

---

---

## Comments on the difficulty of implementing “dynamically combining” model used in the encoding of Tibetan in the UCS - as raised in proposal N2621\* from China.

\*(ISO/IEC 10646 JTC1/SC2/WG2 [N2621](#): "Revised Proposal on Tibetan BrdaRten Character Encoding 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: 2003.09.30

---

### ***Difficulty of implementing “dynamically combining” model used in the encoding of Tibetan script in the UCS.***

On page 2 of its proposal [N2621], the national body of China states:

*“The biggest difficulty for Tibetan information processing is the vertical composition of Tibetan characters. After the composition, each component would be changed greatly in shape and size, especially the vertical composition of the characters would reach to multiple layers where each letter requires different spans in height and width at the same layer that is quite hard to be dealt with. Up to now, there is no report showing any system platform has satisfactorily implemented Tibetan processing system using dynamic combining method.”*

From this statement it seems that China may be unaware that:

- 1) Software for Tibetan script using "dynamically combining" letters has been available at least since the early 1980's when Mr. Pierre Robillard created the LTibetan system & Marpa software for the Macintosh. In 1987 Mr. Steve Hartwell and Mr. Peter Lofting also developed the Druk Mac system for the on Macintosh use by Kuensel, the national newspaper of Bhutan. This software is still used today to publish this newspaper. The Druk Mac system was later developed further and made freely available as the Tibetan Language Kit by Otani University, Kyoto, Japan. Subsequently other developers including Mr. Marvin Moser, myself and the Centre for Development of Advanced Computing, Pune, India created a number of systems for Tibetan using "dynamically combining" letters which ran on Microsoft Windows 3x & '9x operating systems - although none of these early systems used UCS character encoding.

- 2) There are now at least three "smart" font and rendering technologies widely available for rendering complex scripts using dynamically combining glyphs based on underlying UCS data:-
  - a) Apple's ATUSI (Apple Type Services for Unicode Imaging) with AAT (Apple Advanced Typography) fonts.
  - b) Microsoft & Adobe's OpenType font format with "smart" rendering systems such as Microsoft's Uniscribe and Adobe's CoolType OpenType layout engines.
  - c) The Summer Institute of Linguistics' Graphite system for fonts & rendering.
- 3) There are several freely available or inexpensive software tools for creating the so-called "smart" fonts usable with the above technologies. These tools include: -
  - a) FontLab that can be used for creating and hinting glyphs as well as for adding OpenType layout tables. *See:* <http://www.fontlab.com/html/fontlab.html>
  - b) Microsoft's VOLT ["Visual OpenType Layout Tool"] which can be used for adding OpenType tables to existing TrueType fonts and is available at no cost. *See:* <http://www.microsoft.com/typography/developers/volt/>
  - c) Adobe's OpenType Font Development Kit [FDK] - which can be used to add OpenType tables to existing fonts - also freely available. *See:* <http://partners.adobe.com/asn/tech/type/otfdk/index.jsp>
  - d) Apple's free The Font Tool Suite can be used for adding AAT layout tables to fonts. *See:* <http://developer.apple.com/fonts/OSXTools.html>, *also see:* <http://developer.apple.com/fonts/TTRefMan/RM06/Chap6AATIntro.html> *and:* <http://developer.apple.com/intl/atsui.html>
  - e) The Summer Institute of Linguistics, Non Roman Script Initiative has made available Graphite which can be used to create "smart fonts" capable of displaying writing systems with complex behaviours including Tibetan. With respect to the UCS character-encoding model, Graphite can completely handle the "Rendering" aspect of complex script writing system implementation. Graphite is freely available as open source software and is being implemented both under Microsoft Windows and Linux. The Graphite Compiler which can be used to add Graphite tables to fonts is available at: [http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&item\\_id=GraphiteCompilerDownload](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=GraphiteCompilerDownload)
- 4) Modern operating systems and GUI (windowing) environments already provide services or libraries for creating applications capable of rendering complex scripts such as Tibetan dynamically using such "smart" fonts:-
  - a) The API for Apple's Multi Lingual Text Engine can be found at: <http://developer.apple.com/intl/mlte.html>
  - b) Information on Microsoft's Uniscribe engine for complex script shaping can be found at:

Comments on: Difficulty of implementing “dynamically combining” Tibetan.

