1 | Introduction

This book is intended to describe the reading of texts in the classical Tibetan language. It is not a complete survey of all aspects of Tibetan grammar, although I have tried to include everything I thought it was important to know. It is intended to describe the READING of classical Tibetan; therefore it will not attempt to teach the reader to speak either Old Tibetan or some modern Tibetan dialect.[1] Rather my aim has been to provide procedures for the understanding of TEXTS—that is, coherent discourses with literary or philosophical content, whose authors used the resources of their language to convey meaning.

Moreover, this book describes the reading of CLASSICAL Tibetan. I have excluded from systematic consideration the modern newspaper language at one end and the language of translated Indic texts at the other. This leaves a sufficiently immense corpus of written Tibetan material for us to work from.[2]

---

[1] This limitation has made the descriptive task much easier. A RECOGNITION GRAMMAR such as this need not incorporate the output constraints required in a PRODUCTION GRAMMAR, under the assumption that a reader simply will not encounter syntactically ill-formed sequences, whereas a beginning speaker may well produce them. For example, a production grammar of English would require both a rule to produce the sequences *goodness* and *correctness* and a constraint on the same rule to prevent the sequences ?*trueness* or ?*strongness* (as opposed to *truth* or *strength*). But a recognition grammar would need only a rule allowing such sequences as *goodness* or *correctness* to be interpreted when encountered. If *trueness* or *strongness* should turn up in a text, they could be processed by the same rule; if not, the question does not arise.

[2] I think it is fair to say that the translated materials have been more intensively studied than works originally composed in Tibetan, because of the primarily Indological interests of many scholars of Tibetan; most existing textbooks reflect this interest. There can be no doubt, of course, that the classical language, as here defined, is closely related to the translation language. But it is clear too that the language of the translations exhibits its own peculiarities, including occasionally opaque attempts at reproducing Sanskrit syntax. These peculiarities, I firmly believe, are best analyzed from the point of view of the grammar of native Tibetan works, rather than the other way around.

Finally, I hope to introduce classical Tibetan as a LANGUAGE, with a history, with a range of styles, and with ongoing processes of creation and change. Too often the Tibetan language is seen either as a poor substitute for unfortunately vanished Sanskrit texts, or as a spiritual code whose value lies solely in the message it conceals—with the result that the language itself is neglected as a medium of great range, power, subtlety, and humor. I hope to rescue Tibetan from its subordination to Indic criteria, and to help the reader proceed not only with some grammatical confidence but also with an awareness of the individuality and literary potential of the language. I hope to provide the reader with conceptual tools for an intelligent and critical reading of Tibetan texts. I hope to share some of my affection for the Tibetan language.

Figure 1. Dbyaṅs-tśan-ma, goddess of music and speech

2 | Transliteration

1. PHONEMIC SYMBOLS

In this book I will use the following inventory of symbols to transcribe Tibetan of all periods:

	STOPS		FRICATIVES		SONORANTS	
GLOTTAL	ʔ		h			
VELAR	k	g	x	γ	ñ	
PALATAL	c	j	š	ž	ñ	y
RETROFLEX	ṭ	ḍ	ṣ	ẓ	ṇ	r
DENTAL	t	d	s	z	n	l
LABIAL	p	b	f	v	m	w

Table 1. Transliteration of consonants

	FRONT		BACK
HIGH	i	ɨ	u
	e	ə	o
	ɛ	ʌ	ɔ
LOW		a	

Table 2. Transliteration of vowels

2. DIGRAPHS AND DIACRITICS

Combinations of symbols will be used to represent AFFRICATES (such as tš, dz,

3

pf), ASPIRATES (such as *kh, bh, tsh*), PALATALIZED CONSONANTS (such as *khy, zy, my*), and similar modifications. A small subscript circle will indicate that a normally voiced phoneme is VOICELESS, as in New Tibetan (Lhasa) *l̥a* "god" as opposed to *la* "mountain pass." An umlaut will indicate that the marked vowel is articulated at the end of the mouth other than the usual one —that is, that a normally back rounded vowel is a FRONT ROUNDED vowel at the same height, as in New Tibetan (Lhasa) *šü* "tell," where the vowel is fronted as opposed to *šu* "sit" and rounded as opposed to *ši* "destroy"; or that a normally front spread vowel is a BACK SPREAD vowel at the same height, as in New Tibetan (Amdo) *šïñ* "field," where the vowel is backed as opposed to *šiñ* "cloud" and spread as opposed to *šuñ* "protection." The symbol *l̤* will represent a voiced murmured lateral, as in New Tibetan (Ladakh) *l̤ama* "lama," *l̤a* "god," *l̤u* "song." I will use the symbol *N* to indicate both NASALIZATION of a preceding vowel, as in New Tibetan (Dbus) *güN* "grape," and PRENASALIZATION of a following consonant, as in *Ngü* "move."

3. OTHER SYMBOLS

I will use an asterisk to mark an UNATTESTED form which has been historically reconstructed, as in Proto-Tibetan **gryab* "throw." I will use an interrogative to mark a DISALLOWED form which is precluded by the synchronic rules of the language, as in Old Tibetan *?rta-rnams-dag* as opposed to *rta-dag-rnams* "horses." Quotation marks will enclose GLOSSES, as in Old Tibetan *rta* "horse," in order to identify forms and constructions, not to provide their central meaning or best possible translation, although I must confess I think my translations are often quite clever; capitalized glosses are SEMANTIC, as in Old Tibetan *-dag* "MORE THAN ONE."

