

History of the Tibetan Language

TAKEUCHI Tsuguhito

Kobe City University of Foreign Studies, Emeritus

Summary

This paper surveys the historical development of the Tibetan language. This language has two different styles, literary and colloquial, the former of which is documented from the 7th century AD. The founder of the Tibetan Empire *Srong tsen sgam po* hastened to establish the official language and its orthography. Old Literary Tibetan gradually gained literature in many genres, such as annals, law, contracts, and divination. Old Literary Tibetan was used as a lingua franca by neighboring non-native speakers (*Han* and *Khotanese*, for instance) by the 9th century AD.

Literary Tibetan has developed both in orthography and syntax. The complex syllable structure was simplified in colloquial Tibetan, which is reflexed in orthographic change. In syntax, the auxiliary verb *yin* newly emerged and came to function to reflect the speaker's judgment. Literary Tibetan grew more stable by introducing the linguistic change of the colloquial style that occurred in Central Tibet since the 12th century AD, when Old Literary Tibetan was re-established as Classical Tibetan.

In colloquial Tibetan, the simplification of syllable structure caused tonogenesis probably around the 10th century AD, and the distinction between *myi* and *mi* was lost. At the same time, the morphological structure of the colloquial language has changed from synthetic to analytic. After the 11th century, the loss of consonantal codas (*-d*, *-n*, *-l*, *-s*) caused vowel changes in the Lhasa dialect. In Early Modern Colloquial Tibetan, the auxiliary verb *red* was introduced from the *Amdo* dialect and diffused into Central Tibet. This type of linguistic change induced the influence of the Early Modern Literary style.

On the whole, there are many common features attested from the Old Tibetan to Modern Tibetan dialects, which are quite distinct from other Himalayan languages, and after the Old Tibetan Empire was established, the Tibetan language rapidly diffused over the entire Tibetan territory with the Tibetan Empire's abrupt extension, after which the Tibetan dialects evolved. This is what I call the "Big Bang hypothesis."

Key words: Old Literary Tibetan, Old Colloquial Tibetan, Classical Tibetan, Modern Dialects, Big Bang hypothesis

キーワード：古チベット文語，古チベット口語，古典チベット語，現代諸方言，ビッグバン仮説

1. Introduction — What is Tibetan?

“Tibetan” refers to the language of Tibet, but more specifically it usually refers to the spoken (colloquial) and written (literary) languages of Central Tibet. However, the spoken language has a diverse variety of dialects distributed over a vast area, and there are numerous variations in the written language from the 7th century to the present day. Nevertheless, taken as a whole, these dialects can be said to possess a unified Tibetan identity. Establishing the identity of Tibetan can be traced back to the foundation of Tibet in the 7th century.

The founder of Tibet, *Srong brtsan sgam po* (*Khri srong rtsan*), developed a national regime at an astounding speed after founding the country, and his first task was the establishment of an official language (spoken and written). The spoken language of the royal family was used as the official spoken language of the nation, and the written language was established using characters newly created for that purpose. With the expansion of the nation, the spoken and written languages spread throughout the Tibetan Plateau, laying the foundation for Tibetan. The earliest spoken and written languages in this period are collectively referred to as “Old Tibetan.” When making a particular distinction between the written and spoken languages, the terms “Old Literary Tibetan” and “Old Colloquial Tibetan” are used.

Thereafter, the spoken language developed independently in each region and formed the various modern-day dialects. In contrast, the literary language, demonstrating gradual changes compared to changes in the spoken language, continued to the present day with its old forms well preserved. Since it facilitates communication between various dialects, it plays a central role in maintaining a sense of unity in the Tibetan language.

Thus, “Tibetan” can be described as a collective term for diverse spoken dialects distributed over a vast area eight times the size of Japan and a prominent literary language in Asia from the 7th century to the present day.

2. The Formation and Development of Old Literary Tibetan

2.1 The Establishment of Written Characters, Orthography, and Written Forms

The Tibetan script was created based on the Indic Gupta script used in Nepal and Northern India. It is a phonetic script consisting of thirty basic consonant characters and four vowel symbols. While Tibetan characters essentially retained the characteristics of Indic characters, such as the layout of the character tables, it was estimated that Tibetan characters were officially established in the 630s and 640s after some improvements were made, such as the creation of new characters and three-tiered character notations for expressing complex syllables. Ten years later, in 650, yearly events of national importance began to be recorded in the *Old Tibetan Annals* (the Annals), and in 655, a legal code was formulated. Thus, it can be said that Tibetan-language documents were first produced at the beginning of 650.

After the establishment of the script and orthographic rules, the next important task was setting written forms that differed according to the genre of the text and the fixed expressions used in each, for example, for annals, legal codes, contracts, letters, divination, and inscriptions. To illustrate this, in the Annals, each year was expressed in a fixed format such as “the year falling in (i.e., the year that is) (*par lo gchig*) ... the year of the tiger (*stagi lo la bab ste*).” Other texts such as contracts, letters, divination, and inscriptions each have a characteristic form. Recent studies have shown that the existence of such a clear written form is characteristic of Tibetan texts (Takeuchi 2012). As will be described later, Tibetan-language texts came to be used by many ethnic groups in the Tibetan Empire’s territory along the Silk Road, including the Han Chinese, Khotanese, Uyghurs, and Sumpas. Presumably, the existence of prescribed written forms was a key aspect of non-Tibetan peoples’ assimilation of Tibetan-language texts.

Some believed that the Tibetan script was created to transcribe Buddhist scriptures, but in fact it was primarily created to achieve national unification and administrative documentation. Thus, as the Tibetan territory expanded, its official language spread throughout its territories. Buddhism is thought to be first introduced during the reign of King *Srong brtsan sgam po*, and Tibetan Buddhist scriptures would have been produced thereafter (*Śālistamba-sūtra*, for instance, cf. Takeuchi 1997–1998 (OTM): texts 680–693). In fact, many Tibetan Buddhist texts were produced in the latter half of the 8th century, when Buddhism became the state religion. Subsequently, a large-scale project for translating and transcribing Buddhist sutras was carried out, and Buddhist scriptures became the main body of Tibetan-language documents.

2.2 The Diffusion of Old Literary Tibetan — From the Official Language of the Tibetan Empire to an International Lingua Franca

As Tibet rapidly expanded its territory and gained control over the entire Tibetan Plateau, Literary Tibetan spread from Qinghai in the east to Ngari in the west. Moreover, peoples such as the Zhangzhung, Sumpa, and Azha (Tuyuhun) who did not originally speak Tibetan as a mother tongue came to use Literary Tibetan as an official language.

