JTC1/SC2/WG2 N3247

Date: 2007-04-20

(L2/07 114)

Comments on proposal N3240*.

**(ISO/IEC 10646 JTC1/SC2/WG2:N3240 "Proposal on Adding 3 Tibetan Characters and a symbol for ISO/IEC 10646 in BMP"*

Author: Christopher J. Fynn, (individual expert contributor)
4 Chester Court, 84 Salusbury Road,
London NW6 6PA, United Kingdom
+44 207 328 4453
mailto:cfynn@gmx.net
http://www.btinternet.com/~c.fynn/

Date: 2007.04.20

Document ISO/IEC 10646 JTC1/SC2/WG2:N3240 from the national body of China proposes adding three new characters to the Tibetan block:

TIBETAN ABOVE TRANSFORMED LETTER RAGU
TIBETAN ABOVE TRANSFORMED LETTER LA
and TIBETAN ABOVE LETTER SA

The present document addresses the three issues raised in the justification given in the proposal for encoding these three characters.

1. Forming Common Tibetan Ligatures:

The first reason for requesting the addition of these characters is given as:

> *"...these 3 characters are high frequency character in the Modern Tibetan and usually used as a combining element to form Tibetan composite character, such as* ཀྐ, ཁྐ, དྔ, རྔ, ཏྐ, ཐྐ, ཀྵ, ཁྵ, ཤྭ, ཥྭ, ཧྭ, ཨྭ, ཀྻ, ཁྻ, ཤྰ, ཥྰ, ཧྰ, ཨྰ, ཀྼ, ཁྼ, ཤྱ, ཥྱ, ཧྱ, ཨྱ, ཀྵྭ, ཁྵྭ *etc. These are the basis components of Tibetan composite character."*

In the present document all the composite Tibetan character ligatures in the above quote have been re-typed, using a freely available open source word processor *(OpenOffice.org Writer v2.1)* and the existing UCS Tibetan block characters. This I believe adequately demonstrates that there is no need to encode the proposed characters simply to produce correct glyph forms for these frequently occurring Tibetan combinations. For reading purposes all these combinations are represented correctly using the current set of UCS characters, and they are rendered correctly using modern "smart" font formats & associated glyph shaping and display technologies:

ཀྐ	U+0F62 U+0F90	ཁྐ	U+0F62 U+0F92	དྔ	U+0F62 U+0F94
རྔ	U+0F62 U+0F97	ཏྐ	U+0F62 U+0F99	ཐྐ	U+0F62 U+0F9F

ཀྱ	U+0F62 U+0F051	ཁྱ	U+0F62 U+0F53	གྱ	U+0F63 U+0F90
གྲ	U+0F63 U+0F92	གླ	U+0F63 U+0F94	གྭ	U+0F63 U+0F95
གྷ	U+0F63 U+0F97	ཀྵ	U+0F63 U+0F9F	ཀྱ	U+0F63 U+0FA1
ཁྱ	U+0F63 U+0FA4	ཁྲ	U+0F63 U+0FA6	ཀྵ	U+0F66 U+0F90
ཀྲ	U+0F66 U+0F92	ཀླ	U+0F66 U+0F94	ཀྭ	U+0F66 U+0F9F
ཀྱ	U+0F66 U+0FA1	ཀྷ	U+0F66 U+0FA3	ཁྱ	U+0F66 U+0FA4
ཁྲ	U+0F66 U+0FA6	ཁྷ	U+0F66 U+0FA8		

A concern is that encoding the three proposed characters would result in there being two different ways of representing the combinations illustrated in the table above. The resulting glyph forms would be visually identical even though the underlying characters and character ordering would be different. The proposed encoding would also result in two possible base characters for these combinations, a problem which current normalization processes could overcome.

2. Finding the root letter in words

The second reason given for encoding these characters is:

> *"...if added these 3 characters, the code of ISO/IEC 10646 in BMP could describe the characteristics of Modern Tibetan character. The Modern Tibetan character is combining of prefix, head letter, root letter, subjoined letter, vowel, suffix and post suffix. The existing ISO/IEC 10646 can't rightly describe these component, for example, the fist code of* སྦ *and* སྒྱུ *all are U+0F66, but of* སྦ *is head letter and* ས *of* སྒྱུ *is root letter. If added these 3 characters to BMP, from U+0F40 to U+0F69 are the root letters, then it can rightly describe the Modern Tibetan characteristics."*

Tibetan readers who understand the basic rules of Tibetan grammar have of course always been able to determine the bare root letter *(rkyang pa)* in a word. So if glyphs for the characters are rendered correctly on screen or paper a Tibetan reader will have no difficulty identifying the root letter in a word. If any ambiguity is encountered this root may be determined by reading the word in context. As demonstrated in the previous comment (1.) there are no real issues regarding proper display and ligature formation using the existing UCS Tibetan characters - so a Tibetan reader should have no problem determining the root letter.

Thus the reason for the proposed characters can only be to aid the computer processing of Tibetan. To a computer or software application Tibetan words like དབྱ་, སྦྱ་, སྒྲ་, སྒྲུབ་, *etc.* are simply processed as strings of characters in a sequence - the visual position of the resulting glyphs is inconsequential for all underlying processing except rendering.

Determining the root letter in a word *is* important in processes such as collation since the primary collation in a word is based on its root letter. So the computer application needs to either employ some algorithm for determining the root letter in Tibetan words; have the root letter marked or flagged by some sort of marker character; have the characters in words re-ordered so that the sort keys are in sequential order of precedence; or employ a table defining a suitable set of collation weights and elements.

The proposed new characters would *partially* enable one of the possible ways of indicating the root letter in a word as, being combining characters, these super-fixed prefix characters would then have to occur (and be entered) *following* the character for the root letter in a word with which they would combine.

However this only solves half the stated problem. There is also the equal and near identical problem of characters for other prefix characters occurring before the character for the root letter in a word. To solve this other half of the problem is WG2 at some later date going to be presented with a proposal for ◌ག TIBETAN PREFIX LETTER GA; ◌ད TIBETAN PREFIX LETTER DA; ◌བ TIBETAN PREFIX LETTER BA; ◌མ TIBETAN PREFIX LETTER MA; and ◌འ TIBETAN PREFIX LETTER ACHUNG ? These characters would necessarily also be required if one wanted to "solve" the problem of determining the root character in a word by placing it first in the character string representing that word.