The sign > means "changes into" and < means "develops from," as when Proto-Tibetan **gryab* "throw" > Old Tibetan *rgyab* > New Tibetan (Lhasa) *cəp*. The sign ~ means "varies with," as when Old Tibetan *me-tog ~ men-tog* "flower." The sign → in glosses means "is lexicalized as," as in Old Tibetan *rdo-riñ* "long stone → monument," New Tibetan (Dbus) *meNta* "fire arrow → gun." Angle brackets enclose GRAPHS, as when I indicate that New Tibetan (Lhasa) *saNcɛ* "Buddha" has the written form *<sañs-rgyas>*. The graph called *a-tšhuñ* "little *a*" by the Tibetan grammarians will be transcribed, for expository purposes only, by a slash, as when discussing the written form *<be/u>* for Old Tibetan *beu* "calf," but will not otherwise be transcribed, for

reasons that will be made clear in the main body of the text—thus, normally, Old Tibetan *od* <od> "light" rather than </od>, *beu* <beu> "calf" rather than <be/u>, and *mda* <mda> "arrow" rather than <mda/>.

A hyphen will be used to indicate that the syllables which it connects constitute a single WORD, as in Old Tibetan *ndžig-rten* "world," or a single STRESS GROUP, as in Old Tibetan *pad-dkar* "white lotus." It will also be used to indicate that a morpheme is BOUND and must occur with some other form either preceding, as in Old Tibetan *-dag* "MORE THAN ONE," or following, as in Old Tibetan *mI-* "NEGATIVE." A hyphen may also indicate the POSITION of a phoneme in a Tibetan syllable: thus *r-* indicates a leftmost *r*, as in *rgu*, *-r-* indicates a medial *r*, as in *gru*, and *-r* indicates a final *r*, as in *gur*.

I will use a period to distinguish a stop preinitial followed by an initial glide, as in Old Tibetan *g.yon* "left" > New Tibetan (Lhasa) *yön*, from a stop initial followed by a postinitial glide, as in Old Tibetan *gyoñ* "loss" > New Tibetan (Lhasa) *choñ*.

Capitalization of a phoneme will indicate that it undergoes regular MORPHO-PHONEMIC CHANGES according to phonological environment. Such an environ-ment may be across a syllable boundary, as when the Old Tibetan nominal-izer *-Pa* becomes *-ba* after preceding final *-ñ*, *-r*, *-l* and vowels, and *-pa* elsewhere; or within a syllable, as when the Old Tibetan future tense prefix *G-* becomes *g-* before acute consonant initials and *d-* before grave consonant initials. Verb ROOTS will be entirely capitalized, followed, where appropriate, in parentheses, by the tense stems of that root, present and past in the case of intransitive verbs, and present, past, future, and imperative in the case of transitive verbs—for example, KHUM *(ɴkhum/khums)* "become shrunken," TU *(ɴthu/btus/btu/thus)* "gather," SLAB *(slob/bslabs/bslab/slobs)* "teach." Using this convention, we will show the derivation of, say, the present and past stems of GAD "laugh" as *dgod* < G-GAD "laughs" and *bgad* < b-GAD-s "laughed."

The Tibetan vertical stroke or *šad*, marking a reading pause, will be transcribed with a comma.

4. OTHER LANGUAGES

Words in New Chinese, as well as Chinese place names, book titles, and other non-linguistic citations, will be given in Wade-Giles transcription, about

which I am sentimental. Reconstructed forms in Old and Middle Chinese—Karlgren's "Archaic" and "Ancient" Chinese—have been taken from Bernhard Karlgren's *Grammata Serica Recensa*, with several liberties taken with his transcription. Sanskrit is transcribed in the traditional manner, as are, more or less, Mongolian and Burmese. Those familiar with these languages should have no difficulty recognizing the forms. There is nothing even approaching a generally accepted tradition for transcribing the less well-known Tibeto-Burman languages; I have followed, as best I could, the transcriptions of the various authors to whose works I have referred, and I have attempted—probably with little success—to force some consistency upon the various systems.

Figure 2. Charm to bind demons

3 | Tibetan in Context

1. DEFINING TIBETAN

TIBETAN is a language spoken primarily on the high plateau north of the Himalayas. It is related to a number of Himalayan languages, such as Gurung and Magar, whose speakers were a traditional source of recruits for the British Gurkha forces. It is also related to several languages, such as Rgyarong and Minyag, spoken on the great northern plains by nomadic tribes traditionally called "western barbarians" by the Chinese.[1] Tibetan is more distantly related to Burmese; even more distantly to languages spoken by Naked Nagas and other hill tribes of Assam; and more distantly still to Chinese. Tibetan has had a writing system since the seventh century, borrowed from an Indian prototype. India, in fact, has had a massive cultural impact on Tibet; but Tibetan itself is unrelated to Sanskrit or any other Indic language.

We can define Tibetan as that language in which we find the word *bdun* "seven" and its cognates—particularly as opposed to the word **snis* and its cognates found everywhere else among the Tibeto-Burman languages. For example, we find Rgyarong *snyis*, Horpa *zne*, Kanauri *stis*, Garo *sni*, Kachin *sənit*, Burmese *hnats*, Sgaw *nwi*, Taungthu *nöt*, Gurung *ñi*, the ancient Zhangzhung *snis*, and perhaps even Old Chinese **tshyet* "seven."[2] None of these

---

[1] Middle Chinese **bhywan* "barbarian" may in fact be a loanword from Old Tibetan *bon* "shamanic religion" or a related word in one of the Hsi-fan languages.

[2] Another apparent innovation in Tibetan is the word *khyod* "you" and its cognates, as opposed to **nañ* "you" and its cognates found in other related languages—for example, Chepang *nañ*, Kachin *nañ*, Burmese *nañ*, Lushei *nañ*, Sgaw *na*, Pwo *na*, Dhimal *na*, Nung *na*, Phunoi *naʏ*, Bisu *nañ*, Akha *naw*, Mpi *noñ*, Rgyarong *no*, Minyag *na*, and perhaps Old Chinese **ñyo* and **ñia* "you." Compare Old Tibetan *ñid* "you (elegant)" New Tibetan (Sherpa) *ñipo* <*ñid-po*> "you."

