

TIBETAN and BHUTANESE MARKS and SIGNS For INCLUSION INTO TIBETAN UNICODE 3.0

By Tony Duff, 19 August, 2000

The following presents information regarding Tibetan/Bhutanese symbols and punctuation marks that are needed for accurate representation of Tibetan/Bhutanese but which are not included in Unicode 3.0 for Tibetan.

The information is presented three parts:

- 1) Astrological signs
- 2) Punctuation Marks
- 3) Other marks

1. ASTROLOGICAL SIGNS

The most commonly used Tibetan astrological signs are represented already in the Unicode 3 definition for Tibetan. However, one group, the pebble signs, has been incompletely represented.

1. The Pebble (rdel རེལ་) group

In order to determine which pebble signs need to be encoded, I interviewed Bhutanese astrologers and their top calligrapher in Thimpu, Bhutan. Additionally, I interviewed an old Tibetan astrologer here in Kathmandu, a man who is reputed to know the system very well and who was trained in Tibet. I also spoke with a number of other Tibetan and Bhutanese scholars. In conjunction with these interviews I was shown a number of popular astrological almanacs which had (many) clear examples of the pebble group of signs.

Definition: “rdel” means “small object”; in Astrology case it refers to written signs which are the equivalent of small pebbles. In Tibetan astrology, the pebbles are used as value indicators. There are two sorts: “white pebbles” and “black pebbles” (“rdel.dkar” and “rdel.nag” respectively. White pebbles indicate a degree of goodness or a degree of harmony and black pebbles indicate the opposite. A white pebble is written as a small, round circle and a black pebble is written as a small black “x” of size cognate to the white circle. A combination of white and black pebbles can be used to indicate various levels of good and bad as follows:

○ ○ ○ best, total harmony	○ × neutral	× × × worst, total disharmony
○ ○ very good		× × very bad
○ good		× bad

Note that:

- a) the neutral sign is only ever written as ○ × never as × ○ .
- b) the best and worst cases are not written linearly but as triangles. They are ○ ○ three white pebbles (“rdel.nag.dkar”) and × × three black pebbles (“rdel.nag.gsum”) respectively:

With one exception, everyone interviewed agreed on the usage and principles of the pebble signs. The one disagreement came from the old Tibetan astrologer who said that the “three black pebbles” never appears. However, on further questioning, and this was verified by discussion with others, it is not that the combination doesn't exist, but that the “three black pebbles” is so bad that as a matter of decency (or whatever you want to call it) no astrologer would write them down for someone. In fact, a number of the almanacs that I looked at do have the three black signs contained in them. Therefore it does have to be included in any system that seeks to represent the pebble signs

All of the above research is completely in agreement with a very nicely done paper on the subject by Mr. Peter Lofting. (The paper was made available previously on Tibex. I have put it on my site at <http://www.tibet.dk/pktc/onlinepubs>). In accordance with what he says and the examples that he gives, the combinations of pebbles given above are the complete set of combinations required to represent a whole set of basic astrological calculations.

However, there is an additional set of astrological calculations that can be done and which are not mentioned in Peter's paper. These calculations require additional combinations of the white and black pebbles not used in the basic calculations. The additional calculations are called upper and lower doors “sgo.steng.’og”. The results of these calculations are always represented by a triangular group of white and black pebbles, the top pebble of the triangle being the value of the upper door and the bottom two pebbles being the value of the bottom door. In these results, all permutations of white and black pebbles are possible, do appear, and were seen in the almanacs that I was shown (at least in the ones that had these calculations in them—not all almanacs have these additional calculations).

To represent all of the required pebble combinations that are the outcome of standard, Tibetan astrological calculations, the following glyphs are required:

○ × ○○ ○× ×× ○
○ ×
○
○× ×○ ×× ×
×○ ×
○× ×○

Only those glyphs in the top line have been included in the Tibetan 3.0 encoding. To make the set complete, the glyphs in the bottom line must be added.

2. PUNCTUATION AND OTHER MARKS.

The Tibetan (and henceforth unless otherwise stated, Bhutanese, too) writing systems following their Indian heritage do not have a large set of punctuation and related marks. Nearly all of the basic marks of the set have been included in version 3 of Tibetan Unicode. However, a couple of very common signs and a number of less common but important signs are missing and need to be included.

a) Special Bhutanese marks for Dzongkha

In the work leading up to the proposal for version 3.0 of Tibetan Unicode, the group as a whole agreed that the needs of Bhutanese writing (which are almost identical to that of Tibetan) should be supported in the Tibetan Unicode definition.