In passing it might also be noted that there is not universal agreement on what the root letter in some Tibetan words is. Some Tibetan grammarians including Ngag dbang bsTan dar have subscribed to the minority view that the root consonant in words with the combinations ཧྲ and ཧྭ is ཨ rather than ར or ས. [*see:* Ngag dbang bstan dar (1759 - ca. 1840) *Yi ge'i bshad mkhas pa'i kha rgyan*, Beijing: Mi rigs dpe skrun khang, 1982]

3. Collation issues

The third, and would seem most important, reason for requesting the additional characters proposed in N3240 is stated as:

> *"... the sort order of Tibetan character base on the root letter, so the rightly find out root letter of every Tibetan glyph, then Tibetan character has sorted is possible based on BMP."*

Admittedly the collation for the languages using the Tibetan script is particularly complex due to the presence of prefix and head letters before the root letter which is the primary sort key. This prevents culturally correct ordering of Tibetan being accomplished using simple sort routines. However collation of UCS Tibetan can be accomplished using other well documented and widely implemented means. Perhaps members of the China national body are unaware that successful implementations of culturally correct Tibetan and Dzongkha collation (both languages written in Tibetan script follow essentially the same set of rules) already exist for UCS encoded Tibetan in widely available software:

1. The Microsoft Corporation's recently released *Windows Vista* operating system and environment implements culturally correct collation of UCS Tibetan. (Cathy Wisnick of Microsoft using information provided by experts from the Dzongkha Computing Project was largely responsible for this implementation.)

2. The free and open source *OpenOffice.org* software suite since version 2.0 has supported correct collation for Dzongkha which also works for Tibetan. *(Specific "Tibetan" collation has not yet been implemented only because no one has as yet contributed Tibetan locale data to that project.)*

   If we take a random list of Tibetan words typed in UCS characters: དཔལ་, བརྒྱད་, བསྟན་, གཏན་, སྤྱི་, སངས་, དའི་, ཀཱརྨ་, ལལ་, སྒྲུབ་, སྒྲུབ་, ལྡན་, བསྟན་, སྒྲུབ་, གཡང་ they can be successfully sorted in *OpenOffice.org* to obtain the correct Tibetan dictionary ordering:

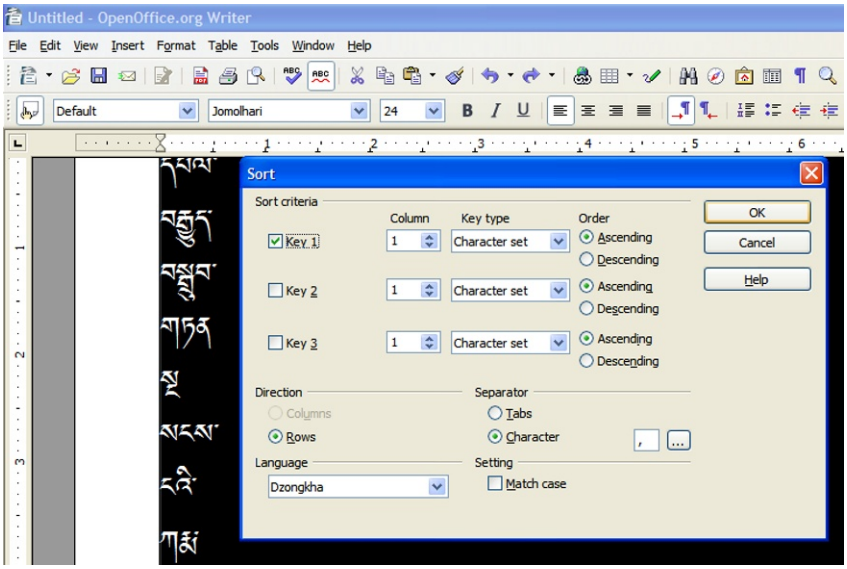


Figure 1

   Resulting ordering: ཀཱརྨ་, སྒྲུབ་, སྒོམ་, སྒྲུབ་, བརྒྱད་, བསྟན་, དའི་, སྤྱི་, གཏན་, བསྟན་, ལྡན་, ལལ་, དཔལ་, གཡང་, སངས་ - which is correct for both Tibetan and Dzongkha.

3. The *Mimer SQL* database engine also successfully implements a multi pass collation for UCS Tibetan using weighted collating elements defined in a table which can be modified if a different collation is desired. See:

   For Tibetan combinations with the root letter letters (KA) and (KHA) this is as follows:

```
'&#0F40#<<#0F88##0F90#'
'       <#0F51##0F40#'
'       <#0F56##0F40#'
'       <#0F62##0F90#'
'       <#0F63##0F90#'
'       <#0F66##0F90#'
'       <#0F56##0F62##0F90#'
'       <#0F56##0F66##0F90#'
'&#0F41#<<#0F88##0F91#'
'       <#0F58##0F41#'
'       <#0F60##0F41#'
```

4. Tibetan collation can also be achieved using routines available as part of the free and open source ICU (International Components for Unicode) library.
   see: <http://icu-project.org/>

5. The Dzongkha Linux project has successfully incorporated Dzongkha collation into the dz_BT locale for glibc. This has now been included in other Linux distributions. Dzongkha Linux is an example of a fully localized operating system, and set of applications using UCS Tibetan script. This collation using ISO 14651 syntax was developed in Bhutan by Mr. Pema Geyleg while he was a student at Sherubtse College. The LC_COLLATE element could be more or less copied straight over into a bo_CN locale for glibc.

Thus it is demonstrable that there is no requirement for the three characters proposed in N3240 order to successfully implement Tibetan collation.

Additional Remarks

The encoding scheme for Tibetan characters in the UCS is based on the natural order of Tibetan characters - this is the order in which Tibetan school children learn to spell out Tibetan words, the order in which Tibetan letters are written on paper and the order in which they are normally typed.

This encoding scheme for Tibetan characters was decided upon only after consideration of all possible alternatives (including that now proposed again in N3240) which included much discussion amongst experts and WG2 members. During these discussions it was realized that each of the possible models had certain strengths and weaknesses but finally a consensus was arrived at that the model, subsequently adopted in the encoding found in ISO/IEC 10646 Amendment 31, was the most suitable. The numerous successful implementations of UCS Tibetan in software applications over the last seven years has confirmed this view.

The current UCS encoding of Tibetan script has already been successfully adopted by all government departments in Bhutan for documents and data in their national language Dzongkha. It has also been adopted by several major Tibetan text input projects.

While the UCS encoding of Tibetan characters does not specifically indicate which is the root letter in Tibetan and Dzongkha words this presents no major difficulties in real world applications including collation. On the other hand, adding the characters proposed in N3240 would have the overall effect of greatly complicating matters with the sole perceived benefit of enabling simplistic sorting in a culturally expected manner (though this would logically require the encoding of an additional four TIBETAN PREFIX LETTER characters as well as the three in the current proposal).

However this would in fact not be the case since characters in the current encoding, which has now existed for as a standard for seven years, would still need to be accommodated - the proposed characters would only introduce multiple ways of representing the same words and a whole new way of ordering Tibetan characters within strings. This would of course mean that existing implementations of UCS Tibetan, several of which I have been closely involved with, would require major modification. Encoding these characters would also introduce many obvious data normalization and security issues which N3240 does not address.

Since there are no real differences in ordering between languages using the Tibetan script a standardized collation for UCS Tibetan script characters following that ordering is plainly desirable. In order to accomplish this, defining a better ordering for Tibetan characters than that currently found in the *Default Unicode Collation Element Table* (DUCET) allkeys.txt and in the ISO/IEC 14651 Standard would be a good start.