[http://msdn.microsoft.com/library/default.asp?url=/library/en-us/intl/uniscrib\\_9xdf.asp](http://msdn.microsoft.com/library/default.asp?url=/library/en-us/intl/uniscrib_9xdf.asp)

- c) Microsoft also licence their OpenType Layout Services Library [OTLS] to application developers for use in and redistribution with applications. See: <http://www.microsoft.com/typography/developers/otls/>
  - d) The IBM sponsored International Components for Unicode [ICU] project provides a library of open source code in both C++ and Java that may be used for rendering both ATUSI/AAT and OpenType fonts for complex scripts. This is freely available from: <http://oss.software.ibm.com/icu/>
  - e) Open source code for Graphite developed by SIL’s Non-Roman Script Initiative and Language Software Development groups to provide rendering capabilities for complex non-Roman writing systems on a cross-platform basis is available from:  
<http://sourceforge.net/projects/silgraphite/>  
Graphite code for rendering complex is currently being integrated into the open source Mozilla web-browser. See: <http://sila.mozdev.org/>
  - f) Sun Microsystems has also been developing STF, a portable and extensible software framework for rendering complex text.. Open source code for STF is publicly available.  
See: <http://sourceforge.net/projects/stsf>
  - g) A comprehensive summary of rendering technologies available for rendering complex scripts can be found at:  
[http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&item\\_id=IWS-Chapter07](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=IWS-Chapter07)
- 5) Using such technologies implementations of Tibetan script have already been developed. And, based on the existing encoding of Tibetan in the UCS, these implementations are all capable of rendering all the pre-composed stacks in the N2621 proposal.
- a) The first implementation of UCS Tibetan has been (freely) available for Omega TeX since 1998 see: <http://www.logic.at/people/preining/tex/>
  - b) From July 2000 to December 2002 the Dzongkha Computing Project of the Royal Government of Bhutan's [Dzongkha Development Commission](#) (for which I was the lead developer), produced three OpenType fonts for Tibetan script (which is also used for writing Dzongkha, the national language of Bhutan). Based on the current encoding of the Tibetan script in the UCS, these fonts are already capable of producing all the combinations in the current N2621 proposal from China. These fonts have been successfully tested under Microsoft Corporation's Windows 2000 and XP operating systems using Microsoft's Uniscribe OpenType shaping engine [usp10.dll].  
  
This project also produced a keyboard (IME) for Tibetan script, UCA & ISO/IEC 14651 compatible tailored collation tables, and locale data for Bhutan all based on the current encoding of Tibetan in the UCS.
  - c) An OpenType font for Tibetan script (called "Ximalaya") that was designed by a foundry in China with OpenType tables added by Mr. Steve

Hartwell, along with a Chinese designed Tibetan IME, is under currently active beta testing by Microsoft and Chinese experts. Based on the current encoding of Tibetan in the UCS, this font is already capable of rendering all the composite Tibetan stacks in the N2621 proposal.

- d) Otani University, Kyoto, Japan is currently beta testing a UCS based Tibetan Language Kit for Apple Macintosh OSX using ATSUI/AAT technology.
- e) In December 2002 Xenotype Technologies introduced a Unicode Tibetan Language Kit and font for Apple Macintosh OSX. See: <http://www.xenotypetech.com/osxTibetan.html>
- f) Two UCS based OpenType fonts for Tibetan script and a UCS Tibetan keyboard have already been developed for Windows XP in Germany by Mr. Gregor Verhufen <<mailto:UZS34B@uni-bonn.de>>. This keyboard and fonts can be already be used to input and render all the stacks in the N2621 proposal based on the existing Tibetan characters in the UCS.

From the information listed above it can be seen that, using widely available technology, it is now feasible and reasonably low cost to successfully develop and implement systems for processing Tibetan text with "dynamic combining method" based on the current encoding of Tibetan in the UCS. Indeed several such implementations already exist and all the composite "characters" in the proposed "Tibetan BrdaRten Character Set" can already be successfully represented in and rendered by such systems.

These comments are intended to point out that the support for complex scripts in all the most widely used operating systems has finally progressed to the stage where the results desired by China can now straightforwardly be achieved with the current encoding of Tibetan in the UCS. An encoding which has been largely developed on the basis of proposals for Tibetan characters put forward by China to WG2 over more than a decade.

Since things have now progressed to this stage, acceptance of China's new N2621 proposal would unfortunately necessitate the re writing of existing applications; the addition of extra large, complex and otherwise unnecessary lookup tables to existing Tibetan script fonts designed to work with UCS Tibetan data; added and unnecessary complexity to IMEs designed for input of UCS Tibetan data and a great deal of additional complexity to tailored collation tables for ISO/IEC 14651.

Furthermore acceptance of China's proposal would also break the assurance given to developers by the Unicode Consortium on the release of Unicode Standard v 3.0, that henceforth no new characters that are composites of already encoded characters will be added to the UCS. I fear that making an exception in the case of proposal N2621 would subsequently make it difficult for both WG2 and the UTC to maintain this pledge and refuse any proposals from other national bodies requesting the addition of composite characters for scripts used in their countries.

*[Christopher J Fynn](#)*

2003.09.30

---