Dzongkha is the name of the official Bhutanese language. In it, there is a set of four signs that are unique to Bhutanese lettering, which are used as the way of starting nearly every piece of correspondence, and hence which have to be included in the the glyphs available for representing Dzongkha.

The four signs are all variations of the common mark sbrul.shad [0F08]. The sbrul.shad is used to mark the beginning of text in general. The four signs are used to open a letter / text but like a sbru.shad but each has a specific function. The four signs are:



From left to right the names and meanings of the Bhutanese signs are:

i) zhu.yig.gi.mgo.rgyan. The “starting flourish for a letter to someone higher” is for writing letters to someone else who is a superior. These days it is used for writing letters to the king.

ii) bka'.shog.gi.mgo.rgyan. The “starting flourish for giving a command (i.e., for talking to someone lower than oneself) is for writing letters to someone who is lower than oneself. A guru, king, high personage would use this when writing this to someone lower than himself.

iii) mnyam.yig.gi.mgo.rgyan. The “starting flourish for a letter to an equal”. The mark indicates that the letter is being written to someone who is equal but moreover who is cared for.

iv) tho.yig or 'ba'.gyan.rgya. This sign is a variant of number three. It is used for personal use only, such as when writing a diary or personal notes. The name means “list” or “a mark just for personal use”.

It is unlikely that one would ever bother with character #4 in the case of using a computer and it is noteworthy that the Bhutanese government insisted that characters #1 to #3 had to be available in the Dzongkha! software that I made for them. They showed me this fourth sign but specifically did not want it included which I take as their statement that in their minds it would never be used in the context of a computer. Should it be included in Unicode or not? Probably. However, the point is to present the information for you to decide.

Two of the four signs needed are already encoded at 0F09 and 0F0A. This leaves these two to be encoded:



b) Variant forms of Shad

The standard punctuation marks of Tibetan include three forms of shad which have the function of dividing text. There is the standard shad [0F0D] and its specialized variant, the rin.chen.spungs.shad [0F11] and there is the sbrul.shad [0F08] which performs a different function from the standard shad and its variant. In that order they look like this (left to right):



In addition to these “official” forms, there are a number of variant forms. They are not “official” and some could only be regarded as sloppy writing (not only my opinion, but the opinion of calligraphy experts here and scholars in general.) Because the others are not “official” I have resisted for fifteen years the attempts of several Tibetans to get me to include variant forms of these shads in my software encoding. However, the last several years of experience with Tibetan text publishing have shown me that at least one of these variant forms *has* to be included in any encoding. And there are a few others which I have chosen to include because of a) popular demand amongst Tibetans, including learned ones, and b) that some texts cannot be reproduced without these alternative forms unless you want to take the liberty of altering what is in the original. Here are the items that need to be included (i) or are worthy of inclusion (ii and on from that)

i) The hooked shad with single tsheg

There are many texts of the Drukpa Kagyu tradition which use this variant form of shad in a particular way and with particular meaning. This is not merely a variant visual form of a shad but the shad with a new function. It is well known inside the texts of this tradition where it has been a feature since the time of the great fourth Drukchen, Padma Karpo (15th century). Chris Fynn reports that it appears in other Kagyu texts, e.g., in the *do.ha.skor.gsum.gyi.Ti.ka.'bring po. sems.kyi.rnam.rthar.ston.pa'i.me.long* of the great scholar of the Karma Kagyu, Karma 'phrin-las-pa. The shad has no special name. It looks like this:



Sometimes the shad and the single tsheg on top tsheg are joined together but the distinctive hook always remains. It is used where a shad is needed grammatically but where there is direct continuation to the next text. E.g., in *sadhanas* the very common construction



followed by text or recitation might require the shad grammatically but the shad breaks the continuity of recitation, therefore this device



is used to reduce the strength of the break introduced by the shad and to give a sense of onward

flow. Its function is different. My advisors insist that it is not the same as a standard shad. I think it does need to be included.

ii) Other variant forms of shad

The following are variant forms of shad that I have felt compelled to encode even though my purist streak for many years prevented me. As I said earlier, encoding them accommodates some people who insist that the “official” forms of shad are not what they want to use and also accommodates the situation where someone wants to preserve a text just as it was written, unusual (but acceptable) punctuation and all.



If you have some reservation about including any of these, please give consideration to this: the glyph 0F10  that has already been encoded appears with a frequency considerably lower than any of the above variants! 0F10 is in fact a quite rare variant of the second from left shad shown above.