- *Christopher J. Fynn*

Appendix 1: Structure of Tibetan "syllables"

The basic unit of meaning or morpheme in Tibetan & Dzongkha is the *tsheg bar* or "name" (*ming)* usually referred to as a "syllable" in English language books on Tibetan. Words consist of one or more these "syllables".

Figure 2

Each syllable contains a root letter *(ming zhi)* which is the primary sort key and may additionally have any/or all of the following parts: prefix, head letter, sub-fixed letter, vowel sign, suffix, and post-suffix. Syllables are normally delimited by a *tsheg* or another punctuation character.

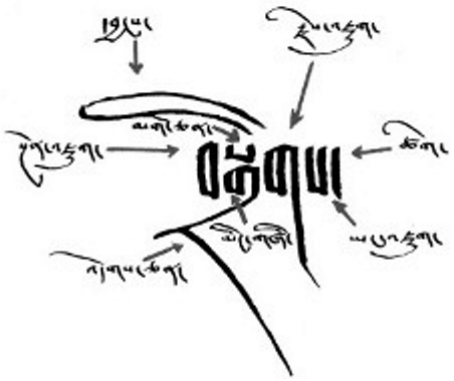


Figure 3

Appendix 2:

Table of standard Tibetan & Dzongkha letter combinations

རྩ་དང་ bare root	འདེགས་ཅན།				མཐོ་ཅན།			འོ་མཐོ་	ས་མཐོ་		
	◌ྭ	◌ྱ	◌ྲ	◌ྭ	ར	ལ	ས	◌ྭ	◌ྭ	◌ྱ	◌ྭ
ཀ	ཀྭ	ཀྱ	ཀྲ	ཀྭ	ཀྠ	ཀླ	ཀྶ	ཀྭ	ཀྭ	ཀྱ	
ཁ	ཁྭ	ཁྱ		ཁྭ							
ག	གྭ	གྱ	གྲ	གྭ	གྡ	གླ	གྷ	གྭ	གྭ		གྲྭ
ང					ངྡ	ངླ	ངྷ				
ཅ						ཅླ					
ཆ											
ཇ					ཇྡ	ཇླ					
ཉ				ཉྭ	ཉྡ		ཉྷ				
ཏ		ཏྱ		ཏྭ	ཏྣ	ཏླ	ཏྷ				
ཐ		ཐྱ									
ད		དྱ		དྭ	ཌྡ		ཌྷ			ཌྷྭ	
ན		ནྱ			ནྡ	ནླ	ནྷ				
པ	པྭ	པྱ									
ཕ	ཕྭ	ཕྱ									ཕྲྭ
བ	བྭ	བྱ	བྲ		བྡ	བླ	བྷ		བྭ	བྷྭ	

རྩ་དཔེ་ bare root	འདོགས་ཅན།				མགོ་ཅན།			རะ་མགོ་	ས་མགོ་		
	◌ྭ	◌ྱ	◌ྲ	◌ྟ	ར	ཡ	ལ	◌ྭ	◌ྭ	◌ྱ	◌ྟ
ཀ	ཀྭ	ཀྱ	ཀྲ	ཀྟ	ཀྐ	ཀྑ	ཀླ	ཀྭ	ཀྭ	ཀྱ	
ཁ	ཁྭ	ཁྱ			ཁྲ		ཁླ	ཁྭ	ཁྭ	ཁྱ	
ག				གཏ	གྲ		གླ				གཏ
ང				ངཀ							
ཅ					ཅྲ						
ཆ											
ཇ				ཇྭ							
ཉ			ཉྲ	ཉྩ							
ཊ											
ཋ											
ཌ			ཌྲ	ཌྜ							
ཎ				ཎྜ							
ཏ		ཏྱ		ཏྟ		ཏྥ					
ཐ		ཐྱ	ཐྲ	ཐྟ							
ད		ད		དྟ		དྷ					
ན											

Note: This table includes only the standard combinations in Tibetan & Dzongkha grammar - i.e. those used in writing native Tibetan & Bhutanese words. It does not include other combinations found in common loan words (e.g. the ཟྭ in ར་ཟླ).

Appendix 3:

Table of Tibetan & Dzongkha prefixes and root letter† combinations

	འཇུག་ཡུལ་གྱི་མིང་གཞི་								
སྔོན་འཇུག་ *prefix*	རྩ་ཙ་ *bare root*	འདོགས་ཅན་			མཚོ་ཅན་			ར་མཚོ་	ས་མཚོ་	
		◌ྱ	◌ྲ	◌ྭ	ར	ལ	ས	◌ྱ	◌ྱ	◌ྲ
ག	ཅ									
	ཉ									
	ཏ									
	ད									
	ན									
	ཙ									
	ཞ									
	ཟ									
	ཡ									
	ཤ									
	ས									

	འཇུག་ཡུལ་གྱི་མིང་གཞི་									
སྔོན་འཇུག་ prefix	རྩ་བ་ bare root	འདོགས་ཅན་			མགོ་ཅན་			ར་མགོ་	ས་མགོ་	
		◌ྭ	◌ྱ	◌ྲ	ར	ལ	ས	◌ྷ	◌ྴ	◌ྵ
ད	ཀ	ཀྭ	ཀྱ							
	ག	གྭ	གྱ							
	ང									
	ཙ	ཙྭ	ཙྱ							
	ཚ	ཚྭ	ཚྱ							
	ཛ	ཛྭ								
ཨ	ཏ	ཏྭ	ཏྱ		ཏྲ		ཏྲས	ཏྷ	ཏྴ	ཏྵ
	ད	དྭ			དྲ		དྲས	དྷ	དྴ	དྵ
	(ན)				ནྲ		ནྲས			
	ཅ									
	(ཉ)				ཉྲ					
	(ཀྵ)				ཀྵྲ		ཀྵྲས			
	ཀ				ཀྲ	ཀྲྀ	ཀྲས			
	ཁ				ཁྲ	ཁྲྀ	ཁྲས			

	འཇུག་ཡུལ་གྱི་མིང་གཞི་								
སྔོན་འཇུག་ *prefix*	རྩ་བ་ *bare root*	འདོགས་ཅན་			མཚོ་ཅན་			ར་མཚོ་	ས་མཚོ་	
		◌ྭ	◌ྱ	◌ྲ	ར	ལ	ས	◌ྭ	◌ྭ	◌ྱ
	(ཀ)				ཀྱ		ཀྱི			
	ཁ				ཁྱ		ཁྱི			
	(ག)				གྱ					
	ང									
	ཅ			ཉྩ						
	(ཆ)			ཉྪ						
	ཇ									
	ཉ		ཉྙ	ཉྲ						
མ	ཏ	ཁྭ	ཁྱ							
	ཐ	གྭ	གྱ							
	ད									
	ན									
	པ									
	ཧ									