Tibet further expanded into Central Asia, and for half a century, it governed the Southern Silk Road to the west from the Wakhan Corridor (Northeast Afghanistan) to Baluchistan (Northeast Pakistan), to the north and east, from Kashgar, Khotan to Dunhuang. During this time, it became clear that various non-Tibetan peoples, such as the Han Chinese residents of Dunhuang and the Iranian-descended Khotans, also learned and used Literary Tibetan.

For example, Tibetan Buddhist scriptures produced in large quantities in Dunhuang were derived from Han Chinese sutra copyists, who produced many other documents besides Buddhist scriptures, such as letters, contract documents, and graffiti (Takeuchi 2013). We also know that Khotan officials, whose native tongue was an Iranian language, wrote contracts and many other documents in Tibetan (Takeuchi 1995: 169–174).

Thus, Literary Tibetan, established as the official language of Tibet, came to be widely used by many ethnic groups who did not speak Tibetan as a native language. As a result, Literary Tibetan took on a universal character and transitioned into being an international lingua franca suitable for intercultural communication. This gave rise to a situation in which Literary Tibetan continued to be used from Central Asia to Mongolia even after the end of the Tibetan Empire, mainly by non-Tibetan Buddhists.

The Tibetan Empire collapsed in 842, and Tibet's rule over Central Asia ceased around 850. It was impossible that Literary Tibetan texts would continue to be used in Central Asia, where there were no more Tibetans. However, Hungarian scholar Uray (1981) identified the fact that letters written in Tibetan were exchanged between the Han administration (the Guiyi Circuit) in Dunhuang, the Uyghur state established in Ganzhou, and the Kingdom of Khotan, a country that was independent from Tibet. Countries that spoke different languages—Chinese, Uyghur, and Khotan—used Tibetan as a lingua franca, as it was the language that they had all learned. It also became apparent that, thereafter, an unexpected situation arose in which Literary Tibetan documents were used not only as diplomatic documents, but also in various other fields, such as letters and litigation documents used by the Hans and Khotans. Moreover, during the same period, Tibetan Buddhism flourished in Central Asia, and many Tibetan Buddhist texts were produced. Tibetan was the international lingua franca in Central Asia in the 10th century (Takeuchi 2002, 2012).

Thus, Old Literary Tibetan, which was established in the 7th century, was used continually for five centuries, changing its shape and character gradually. Over this long period, Old Literary Tibetan can be divided into the following three phases:

- (1) **Early Old Tibetan** Mid-7th century to late 8th century
The formative period of Old Literary Tibetan. Documents during this period were transmitted to the present day only in the form of copied manuscripts, such as the Annals.
- (2) **Middle Old Tibetan** Late 8th century to early 9th century
The period of diffusion of Old Literary Tibetan. At the zenith of the Tibetan Empire, Literary Tibetan was widely used throughout its territories. Many paper documents and wooden tablets during this period were excavated in Central Asia, which was controlled by Tibet, and stone inscriptions survived in the Tibetan heartlands.
- (3) **Late Old Tibetan** Late 9th century to 11th century
The international lingua franca of Central Asia after the collapse of the Empire. It was used as a lingua franca by non-Tibetan peoples, such as the Hans, Khotans, and Uighurs, and became popular as the sacred language of Tibetan Buddhism. It has features in terms of the written form and style that are referred to as the Post Imperial Style.

3. The Development and Diffusion of Old Colloquial Tibetan

3.1 The Establishment and Diffusion of an Official Language

The core peoples of Tibet when it was first founded were *spu rgyal*, who migrated from Yarlung to Central Tibet, the *rkong po*, who were the royal family and their relatives, the *myang po*, and the *dvags po*, those from southeast regions. It is believed that the official language of the state was based on the spoken language of these peoples. With the rapid expansion of the nation, people who spoke the official language as a native language or a second language spread throughout the Tibetan Plateau, conquering and absorbing people who spoke various other languages in the process. Peoples who fall under this category include the Zhangzhungs in the west, who spoke a Tibeto-Burman language, and, to the east and north, the Sumpa and Nampa, who spoke languages similar to the Tibeto-Burman language, the Turkic Bailan (白蘭), and the Mongolic Azha (Tuyuhun 吐谷渾). It is thought that they were incorporated into Tibet's military regime and were assimilated into the cultural mainstream. They continued to retain their languages to some extent, but they were eventually assimilated into the Tibetan culture. This process may have influenced Tibetan as a substrate language. The wide variety of dialects that persist to this day, particularly in the West Kham region where Tibetan or Tibeto-Burman languages were spoken, may be due to the influence of substrate languages sharing a common ancestry. Meanwhile, in areas where there were no Tibetic languages, such as Baluchistan, Amdo, and East Kham, the linguistic forms of Old Tibetan are still well preserved.

3.2 Sound Forms of Early Old Colloquial Tibetan (7th Century)

The spellings of words in Old Literary Tibetan established in the 7th century were based on the pronunciation of the spoken language. Thus, pronunciation can be estimated as follows from a comparison between the spelling forms of Old Tibetan and the various dialects preserving the old forms described above.

Syllable structure $C_1C_2C_iRVC_3C_4$ (Example: *bsgrubs* 'completed')

Before the initial consonant C_i , the enclitic consonants C_1 , C_2 , and C_i may be followed by an intervening sound R (optional).

C_1 : *b-*
 C_2 : *r-, l-, s-, g-, d, b-, m-, ' [fi-]*
 R: *-r-, -y-*

There are spellings with the medial sound *-l-* in them (*kl-, gl-, bl-, rl-, sl-, zl-*), but *l* should be considered the initial consonant preceded by an enclitic consonant. This will

be explained later using the example of ‘moon’. The medial sound *-w-* (*wa zur*) was also present in the spelling and was used in the transliteration of sounds from foreign languages (e.g., *kwa cu* for Chinese Guazhou 瓜州), but it was not present in the spoken Tibetan language.

Initial consonant C₁

Stops and Affricates: In the Tibetan alphabet, we can observe a triadic contrast of voiceless/unaspirated, voiceless/aspirated, and voiced for the following sound groups: *k, kh, g; c, ch, j; t, th, d; p, ph, b; ts, tsh, dz*. Since the unaspirated/aspirated contrast can be regarded as a complementary distribution, it is probable that they made a binary opposition in its phonological level. A voiceless glottal stop [ʔ] stands alone.

