (1617–82), as part of his strategy of legitimization (Hazod 2009: 181–4). A bell discovered in Dpaḫ-ris may precede the Źol inscription, as it appears to date to the reign of Khri Lde gtsug brtsan (704–55) (Lha mčhog skyabs 2011).

Van Schaik divides the epigraphic monuments from the period of the Old Tibetan empire into four categories: pillar inscriptions from central Tibet, religious inscriptions from north-eastern Tibet, graffiti from Ladakh and adjoining areas of western Tibet, and inscriptions on bells (2013: 120 note 4). Li and Coblin (1987) and Iwao et al. (2009) anthologize most Old Tibetan inscriptions.

The other cache of Tibetan documents securely assignable to the Tibetan imperial period are wood slips and paper documents excavated at the fort of Miran, which date from the Tibetan occupation (van Schaik 2013: 119). Takeuchi (1997–8) has published a catalogue of the paper manuscripts from Miran; the wood slips await thorough cataloguing and publication.

§3. Old Tibetan continued to function as a lingua franca of commerce and administration in the oasis cities of the Silk Road for some decades after the Tibetan empire lost control of these territories in 850 (Uray 1981, 1988, Takeuchi 1990: 187–9; Takeuchi 2012a: 7–9, 2012b). Thus, non-native speakers of Old Tibetan composed some of the extant Old Tibetan documents found at Dunhuang and in the deserts of Central Asia.

The paper manuscripts preserved in the library cave of Dunhuang generally date from the post-imperial period, although some are of imperial provenance. The library cave was closed during the first half of the eleventh century (Imaeda 2008), and this event, whatever its exact date, serves as a convenient *terminus ante quem* of Old Tibetan literature as a language and textual corpus.

The documents from Dunhuang include historical texts, official documents, foreign literature in translation, divination texts, and a sizable number of canonical Buddhist texts. Both the collections of the Bibliothèque nationale de France (Lalou 1939–61) and the India Office Library, now held at the British Library (de la Vallée Poussin 1962) have been catalogued. Tantric manuscripts held at the British Library were more recently re-catalogued in greater depth (Dalton and van Schaik 2006) and the Stein collection held at the British Library was also recently catalogued (Iwao, van Schaik and Takeuchi 2012). Imaeda et al. (2007) provides the sixty-five best-known and best-studied Tibetan Dunhuang documents in transliteration with complete bibliographic references.

§4. Many texts in the Tibetan Buddhist canon (Bkaḥ ḥgyur and Bstan ḥgyur) were translated during the period of the Tibetan empire, but because they have been subjected to editorial changes in subsequent centuries they are not generally regarded as within the purview of Old Tibetan studies (cf. Harrison 1996). Of comparable status are three edicts from imperial times that are quoted in a historical work, the *Mkhas-paḥi-dgaḥ-ston* by the sixteenth-century historian Dpaḥo gtsug-lag phreṅ-ba (1504–66) (Coblin 1990).

§5. The use of Old Tibetan across and beyond the plateau at the height of the empire in the ninth century set the stage for the break-up into the spoken languages of today, which together constitute the Tibetan language family.

The *Stammbaum* of the Tibetan family is poorly understood. There is a typological tendency for the languages of central Tibet to have phonemic tone as well as relatively simplified syllable structure, whereas the dialects of the periphery lack tone and have complex syllable structure. Individual Tibetan languages are usually classified into groups based on the provinces of Tibet they are spoken in; one reads therefore of 'Khams dialects', 'Amdo dialects', etc. (Denwood 1999: 23–36, Tournadre and Dorje 2009: 17–20). The identity of such geographic groupings with genetic groupings has never been demonstrated through the tracing of shared innovations, and is best regarded as a heuristic.[3]

Because Tibetan languages began to diverge from each other some centuries after Old Tibetan was committed to writing, the written system represents an *état de langue* (Old Tibetan) older than that reconstructible via the comparative method (Common Tibetan). As a consequence, for the purposes of this study there is no need to consult data from the Tibetan languages of today.