	འཇུག་ཡུལ་གྱི་མིང་གཞི་								
སྔོན་འཇུག་ prefix	རྩ་བ་ bare root	འདོགས་ཅན་			མཚོ་ཅན་			ར་མཚོ་	ས་མཚོ་	
		◌ྭ	◌ྱ	◌ྲ	ར	ལ	ས	◌ྭ	◌ྭ	◌ྱ
	ཀ									
	ཁ									
	ག									
	ང									
	ཅ									
ད	ཆ	ཁྭ	ཁྱ							
	ཇ	གྭ	གྱ							
	ཉ									
	ཊ									
	ཋ									
	ཌ		ཉྱ							
	ཎ	ཏྭ	ཏྱ							
	ཏ	ཐྭ	ཐྱ							
	ཐ									

	འཇུག་ཡུལ་གྱི་མིང་གཞི་								
སྔོན་འཇུག་ *prefix*	རྩ་པ་ *bare root*	འདོགས་ཅན་			མཚོ་ཅན་			ར་མཚོ་	ས་མཚོ་	
		◌ྒ	◌ྔ	◌ྦྷ	ར	ལ	ས	◌ྒ	◌ྒ	◌ྔ
	ཀ									

† མིང་གཞི *ming gzhi* - name-base. This is the letter which forms the basis of a name (T. *ming*) or morpheme. It is one of the thirty consonants *either alone (kyang pa) or with modifications*. It can be modified with one or more of the following additions to produce a name: a suffix, post-suffix, prefix, super-fix letter, sub-joined letter, or one of the four vowel signs. It is the primary letter upon which alphabetization is carried out. It is the only letter in a name or syllable (with a few exceptions) which can have a vowel sound joined to it. In Tibetan there can only be one name-base per name. Some Dzongkha words contain an additional name base or *ming mtha'* (མིང་མཐའ་).

Appendix 4. Encoding model of Tibetan in the UCS

It is important to understand the model used for encoding Tibetan script in the UCS following the adoption of ISO/IEC 10646 Amendment 31, in 1999.

Regular & Combining Consonants

Vertically combined conjuncts of consonants and vowels occur frequently in Tibetan script text. However whether or not two neighbouring characters should stack vertically or be written left to right, one following the next, cannot always be determined simply by applying contextual or grammatical rules. For this reason, as well as the frequency and complexity of these vertical conjuncts in Tibetan text, experts concurred that a model somewhat different from that adopted for Devanagari and other many other Indic scripts was necessary.

The model used for encoding Tibetan in the UCS is an explicitly stacking model based on Tibetan orthography or the layout of Tibetan letters - *not on the rules or logic of Tibetan grammar*. In the UCS two complete sets of consonants were encoded as separate characters: one set of "headline" consonant characters (from U+0F40 to U+0F6A), to be used for single consonants or for the consonant occurring in the topmost position of any conjunct stack; and a second a set of combining consonant characters (from U+0F90 to U+0FBC) to be used for all additional consonants which occur in a stack. Characters for Tibetan vowels, usually written as marks combining with or dependant on consonants or consonant stacks, are encoded between these two sets of consonants from U+0F71 to U+0FB1. Strictly speaking modern Tibetan and Dzongkha proper use only four vowel signs U+0F72, U+0F74, U+0F7A and U+0F7C. The other vowel signs are used only in transliteration of Sanskrit and other foreign words in Tibetan script.

Character Order

Conjunct stacks are usually encoded in the order which the parts are written, first the character for the consonant in the topmost or headline position, followed by characters for any combining consonants and then by the character(s) for any vowel(s):

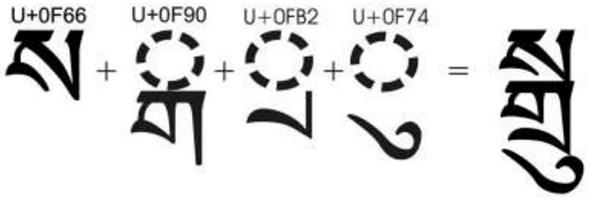


Figure 4

This follows the order in which Tibetan school children learn to write and spell out words.

In this way it is easy to represent even the very long stacks found in some religious texts:

Figure 5

After the character for the final below base consonant in a stack, characters for vowels or marks occurring below the final consonant are entered followed by those for vowels or marks occurring above the glyph for the base character (from the top of the first consonant upwards):

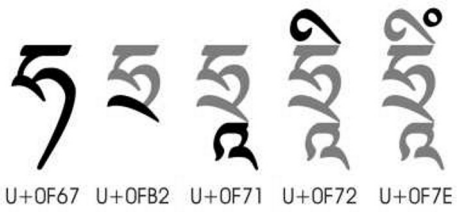


Figure 6

Syllables & Encoding

As mentioned above (Appendix 1) the basic unit in Tibetan & Dzongkha is the *tsheg bar* usually referred to as a "syllable" in English language books on Tibetan. The character encoding for the complex syllable *sgrubs* which contains a prefix, head letter, root letter sub-fixed letter, sub fixed vowel sign, suffix and post-suffix is as follows:

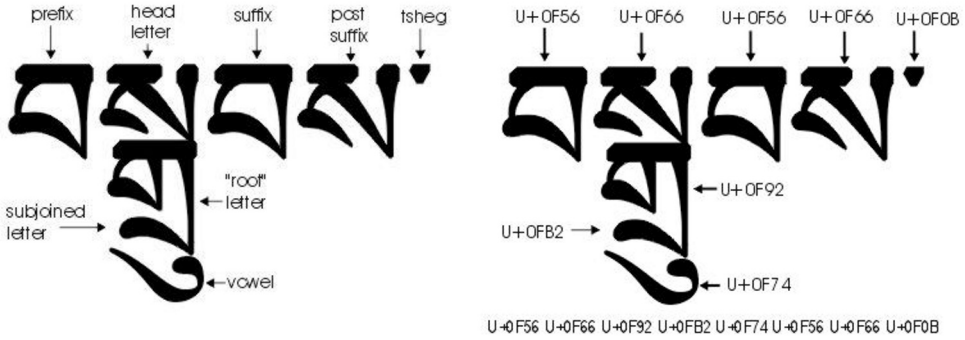


Figure 7

The base character in a Tibetan stack talked about in the character encoding for Tibetan should *not* be confused with the grammatical root letter (ming zhi) in a Tibetan syllable.

Also see: Encoding model of the Tibetan script in the UCS *by* Christopher J Fynn *(March 5, 2004)*

A table of Tibetan collation elements based on the UCS encoding is illustrated in the following section.