Fricatives:	<i>s-</i> , <i>z-</i> , <i>sh</i> [ɕ], <i>zh</i> [ʒ], <i>h</i> , ' [ɦ]
Nasals:	<i>ng</i> [ŋ], <i>ny</i> [ɲ], <i>n</i> , <i>m</i>
Liquids:	<i>l</i> , <i>lh</i> , <i>r</i>
Semi-vowels:	<i>y</i> [j], <i>w</i>

The liquid *lh* is represented by a compound letter, but, in Old Colloquial Tibetan, it is regarded as a single phoneme. It is a voiceless alveolar lateral fricative. The 9th century phonemic transcription suggests that it may have had a rather stronger voicing (e.g., *lha* = 他, *lho* = 士; cf. Li Fang-kuei 1979).

Vowels

Vowels are *i*, *e*, *a*, *o*, and *u*.

We have a mirrored or backward letter for *i*. According to this fact, we may be able to hypothesize the existence of a central unrounded *i*. However, our observation of the Old Tibetan texts after the 8th century demonstrate that it is hard to consider that the mirrored or backward letter for *i* stood for another sound than *i*.

Final Consonants

C ₃ :	<i>g, d, b, s, l, r, m, n, ng</i> [ŋ]
C ₄ :	<i>s, d</i>

As mentioned above, Old Colloquial Tibetan seems to have had a complicated syllable structure. Its pronunciation mostly corresponds to the orthographical shape with some exceptions. Let us examine *l* and *lh* as an example.

The word ‘moon’ is spelled as *zla* and pronounced [ta, da, nda, ldza] depending upon dialects. This fact reflects particular sound changes in Tibetan. In the proto-Tibeto-Burman,

the form *s-la is reconstructed (STC #144, Benedict 1972: 42), while, in the modern Burmese, rGyalrong, and Nung, we have [la], [la], [sala], respectively. Moreover, in Old Tibetan texts, ‘moon’ was *sla*. The prefix *s-* was assimilated by the voiced *l-* that follows it, becoming *z-*. If the prefix *s-* is followed by a voiceless *lh*, it is not voiced. For example, if *lhan* ‘patch’ is prefixed by *s-* (causative), it becomes *s-lhan* > *slan* ‘to stick on’.

Let us go back to the relationship between pronunciation and orthography. In terms of orthography, in the word *zla* ‘moon’, the root consonant letter is *z* and *l* suffixes it. However, in terms of pronunciation in the colloquial language, *l* is the initial consonant C_1 and *z* is C_2 . Nevertheless, as mentioned earlier, the consonant that occupies the C_2 slot should be *s-*; *z-* was a voiced *s-*. In Old Tibetan texts, we have both *sla* and *zla*, which testify the sound change *sla* > *zla*.

When the alveolar lateral fricatives, voiceless *lh-* and voiced *l-*, are prefixed by *g-*, *d-*, and *s-*, the *s-* before the voiced *l* is voiced and the *g-* and *d-* before the voiceless *lh-* are devoiced.

<i>s-l</i> > <i>zl</i>	<i>g-l</i> > <i>gl</i>	<i>d-l</i> > <i>dl</i>
<i>s-lh</i> > <i>sl</i>	<i>g-lh</i> > <i>kl</i>	<i>d-lh</i> > <i>tl</i>

Furthermore, a metathesis occurred; the *dl* and *tl* became *ld* (*dl* > *ld*) and *lt* (*tl* > *lt*), respectively.

<i>d-log</i>	>	<i>dlog</i>	>	<i>ldog</i>	‘to return’
<i>d-lhung</i>	>	<i>tlung</i>	>	<i>ltung</i>	‘to drop’

The ‘moon’ discussed above reached the present-day shape after further changes in, for example, *zla* > *dzla* > *dla* > *lda* > *da*.

Thus, there seems to be some discrepancy between Old Colloquial Tibetan and the written forms, but its pronunciation fundamentally corresponded to the orthographical shape.

3.3 Changes in Pronunciation (7th to 10th Centuries)

The sound forms of the spoken language in the 7th century, when Old Literary Tibetan was established, can be assumed as described above. Later, the literary language stabilized, spread throughout the Tibetan Empire, and continued to be used. However, materials with corresponding sounds in other languages, such as Khotan and Chinese, indicate that the spoken language had already changed in the later period of the Empire in the 9th century.

(1) Simplification of the Syllable Structure

Enclitic consonants (*b-*, *d-*, *g-*, *r-*, *m-*) preceding the initial consonant were no longer pronounced at the beginning of the word except for *s-*. For example, the name of the Khotan official Sudārjām, who appears in Tibetan-language contract documents, is expressed variously as *sar-zhong*, *bsar-gzhong*, and *ksar-gzhong* in Tibetan scripts (Takeuchi 1995: 272). It seems that the deliberate addition of *b-* and *g-*, which were no longer pronounced at that time, may have made the Khotan name seem more Tibetan. In the Sino-Tibetan Treaty Inscription, written in Tibetan and Chinese, Tibetan names are transcribed in Chinese characters.

In this case, the enclitic *s-* is transcribed with the Chinese character 悉 — for example, *stag* (悉諾), and *snam* (悉南). Conversely, we can see that the enclitic *r-* such as *rma* (摩) and *rgan* (幹) had no transcribed equivalent in Chinese characters, thus indicating that it was already not pronounced (Li 1979: 235–237).

(2) The Devoicing of Initial Voiced Consonants

In Early Old Tibetan, as mentioned earlier, the initial consonants were stops, affricates, and fricatives, with opposing pairs of unvoiced/voiced sounds. However, as the voiced sounds gradually became unvoiced, the unvoiced/voiced pairing was replaced by a pitch difference of high/low pitch.

(3) The Development of Tones

This way, the unvoiced/voiced pairs were replaced by a difference in terms of the high/low, and enclitic consonants were no longer pronounced, thereby leading to a difference in pitch as a pairing of high tone/low tone.

<i>rtā</i>	>	tā	‘horse’	<i>mthā</i>	>	thā	‘edge’	(High)
<i>mdā</i>	>	tā	‘arrow’	<i>dā</i>	>	thā	‘now’	(Low)

In the case of initial nasal sounds, liquid sounds, and semivowels, it is believed that the disappearance of the enclitic consonant directly led to the development of tones.