1.1 Old Tibetan

§6. Old Tibetan has the following consonants: k, g, ṅ, t, d, n, s, z, p, b, m, ts, dz, y, r̥, r, l̥, l, h̥, h, w, ʸ (cf. Hill 2010b). The characters of the Tibetan script do not quite match these sounds one for one. Thus, the palatalized consonants /tʸ/, /dʸ/, /nʸ/, /sʸ/, and /zʸ/ have their own unitary characters <č>, <ǰ>, <ñ>, <ś>, and <ź>, whereas otherwise /ʸ/ is represented with a separate character <ྱ>, e.g. /bʸ/ is written <by>. The two phonemes /l̥/ and /r̥/ are spelled with the digraphs <lh> and <hr>. Finally, the script distinguishes a series of voiceless aspirated obstruents distinct from the plain voiceless obstruents. Originally this distinction was sub-phonemic (Hill 2007). Thus, the script contains the following letters: k, kh, g, ṅ, č, čh, ǰ, ñ, t, th, d, n, p, ph, b, m, ts, tsh, dz, w, ź, z, h̥, y, r, l, ś, s, h (cf. Figure 1.1). With the exception of the letter '*h̥*', the phonetic value of the these letters is uncontroversial. I have argued that '*h̥*' represents [ɣ] (Hill 2005: 115–18, 2009: 129–31).[4] The letter <w> originally occurred only as a medial (Uray 1955).

[3] J. Sun (2003: 794–7) argues strongly against the value of this heuristic.

[4] On the phonetic value of this letter also see Preiswerk (2014: 76) and X. Gong (2016b: 143 note 16). Although its reality has received acknowledgement at least since Simon (1938: 272), some scholars fail to transliterate a syllable final *-h̥*, believing that it represents an orthographic device of no phonetic meaning (e.g. Matisoff 2003: 50, 486 *et passim*, Jacques 2012c: 92). The work of these researchers makes clear that the evidence for the phonetic reality of syllable final *-h̥* merits repeating. An orthographic final *-h̥* in WTib. corresponds to a long vowel in Common Tibetan. These long vowels were subsequently lost in most Tibetan languages, but are sporadically reported across the Tibetan linguistic area. Bell writes concerning Central Tibetan that, as a final, 'ཧ [-h̥] is not itself pronounced but lengthens the sound of the vowel preceding it' (1905: 7). De Roerich also describes this phenomenon in two Tibetan languages: for Central Tibetan he offers the four

Consonants

velars	ཀ *k*	ཁ *kh*	ག *g*	ང *ṅ* [ŋ]
palatals	ཅ *č* [tʃ]	ཆ *čh* [tʃh]	ཇ *ǰ* [dʒ]	ཉ *ñ* [ɲ]
dentals	ཏ *t*	ཐ *th*	ད d	ན *n*
labials	པ *p*	ཕ *ph*	བ b	མ *m*
dental affricates	ཙ *ts*	ཚ *tsh*	ཛ dz	
voiced fricatives	ཞ *ź* [ʒ]	ཟ *z*	འ *ḫ* [ɣ]	
glides	ཡ j	ར *r*	ལ *l*	w
voiceless fricatives	ཤ *ś* [ʃ]	ས *s*	ཧ *h*	
null consonant	ཨ *q* [Ø] or [ʔ]			

Vowels ཨ *a* ཨི *i* ཨཱི *ī* [ɨ] ཨུ *u* ཨེ *e* ཨོ *o*

Figure 1.1 The Tibetan alphabet

The Tibetan alphabet distinguishes five vowels: *a, e, i, o, u*. As in most Indic-derived scripts, the vowel '*a*' is implicitly present in any *akṣara*, whereas other vowels are explicitly marked by diacritics.[5]