Basic Collation Table for UCS Tibetan Characters

ཀ 0F40	ཀྐ 0F40 0F90	ཀྑ 0F40 0F91	ཀྒ 0F40 0F92	ཀྒྷ 0F40 0F93	ཀྔ 0F40 0F94	ཀྕ 0F40 0F95	ཀྖ 0F40 0F96	ཀྗ 0F40 0F97
ཁ 0F41	ཁྐ 0F41 0F90	ཁྑ 0F41 0F91	ཁྒ 0F41 0F92	ཁྒྷ 0F41 0F93	ཁྔ 0F41 0F94	ཁྕ 0F41 0F95	ཁྖ 0F41 0F96	ཁྗ 0F41 0F97
ག 0F42	གྐ 0F42 0F90	གྑ 0F42 0F91	གྒ 0F42 0F92	གྒྷ 0F42 0F93	གྔ 0F42 0F94	གྕ 0F42 0F95	གྖ 0F42 0F96	གྗ 0F42 0F97
ང 0F44	ངྐ 0F44 0F90	ངྑ 0F44 0F91	ངྒ 0F44 0F92	ངྒྷ 0F44 0F93	ངྔ 0F44 0F94	ངྕ 0F44 0F95	ངྖ 0F44 0F96	ངྗ 0F44 0F97
ཅ 0F45	ཅྐ 0F45 0F90	ཅྑ 0F45 0F91	ཅྒ 0F45 0F92	ཅྒྷ 0F45 0F93	ཅྔ 0F45 0F94	ཅྕ 0F45 0F95	ཅྖ 0F45 0F96	ཅྗ 0F45 0F97
ཆ 0F46	ཆྐ 0F46 0F90	ཆྑ 0F46 0F91	ཆྒ 0F46 0F92	ཆྒྷ 0F46 0F93	ཆྔ 0F46 0F94	ཆྕ 0F46 0F95	ཆྖ 0F46 0F96	ཆྗ 0F46 0F97
ཇ 0F47	ཇྐ 0F47 0F90	ཇྑ 0F47 0F91	ཇྒ 0F47 0F92	ཇྒྷ 0F47 0F93	ཇྔ 0F47 0F94	ཇྕ 0F47 0F95	ཇྖ 0F47 0F96	ཇྗ 0F47 0F97
ཉ 0F49	ཉྐ 0F49 0F90	ཉྑ 0F49 0F91	ཉྒ 0F49 0F92	ཉྒྷ 0F49 0F93	ཉྔ 0F49 0F94	ཉྕ 0F49 0F95	ཉྖ 0F49 0F96	ཉྗ 0F49 0F97
ཊ 0F4A	ཊྐ 0F4A 0F90	ཊྑ 0F4A 0F91	ཊྒ 0F4A 0F92	ཊྒྷ 0F4A 0F93	ཊྔ 0F4A 0F94	ཊྕ 0F4A 0F95	ཊྖ 0F4A 0F96	ཊྗ 0F4A 0F97
ཋ 0F4B	ཋྐ 0F4B 0F90	ཋྑ 0F4B 0F91	ཋྒ 0F4B 0F92	ཋྒྷ 0F4B 0F93	ཋྔ 0F4B 0F94	ཋྕ 0F4B 0F95	ཋྖ 0F4B 0F96	ཋྗ 0F4B 0F97
ཌ 0F4C	ཌྐ 0F4C 0F90	ཌྑ 0F4C 0F91	ཌྒ 0F4C 0F92	ཌྒྷ 0F4C 0F93	ཌྔ 0F4C 0F94	ཌྕ 0F4C 0F95	ཌྖ 0F4C 0F96	ཌྗ 0F4C 0F97
ཎ 0F4E	ཎྐ 0F4E 0F90	ཎྑ 0F4E 0F91	ཎྒ 0F4E 0F92	ཎྒྷ 0F4E 0F93	ཎྔ 0F4E 0F94	ཎྕ 0F4E 0F95	ཎྖ 0F4E 0F96	ཎྗ 0F4E 0F97
ཏ 0F4F	ཏྐ 0F4F 0F90	ཏྑ 0F4F 0F91	ཏྒ 0F4F 0F92	ཏྒྷ 0F4F 0F93	ཏྔ 0F4F 0F94	ཏྕ 0F4F 0F95	ཏྖ 0F4F 0F96	ཏྗ 0F4F 0F97
ཐ 0F50	ཐྐ 0F50 0F90	ཐྑ 0F50 0F91	ཐྒ 0F50 0F92	ཐྒྷ 0F50 0F93	ཐྔ 0F50 0F94	ཐྕ 0F50 0F95	ཐྖ 0F50 0F96	ཐྗ 0F50 0F97
ད 0F51	དྐ 0F51 0F90	དྑ 0F51 0F91	དྒ 0F51 0F92	དྒྷ 0F51 0F93	དྔ 0F51 0F94	དྕ 0F51 0F95	དྖ 0F51 0F96	དྗ 0F51 0F97
ན 0F53	ནྐ 0F53 0F90	ནྑ 0F53 0F91	ནྒ 0F53 0F92	ནྒྷ 0F53 0F93	ནྔ 0F53 0F94	ནྕ 0F53 0F95	ནྖ 0F53 0F96	ནྗ 0F53 0F97
པ 0F54	པྐ 0F54 0F90	པྑ 0F54 0F91	པྒ 0F54 0F92	པྒྷ 0F54 0F93	པྔ 0F54 0F94	པྕ 0F54 0F95	པྖ 0F54 0F96	པྗ 0F54 0F97
ཕ 0F55	ཕྐ 0F55 0F90	ཕྑ 0F55 0F91	ཕྒ 0F55 0F92	ཕྒྷ 0F55 0F93	ཕྔ 0F55 0F94	ཕྕ 0F55 0F95	ཕྖ 0F55 0F96	ཕྗ 0F55 0F97
བ 0F56	བྐ 0F56 0F90	བྑ 0F56 0F91	བྒ 0F56 0F92	བྒྷ 0F56 0F93	བྔ 0F56 0F94	བྕ 0F56 0F95	བྖ 0F56 0F96	བྗ 0F56 0F97
མ 0F58	མྐ 0F58 0F90	མྑ 0F58 0F91	མྒ 0F58 0F92	མྒྷ 0F58 0F93	མྔ 0F58 0F94	མྕ 0F58 0F95	མྖ 0F58 0F96	མྗ 0F58 0F97
ཙ 0F59	ཙྐ 0F59 0F90	ཙྑ 0F59 0F91	ཙྒ 0F59 0F92	ཙྒྷ 0F59 0F93	ཙྔ 0F59 0F94	ཙྕ 0F59 0F95	ཙྖ 0F59 0F96	ཙྗ 0F59 0F97
ཚ 0F5A	ཚྐ 0F5A 0F90	ཚྑ 0F5A 0F91	ཚྒ 0F5A 0F92	ཚྒྷ 0F5A 0F93	ཚྔ 0F5A 0F94	ཚྕ 0F5A 0F95	ཚྖ 0F5A 0F96	ཚྗ 0F5A 0F97
ཛ 0F5B	ཛྐ 0F5B 0F90	ཛྑ 0F5B 0F91	ཛྒ 0F5B 0F92	ཛྒྷ 0F5B 0F93	ཛྔ 0F5B 0F94	ཛྕ 0F5B 0F95	ཛྖ 0F5B 0F96	ཛྗ 0F5B 0F97
ཝ 0F5D	ཝྐ 0F5D 0F90	ཝྑ 0F5D 0F91	ཝྒ 0F5D 0F92	ཝྒྷ 0F5D 0F93	ཝྔ 0F5D 0F94	ཝྕ 0F5D 0F95	ཝྖ 0F5D 0F96	ཝྗ 0F5D 0F97
ཞ 0F5E	ཞྐ 0F5E 0F90	ཞྑ 0F5E 0F91	ཞྒ 0F5E 0F92	ཞྒྷ 0F5E 0F93	ཞྔ 0F5E 0F94	ཞྕ 0F5E 0F95	ཞྖ 0F5E 0F96	ཞྗ 0F5E 0F97
ཟ 0F5F	ཟྐ 0F5F 0F90	ཟྑ 0F5F 0F91	ཟྒ 0F5F 0F92	ཟྒྷ 0F5F 0F93	ཟྔ 0F5F 0F94	ཟྕ 0F5F 0F95	ཟྖ 0F5F 0F96	ཟྗ 0F5F 0F97
འ 0F60	འྐ 0F60 0F90	འྑ 0F60 0F91	འྒ 0F60 0F92	འྒྷ 0F60 0F93	འྔ 0F60 0F94	འྕ 0F60 0F95	འྖ 0F60 0F96	འྗ 0F60 0F97
ཡ 0F61	ཡྐ 0F61 0F90	ཡྑ 0F61 0F91	ཡྒ 0F61 0F92	ཡྒྷ 0F61 0F93	ཡྔ 0F61 0F94	ཡྕ 0F61 0F95	ཡྖ 0F61 0F96	ཡྗ 0F61 0F97
ར 0F62	རྐ 0F62 0F90	རྑ 0F62 0F91	རྒ 0F62 0F92	རྒྷ 0F62 0F93	རྔ 0F62 0F94	རྕ 0F62 0F95	རྖ 0F62 0F96	རྗ 0F62 0F97
ལ 0F63	ལྐ 0F63 0F90	ལྑ 0F63 0F91	ལྒ 0F63 0F92	ལྒྷ 0F63 0F93	ལྔ 0F63 0F94	ལྕ 0F63 0F95	ལྖ 0F63 0F96	ལྗ 0F63 0F97
ཤ 0F64	ཤྐ 0F64 0F90	ཤྑ 0F64 0F91	ཤྒ 0F64 0F92	ཤྒྷ 0F64 0F93	ཤྔ 0F64 0F94	ཤྕ 0F64 0F95	ཤྖ 0F64 0F96	ཤྗ 0F64 0F97
ཥ 0F65	ཥྐ 0F65 0F90	ཥྑ 0F65 0F91	ཥྒ 0F65 0F92	ཥྒྷ 0F65 0F93	ཥྔ 0F65 0F94	ཥྕ 0F65 0F95	ཥྖ 0F65 0F96	ཥྗ 0F65 0F97
ས 0F66	སྐ 0F66 0F90	སྑ 0F66 0F91	སྒ 0F66 0F92	སྒྷ 0F66 0F93	སྔ 0F66 0F94	སྕ 0F66 0F95	སྖ 0F66 0F96	སྗ 0F66 0F97
ཧ 0F67	ཧྐ 0F67 0F90	ཧྑ 0F67 0F91	ཧྒ 0F67 0F92	ཧྒྷ 0F67 0F93	ཧྔ 0F67 0F94	ཧྕ 0F67 0F95	ཧྖ 0F67 0F96	ཧྗ 0F67 0F97
ཨ 0F68	ཨྐ 0F68 0F90	ཨྑ 0F68 0F91	ཨྒ 0F68 0F92	ཨྒྷ 0F68 0F93	ཨྔ 0F68 0F94	ཨྕ 0F68 0F95	ཨྖ 0F68 0F96	ཨྗ 0F68 0F97