<i>lṅga</i>	>	ṅga	‘five’	<i>sna</i>	>	na	‘nose’	<i>bla</i>	>	la	‘up’	<i>gyu</i>	>	yu	‘turquoise’
<i>ṅga</i>	>	ṅga	‘I’	<i>na</i>	>	nā	‘illness’	<i>la</i>	>	lā	‘pass’	<i>yug</i>	>	yū	‘stuff’

The development of tones in Tibetan is presumed to have taken place around the 10th century, based on materials with corresponding sounds in Chinese (Takata 1988: 82).

(4) Changes in Initial *m-/my-*

In the 7th-century spoken language, it appears that there were two series of non-palatal and palatal sounds with the initial consonant *m* plus a vowel.

Non-palatal	<i>m-</i> :	<i>mi</i>	<i>me</i>	<i>ma</i>	<i>mu</i>	<i>mo</i>
Palatal	<i>my-</i> :	<i>myi</i>	<i>mye</i>	<i>mya</i>	<i>myu</i>	<i>myo</i>

However, the distinction between *mi* and *myi* was already unclear, with the majority of cases written *myi*. Regarding the distinction between *me* and *mye*, *me* was used in *men-tog* (flower) and *men-tri* (silk), while in other cases, it was written as *mye*. This distinction gradually disappeared, and *myi* and *mye* were unified with *mi* and *me*, respectively.

However, *mya*, *myu*, and *myo* were pronounced as the palatalized [ɲa], [ɲu], and [ɲo] and were distinguished from the non-palatal sounds [ma], [mu], and [mo].

These phonological changes occurred in the official spoken language, especially in Central Tibet, but the opposite change occurred in the Amdo dialect in the northeast. *myi* and *mye* became the palatal sounds [ɲi] and [ɲe]. The enclitic consonants changed but they were preserved, and tones did not develop.

For example, the Old Literary Tibetan form for ‘eye’ is *dmyig*, and it is believed that the spoken language was pronounced almost the same way. However, as mentioned earlier, the enclitic consonant *d-* was no longer pronounced in the 9th century, and the form *myig* also appears in the Old Tibetan literature. Furthermore, as a result of the change *myi* > *mi*, the literary form of *mig* developed, which was established as the form of Classical Tibetan as described in the next section. In the Modern Tibetan central dialect, it is pronounced [mik] (high tone; -k is an unreleased stop). Meanwhile, in the Amdo dialect of Qinghai, it is pronounced [ʰɲɿx]. *myi* changed to [ɲɿ], and the enclitic consonant *d-* was retained as [ʰ].

These phonological differences suggest that in the latter half of the 9th century, the dialects of Central Tibet, the official language (standard language) based thereon, and the Amdo dialect had already diverged from one another.

3.4 Changes in Morphology and Syntax (7th to 10th centuries)

The basic sentence structure of the 7th century colloquial and literary Old Tibetan consists of multiple nouns (or noun phrases) and a predicate comprised of either a noun (N), an adjective (A), or a verb (V), followed by a sentence final marker *-o* (*rdzogs tshig*).

Stage 1 S → (N)+(N) + Pred (N/A/V) + -'o (a sentence final marker)

- (1) *mo di' ci la btab kyang bzang rab 'o/*. [IOL Tib J 738: lv line 39]
 'This divination, for whatever it is cast, (the result is) very good.'
- (2) *zu tse glo ba nye 'o*. [P.t. 1287: line 199]
 'Zu-tse is a man of faithfulness.'
- (3) *slan cad gyang nye zho dang myed par smon to*. (M.Tagh.b.i.0096)
 'Hereafter as well, [we] wish you to be free from illness.'

In this structure, it was difficult to add any negative marker (*ma/myi*). To solve the problem, an auxiliary verb *yin* was introduced to support the negative marker *ma*.

Stage 2 Introduction of the auxiliary verb *yin*

with negation: *ma yin*

- (4) *bde ba'i skabs/ sdug bsngal ba'i gnas skabs/ sdug bsgal yang ma yin bde ba yang ma yin ba'i gnas skabs/* [P.t. 1261: line 70]
 'the state of happiness, the state of misery, the state of neither misery nor happiness'

In this example, the auxiliary verb *yin* is used only when it is in the negative form. *yin* further began to have an effect on emphasizing the speaker's will or judgment as shown in the following examples.

speaker's will

- (5) *sman snga ma skur ba / yin no /* (BTT: 1.2)
 '[I] will send [you] medicine immediately.'
- (6) *gdod 'jug par gsol ba yin no /* (P.t. 1287: 212)
 '[I] am requesting to be put [into the bag] for the first time' thus [Myi-chen] said.'

Furthermore, the syllable structure was simplified as a result of the sound changes described in the previous section. This affected verbs with complex syllable structures. As illustrated by *gsad* ‘to kill (imperfect)’ and *bsad* ‘killed (perfect)’, the difference in temporal phase was indicated by the enclitic consonants *g-* and *b-*. However, the disappearance of the enclitic consonants made them indistinguishable. To make up for this, a verb was suffixed with the particle *pa* to form a noun phrase, thus deriving the more analytic form of *V + pa yin* with the auxiliary verb *yin*.

Stage 3 Verb + *pa yin*

(7) *nga rang gyis bris pa yin*. [IOL Tib J 773: line 2]

‘(I) myself wrote (it).’

Phonetic changes in the spoken language, namely, the simplification of the syllable structure, triggered changes in the morphological and syntactic structure from a synthetic structure to an analytic structure over the course of 200 years. This gradually influenced the literary form, resulting in a shift from Old Literary Tibetan to Classical Tibetan.

4. The Establishment of Classical Literary Tibetan (11th to 14th Centuries)

As we have seen, in the first half of the 7th century when Old Literary Tibetan was established, the spoken and written languages matched. However, both spoken and written languages spread throughout the Tibetan Empire, such that 200 years later in the second half of the 9th century, regional differences (dialects) emerged in the spoken language. In particular, the spoken language in Central Tibet underwent major changes, as described in Section 3. Meanwhile, in the first half of the 9th century, reforms were introduced, such as a revision of spellings in the literary language and a revision of the language used for the translation of Buddhist scriptures. However, this did not have the expected effect, and the result was that the gap between the written and spoken languages widened.

In the late 9th and 10th centuries, the only extant Old Tibetan documents were those written by non-Tibetans such as the Han Chinese, Khotans, and Uighurs. It is assumed that Colloquial Tibetan was not their native language but a second language. The Old Tibetan literature in the 10th century includes vocabulary and expressions that seem to be influenced by Chinese and Khotanese, but it essentially follows the Literary Tibetan of the Imperial Period. The turning point in the written language was in the 12th century.