examples བཀའ་ *bkaḫ* /kā/ 'order', རྣམ་མཁའ་ *nam mkhaḫ* /nam-kʰā/ 'sky' (1931: 299), དགའ་ *dgaḫ* /gā/ 'delight', and དམའ་ *dmaḫ* /mā/ 'low' (1933: 17); for Lahul he cites the three examples རྣམ་མཁའ་ *nam mkhaḫ* /nam-kʰā/ 'sky', དགའ་ *dgaḫ* /gā/ 'delight', and དམའ་ *dmaḫ* /mā/ 'low' (1933: 17). Migot draws attention to the same correspondence between a written final *-ḫ* and a spoken long vowel in dialects of Khams (1957: 455). Sedláček discusses the complicated effects of original final *-ḫ* on tone in Lhasa dialect, and separates this discussion clearly from his treatment of original open syllables (1959: 216–19). Sedláček additionally implies that final *-ḫ* has a segmental realization which he symbolizes in his phonetic transcriptions as [·], for example སྣའ་ *mṅaḫ* 'might, power' [ŋa· 1155] (1959: 219). Jin confirms the existence of long vowels in Lhasa Tibetan citing the word སྨད་ *mdaḫ* [da:³] 'arrow' (1958: 12). Since in OTib. this letter reflects a velar fricative in other syllable positions and its reflex in Common Tibetan as a syllable final is vowel length, it is reasonable to explain that the loss of the velar fricative in syllable-final position led to compensatory lengthening of the preceding vowel (Hill 2009). As a final piece of evidence in favour of its reality, final *-ḫ* has a correspondence in OChi. which is distinct from open syllables (cf. §38).

[5] OTib. has two graphic forms of the vowel which is called *gi-gu* in WTib. One of these characters is the same as the WTib. *gi-gu* ི <i>. The other is the mirror image ཱི <ī>, and is thus called the *gi-gu inversé*. Whether this character represents a phonetic reality or not remains controversial (Hill 2010: 116). In this study the *gi-gu inversé* is reflected in the transliteration of textual passages, but is not otherwise considered.

§7. Hill (2010b: 121–2) provides a rudimentary discussion of Old Tibetan phonotactics. In general terms, Old Tibetan requires a minimal CV syllable, permits complex onsets of up to four consonants, and has some tolerance for syllable-final clusters of two consonants. Voicing is distinguished only in immediate pre-vowel position (cf. Sprigg 1974: 261).

1.2 Classical Tibetan

§8. Classical Tibetan differs little phonologically from Old Tibetan. Aspiration emerges as a phonemic contrast, but is still a contrast with a small functional load (Hill 2007). One reflection of the emergence of aspiration as a phonemic contrast is that the distribution of aspirates becomes orthographically stable. The most substantial sound change is *sts-* > *s-* (OTib. ལ་སྩོགས་པ་ *la stsogs pa*, WTib. ལ་སོགས་པ་ *la sogs pa* 'etc.', OTib. གསྩང་ *gstsaṅ*, WTib. གསང་ *gsaṅ* 'secret', OTib. གསྩན་ *gstsan*, WTib. གསན་ *gsan* 'listen'). The *gigu inversé* disappears from use. The sporadic final *-ḫ* of Old Tibetan disappears in Written Tibetan (OTib. ལའ་ *laḫ*, WTib. ལ་ *la* allative case marker, OTib. བརླའ་ *brlaḫ*, WTib. བླ་ *bla* 'soul', OTib. འདའས་ *ḫdaḫs*, WTib. འདས་ *ḫdas* 'passed'), but continues into Written Tibetan in words of the syllable structure CCV (བཀའ་ *bkaḫ* 'speech', མདའ་ *mdaḫ* 'arrow').

1.3 The Bodish Languages

§9. Agreement prevails that Tibetan is on the Bodish branch of Trans-Himalayan. Robert Shafer, who introduced this terminology, imagined the Bodish family as consisting of four inner branches (West Bodish, Central Bodish, Southern Bodish, and East Bodish) and three outer branches (Tsangla, Gurung, and Rgyalrong) (1966: 78–123). Shafer considers both Central Bodish and Southern Bodish to be descendants of Old Tibetan (1966: 87). It is unclear whether he believes that Central Bodish and Southern Bodish are genuine subgroups in their own right, or whether they are convenient geographical labels for discussing the daughter languages of Old Tibetan (cf. Figure 1.2).