Interestingly, another apparent Tibetan innovation is *rta* "horse" and its cognates, as opposed to **srañ~mrañ* found elsewhere—for example, Chepang *sərañ*, Kachin *kumrañ*.

7

languages is a Tibetan dialect, however closely related it may otherwise be to Tibetan. But when we find Balti *bdun,* Purig *rdun,* Ladakh *dun,* Golok *wdən,* Amdo *dïn,* and Lhasa City *tüN* "seven," we know we are dealing with a series of DIALECTS within the Tibetan language.

2. TIBETAN AND RELATED LANGUAGES

When we say Tibetan is related to another language—say, Burmese or Chinese—we mean that the languages are both descendants of an earlier language no longer in existence. Frequently such a hypothetical ancestor is proposed to account for many such offspring; this common ancestor is then often named after those two of its descendants with the oldest written records—for example, Proto-Tibeto-Burman, which is the hypothetical language from which all Tibeto-Burman languages have come, or Proto-Sino-Tibetan, which is the hypothetical language from which have come not only the Tibeto-Burman languages but Chinese as well.

When comparing languages to see if they are related, random correspondences of words of course prove nothing. The apparent cognates could just be accidental: compare Tibetan *rgyal* "king" with English *royal~regal.* More frequently such apparent cognates are loan words, in one direction or the

---

Burmese *mrañ,* Kanauri *rañ,* Manchati *hrañ,* Bunan *šrañs,* Haka *rañ,* Lisu *amu,* Phunoi *moN,* Bisu *ʔamɔñ,* Akha *mah,* Mpi *myuñ,* Rgyarong *bro~mbro,* and perhaps Old Chinese **mɔ* "horse." However, in several archaic texts from Central Asia, we find, to our delight, alongside Old Tibetan *rta* "horse," the word *rmañ,* which apparently means something very much like "horse"—for example, in a mythological text from the caves near Tun-huang, in the couplet *rta-skad ni tsher-tsher, rmañ-skad ni tsher-tsher* "In horse language, yes, *tsher-tsher!* in steed language, yes, *tsher-tsher!*" or in the couplet *rta bžugs ni gnam-la bžugs, rmañ bžugs ni dguñ-la bžugs* "The horse dwells, yes, dwells in the sky; the steed dwells, yes, dwells in the heavens," or, again, *rta ni log-pa tše, rmañ ni mkhris-pa tše* "The horse, yes, his revulsion was great; the steed, yes, his bile was great." In the administrative correspondence recovered from the Tibetan garrisons in Central Asia, we find, as we might expect, frequent references to horses, but almost always as *rta.* Still, the word *rmañ* lingers in the collocation *rmañ-rogs* "horse attendant → groom," and, perhaps, in the proper name *rmañ-sbyin* "Horse Giver." By the time of the classical texts the word *rmañ* has disappeared entirely. It seems clear that this Old Tibetan *rmañ* is related to Proto-Tibeto-Burman **mraň* "horse," and was replaced by Old Tibetan *rta* "horse" during the seventh century—in effect, before our eyes.

other: this might be the case with words such as classical Tibetan *dža* Middle Chinese **jha* "tea," or classical Tibetan *džag* Middle Chinese **dzhək* "robbery," where, as one Sinologist has put it, too close a likeness is even more suspect than too distant a one. But what makes it likely that, say, Tibetan and Chinese are related languages is a SYSTEMATIC correspondence among their words—for example, the fact that in both languages the word for "I" (Old Tibetan *ña,* Old Chinese **ña*) and "five" (Old Tibetan *l-ña,* Old Chinese **ño*) both have velar nasal initials, or the word for "three" (Old Tibetan *g-sum,* Old Chinese **səm*) and "kill" (Old Tibetan *SAD,* Old Chinese **sat*) both begin with a dental fricative. It is only on such a systematic basis that we are justified in assuming that Tibetan and Chinese derive from a common ancestor.

Technically speaking, the only way actually to demonstrate that two or more languages are cognate descendants of a common ancestral language is to reconstruct the common language from which they descended. Such reconstructions have been cited as the most triumphal vindication of Indo-European comparative linguistics. Yet similar attempts to reconstruct earlier stages of Tibetan and related languages have encountered serious difficulties.

> *You know ek, that in forme of speche is chaunge*
> *Withinne a thousand yer, and wordes tho*
> *That hadden prys now wonder nyce and straunge*
> *Us thenketh hem, and yet they spake hem so,*
> *And spedde as wel in love as men now do.*
>
> —Geoffrey Chaucer,
> *Troylus and Criseyde*

For one thing, such reconstructions must take account of literally hundreds of related languages—overwhelmingly unwritten and, until recently, poorly recorded and described. For another, the words which are being compared in these language are remarkably compact. For example, we find classical Tibetan *grog-ma,* Burmese *pərwak* "ant." Are these words cognate? Additional comparisons from other languages do not seem immediately helpful: Rgyarong *korok* and Kiranti *khorok* seem related to the Tibetan *grog,* while Lahu *puɣɔʔ,* Lisu *bawlaw,* and Mpi *pihoʔ* seem related to the Burmese

pərwak. Yet we also find Miri *tǝruk,* Dafla *tǝrub,* and Nung *sǝrɔ.* What are we to make of this?

One proposed solution postulates a Tibeto-Burman word **rwak* "ant," to which Lahu and Burmese added a prefix **p-* related to the word for "insect" (compare, for example, classical Tibetan *ɴbu,* Burmese *pui,* Mpi *pi* "insect"); to which Tibetan, Rgyarong and Kiranti added a **k-* "ANIMAL" prefix; to which Nung added an **s-* "ANIMAL" prefix (compare, for example, Old Tibetan *šwa,* Burmese *sa,* Kachin *šan,* Nung *ša,* Kanauri *šya* "deer"); and to which Miri and Dafla added a late **d-* prefix of uncertain signification. Now such explanations can quickly become uncomfortably *ad hoc,* and there is often an unexplained residue in any event; for example, we are still left to account for Gurung *nabbru* "ant."[3] But such are the challenges faced by the Tibeto-Burman comparativist.