Tibetan Collation Table *cont.*

ཅ 0F45	གཅ 0F42 0F45	བཅ 0F56 0F45	ལྕ 0F63 0F95	བལྕ 0F56 0F63 0F95						
ཆ 0F46	མཆ 0F58 0F46	འཆ 0F60 0F46								
ཇ 0F47	མཇ 0F58 0F47	འཇ 0F60 0F47	ཇྲ 0F62 0F97	ལྗ 0F63 0F97	བཇྲ 0F56 0F62 0F97					
ཉ 0F49	ཉྭ 0F8B 0F99	གཉ 0F42 0F49	མཉ 0F58 0F49	ཉྲ 0F62 0F99	སྙ 0F66 0F99	བཉྲ 0F56 0F62 0F99	བསྙ 0F56 0F66 0F99			
ཏ 0F4F	ཊ 0F4A	གཏ 0F42 0F4F	བཏ 0F56 0F4F	ཏྲ 0F62 0F9F	ལྟ 0F63 0F9F	སྟ 0F66 0F9F	བཏྲ 0F56 0F62 0F9F	བལྟ 0F56 0F63 0F9F	བསྟ 0F56 0F66 0F9F	
ཐ 0F50	ཋ 0F4B	མཐ 0F58 0F50	འཐ 0F60 0F50							
ད 0F51	ཌ 0F4C	དག 0F51 0F42	དང 0F51 0F44	དལ 0F51 0F54	དཡོ 0F51 0F54 0F7C	དབ 0F51 0F56	དམ 0F51 0F58	དམོ 0F51 0F58 0F7C	གདག 0F42 0F51 0F42	གདང 0F42 0F51 0F44	གདན 0F42 0F51 0F53
	གདབ 0F42 0F51 0F56	གདམ 0F42 0F51 0F58	གཌོ 0F42 0F51 0F7E	གདའ 0F42 0F51 0F60	གདར 0F42 0F51 0F62	གདལ 0F42 0F51 0F63	གདས 0F42 0F51 0F66	གདི 0F42 0F51 0F72	གདུ 0F42 0F51 0F74	གདེ 0F42 0F51 0F7A	གདོ 0F42 0F51 0F7C
	བདག 0F56 0F51 0F42	བདམ 0F56 0F51 0F58	བཌོ 0F56 0F51 0F7E	བདའ 0F56 0F51 0F60	བདར 0F56 0F51 0F62	བདལ 0F56 0F51 0F63	བདས 0F56 0F51 0F66	བདུ 0F56 0F51 0F74	བདེ 0F56 0F51 0F7A	བདོ 0F56 0F51 0F7C	མདག 0F58 0F51 0F42
	མདང 0F58 0F51 0F44	མདན 0F58 0F51 0F53	མདའ 0F58 0F51 0F60	མདར 0F58 0F51 0F62	མདུ 0F58 0F51 0F74	མདེ 0F58 0F51 0F7A	མདོ 0F58 0F51 0F7C	འདག 0F60 0F51 0F42	འདང 0F60 0F51 0F44	འདད 0F60 0F51 0F51	འདན 0F60 0F51 0F53