The four shad to the left are variations usually of a rin.chen.spungs.shad. However, the leftmost is sometimes seen as a variation of the hooked shad mentioned above. The centre shad and fourth from the left are also occasionally used as alternatives to a sbrul.shad though this is less common. The rightmost shad is a perfectly acceptable alternative to a sbrul.shad which is frequently seen (the difference is that a standars sbrul shad has two squiggles in the intervening space rather than the one shown). All of the above are seen with moderate to high frequency in texts when you take a proper cross-section.

Of these variant shad’s, the hooked shad with single tsheg has to be encoded if the texts of the Drukpa Kagyu tradition are to be rendered correctly. The remaining shads, whilst perhaps harder to justify from a certain perspective, are **required** by some Tibetans and Bhutanese—including learned ones—and it seems to me, after a lot of consideration, that there is no choice but to encode these. Certainly, they should not be just brushed off. Inclusion of them will complete the representation of the Tibetan punctuation set fully.

c) Treasure revealers treasure tags (“gter.ston.gter.btags”)

In the Tibetan Buddhist system there is a thing called Concealed treasure (gter.ma). Those who withdraw treasures from concealment are called Treasure revealers (gter.ston). There are various

kinds of treasure, including written texts of dharma. The written texts are written in a special code called “Dakini Script” or “Code script” (dak.ki.yi.ge or brda.yig). Only the specific revealer of the texts can decipher that script. The revealer has the job of revealing the treasure and, if it is a text, of deciphering it. Such texts have a yig.mgo as has been written up in the block introduction to version 3 of Tibetan Unicode. However, whereas a normal yig.mgo looks like this:



The yig.mgo in a treasure has the shad’s replaced by a terma break ཅོ་མོ་ (“gter.tsheg”) and has the zla.tshe.gnyis replaced by a sign that is specific to the treasure itself. Compare the above with a yig.mgo from a treasure text of Ratnalingpa:



Since only the treasure revealer can read the treasure, it is then said that the sign specific to the treasure is the treasure revealer’s treasure tags (gter.ston.gter.btags). However, in spite of the name, it is actually the sign that was in the treasure itself and not the sign of the treasure revealer per se. If they were the latter it might be argued that they should not be included in the Unicode definition, but they are not, they are an integral feature of the texts themselves. They are punctuation marks because they are direct replacements for the zla.tshe.gnyis (0F05+0F04) punctuation marks and are not used for any other purpose (unlike someone’s logo or personal seal).

There are many different treasure revealer’s marks. I have seen a book containing reprints of the pages used at the famous Derge printery containing the samples of treasure revealer’s marks to be used by wood-block carvers to do their work. Chris Fynn has also seen this and has a copy of it. Here are the treasure revealer’s marks that I have had to encode for our work and the work of a few other people here:



Ratna lingpa



Terton Mingyur dorje



Terton Dongag lingpa (Dilgo Khyentse)

For Unicode to be able to publish any and every Tibetan text, all of these treasure revealer's marks must be available. Great chunks of Nyingma Literature and Kagyu literature, too, depend on these marks. This brings an unsavoury point but it must be mentioned. Tibetan Buddhist history contains lots of acrimony. One major point is that influential people in the Gelugpa tradition have from time to time decried the Nyingma tradition, and especially its system of treasures and treasure revealers as bunk. This position has also been taken up by some Western scholars. However, regardless of the status of this tradition of treasures and treasure revealers, there is a huge corpus of treasure texts which a very large portion of the Tibetan population (and other cultures now, too) regard as authentic. These people want to have these texts. If the basic letters which are the special features of these texts—the treasure revealers' marks—are not available then the texts will not be reproducible, electronically preservable, etc., etc. and that would mean that a very significant portion of Tibetan literature would not be reproducible, etc., etc.

3. OTHER MARKS

a) Music: The peaceful cymbal cadence “zhi.rol”.

There is a system for writing music in Tibetan just as there is in European countries. Most of this system (as I understand it) involves the use of flourishes of the pen which I do not think can be converted into a computerized script. I leave this for others to decide upon. However, there are a several pre-defined marks in common used in classical texts (sadhanas and liturgies). The comment from our advisors has always been that these signs *have* to be made available for the production of the texts.