In the Tibetan heartland, which was divided after the collapse of the Empire, a new movement to revive Buddhism emerged at the beginning of the 11th century in Kham in Eastern Tibet and Ngari in Western Tibet. Particularly in Western Tibet, which was

geographically close to India, new sutra translations and missionaries flourished with the arrival of high priests such as Atiṣa from India. In this Buddhist revival movement, new orthography was established to correct the spelling variations found in Old Literary Tibetan, especially in the translation of Buddhist texts, and to adapt to changes in the spoken language in Central and Western Tibet. This led to the establishment of Classical Literary Tibetan.

The major revisions made in orthography in the transition from Old Literary Tibetan to Classical Tibetan were as follows:

- (1) Determining aspirated and unaspirated sounds
- (2) *myi* > *mi*; *mye* > *me* (cf. Section 3.3)
- (3) Disappearance of the reverse *i* symbol (*gi gu*)
- (4) Disappearance of re-addition suffix (*da drag*), e.g., *gyrud* > *gyur*, *phyind* > *phyin*
The re-addition suffix (*da drag*), which indicates a verb's perfect form, ceased to be pronounced in the 9th century and disappeared from the spelling. However, thereafter when the final sentence *-o* was used, a trace of *da drag* remained in the form of *to* (*gyurd-o* > *gyur to*).
- (5) *sts-* > *s-*: *stsel* > *sel*; *lastsogste* > *la sogs te*
- (6) Thorough separation of syllables: *lastsogste* > *la sogs te*

However, these reforms were not put into practice simultaneously, but they gradually spread from the 12th to the 14th centuries. The actual process of the transition from Old Literary Tibetan to Classical Tibetan was substantiated by recently discovered documents such as the Tabo Monastery documents found in Western Tibet (Scherrer-Schaub and Steinkellner eds. 1999), the Pu Ri manuscripts (Xizang Daxue Zangwen Guji Yanjiusuo eds. 2018), the Book of Kadam in Central Tibet (Kano 2021), the Khara-Khoto documents in the northeast (Takeuchi and Iuchi 2016), and the documents of the northern grottoes of Mogaoku, Dunhuang (Peng, Wang, and Dunhuang Yanjiuyuan eds. 2004). These are a valuable group of manuscripts that can verify the transitions in the written form before woodcut printed books became mainstream in the 15th century, and further research into these documents should be fruitful. Since the 15th century, woodcut printed books became mainstream, and Classical Tibetan became more established and more widespread.

5. Changes in the Spoken and Written Languages (11th Century Onwards)

5.1 Changes in Vowels and Final Consonants

In Section 3.3, we mentioned that the syllable structure $C_1C_2CiRVC_3C_4$ of Old Colloquial Tibetan changed, the enclitic consonants C_1C_2 disappeared, the initial consonant C_i became

unvoiced, and the high and low tones developed instead of this. These phonetic changes occurred in the Central Tibetan colloquial language in the 9th and 10th centuries. We also mentioned that, at that time, there was already a difference in pronunciation between Central Tibet and Amdo in the northeast; that is, a dialectical difference emerged. After the establishment of Classical Tibetan, the spelling was fixed to reflect the changes in the spoken language of Central Tibet, and changes in the spoken language were not further reflected in the written language until the early modern period. For this reason, changes in the spoken language from the 12th century to the present day cannot be easily determined from the literary form.

Materials with corresponding sounds that capture the pronunciations used in the past include the *Pentaglot Dictionary* (*Wuti Qing Wenjian* 五體清文鑑) and *Tibetan-Chinese Vocabulary* (*Hsi-Fan-Kuan I-yu* 西番館譯語). The former, written in the late 18th century, transcribed Tibetan pronunciations into Manchu characters. It is worth noting that while this recording of Tibetan in the phonetic Manchu script is valuable, observation of its notation reveals that it is likely a transcription of the sound of reading by a Mongolian Buddhist monk rather than a specific dialect such as the Amdo dialect. The *Tibetan-Chinese Vocabulary* was written during the reign of the Ming and Qing dynasties and existed in multiple forms. Transcriptions of the East Tibetan dialects into Chinese characters in the 17th and 18th centuries are valuable linguistic materials, and research is already advancing in this regard. With more estimation of the sounds of Chinese characters and detailed description of modern related dialects, the value of these materials increases (Nishida 1970, Suzuki 2009, Matsukawa and Miyake 2015).

Apart from this, if we compare modern dialects with literary forms, we can estimate the changes that occurred during that time. We will now describe changes in the vowel V and the final consonants C₃ C₄, which comprise the latter half of the syllable C₁C₂C₁RVC₃C₄. The simplification of the syllable extended to the latter half of the syllable. C₄ was the first to disappear, and C₃ was only partially pronounced. Instead, a vowel before an alveolar sound (-n, -d, -s, -l) was transformed into a front (long) vowel.

V > \ddot{V} (\ddot{V}) / _____ n, d, s, l

<i>bdun</i>	>	t \ddot{u} n	‘seven’	<i>skad</i>	>	k \ddot{a}	‘language’
<i>chos</i>	>	ch \ddot{o}	‘dharma’	<i>bal</i>	>	ph \ddot{a}	‘wool’

As a result, in the Central Tibetan dialect, the vowels ä [ɛ], ö [ø], and ü [y] were added to *i*, *e*, *a*, *o*, and *u*, resulting in eight vowels. The disappearance of the final consonants C₃ and C₄ further resulted in a tonal difference in terms of flat/falling in addition to high/low. This is a phonetic change characteristic of Central Tibet, especially the Lhasa dialect. An analysis of the Lhasa dialect suggests that the number of vowels is 12 and that they are in

vowel harmony (Chang and Chang 1968). This is the most significant example of a change in the forms of Old Tibetan.

In dialects where the original syllable structure was relatively well preserved, the number of vowels did not greatly change. In the Amdo dialect of the northeast, *i* and *e* fused with the central vowel [ɤ]. As a result of a chain shift in which *e* and *o* moved to the place where *i* and *u* were originally articulated, a system of six vowels [i, e, a, o, u, ɤ] now exists. The western Balti dialect retains the original five-vowel system.