2.1. TIBETAN AND CHINESE

Scholars have long suspected that Tibetan is related to Chinese, and have postulated a Sino-Tibetan family of languages descended from a hypothetical Proto-Sino-Tibetan ancestor. The relationship between Tibetan and Chinese, however, is certainly not obvious if we compare contemporary Tibetan with contemporary Chinese. In Peking city the old word for "dog" is pronounced *tšhüan* but in Lhasa city is pronounced *chi,* while a Peking fish is called *yü* but a Lhasa fish is called *ña.*

But thanks to the extraordinary conservatism of Tibetan writing on the one hand, and the scholarly detective work of such Sinologists as Bernhard Karlgren on the other, we can reconstruct these same words in Old Tibetan and Old Chinese:[4] when we compare Old Tibetan *khyi* with Old Chinese

---

[3] Unlike many words in Gurung, *nabbru* does not appear to be a loan word from Nepali, where the word for "ant" is *kamilo.* Nepali is an Indo-Aryan language not very distantly related to Hindi.

[4] Languages are dated from the first appearance of writing: thus the earliest Tibetan records are said to be in Old Tibetan, and the earliest Chinese records in Old Chinese. (Forms reconstructed for a period prior to the appearance of writing are said to be in the proto-language—Proto-Tibetan, say, or Proto-Chinese.) But Chinese was first written much earlier than Tibetan was, so Old Chinese is older than Old Tibetan; in fact, Old Tibetan is

khywen "dog" and Old Tibetan *ña* (from an even older Proto-Tibetan **ñya*) with Old Chinese **ñyo* "fish," the similarities between the two languages become much more striking. In the same way, other correspondences have been proposed—for example, Old Tibetan *ñi* Old Chinese **ñyet* "sun," Old Tibetan *mig* Old Chinese **myok* "eye," Old Tibetan *ma* Old Chinese **ñyag* "ear," Old Tibetan *śiñ* Old Chinese **syen* "firewood," Old Tibetan *lña* Old Chinese **ño* "five," Old Tibetan *gsum* Old Chinese **sam* "three."

Let us assume, then, on the basis of such partial evidence, that Tibetan and Chinese are descended from a common ancestor. Is there any way of telling how long ago it was that Tibetan and Chinese were, in some sense, the same language? Archeological finds indicate that human beings first appeared in northern China around 10,000 BC, in all likelihood having come eastward from the frozen tundras of Siberia, where they had survived and adapted through the most recent of the recurrent ice ages; by 5,000 BC neolithic culture had appeared on the fertile northern plains of China, which then developed with remarkable continuity and coherence directly into historical times, with a language we have every reason to believe was already distinctively Chinese. These speakers of Chinese continued to spread from the middle Yellow River area toward the southern and eastern coasts—an extension even now in progress.

If the Chinese language split off from the common stock somewhere between 10,000 and 5,000 BC, then the Chinese and Tibeto-Burman language groups may simply have been separated too long, and their descendants simply have changed too much, to permit any convincing reconstruction of their common source; but a reconstruction of Proto-Sino-Tibetan is so challenging a prospect that such efforts are not likely to stop.

2.2. TIBETAN AND BURMESE

The Burmese language was first written down, on inscriptions, using an adaptation of the Mon orthography, around 1112. This writing system was later replaced by a form of Sinhalese script, also derived from an Indic prototype, and by about 1500 the Burmese writing system had taken on more

---

contemporaneous with Middle Chinese, so that we speak, say, of seventh-century Middle Chinese words borrowed into Old Tibetan. Here we compare Old Tibetan with Old Chinese, since we want to compare the earliest available forms in each language.

or less its present form. The writing, like that of Tibetan, is conservative, and presumably reflects the phonological state of the language at about the time the orthography was fixed; that language in turn differs in some significant ways from modern "standard" Burmese, spoken throughout the Irrawaddy plain and delta, in Upper and Lower Burma, by more than thirty million people.[5]

The relationship of Tibetan and Burmese—and closely related languages such as those grouped together as Lolo—is only slightly more apparent than the relationship of Tibetan and Chinese. A dog in Rangoon is *khwei*, and, as we travel through Southeast Asia, we find Lahu *kwe*, Phunoi *khɔ*, Bisu *khɔ*, Akha *akui*, Mpi *khɔ*, but in Lhasa city a dog is *chi*. Similarly, a Rangoon fish is *ŋa*, and we find Lahu *ŋa*, Lisu *ŋwa*, Akha *ŋa*, Mpi *ŋo*, but in Lhasa city a fish is *ña*—a nasal initial, but, apparently, in the wrong part of the mouth.

However, when we compare the older written forms in Tibetan and Burmese, even a cursory inspection reveals systematic correspondences between the two languages much more extensive than those between either language and Chinese. Thus we can, again, compare Old Tibetan *khyi* "dog" to Proto-Burmese **khuy*, and Proto-Tibetan **ŋya* "fish" to Burmese *ŋa*. Among the many cognates that have been proposed, we may note Old Tibetan *ñi* Burmese *ne* "sun," Old Tibetan *myig* Burmese *myak* "eye," Old Tibetan *rna* Burmese *na* "ear," Old Tibetan *šiñ* Burmese *sats* "firewood," Old Tibetan *lña* Burmese *ña* "five," Old Tibetan *gsum* Burmese *sum* "three."[6]

---

[5] For the divergence of the spoken and written forms, note—randomly—modern standard Burmese *mwei* <mrwe> "snake," *šwei* <hrwe> "gold," *an* <sam> "iron," *myiʔ* <mrats> "river," *thamin* <htaʔmañ> "cooked rice," *shan* <htsan> "husked rice," *nei* <neʔ> "sun," *myin* <mrañ> "horse," as well as *pan* <pan> "flower," *mi* <mi> "fire," *na* <na> "ear," *hna* <hna> "nose," *ñwei* <ñwe> "silver."

