TO: **Unicode Technical Committee**
From: **Andrew West and Peter Lofting (via Debbie Anderson, SEI, UC Berkeley)**
Subject: Comments on Tibetan DOUBLE SHAY (response to Error report by D. Corbett)
Date: **3 October 2019**

This document includes comments from Peter Lofting (Apple) and Andrew West in response to the Error report from David Corbett (below).

1. Error report *from David Corbett (from L2/19-124 Public Review Issues, dated Tue Apr 2 2019)*

> *Regarding Tibetan punctuation, chapter 13 says "Because some writers use the double shay with a different spacing than would be obtained by coding two adjacent occurrences of U+0F0D, the double shay has been coded at U+0F0E with the intent that it would have a larger spacing between component shays than if two shays were simply written together. However, most writers do not use an unusual spacing between the double shay, so the application should allow the user to write two U+0F0D codes one after the other. Additionally, font designers will have to decide whether to implement these shays with a larger than normal gap."*
>
> *I've downloaded a bunch of Tibetan fonts and most of them display U+0F0E as slightly narrower than two U+0F0Ds. Many make them the same width. A few of the Qomolangma fonts make U+0F0E slightly wider. The code chart glyph for U+0F0E consists of two shays so close together there is barely any space between them. If the standard is correct, the code chart glyph is misleading, if not wrong, and should have more space between the shays. If the majority of my test's fonts are correct, chapter 13 should not imply their spacing is wrong.*

2 Comments from Peter Lofting (Apple)

The use of two consecutive single shays is equivalent to one double shay character.

Quite a common practice is to have two pairs of double shays separated by a large white space like this:

།།　　།།

as in the start of a text for example...

༄།།　　།།

The intent of the encoding of both single and double shays was for document structuring – not display variation: It was to have a series of Tibetan punctuation characters that could be used in order of ascending strength, analogous to the series: comma, semicolon, colon, period. Double shay is the most common escalation above a single shay, so it deserved it's own encoding. Beyond that you get the single, double and triple-dot shays, the snake (drul) shay and the chapter section shay.

The large space between two pairs does have honorific connotations. But there is no semantic I know of associated with the spacing within the pair of shays. That is just calligraphic style variation that should be left to individual font designs.

Therefore I don't believe that the Unicode Tibetan chapter should be commenting on spacing of two single versus one double shay at all. That is like commenting on the spacing of three periods versus one horizontal ellipsis… … . . . ……

Afterthoughts:

The widely spaced pair of single shads could be argued to be part of the hierarchy of punctuation characters and could take it's place in an ordering of ascending strength as it does appear in the context of a sentence delimiter at the end of verses and within long texts, as illustrated by Andrew's examples 1 and 3.

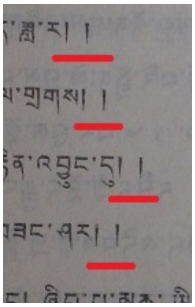
However there is another usage within manuscripts, which is at the start of texts. Furthermore, the widely spaced shads at the start of a text can vary in width of spacing. I was told this in Bhutan. They said that the amount of space given was in proportion to how important the subject or addressee letter is. I don't know if this convention applies just to correspondence letters or also applies to books.

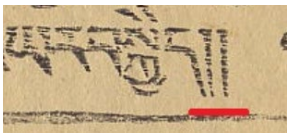
I've seen both single and double shads used in the start of texts. Below is a hierarchy of descending elaborateness/honorific levels.

A text representation of this variation is more intuitive and more explicit than width variants of a single character glyph, which actually hides important formatting data in font variations. Because of the two different uses – at the front and end of sentences, you would need to have two encodings. This is spawning non-intuitive homoglyphs in the representation for very little benefit and introduces the risk of confusing typists and generating mixed-up data (delimiters at start of sentences and head shads at end of sentences).

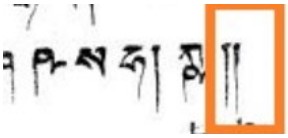
3. Comments from Andrew West

A very *provisional* answer is that there seem to be two different usages: 1) widely-spaced shad + shad (see attached examples 1 and 3); 2) close-together double shad (see attached examples 2 and 4). I do not know if both usages can occur in the same context.


Example 1


Example 2


Example 3


Example 4

My feeling is that the widely-spaced double shad should be represented as shad + space + shad, and the Unicode double shad character should only be used for the cases where two shad occur closely together as a single unit. If my interpretation is correct then this will require a rewrite of the TUS text, but no change to the code chart glyph.

I will try to find more examples from a wider range of materials, and if I possible I can write up a document for the UTC