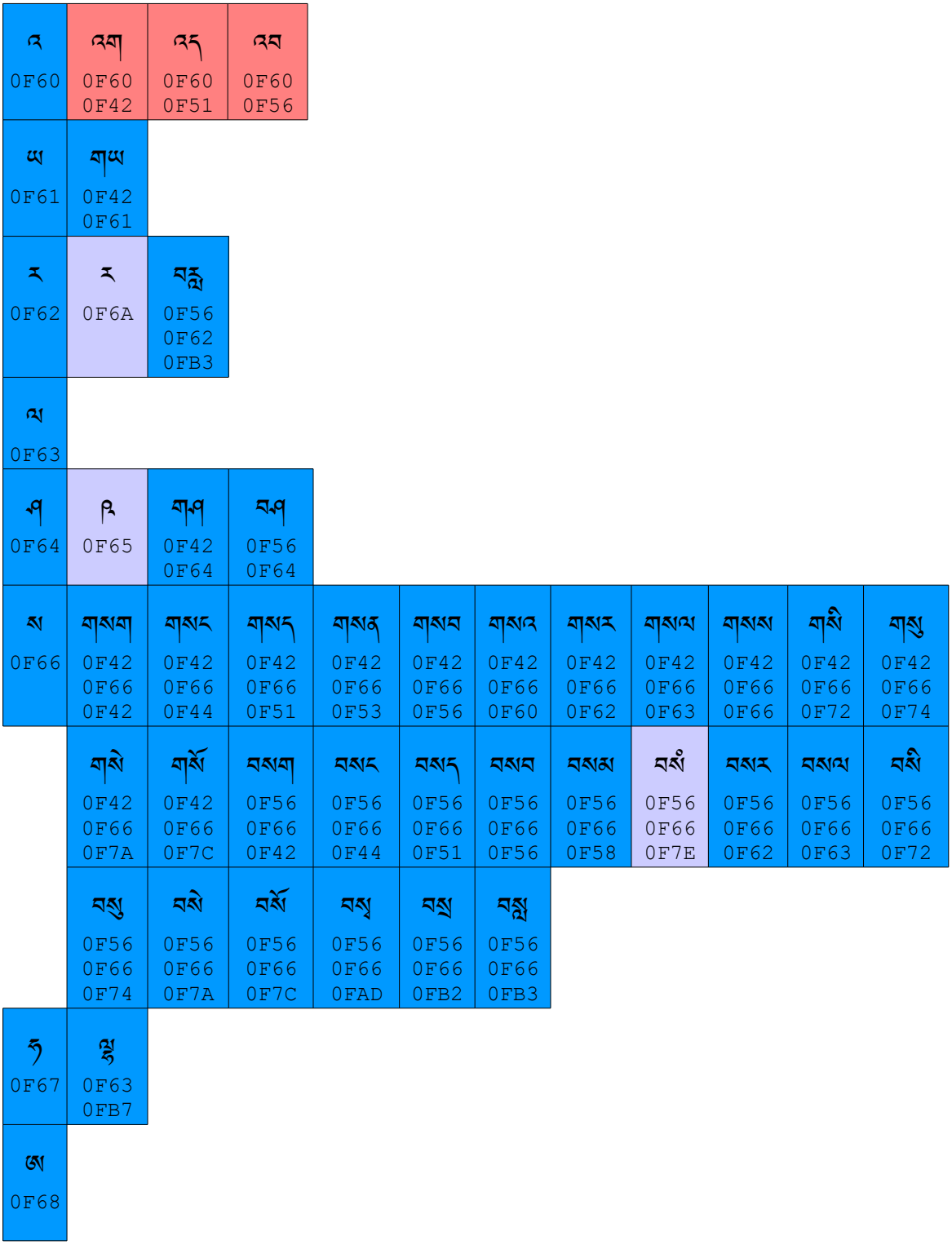
Tibetan Collation Table *cont.*

	ཨདྷ 0F60 0F51 0F56	ཨདྶ 0F60 0F51 0F58	ཨཱ 0F60 0F51 0F7E	ཨཁྱ 0F60 0F51 0F5D	ཨདྲ 0F60 0F51 0F60	ཨདྲ 0F60 0F51 0F62	ཨཧྱ 0F60 0F51 0F63	ཨཧྲ 0F60 0F51 0F66	ཨཱི 0F60 0F51 0F72	ཨཱུ 0F60 0F51 0F74	ཨེ 0F60 0F51 0F7A
	ཨེྃ 0F60 0F51 0F7C	ཨཱྀ 0F60 0F51 0FB2	ཀ 0F62 0FA1	ཁྱ 0F63 0FA1	ཁྲ 0F66 0FA1	གཀ 0F56 0F62 0FA1	གཁྱ 0F56 0F63 0FA1	གཁྲ 0F56 0F66 0FA1			
ཉ 0F53	ཾ 0F4E	གཉག 0F42 0F53 0F42	གཉང 0F42 0F53 0F44	གཉད 0F42 0F53 0F51	གཉན 0F42 0F53 0F53	གཉབ 0F42 0F53 0F58	གཉི 0F42 0F53 0F7E	གཉཁ 0F42 0F53 0F5D	གཉའ 0F42 0F53 0F60	གཉཇ 0F42 0F53 0F66	གཉཉ 0F42 0F53 0F74
	གཉེ 0F42 0F53 0F7C	ཉཉག 0F58 0F53 0F42	ཉཉང 0F58 0F53 0F44	ཉཉད 0F58 0F53 0F53	ཉཉན 0F58 0F53 0F56	ཉཉབ 0F58 0F53 0F58	ཉཉི 0F58 0F53 0F7E	ཉཉའ 0F58 0F53 0F60	ཉཉཇ 0F58 0F53 0F62	ཉཉཀ 0F58 0F53 0F63	ཉཉཉ 0F58 0F53 0F74
	ཉཉེ 0F58 0F53 0F7A	ཉཉེ 0F58 0F53 0F7C	ཊ 0F62 0FA3	ཋ 0F66 0FA3	ཏཊ 0F56 0F62 0FA3	ཏཋ 0F56 0F66 0FA3					
ཋ 0F54	ྐྵ 0F89 0FA4	དཀག 0F51 0F54 0F42	དཀང 0F51 0F54 0F44	དཀད 0F51 0F54 0F51	དཀན 0F51 0F54 0F60	དཀབ 0F51 0F54 0F62	དཀི 0F51 0F54 0F63	དཀཁ 0F51 0F54 0F66	དཀའ 0F51 0F54 0F74	དཀཇ 0F51 0F54 0F7A	དཀཀ 0F51 0F54 0F7C 0F42
	དཀོད 0F51 0F54 0F7C 0F44	དཀོད 0F51 0F54 0F7C 0F51	དཀོན 0F51 0F54 0F7C 0F53	དཀོབ 0F51 0F54 0F7C 0F62	དཀྵ 0F51 0F54 0FB1	དཀྵ 0F51 0F54 0FB2	ཁྵ 0F63 0FA4	ཁྵ 0F66 0FA4			
ཌ 0F55	ྐྵ 0F89 0FA5	ཨཌ 0F60 0F55									
ག 0F56	གཁ 0F56 0F42	གད 0F56 0F51	གན 0F56 0F62	གབ 0F56 0F66	དགཁ 0F51 0F56 0F42	དགད 0F51 0F56 0F44	དགན 0F51 0F56 0F51	དགབ 0F51 0F56 0F53	དགཁ 0F51 0F56 0F56	དགའ 0F51 0F56 0F60	དགཇ 0F51 0F56 0F62

