

CCMP Feature in Tibetan Script Fonts

by [Christopher Fynn](#)





ccmp Glyph Composition/Decomposition



ccmp lookups are used to compose a sequence of glyphs to a single ligature glyph (GSUB lookup type 4); or do decompose a single glyph to a number of glyphs (GSUB lookup type 2 - multiple substitution). Lookups under the **ccmp** feature should be executed before lookups under any other feature. This means that the first lookup under the **ccmp** feature is applied to the glyph string of the underlying UCS characters mapped directly to their nominal glyph forms.

Canonical Decomposition

The first lookup in a Tibetan font will normally be a decomposition (GSUB type2) lookup under the **ccmp** feature to decompose glyphs for the combining compound characters (U+0F73, U+0F76, U+0F77, U+0F8, U+0F79, U+0F81) into their component parts. The glyphs for these characters need to be decomposed since they have elements which need to be placed both above and below the base stack to which they are applied which may be of different heights.

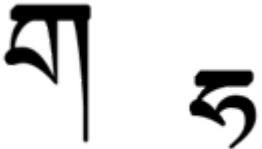

ccmp Decomposition Examples:

original glyph	ccmp decomposed glyphs
	
 uni0F73	 uni0F71 uni0F72
Decomposing the uni0F73 combining vowel glyph into two glyphs, uni0F71 (<i>achung</i>) below and uni0F72 (<i>gigu</i>) above, allows the two parts to be separately positioned above and below any stack without having to make many forms of uni0F73.	

original glyph	ccmp decomposed glyphs
 uni0F79	 uni0FB3 uni0F71 uni0F80
Decomposing the uni0F73 combining vowel glyph into three glyphs, uni0FB3 & uni0F71 below, and uni0F80 above allows uni0FB3 to be used in ligature composition, and uni0F71 and uni0F80 to be positioned above and below the final stack.	

Canonical Composition

The nextlookup under the **ccmp** feature may be a lookup to compose glyphs for certain compound characters (U+0F43, U+0F4D, U+0F52, U+0F57, U+0F5C, U+0F69 U+0F75, U+0F93, U+0F9D, U+0FA2, U+0FA7, U+0FAC, U+0FA9) from their component parts:

original glyph string	ccmp substituted ligature
 uni0F42 uni0FB7	 uni0F43
Substituting the glyphs uni0F42 (<i>ga</i>) and uni0FB7 (<i>ha</i>) for a single glyph uni0F43 (<i>gha</i>) at this point ensures that later substitution lookups (under blws feature) involving this combination are simplified.	

If such a lookup is not present under the **ccmp** feature it will be necessary to include multiple mappings to a single ligature in subsequent lookups. For example you would need to include four possible sequences for the ligature uni0F430F93 in a subsequent lookup:

```
sub uni0F42 uni0FB7 uni0F92 uni0FB7 by uni0F430F93;
sub uni0F42 uni0FB7 uni0F93 by uni0F430F93;
sub uni0F43 uni0F92 uni0FB7 by uni0F430F93;
sub uni0F43 uni0F93 by uni0F430F93;
```

instead of only the single line:

```
sub uni0F43 uni0F93 by uni0F430F93;
```

Ligature Composition

Under **ccmp** some font developers may also wish to include a lookup to compose ligatures for common compounds of vowels - or to substitute any other series of glyphs in the current glyph string for a ligature. Provided these are in a simple GSUB Type 4 lookup they may be included in **ccmp**.

Note: composition lookups under **ccmp** are a GSUB lookup type 4, and in this type of lookup sequences to be substituted which are made up of a larger number of components must be placed before those with a fewer number of components since the lookup stops when it meets the first matching pattern in a glyph string.

Sample lookups under **ccmp** feature:

```
feature ccmp { # Glyph Composition/Decomposition
script tib; # Tibetan
lookup decompose {
    sub uni0F73 by uni0F71 uni0F72;
    sub uni0F76 by uni0FB2 uni0F80;
    sub uni0F77 by uni0FB2 uni0F71 uni0F80;
    sub uni0F78 by uni0FB3 uni0F80;
    sub uni0F79 by uni0FB3 uni0F71 uni0F80;
    sub uni0F81 by uni0F71 uni0F80;
} decompose;
lookup compose {
    sub uni0F40 uni0FB4 by uni0F69;
    sub uni0F42 uni0FB7 by uni0F43;
    sub uni0F4C uni0FB7 by uni0F4D;
    sub uni0F51 uni0FB7 by uni0F52;
    sub uni0F58 uni0FB7 by uni0F59;
    sub uni0F7A uni0F7A by uni0F7B;
    sub uni0F7C uni0F7C by uni0F7D;
    sub uni0F90 uni0FB4 by uni0FB9;
    sub uni0F92 uni0FB7 by uni0F93;
    sub uni0F9C uni0FB7 by uni0F9D;
    sub uni0FA1 uni0FB7 by uni0FA2;
    sub uni0FA8 uni0FB7 by uni0FA9;
} compose;
} ccmp;
```

The **ccmp** feature is crucial for Tibetan script fonts, since without this it is much more difficult to deal with characters like U+0F73 U+0F76, U+0F77 U+0F78 U+0F79 & U+0F81.

17, December 2006