Shafer's scheme contains two major errors: Rgyalrong is now widely recognized to be a sub-branch of the Qiangic family (Jacques 2004b: 3) and Balti and Ladakhi are direct daughter languages of Old Tibetan, just like the other Tibetan languages (*pace* Bielmeier 2004).[6] Because the relationship of Tsangla and Gurung to Tibetan is not well investigated, these languages are best omitted from the Bodish family. With these adjustments, I previously proposed the *Stammbaum* shown in Figure 1.3 (Hill 2010b: 111). However, this *Stammbaum*

[6] Because Baltistan and Ladakh were Dardic-speaking before the invasion of the Tibetan empire, Shafer's West Bodish hypothesis is a historic impossibility (Petech 1977: 5–13).

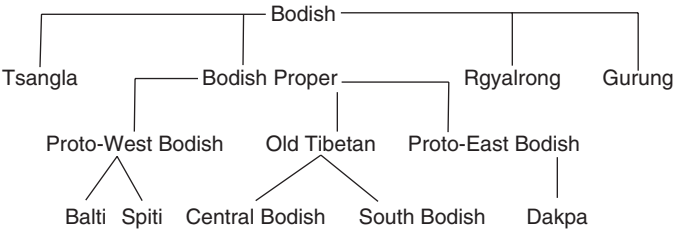

Figure 1.2 Bodish *Stammbaum* (after Shafer 1966: 113)

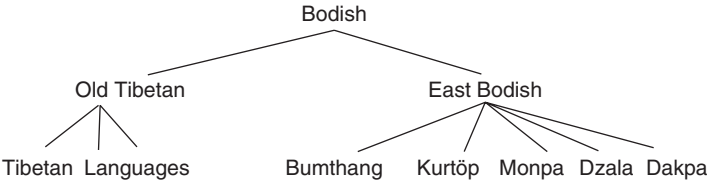

Figure 1.3 Bodish *Stammbaum* (after Hill 2010b: 111)

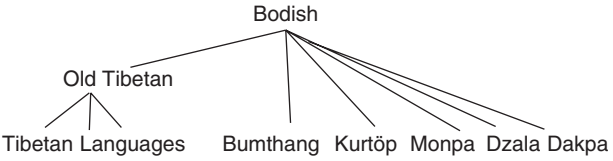

Figure 1.4 *Stammbaum* of the Bodish family proposed here

was arrived at only by correcting Shafer's major errors. It implies that all of the East Bodish languages share common innovations that Old Tibetan does not share. No one has proposed any such common innovation. Until such a change is proposed, the most reasonable *Stammbaum* is simply to derive the various 'East Bodish' languages and Old Tibetan itself from the Bodish proto-language (cf. Figure 1.4).

Based on geographic considerations, one might suggest that languages such as Źaṅ-źuṅ, Kinnauri, Darma, etc. have a particularly close relationship to Tibetan, but this has yet to be demonstrated.

1.4 Tibetan Diachronic Phonology

§10. Tibetan shares innovations with the East Bodish languages; these shared innovations allow us to divide the history of Tibetan into two phases: a more

recent phase, during which its fate was independent of the East Bodish languages, and an early phase when together with the East Bodish languages it was a single tongue.

It is not possible in every case to determine whether or not an East Bodish language underwent the same change as Tibetan. All changes which happened after the earliest change not shared by the East Bodish languages must be independent of the changes in the East Bodish languages. I use evidence from Kurtöp and Mstho-sna Monpa (Wenlang dialect) as representatives of the East Bodish family. The internal phylogeny of the East Bodish family and this family's historical phonology is not a concern here (cf. Hyslop 2008, 2013).

1.4.1 From Old Tibetan to Proto-Bodish

§11. The following changes, presented in reverse chronological order, all post-date the break-up of proto-Bodish, i.e. they are innovations unique to Tibetan.

