Title: Comments on proposal N2621 (Tibetan encoding additions)
Source: Steve Hartwell, individual contributor
Date: 2003-09-25

I have read from a recent posting by Rick McGowan on the Tibex list
that comments on proposal N2621, a set of proposed additions to the
Tibetan encoding by the Chinese government to be reviewed by the WG2
meeting in October, are to be sent to you.

I anticipate that several others involved in the Tibetan encoding
discussions, including Michael Everson, Chris Fynn, Robert Chilton, and
Andrew West, will represent their respective organizations and provide
you with compelling rationale as to why this proposal should not be
accepted. Having followed their discussion on the Tibex mailing list,
I entirely agree with their views, and as an implementor of the Tibetan
script on computer platforms dating back to 1987, to their reasoned
arguments I would like to add my concerns about the considerable
complexity of implementation that would be introduced if proposal N2621
were to be accepted.

I first developed operating system support for the complex Tibetan
script in 1987, and have continued the development of this software
every year to the present, which is in common use today by Tibetan
scholars worldwide. Recently, I adapted a Tibetan font designed by a
Chinese font foundry on behalf of the Microsoft Typography group for
use in support of the Tibetan Uniscribe implementation which is to be
released shortly for Windows XP. Therefore, the encoding issues
related to the implementation of the Tibetan script are very familiar
to me.

At the request of Paul Nelson, the development lead for complex script
support in Uniscribe, I reviewed the N2621 proposal, as well as the
related formerly rejected proposal N964. This proposal, like its
predecessor, consists entirely of pre-composed renderings of characters
which already exist in the current Unicode Tibetan encoding.

Although some of these pre-composed sequences are ligature renderings
which might aid an implementation to render the basic Tibetan character
sequences more easily, a very large number of them are nothing more
than renderings of a base consonant or a ligature with a vowel added
above. Since even simple font rendering systems are capable of
displaying these sequences correctly without complex glyph processing,
there is no benefit to pre-composing these sequences even as
presentation forms, and certainly not as code points.

Therefore, apart from the added encoding complexity to support these
non-characters (an immense issue that has been elaborated in detail by
my colleagues), an important implementation consideration, should this
proposal be adopted, is that either these redundant compounds will have

to be added to the glyph repertoire of every Unicode-conforming Tibetan font (despite the fact that many have no distinct rendering value), or large and complex tables will have to be constructed to decompose them into the basic sequences which they actually represent. Note that the latter would require operating system support, as not all complex font formats can decompose a single 'character' into many component glyphs.

As an implementor of complex Tibetan script support for various platforms, I would have considerable difficulty in justifying the burden of adding many hundred unneeded glyphs to existing Tibetan fonts, or the construction of complex decomposition tables to handle these redundant encodings. On the other side, implementors who would develop fonts to support the precomposed 'characters' in this proposal would not likely bother to support the complex features already defined in the Tibetan script. Therefore the adoption of this proposal would have the effect of creating two distinct encodings for the Tibetan script, and with it, two distinct, incompatible implementations. In my opinion, this could seriously impede the interchange of Tibetan text between China and the rest of the world. This is certainly the most undesirable of the consequences of accepting this proposal.

In addition, the proposal's reasoning that these precomposed 'characters' are intended to support systems that are not capable of complex rendering is without basis, in view of the capabilities provided by the major operating systems: Microsoft is actively beta-testing its Windows XP support of the Tibetan script in Uniscribe; Otani University is doing the same for its Unicode Tibetan Language Kit for MacOS X, and IBM's complex layout support for OpenType and AAT in its ICU package, incorporated both in Java text rendering on Sun Microsystems platforms and also in the open-source Pango project for Linux, is capable of Tibetan script support of these "smart" font formats.

I hope that the many reasoned arguments against this proposal are not misunderstood as criticism of the Chinese government's indispensable participation in the development of the Tibetan script, and wish to reassure them that the support for complex scripts in use together with ideographic scripts has advanced on most widely-used operating systems to the stage where very good results can be achieved with the current Unicode Tibetan encoding.

Very best regards,

Steve Hartwell
MultiScript Solutions, International
Visiting professor, Otani University