5.2 Changes in Word Form and Vocabulary

Although the pronunciation of words changed greatly as described in the previous section, the spellings of words essentially maintained their original written form. Changes in the vocabulary may be attributed to an increase in loanwords. Tibetan originally tended to favor translation over loanwords. Although a huge amount of translations of Buddhist scriptures was carried out after the Old Tibetan era, many Buddhist terms were replaced by translated words comprised of combinations of Tibetan vocabulary, with limited use of loanwords from pure Sanskrit. This is remarkable in comparison with the percentage of words borrowed from Pali in Thai and Burmese, for example.

With the establishment of what is referred to as the “Lama-Patron” relationship between Tibet and Mongolia in the 13th century, many Buddhist terms flowed from Tibetan, which was predominantly religious in nature, into Mongolian. In contrast, words which were borrowed from Mongolian, which was predominantly secular, into Tibetan included titles such as *tar han* (Mongol: *darqan*), *hu thog thu* (Mongol: *xutuɣtu*), and *no mon han* (Mongol: *nom-un-qan*), as well as secular vocabulary such as *thob-ci* ‘button.’ The word for ‘doctor’ is an interesting case. Despite the existence of the word *smān pa* in traditional Tibetan vocabulary, it coexisted with the Mongolian loanword *am ci*, and the latter came to be used as an honorific. This reflects the relationship between Tibet and Mongolia at that time (Norbu and Takeuchi 1991: 383–386). Thereafter, since Tibet was under the Chinese control, loanwords from Chinese and English increased.

Apart from words with the above lexical content, words with grammatical functions were also newly formulated, and the structure of sentences changed.

5.3 Syntactic Change and Formation of the Modern Written Tibetan

I have discussed the introduction of an auxiliary verb *yin* and the development of the predicate shape verb + *pa yin*. The tendency of change in the predicate from its integral structure to an analytic structure was further accelerated by the introduction of a new auxiliary verb *red*. We can demonstrate the following four schematic stages of change in predicates:

- Stage 1** Early Old Tibetan (colloquial/literary)
 -o copula + sentence final particle

Stage 2 Middle Old Tibetan (colloquial/literary)

-o copula + sentence final particle

yin-o the speaker's intent**Stage 3** Late Old Tibetan (colloquial/literary)

-o copula + out of the speaker's intent

yin copula + within the speaker's intent

After this stage, *red*, being an innovative change in the Amdo colloquial dialect, appeared and spread to Central Tibet.

Stage 4 Modern Central Tibetan (colloquial/literary)*red* copula + out of the speaker's intent*yin* copula + within the speaker's intent (Hoshi 2021)

A new auxiliary verb *red* spread from the Amdo dialect to Central Tibetan, but it was not so popular among the western and southern Tibetans. Because it became common in the Lhasa Tibetan, it influenced the modern literary Tibetan.

That new auxiliary verb was applied to the predicate shape, verb root + *pa yin*, resulting in a new formula, verb root + *pa red*.

The particle *pa* nominalizes the perfective root of the verb, meaning past/perfect as a whole, while the genitive particle *kyi* nominalizes the imperfective root of the verb, resulting in new formulae, verb root + *kyi yin* and verb root + *kyi red*. The framework of the analytic predicate forms representing tense and aspect was established, being adopted to the modern literary Tibetan.

Present/Future	Past	
V (imperfective) + <i>kyi yin</i>	V (perfective) + <i>pa yin</i>	within the speaker's intent
V (imperfective) + <i>kyi red</i>	V (perfective) + <i>pa red</i>	out of the speaker's intent

After the 15th century, some forms originating from Amdo were imported into Central Tibetan as well as the literary Tibetan, for instance, the interrogative particle *ʼe* (Hoshi 2012, 2021). The distinction between *ʼa chung* (small *a*) and *ʼa chen* (big *a*) originated in Amdo. *ʼɿ* and *ʼɿʰ* primarily represented different sounds, [ɦa] and [ʔa], respectively, and were distinguished in the Modern Central dialect, but the initial consonants became zero in the Amdo dialect. In order to avoid cohesion, Amdo coined the new terms.

6. Changes in Characters and Script

Tibetan characters, with their long tradition, come in a variety of scripts and typefaces. In Tibet, traditionally, they are roughly divided into *uchen* ‘headed’ (*dbu can*) and *ume* ‘headless’ (*dbu med*) characters. The *u* ‘head’ (*dbu*) refers to the top line of the characters, that is, the horizontal line at the head. This is characteristic of Indic scripts, in which it has seen full expression in the Nagari script, which connects words with horizontal bars. Headed characters are characters in a square style that start with a horizontal line at the head and are written with clear brushstrokes. Documents from the first half of the 7th century, when the Tibetan script was founded, no longer exist. However, manuscripts transcribing works from the earlier period and steles from the 8th century are all written in headed characters. In addition to inscriptions and Buddhist texts, official documents such as notifications and legal documents were also written in a neat, square script (i.e., headed characters).

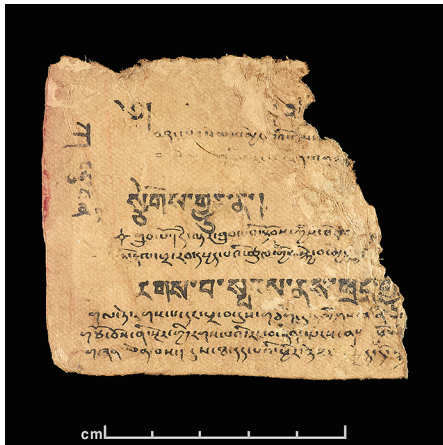
However, some private documents, such as personal correspondence, and paper documents, such as marginalia, have slightly broken from the standard style and written characters as a single stroke. In documents in the post-Imperial period as well, distinctive scripts that could be called ‘semi-square scripts’ became widely used in Central Asia. This is what I refer to as the ‘Post-Imperial Style’.

In the Buddhist scriptures of the late 9th and 10th centuries, the main text is in a square script, but observations and notes written in the spaces between lines and in margins are written in scribbled characters without heads (headless characters). This tendency is notable in the subsequent 12th to 15th century Buddhist texts excavated in Khara-Khoto. The *ume* ‘headless’ (*dbu med*) characters were likely established at this time (see Figure 1).