Tibetan Collation Table *cont.*

	ད་ག་ལ 0F51 0F56 0F63	ད་ག་ས 0F51 0F56 0F66	ད་ག་ཤ 0F51 0F56 0F74	ད་ག་ཧ 0F51 0F56 0F7A	ད་ག་ཨ 0F51 0F56 0F7C	ད་ག་ཀྵ 0F51 0F56 0FB1	ད་ག་ཁ 0F51 0F56 0FB2	འ་ག་ཀ 0F60 0F56 0F42	འ་ག་ཁ 0F60 0F56 0F44	འ་ག་ག 0F60 0F56 0F51	འ་ག་ང 0F60 0F56 0F53
	འ་ག་ཅ 0F60 0F56 0F56	འ་ག་ཇ 0F60 0F56 0F58	འ་ག་ཉ 0F60 0F56 0F7E	འ་ག་ཏ 0F60 0F56 0F60	འ་ག་ད 0F60 0F56 0F62	འ་ག་ན 0F60 0F56 0F63	འ་ག་པ 0F60 0F56 0F72	འ་ག་ཕ 0F60 0F56 0F74	འ་ག་བ 0F60 0F56 0F7A	འ་ག་མ 0F60 0F56 0F7C	འ་ག་ཙ 0F60 0F56 0FB1
	འ་ག་ཚ 0F60 0F56 0FB2	ཛ 0F62 0FA6	ཛྷ 0F63 0FA6	ཉ 0F66 0FA6							
ཨ 0F58	◌ཾ 0F7E	◌ཿ 0F82	◌ྃ 0F83	ཨ་ཀ 0F58 0F42	ཨ་ཁ 0F58 0F44	ཨ་ག 0F58 0F51	ཨ་ང 0F58 0F53	ད་ཨ་ཀ 0F51 0F58 0F42	ད་ཨ་ཁ 0F51 0F58 0F44	ད་ཨ་ག 0F51 0F58 0F53	ད་ཨ་ཚ 0F51 0F58 0F5D
	ད་ཨ་ད 0F51 0F58 0F60	ད་ཨ་ཇ 0F51 0F58 0F62	ད་ཨ་ཉ 0F51 0F58 0F66	ད་ཨི 0F51 0F58 0F72	ད་ཨུ 0F51 0F58 0F74	ད་ཨེ 0F51 0F58 0F7A	ད་ཨོ 0F51 0F58 0F7C 0F51	ད་ཨྃ 0F51 0F58 0FB1	ཨཿ 0F62 0FA8	ཨྈ 0F66 0FA8	
ཀ 0F59	ཀ་ཅ 0F42 0F59	བ་ཅ 0F56 0F59	ཛ 0F62 0FA9	ཉ 0F66 0FA9	བ་ཛ 0F56 0F62 0FA9	བ་ཉ 0F56 0F66 0FA9					
ཧ 0F5A	ཨ་ཧ 0F58 0F5A	འ་ཧ 0F60 0F5A									
ཌ 0F5B	ཨ་ཌ 0F58 0F5B	འ་ཌ 0F60 0F5B	ཌྷ 0F62 0FAB	བ་ཌྷ 0F56 0F62 0FAB							
ཥ 0F5D											
ཎ 0F5E	ཀ་ཎ 0F42 0F5E	བ་ཎ 0F56 0F5E									

Tibetan Collation Table *cont.*

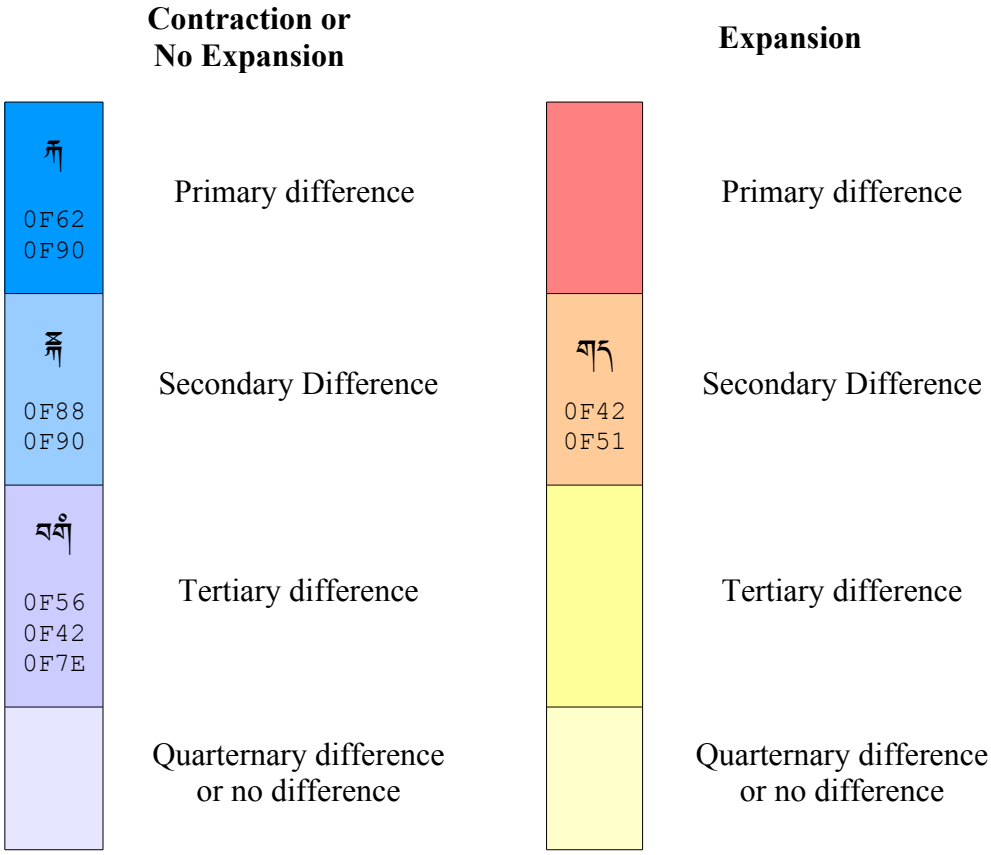


By combining the rules of Tibetan syllable formation (Appendix 1, 2, 3) with the encoding for Tibetan in the UCS which follows the principles described above (Appendix 4) a table of collation elements for use with ISO 14651 or the UCA can be derived. Such a table is illustrated above.

Key to collation table

The table on the preceding pages lists the basic UCS Tibetan characters & combinations used in words in ascending collation order. It is based on the usual (dictionary) order for collating Tibetan & Dzongkha words. (Any corrections or improvements that can be suggested by Tibetan experts in China would be most welcome.)

The colour of the cell the UCS characters are in indicates the strength of the difference between that element and the previous element in the chart, as follows:



Also see:

Sorting Unicode Tibetan using a Multi-Weight Collation Algorithim - by Robert Chilton.