Some of the marks are already included in Tibetan Unicode 3.0. For example, the most common mark used, the “cantillation beat” [0FC0] and [0FC1] is already included. (I think the information provided about the use of this mark in the Unicode def. is that it goes over the text. However, it also goes down and between sections of text as shown in the following example.) The other common mark in use but not included in Tibetan 3.0 is the “peaceful cadence with cymbals”, “zhi.rol”:



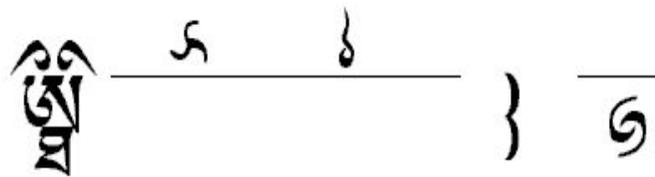
Zhi means “peaceful cadence” and rol means music made with with cymbals. The sign is written above the text where the music is to be made. It can also be written in between two sections of text like the cantillation beats shown to indicate music at that point. For example, here is a single cantillation beat between shad’s, followed by a double cantillation beat between shad’s followed by a peaceful cadence over a word:

འཕོ་མེད་ཡེ་ཤེས་སྐྱབ་བཞེས་བདེ་བར་གསལ་གསལ། །བསྐྱུ་མེད་
མཚོ་གཤམ་གསུམ་རབ་འབྱུམ་ས་རྒྱལ་བ་དང་། ° །འཚི་མེད་ཚོ་ཡི་
དཔལ་སྟེར་མཚོ་གཤམ་རྣམས་གྱིས། °° །འཁྱོར་མེད་རྩི་རྩི་གྱིན་

The peaceful cymbal cadence needs to be included. The information here should be sufficient to prove the need for its inclusion and to write a description of it in a later version of the block introduction.

b) Special tantric marks

There are a number of special marks and letters needed for representation of tantric materials beyond the basic Tibetan character set. Some of these have been included in the Tibetan Unicode definition. The following have not. (The line is an artificial baseline to give the correct sense of positioning).



i) The Secret Sign “ithi” །འཁྱོར་

This sign is a Tibetan corruption of the transliteration of a sign used in Sanskrit, Buddhist tantric

texts. Its name is simply the pronunciation of the sign. It placed at the end of a text to seal the text with an indication that the text has a “transmission secrecy setting”. It shows that the text may be given only to a specified number of students (usually one but not always) for a certain number of generations, often 13. There is a very significant portion of Tibetan tantric literature which needs this sign.

ii) The Bliss swirl “dga’.dkyil”

This sign is not found often, though in certain cycles of literature its use is widespread and its meaning is crucial. For example, the *New treasures of Choling* “Chos.gling.gter.sar” constitutes about forty volumes of texts and a significant number of the them rely on its use for transmitting a certain meaning. It is always used centred over another letter; e.g., like this with the half-a:



iii) I cannot find a name for the third for the third mark from the left. It is simply called a “mark” “rtags”. It is used as an ornament for letters and punctuation marks in some tantric texts. It is definitely not the editor’s mark referred to below in the section on editor’s marks. It is smaller, different in shape (look carefully!), and very different in use and meaning. The editor’s mark could not be used as a replacement for it or vice-versa. We have found it several times in our texts at Drukpa Kagyu Heritage Project and have verified that it is a discrete and needed mark with our scholars. It is frequently used over an opening yig.mgo like this:



iv) The Kaya mark

The fourth sign from the left appears in a variety of treasures and is crucial to the full representation of the meaning contained in the treasure. It is for example, found through the entire collection of gter.ston mi.’gyur rdo.rje’s treasures (which are very extensive, given that he is now in his eighth incarnation and all incarnations have contributed to the total corpus of his works). It is also found in various other places in Nyingma and Kagyu treasure texts. There is no name for the mark but when placed it is always placed in a group of three and means the three kayas:



v) **The fifth sign** is used both as a section divider, and also used in tantric texts to provide a further level of esoteric meaning. I cannot find an official name for it though some of my scholars think it should be called a *dga'.dkyil* (“bliss wheel” not “bliss swirl”—there is a slight difference). It is found between sections of text marked by *shad* where it performs one of the two functions just mentioned.



It is also found in conjunction with a *sbrul.shad* where it is used to mark the end of a major section of text:



Over the several years of experience that I’ve gained with Tibetan texts at the Drukpa Kagyu Heritage project, I’ve seen all of the above signs, some very frequently, some infrequently but all more than once. I have verified that they all are meaningful and not just merely unusual forms of some other letter, usually with a number of local scholars. I maintain that they all definitely need to be included in the Unicode definition.

c) Annotation marks “*mchan.btags*”

Tibetan writing does not use footnotes or endnotes. However, there is a system for annotating text with notes. Annotations done Tibetan-style are called *mchan* and the mark for them is called an annotation mark “*mchan.btags*”. Annotation marks join the text being annotated to the annotation. They function like the superscripted small numeral in an English text which directs the reader’s attention to the actual footnote or endnote. However, the Tibetan annotation mark always connects the annotation directly to the annotated text. Annotations in Tibetan text, are always made in the smaller of the two font sizes in use in a text and are connected to the text that they annotate with the annotation mark itself. I have enlarged a portion of the above text to show the annotation mark itself more clearly and put it in red:



The annotation mark shown above is a “trailing” annotation mark. It is the most commonly used one of two versions, the other being a “leading” annotation mark. A “trailing” annotation mark is used when the annotation follows the text being annotated. A “leading” annotation mark is used when the annotation precedes the text being annotated. Here is an example of a “leading” annotation mark.