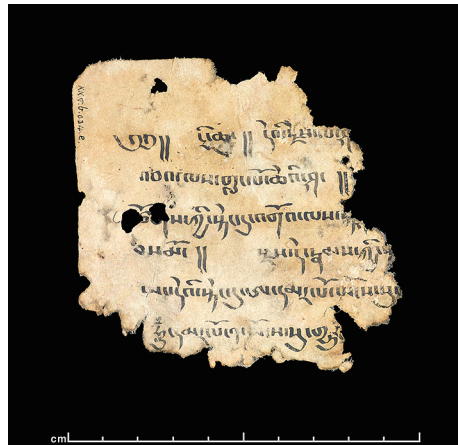
Headless characters gradually evolved from their initial scribbles and became stylized, which also developed into calligraphy. As a result, various script styles were developed, but the typical scripts were the *peik* ‘semi-cursive script’ (*dpe yig*) and the *chuyig* ‘cursive script’ (*khyug yig*) (see Figure 2). The semi-cursive script is used in handwriting and woodcut printing and has a wide range of uses, including administrative documents, edicts, and almanacs. More decorative scripts include variations such as the vertical *bamik* (*bam yig*) and the somewhat square-shaped *drutza* (*bru tsha*). The cursive script is a flowing and elegant script that is often used in letters. It is interesting to note that the various scripts that were created were at least as many as Japanese *kana*.

Although the semi-cursive and cursive scripts were used for woodblock prints, post-woodcut modern printing extensively used the square style. The oldest print-like text in the Tibetan script was a *dharani* incantation called *Fashen sheli ji* (法身舍利偈) excavated from Central Asia, which was made with a stamp pressed down on the paper. Printing began with the appearance of woodblock prints in the 15th century, and sutra printing temples (sutra printers) were established in various places between the 17th and 18th centuries. The complete Buddhist canon was printed over time, and, as a result, print-

Fragments of Buddhist text from Khara-Khoto



K.K.V.b.033.j (IOL Tib M279)
Main text is written in *dbu can* 'headed characters', but notations are written in *dbu med* 'headless characters'.
©British Library



K.K.V.b.034.e (IOL Tib M824)
A fragment of Tantra. Main text is written in *dbu med* 'headless characters'.
©British Library

Figure 1 Style Change in the Tibetan Script (*dbu can* to *dbu med*)

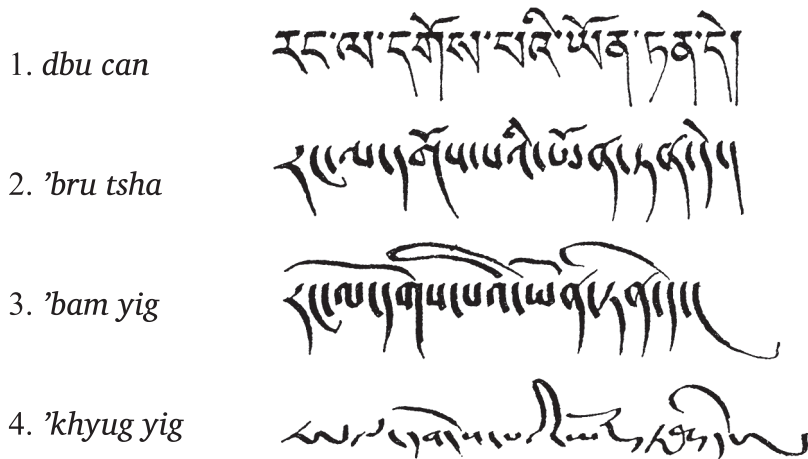


Figure 2 Variety of styles in the Tibetan Script

ing flourished (Onoda 2021). However, with the introduction of letterpress printing in the mid-20th century, modern books began to be printed, and this further transitioned into computer editing and printing.

In the 1980s, the current author, working with Yoshiro Imaeda and Steve Hartwell, a former Apple programmer, developed Apple's Tibetan language system. The system was first developed with the assistance of the government of Bhutan and with the support of Otani University, and it was approved by Apple. Furthermore, in Windows, Tibetan characters and keyboards have been available since Windows Vista, and the fonts have also become part of Unicode. As a result, Tibetan language and characters can be used online, creating a free-use environment for Apple, Windows, and smartphones. It is now possible to have online chats and launch websites in Tibetan in the same way as English and Japanese. Thus, a favorable environment is provided for Tibetan children to continue using the Tibetan language. Children of nomadic peoples are already enjoying surfing the Internet in Tibetan, using their iPhones while riding horseback.

7. Pre-Tibetan and the Historical Classification of Tibetan

I divided the history of the Tibetan language from the 7th century to the present day as follows:

Literary

Old Literary Tibetan	7th to 11th centuries
Classical Tibetan	12th to 17th centuries
Modern Literary Tibetan	18th to 21st centuries

Colloquial

Old Colloquial Tibetan
Middle Colloquial Tibetan
Modern Colloquial Tibetan

The above classification is only a guide. In particular, changes in the spoken language are difficult to classify because of the significant dialectical differences and lack of written records. Thus, it is not possible to indicate the century.

Additionally, a spoken language would naturally have existed even before Old Tibetan. However, since there are no written records of this, its form has been reconstructed in comparison with various Himalayan languages (Monpa, Bumthang, Tamang, etc.), which are thought to have diverged before the formation of Old Tibetan, its origins being traced back further to the Tibeto-Burman languages and ultimately the Sino-Tibetan languages (Ikeda 2021).

However, Tibetan dialects share a number of characteristics, from Old Tibetan to the modern dialects, which give the Tibetan language an overall distinct character compared to the other Himalayan languages. As I have stated above, I believe that the reason for this is that, with the rapid expansion of the state after the establishment of Old Tibetan, the Tibetan language spread equally rapidly, and this gave rise to the emergence of dialects, in other words, a “Big Bang” hypothesis. This paper describes the history of Tibetan based on this hypothesis.

8. Conclusion — A Summary of the Historical Changes in Tibetan

The transition from Old Colloquial Tibetan to the modern dialects follows a path from “synthetic to analytic.” The complexity of the syllable structure was simplified, and tones developed instead. Verbs for which the distinction of the temporal phase by word form became unclear as a result of this simplification were converted into noun phrases with particles (*pa* ‘thing’; *kyi* ‘of’), and subsequently an analytic syntactic form with auxiliary verbs (*yin* and *red*) was developed. Nouns (and noun phrases) that precede predicates are consistently structured with case markers in the form of final consonants or particles (e.g., *nga* → *ngar* ‘to me’; *khong* → *khong la* ‘to him’). However, since in this example the final consonant is no longer pronounced, there is an ongoing shift toward analytic expressions (e.g., *ngar* → *nga la* ‘to me’). These changes most radically took place in Central Tibet, the political and cultural center of Tibet, especially in the Lhasa dialect.