[6] Apparent cognates in Old Tibetan and Burmese are not hard to find. A few minutes with a dictionary will turn up Old Tibetan *tshwa* Burmese *tsha* "salt," Old Tibetan *sku* Burmese *kuiy* "body," Old Tibetan *dgu* Burmese *kui* "nine," Old Tibetan *gri* "knife" Burmese *kre* "copper," Old Tibetan BGRE "grow old" Burmese *kri* "be great," Old Tibetan *byi* Burmese *pwe* "rat," Old Tibetan *bya* "bird" Burmese *pya* "bee," Old Tibetan *gtšig* Burmese *tats* "one," Old Tibetan *ña* Burmese *ña* "I," Old Tibetan *gñis* Burmese *hnats* "two," Old Tibetan *sna* Burmese *hna* "nose," Old Tibetan *mye* Burmese *mi* "fire," Old Tibetan *smyig* Burmese *hmyats* "bamboo," Old Tibetan SAD Burmese *sat* "kill," Old Tibetan *gzigs* Burmese *sats* "leopard," Old Tibetan *zla* Burmese *la* "moon," Old Tibetan *ri* "painting" Burmese *ri* "write."

There can be no doubt that Tibetan and Burmese are related, or that
Burmese in turn is related to a number of other Southeast Asian languages,
in what is commonly called the Tibeto-Burman family—here, once again,
named after the two members of the family with the oldest written records.
In this family, in addition to Tibetan and Burmese, there is in fact a vast
complex of languages, stretching from the northern reaches of Assam and
Burma westward along the Himalayas, eastward into southern China, and
southward along the Salween and Irrawaddy Rivers to the Bay of Bengal.
These regions constitute one of the most linguistically diverse areas of the
world; it is still very difficult to get a clear picture of the relationships of the
various languages and dialects, not only within the Tibeto-Burman family, but
also in terms of the areal and borrowing relationships between the Tibeto-
Burman languages and the unrelated Thai and Mon-Khmer languages with
which they have long been in contact.

The cultural diversity of this area is equally striking. Speakers of Tibeto-
Burman languages include goat herders in the mountains of Nepal, former
head hunters along the Indo-Burmese frontier, naked tribes in the jungle hills
of Assam, as well as the Tibetans and Burmese, who built successful Buddhist
kingdoms and literate cultures which have survived to the present day.[7] The

---

Some of these Tibetan and Burmese forms correspond even more closely if we look at the
orthography of the oldest dated Burmese inscription—the inscription of Prince Rajkumar,
dated 1112, often called the Myazedi Inscription because it was found on the *mya zedi*
"Emerald Pagoda." Here we find Burmese *se* "die" written *<siy>*, *ri* "water" written *<riy>*,
and *pe* "give" written *<piy>*. Presumably Proto-Burmese **-iy* > Burmese *-e*, and we can
compare Old Tibetan *ñi* "sun" with Proto-Burmese **niy* rather than with Burmese *ne*, Old
Tibetan *gri* "knife" with Proto-Burmese **kriy* rather than with Burmese *kre*, Old Tibetan *ʦi*
"die" with Proto-Burmese **siy* rather than with Burmese *se*, and Old Tibetan SBYIN "give"
with Proto-Burmese **piy* rather than with Burmese *pe*. Similarly, we find Burmese *hrwe*
"gold" written *<hruy>*, *mwe* "nourish" written *<muy>*, and *ʔaʦhe* "kin" written *< ʔaʦuy>*.
Presumably Proto-Burmese **-uy* > Burmese *-we*, and we can compare Old Tibetan *dñul*
"silver" with Proto-Burmese **ñuy* rather than with Burmese *ñwe*, Old Tibetan *sbrul* "snake"
with Proto-Burmese **mruy* rather than with Burmese *mrwe*, and Old Tibetan *dgur* "crooked"
with Proto-Burmese **kuy* rather than with Burmese *kwe*.

[7] Other speakers of Tibeto-Burman languages also had states, primarily on the Hindu
model, in the valleys around the edges of South Asia—the Newari in Kathmandu; the
Meithlei in Manipur; the Lushei in the Mizo area; the Tripuri in Tripura; the Pyu in Burma,
conquered by the Burmese; and the Bodo or Kachari in central Assam, conquered by the Dai
Ahom.

generally accepted picture is that this entire area was occupied by an originally southward movement of Tibeto-Burman-speaking peoples along the great Irrawaddy and Salween river basins, which carry the waters of the Himalayas to the sea. Such southward migrations, perhaps prompted by periodic dessication of the Inner Asian plains, presumably began from the same point from which another group had moved eastward into the fertile plains of north China; and from secondary diffusion centers along the way there occurred further migrations westward along the great arc of the Himalayas, southward deeper into Burma, and eastward into northern Thailand and Laos, with the languages diverging, interacting, and borrowing from each other, and interacting as well with the unrelated Mon-Khmer and Thai languages whose speakers were both being displaced and migrating themselves.