བས་ འཛིན་

Note that there is no distinction of name in Tibetan for the leading and trailing versions of a mchan btags—they are both just annotation marks.

Annotation marks are heavily used in classical Tibet texts. For example, approximately 15% of our texts have annotation marks in them and many of those have them in very heavy used (nearly every phrase is annotated). These two marks are glaring omissions from Tibetan 3.0 Unicode.

d) Editor’s marks “zhu.dag.mkhan.btags.”

Just as there is a whole series of editorial marks in English (caret, paragraph here, put into caps, lower the case, etc., etc.) there is also a series of editorial marks in Tibetan. When they are being use as editor’s marks, I have some doubt about their need for inclusion in computerized character sets for Tibetan. For example, although our editors (I have 12 of them at present working full-time in my office) constantly use these on the paperwork that they are editing and correcting, the marks themselves never need to be included in the computer input much as editorial marks in English affect the computer input but do not appear in the input itself. However, they are sometimes as markings in a text that are more than editorial marks. When used like that they are being used in a way that becomes important to the meaning of the text. Because of that, I think that these signs should be included in the Tibetan Unicode standard.

The four most common marks are as follows:

༠༠ ༠ ༡ ༢

These marks do not have names per se though when pressed some of my older (trained in Tibet) and very learned scholars came up with names which I could and would have come up with myself. In short, I can’t find names for them though anyone can make up a descriptive one that

fits. I think it best that the following just be called Editor's marks.

These marks are all variations on a theme. They are used in a number of ways, including use as an English caret, numbering or marking of divisions or special points, etc. Even though they are variations on a theme, it is important to note that an editor will use more than one of them in a document, just as we might use A. B. C. to mark up something in English. Therefore including only one of them is not sufficient.

Note:

1) These are not variations on the pebble astrological signs shown earlier. The open circles and the x are not the same as the black and white pebbles and I do not think that one should attempt to create these glyphs from them.. If nothing else, the circles and x's used here are slightly smaller than the the circles and x's used for the pebbles. If one were to use them in the same text, it wouldn't look right.

2) The top parts of all four are definitely not the mark used for tantric purposes mentioned earlier. If you look closely, you will see that the tails are different. The tail on the editor's mark is much longer, leans to the right, and does not have a series of wiggles in it. The tantric mark is shorter, does have the series of wiggles (an important part of its meaning) and goes straight up and not to the right. One could not be used as a substitute for the other. Finally, the editors marks and that tantric mark have totally different functions. When you see the two side by side in a real text, there is no possibility of mistaking one for the other.

END OF ITEMS PROPOSED FOR INCLUSION IN UNICODE 3.0

INFORMATION ON OTHER SIGNS AND MARKS

a) Name tags (mtshan.btags)

The following have been included in version 3 of the Tibetan code already.

0F35;TIBETAN MARK NGAS BZUNG NYI ZLA, TIBETAN HONORIFIC UNDER RING

0F37;TIBETAN MARK NGAS BZUNG SGOR RTAGS; TIBETAN UNDER RING

There is a system in place for marking someone's name in text. The name is marked by affixing a small circle below the parts of the name. The circle is centred under each inter tsheg (or slightly to the left of centre where the descender of a letter obstructs the affixing of the mark).

There are two variations of the mark: non-honorific and honorific. The non-honorific is a single circle, the honorific is the same circle but with a crescent moon (zla.tshe) placed below it, as follows:



Why would there be need for such a thing? Well, for example, long-life prayers for people of importance are an integral part of Tibetan culture. The prayers are usually very poetic with lots of metaphor and alliteration. The person's name is usually cleverly woven into the text so that the name is the name but is also part of the text which is a commentary on the meaning of the name. In this case, the name tag shows which words are the actual name of the person being honoured. Usage is not confined solely to long-life prayers. There are many different texts, even in government service for example, in which a person's name needs to be de-marcated from the rest of the text and this method serves that purpose.

Here is an example of a real long-life prayer (from the archives at Drukpa Kagyu Heritage Project) that uses many honorific name tags. I have highlighted two of several to assist with their visual identification. The name tags here show the name of the Lama who is being supplicated to stay.