The literary language was re-established as Classical Tibetan, incorporating the changes in the spoken language that had occurred in Central Tibet after Old Literary Tibetan had been used for five centuries. Thereafter, the original spellings were maintained despite changes in the spoken language. As a result, there is a large gap between spelling and pronunciation, which learners of Tibetan continue to struggle with till now. However, considering its connection with a long and rich literary tradition and how it facilitates communication between dialects, unifying the spelling into a standard spoken language, for example, based on the Lhasa dialect, would be a profound upheaval. In the case of Dzongkha, the national language of Bhutan, there has been an ongoing struggle to reformulate the spelling to match the spoken language in order to give it the identity of a national language.

We do not know how Tibetan will change in the future, but with the production of modern literary works and films and the use of the Tibetan language and characters on the Internet and smartphones, it is highly likely that both the spoken and written languages will live on as a language supporting Tibetans and the Tibetan culture.

As mentioned in Section 3, the Tibetan language has long gained universality by permeating into and being continually used by non-Tibetan peoples. Today, devotees of Tibetan Buddhism are spreading throughout Europe and the United States, and they, together with

younger generations of researchers of the Tibetan language and culture, have become freely able to use Tibetan. Here lies the potential for the dawn of a new era in Tibetan language and culture.

References

[English]

Benedict, Paul

1972 *Sino-Tibetan: A Conspectus*. Cambridge University Press.

Chang Kun and Betty Sheft Chang

1968 Vowel Harmony in Spoken Tibetan. *Bulletin of the Institute of History and Philology* 40: 53–132. Academia Sinica.

Hoshi, Izumi

2012 The Flow of Eastern Tibetan Colloquial [^]e into Middle Tibetan. In Tsuguhito Takeuchi and Norihiko Hayashi (eds.), *Historical Development of the Tibetan Languages*. (Journal of Research Institute 49): 71–83. Kobe City University of Foreign Studies.

Li, Fang-kuei

1979 The Chinese Transcription of Tibetan Consonant Clusters. *Bulletin of the Institute of History and Philology* 50 (2) Academia Sinica: 231–240.

Norbu, Thubten J. and Tsuguhito Takeuchi

1991 Mongolian Loan-words in Tibetan and their Socio-cultural Implications. In E. Steinkellner (ed.) *Tibetan History and Language. Arbeitskreis für Tibetische und Biddhistische Studien*. Universtät Wien: 383–386.

Scherrer-Schaub, Cristina A. and Ernst Steinkellner (eds.)

1999 *Tabo Studies II*. Roma. (Serie Orientale Roma LXXXVII).

Takeuchi, Tsuguhito

1995 *Old Tibetan Contracts from Central Asia*. Tokyo: Daizō-syuppan.

1997–1998 (OTM) *Old Tibetan Manuscripts from East Turkestan in the Stein Collection of the British Library*. Volume 1 (Plates), Volume 2 (Descriptive Catalogue), Supplementary Volume (Syllabic Index) Joint Publication of the Centre for East Asian Cultural Studies for UNESCO and The British Library.

2012 Formation and Transformation of Old Tibetan. In Tsuguhito Takeuchi and Norihiko Hayashi (eds.) *Historical Development of the Tibetan Languages*. Journal of Research Institute vol. 49. Kobe City University of Foreign Studies: 3–17.

2013 *Glegs-tshas: Writing Boards of Chinese Scribes in Tibetan-ruled Dunhuang*. In Brandon Dotson, Kazushi Iwao and Tsuguhito Takeuchi (eds.) *Scribes, Texts and Rituals in Early Tibet and Dunhuang*. Wiesbaden: Dr. Ludwig Reichert Verlag: 101–110.

Takeuchi, Tsuguhito and Maho Iuchi

2016 *Tibetan Texts from Khara-khoto in the Stein Collection of the British Library*. *Studia Tibetica* 48. Toyo Bunko.

Uray, Géza

1981 L'emploi du tibétain dans les chancelleries des Etas du Kan-sou et de Khotan postérieurs à la domination tibétaine, *Journal Asiatique* 269: 81–90.

[Japanese]

星泉 (Hoshi, Izumi)

- 2021 チベット語の文法を動的に見る. 岩尾一史, 池田巧 (編) 『チベットの歴史と社会』下: 161–195. 京都: 臨川書店.

池田巧 (Ikeda, Takumi)

- 2021 類型論と系統論からみたチベット語. 岩尾一史, 池田巧 (編) 『チベットの歴史と社会』下: 227–261. 京都: 臨川書店.

加納和雄 (Kano, Kazuo)

- 2021 チベット仏教史の再構築に向けて—新出資料カダム派全集—. 岩尾一史, 池田巧 (編) 『チベットの歴史と社会』上: 208–232. 京都: 臨川書店.

松川節・三宅伸一郎 (Matsukawa, Takashi / Miyake, Shin'ichiro)

- 2015 『華夷譯語 (西番譯語四種 猻羅譯語一種)』京都: 松香堂書店.

西田龍雄 (Nishida, Tatuso)

- 1970 『西番館譯語の研究: チベット言語學序説』京都: 松香堂書店.

小野田俊蔵 (Onoda, Shunzou)

- 2021 チベット語仏典の諸相. 岩尾一史, 池田巧 (編) 『チベットの歴史と社会』上: 261–275. 京都: 臨川書店.

鈴木博之 (Suzuki, Hiroyuki)

- 2009 《西番譯語》〈川七〉18世紀チベット語打箭爐方言の性格について. 『京都大学言語学研究』28: 33–63.

高田時雄 (Takata, Tokio)

- 1988 『敦煌資料による中国語史の研究』東京: 創文社.

武内紹人 (Takeuchi, Tsuguhito)

- 2002 帰義軍期から西夏時代のチベット語文書とチベット語使用. 『東方學』104: 106–124.

- 2021 チベット語の歴史. 岩尾一史, 池田巧 (編) 『チベットの歴史と社会』下: 133–156. 京都: 臨川書店.

[Chinese]

彭金章、王建軍、敦煌研究院 (Peng, Jinzhang, Wang, Jianjun and Dunhuang Yanjiuyuan)

- 2004 《敦煌莫高窟北區石窟》第3卷 北京: 文物出版社.

西藏大学藏文古籍研究所 (編) Xizang Daxue Zangwen Guji Yanjiusuo (eds.)

- 2018 《普日文献》西藏藏文古籍出版社.