2.3. TIBETAN AND THE HIMALAYAN LANGUAGES

Scattered along the arc of the Himalayas, like beads on a string, are a number of more or less related languages, usually called—for want of any more informative name—the Himalayan languages. The relations among these languages are not at all clear; for example, Newari, the historically important language of the old kings of Kathmandu, apparently cannot be grouped directly with any other of these Himalayan languages; and the remainder tend to be classified in primarily geographic groupings, with names like West Central Himalayish, on the assumption that human occupation of the Himalayan valleys proceeded linearly, from east to west, so that more closely related languages would tend to cluster geographically as well. I am not at all persuaded that this picture is correct; but I certainly have nothing better to offer.[8] What does seem clear is that, among these Himalayan languages, some—Tamang, Gurung, Thakali, Magar, Kiranti—seem quite closely related to Tibetan.

---

[8] Speakers of several of these Himalayan languages have traditionally been the source of recruits for the British Gurkha forces. Nepali, the dominant language of Nepal, came to be the *lingua franca* of the Gurkha forces at brigade posts of the Indian Army throughout India and of the British Army in Hong Kong and Malaysia. Glover has noted that Gurung children returning to Nepal from military posts can converse with their village relatives only in Nepali while the latter converse among themselves in Gurung. Nepali is an Indo-Aryan language related to Hindi; and, although unrelated to the Himalayan languages, Nepali loanwords have thoroughly infiltrated the Himalayan lexicons.

In 1927, Sten Konow, of the Linguistic Survey of India, distinguished between "complex pronominalized" and "non-pronominalized" Himalayan languages, with the pronominalized languages further subdivided into eastern and western branches. The pronominalized languages fuse subject and object pronouns to the verb, where they appear as prefixes and suffixes, yielding in effect a verbal inflection for both subject and object: for example, in Limbu, the language of the principal tribal people of eastern Nepal, we find the verb forms *hiptūñ* "I hit him," *hipne* "I hit you," *khiptū* "You hit him," *hiptū* "He hits him," *khiptī* "He hits you," *āhiptūm* "We hit him," *khiptūm* "You all hit him," and so on.[9]

Konow believed that the pronominalized languages had borrowed this syntactic device from neighboring speakers of the entirely unrelated Munda languages. Such syntactic borrowing is not in itself impossible; in this case, however, it seems unlikely, for two reasons. First, the Munda verbal inflection system is very different in its basic structure from that of the pronominalized Himalayan languages studied by Konow; one would expect a greater similarity in structure—even if not in content—if the syntactic device had in fact been borrowed. Second, the Himalayan verbal inflections are quite similar among themselves, even between geographically distant languages, to the point where it appears possible to reconstruct a Proto-Himalayan verb system.[10] Such a proposed reconstruction would presumably place the development of the proto-inflectional system prior to any contact with Munda speakers. In any event, it seems both possible and plausible that the development was an internal one.

Many of these Himalayan languages, such as Newari, do not at first glance seem closely related to Tibetan; others—especially those in the Tamang-Gurung-Thakali group—appear strikingly similar not only in basic portions

---

9   This distinction cuts across geographical lines. Eastern pronominalized languages include Limbu, Rai, Chepang, and other groups in eastern and central Nepal; western pronominalized languages such as Kanauri are spoken primarily in the mountain areas of northwestern India outside Kashmir. Nonpronominalized languages include Gurung, Magar, Newari, and Lepcha or Rong, among others; these are distributed from the north of western central Nepal across to eastern Nepal and adjacent areas of India.

10   Indeed, some scholars have gone so far as to suggest not that the pronominalized Himalayan languages idiosyncratically acquired their inflections, but rather that the non-pronominalized languages lost the inflections they once had.

of their vocabulary but in syntax as well. For example, compare Gurung *khi-e tshami ña-e tshai-lai piɴ-n* "Give your daughter to my son" with Old Tibetan *khyod-kyi tsha-mo ñai tsha-la sbyin* New Tibetan (Lhasa) *khör-ɛ tshamo ñɛ tsha-lə ciɴ* "Give your niece to my nephew."[11] Note too the following apparent cognates—Old Tibetan *ñi* Kham *nimi* Limbu *nam* Gurung *din* "sun," Old Tibetan *mig* Kham *mi* Limbu *mik* Gurung *miɴ* "eye," Old Tibetan *rna* Kham *na* Limbu *nekho* Gurung *ŋa* "ear," Old Tibetan *śiñ* Kham *sin* Limbu *siñg* Gurung *siɴ* "firewood," Old Tibetan *gsum* Kham *sohn* Limbu *sumsi* Gurung *soɴ* "three," Old Tibetan *ltśe* Kham *se* Limbu *lesot* Gurung *le* "tongue."[12]

Such cognates must, of course, be distinguished—somehow—from loanwords. There is every reason to believe, for example, that Old Tibetan *tśhos* New Tibetan (Dbus) *tśhö* "dharma" Gurung *tśhyoe* "religious book" is a relatively recent loan. Note also other apparent loans in the same cultural sphere—Old Tibetan *bla-ma* New Tibetan (Dbus) *lama* Gurung *lamaa* "lama" (compared to the apparently genuine cognate Old Tibetan *bla* Gurung *pḷa* "soul"), Old Tibetan *rluñ-rta* New Tibetan (Dbus) *luñta* Gurung *luñta* "prayer flag," Old Tibetan *sño* New Tibetan (Dbus) *ño* "bless, pronounce benediction" Gurung *ño* "blow upon a sick person (by religious officiant)," Old Tibetan *rña* New Tibetan (Dbus) *ña* Gurung *ña* "shaman's drum."

2.4. TIBETAN AND THE WESTERN BARBARIANS

The Chinese historical records speak of nomadic and barbarian inhabitants of the high plains to the west, called, first, **khyañ > ch'iang* "sheepherders" (the graph represents a man and a sheep), and, later, **bhywan > fan* "barbarians," a word which may in fact have been borrowed from Old Tibetan *bon* "shamanic religion" or a related word in one of the Hsi-fan

---

[11] Gurung *tsha* "son" Old Tibetan *tsha* "nephew, grandchild" appear to be genuinely old Sino-Tibetan words: we find, for example, Dĩ·ʾnal *tśan* "son," Tsangla *za-sa* "child, baby," Burmese *sa* "child," Lushei *tu* "grandchild, nephew," and perhaps also Old Chinese **tsyəg* "son, daughter, child." Note also Old Tibetan *BTSA* "bear, bring forth," and perhaps Old Chinese **dzyəg*, which Karlgren interprets as having originally meant "foetus."