§12. *Chang's Law: Assimilation of b- before Nasals*. Betty Shefts Chang (1971: 738) discovered that cluster initial *b-* assimilates to the labial nasal *m-* before nasals. The seemingly anomalous *m-* in the past stem of verbal roots beginning with nasals becomes thereby a sub-case of the nearly ubiquitous *b-* prefix of the past stem (compare √sad 'kill', past བསད་ *bsad*).

*bn > mn, e.g. √nan 'suppress', past *bnans > མནན་ *mnand*
*bñ > mñ, e.g. √ñan 'listen', past *bñans > མཉན་ *mñand*

§13. *Coblin's Law: Loss of Prefixes*. Prefixes are lost when the resulting cluster is not phonotactically possible (Coblin 1976). This law greatly facilitates the internal reconstruction of the Tibetan verbal system. Coblin himself proposed three specific changes that fall under this rubric, and Jacques (2014c) adds a fourth.

Change 1: *bb- > *b-*, *bp- > *p-* (cf. §13a)
Change 2: *ɧCC- > CC- (cf. §13b)
Change 3: *gCa > *g*C*o*, *gCCa > C*o* (cf. §13c)
Change 4: *sNC > *s*C (cf. §13e)

Each of these changes is motivated by cases in which positing of a lost prefix resolves some anomaly in a verb's paradigm, and renders the verb in question an example of a paradigm type which is otherwise well attested.[7]

§13a. The usual prefix met in the past stem is *b-* (e.g. cf. √sad 'kill', past བསད་ *bsad*), but this *b-* prefix does not appear before verbs with bilabial root initials (√bya 'do', past བྱས་ *byas*). The supposition of a phonotactic constraint that assimilates the past prefix *b-* to a following bilabial permits the analysis of bilabial roots as regular outcomes of a paradigm involving a prefix *b-*. In other words, a phonetic regularity (assimilation of *b-* before bilabials) takes the place of a morphological irregularity (a conjugation class without a *b-* prefix). The following examples show the effect of Coblin's law on the past prefix *b-*.

*bb- > *b-*, e.g. √bya 'do', past *bbyas > བྱས་ *byas*
*bp- > *p-*, e.g. √pyag 'sweep', past *bpyags > ཕྱགས་ *phyags*

§13b. One of two prefixes met in the present stem is *h-* (e.g. cf. √kru 'bathe', present འཁྲུད་ *hkhrud*), but this *h-* prefix does not appear in verbs with roots that have complex onsets (√rkam 'long for', present རྐམ་ *rkam)*. The supposition of a phonotactic constraint that removes *h-* when its application would produce a series of three consonants permits the analysis of roots with complex onsets as regular outcomes of a paradigm involving a prefix *h-*. In other words, a phonetic regularity (loss of *h-* before complex onsets) takes the place of a morphological irregularity (a conjugation class without a *h-* prefix). The following examples show the effect of Coblin's law on the present prefix h̲-.

*h̲rk > rk, e.g. √rkam 'long for', present *h̲rkam > རྐམ་ *rkam*
*h̲rŋ > rṅ, e.g. √rṅa 'mow', present *h̲rṅa > རྔ་ *rṅa*

§13c. Many researchers, including formerly the author of these lines (Hill 2010a: xvii), follow Coblin in positing the loss of a present prefix *g- in order to explain o-ablaut in the present stem of Tibetan verbs. This proposal relies on the observation that o-ablaut usually coincides with a *g-* prefix where the latter is a phonotactic possibility (e.g. cf. √sad 'kill, present གསོད་ *gsod*), but in verbs with o-ablaut in the present that have a root beginning with a complex onset this *g-* prefix does not appear (√skaṅ 'fulfill', present སྐོང་ *skoṅ*). The supposition of a phonotactic constraint that removes *g-* when its application would produce a series of three consonants, in the view of these authors, permits the analysis of roots with complex onsets as regular outcomes of a paradigm involving a prefix *g-*.

[7] For a complete discussion of Tibetan verb morphology see Hill (2010a: xv–xxi) and the more recent contributions Jacques (2012b), Hill (2014c), and Hill (2015b).