[12] Kham (not to be confused with the Khams dialects of Tibetan) is spoken in west Nepal by Magars of the Buda, Gharti, Pun, and Rokha subtribes; Limbu is spoken in east Nepal; and Gurung is spoken in the Gandaki zone in central west Nepal.

languages. The annals of the Han Dynasty note the existence of one group of *khyañ, located far from China, called the *pywat-khyañ, a term in which we may perhaps see a relationship with Old Tibetan bod "Tibet."[13] Later, during the T'ang dynasty, the Chinese distinguished between the *tho-bhywan > t'u-fan "agricultural barbarians," a term which came to be used regularly for the Tibetans, and the *syər-bhywan > hsi-fan "western barbarians," a loosely defined group of nomadic tribes ranging the plains in what is now the Amdo region.[14] The Tibetans drew the same distinction between themselves and these other nomads, even though the ways of life of the Tibetan and Hsi-fan nomads were basically the same; the Tibetans speak of the Horpa, Minyag, the eighteen tribes of the Rgyarong as not speaking the Tibetan language, although these languages have clearly borrowed a large number of words from central Tibetan.

In fact, the influence of central Tibetan on these languages has been so great that they have frequently been considered to be Tibetan themselves; note, however, Rgyarong snyis, Horpa zne, Minyag šan "seven" instead of Old Tibetan bdun "seven" and its New Tibetan reflexes, Rgyarong no, Horpa ni, Minyag na "you" instead of Old Tibetan khyod "you" and its New Tibetan reflexes, Rgyarong pram, Horpa phrū-phrū, Minyag phri "white" instead of Old Tibetan dkar "white" and its reflexes. The relationship between these languages and Tibetan, however, is clearly a close one: for example, in the Rgyarong dialect of Lcog-rtse, we find Old Tibetan rna Rgyarong rna "ear," Old Tibetan šiñ Rgyarong sying "firewood," Old Tibetan mig Rgyarong mnyak "eye," Old Tibetan dñul Rgyarong ngiy "silver," Old Tibetan gñis Rgyarong nis "two." But we must, as always, be wary of possible loan words, especially, here, from written Tibetan texts. A correspondence as close as Old Tibetan ldžañ-gu Rgyarong ldžañ-ku "green" is sufficient to arouse suspicion; but

---

[13] I think it is pretty clear that bod "Tibet," bon "shamanic religion," BO "call out, cry out," and perhaps PHO "change place, migrate," SPO "remove, shift, migrate," dbon/ʌbon "descendant, nephew, grandchild," NPHYO "range, roam about," form what we will later in this text call a word family.

[14] The Old T'ang History has a chapter on T'u-fan, and cites the recognizably Old Tibetan words btsan-po "king," transcribed as Middle Chinese *tsan-phwo, and blon "minister," transcribed as Middle Chinese *lywen, as native T'u-fan words. The New T'ang History, in its parallel chapter on T'u-fan, gives a more extensive vocabulary list of Tibetan government officials, and similarly transcribes Old Tibetan btsan-po "king" as Middle Chinese *tsan-phwo and Old Tibetan blon "minister" as Middle Chinese *lywen, using the same Chinese characters; but here the language is cited as being that of the *khyañ.

when we find Old Tibetan *stag* Rgyarong *khuñ* "tiger" but *stak* "tiger" in the Lcog-rtse dialect, or Old Tibetan *dbyar-ka* Rgyarong *tsar* "summer" but *dbyar-ke* in the Chos-kia dialect, it is reasonable to believe we have found an informant with a literary education.

3. VARIATION IN TIBETAN

If Tibetans from different parts of Tibet are asked to give their word for "hair," a Tibetan from Purik will say *skra,* one from Amdo will say *škya,* one from Kham will say *štra,* one from Tao-fu will say *ṣṭra,* and one from Bhutan or Sikkim will say *kya.* Similarly, a Ladakhi will say *ṣa,* a rural Central Tibetan will say *ṭa,* and an upper-class resident of Lhasa City will say *ṭṣa.* But if these Tibetans are literate, and are asked to write the word they had just spoken, they will all produce the same written form, which we here transcribe as <*skra*>. And, if they are shown the written form <*skra*>, they will, again, pronounce the word differently, but they will all recognize the form and agree that—however it is pronounced—it means "hair."

One reason for this is the remarkable conservatism of the Tibetan writing system. The written form <*skra*>, for example, with the same meaning "hair," can be found in manuscripts more than a thousand years old, preserved in the deserts of Central Asia, which can still be read—in some sense of that term—by any literate Tibetan.[15] The written form has remained unchanged; the word represented by that form has come to be—or has continued to be—different in different dialects. The advantage of such uniform orthography is its transcendence of regionalism: all literate Tibetans share a single written language, however different their spoken dialects may be. The disadvantage is the divorce between the written and spoken languages, making literacy an increasingly difficult and elite accomplishment.

Now when Tibetan was first reduced to writing, it seems reasonable to assume that the written form <*skra*> was, in fact, an attempt to render a word pronounced something very much like *skra.* We thus find variation in

---

[15] For example, in a mythic text from the caves near Tun-huang we find *dbu-skra bdun . . . ɴbrog-srin dre-das ɴtsald* "The fiend of the wastes, Dre-da, demands seven hairs from his head." Again, in a prosaic administrative memorandum from a Tibetan oasis garrison in Central Asia, reporting on the collection of animal hair for rope-making, we read *bžer-gis phyugs ɴga skra srañ phyed gyis* "Bžer has made a half *srañ* of hair of some animals."

the Tibetan language along two dimensions. The language varies along a
DIACHRONIC dimension, wherein a word pronounced *skra* in the ninth century
has come to be pronounced, say, *ṭa* in the twentieth; and the language varies
along a SYNCHRONIC dimension, wherein a word now pronounced *ṣa* in
Ladakh is pronounced *štra* in Kham, or pronounced *ṭa* by a Lhasa City
storekeeper is pronounced *tṣa* by a Lhasa City aristocrat.[16]

When we speak of the history of the Tibetan language, we will use the term
PROTO-TIBETAN to refer to the Tibetan language spoken before the existence
of any written records. We will use the term OLD TIBETAN for the language
spoken during the earliest period for which written records exist—that is,
more or less arbitrarily, for the language spoken, say, from the seventh to the
tenth centuries, which is the language upon which those earliest written texts
were based. The term MIDDLE TIBETAN will refer to the language spoken
from the tenth to the nineteenth centuries, a period for which we have an
awesome quantity of written materials, but about whose spoken language we
can make only scattered inferences. Finally, the term NEW TIBETAN will refer
to the spoken language for which we have modern contemporaneous tran-
scriptions and analyses, beginning in the nineteenth century with the first
European explorers and missionaries.[17] When we speak of synchronic
variation, we will adopt the convention of citing forms by historical period
followed by a parenthetical indication of dialect where such information is
available—for example, Old Tibetan *bdun* but New Tibetan (Dbus) *dün*
"seven," Old Tibetan *myi* but Old Tibetan (Sumpa) *mu* "man."

3.1. VARIATION IN NEW TIBETAN

When a Tibetan from Ladakh and a Tibetan from Lhasa City go to the

---

[16] Diachronic variation, of course, occurred as well before the earliest written texts.
Where we find Kanauri *kra* and Kachin *kəra,* for example, we can hypothesize an earlier
Proto-Tibetan **kra* "hair," to which was prefixed the formative **śa* "ANIMAL."—thus Proto-
Tibetan **śa-kra* "ANIMAL hair" > **s-kra* > Old Tibetan *skra* "hair." And synchronic
variation occurred at historical periods other than the modern. There is some textual
evidence, for example, that, alongside Old Tibetan (Lhasa) *myi* "man," there was an Old
Tibetan (Sumpa) *mu* "man" as well.

[17] Note that these terms are really methodological rather than properly linguistic. The
classification depends upon the fortuitous existence of written records on the one hand and
modern transcriptions on the other. Middle Tibetan simply includes everything in between.

market together to buy vegetables, the Ladakhi is shopping for *tshodma* but the Lhasan for *tshɛ*. If they buy spinach, the Ladakhi calls it *palak* and the Lhasan calls it *potsɛ*. If they buy peas, the Ladakhi calls them *ṣanma* and the Lhasan calls them *ṭɛɴma*. When they pay, the Ladakhi calls the rupee coin *kyirmo* and the Lhasan calls it *kɔmo*. Are they speaking the same language? They will both say they are speaking Tibetan; but the Ladakhi will call the language *potskat* and the Lhasan will call it *phökɛ*.

Even if we look just at the lexicon, leaving grammar aside, the relationship between the two dialects is complex. For example, continuing with vegetables, we find Ladakhi *labuk* and Lhasa *ləpu* "radish" < Old Tibetan *la-phug*, where the word is recognizably the same in both dialects;[18] and, similarly, we find Ladakhi *tsoñ* and Lhasa *tsoɴ* "onion," although in this case the word gives every appearance of having come into Middle Tibetan from Middle Chinese *tshuñ "onion," rather than of being a native Tibetan word. On the other hand, we find Ladakhi *ṣanma* and Lhasa *ṭɛɴma* "peas" < Old Tibetan *sran-ma*, where a common origin of the word in Old Tibetan is less obvious, but the differences in pronunciation are the result of more or less regular phonological changes in each dialect.[19] Ladakhi *gobi* and Lhasa *kopi* "cauliflower" appear alike not because the words have a common Old Tibetan origin, but because the two dialects have recently—and apparently independently—borrowed the Hindi word *phul gobhi* "cauliflower." The Lhasa dialect uses the compound *koɴləpu* <*guñ la-phug>* "middle finger radish → carrot," while Ladakhi uses the term *sarakturman* "carrot," almost certainly borrowed in part from Urdu *zardak* "carrot."[20]

---

[18] The radish—although certainly not a literary staple—does crop up in some genuinely ancient texts. For example, a ninth-century administrative memorandum from a Tibetan garrison in Central Asia, written on a strip of wood, apparently listing expenses incurred, includes *spreu lo-la sku-bla gsol-bai lha-phug rin* "for the monkey year, the cost of radishes for entertaining the nobles."

[19] Thus we find not only Old Tibetan *sran-ma* "peas" > New Tibetan (Ladakh) *ṣanma* (Lhasa) *ṭɛɴma*, but also Old Tibetan *skra* "hair" > New Tibetan (Ladakh) *ṣa* (Lhasa) *ṭa*, Old Tibetan *sprin* "cloud" > New Tibetan (Ladakh) *ṣin* (Lhasa) *ṭiɴ*. Similarly, Old Tibetan *skad* "language" > New Tibetan (Ladakh) *skat* (Lhasa) *kɛ*, Old Tibetan *žal* "face" > New Tibetan (Ladakh) *žal* (Lhasa) *šɛ*, Old Tibetan *mtshan* "name" > New Tibetan (Ladakh) *tshan* (Lhasa) *tshɛɴ*.

[20] The second half of the Ladakhi *sarakturman* "carrot" is not so easy. My best guess is that it is derived from Urdu *darman* "medicine," but the semantics are certainly not obvious.