Survey of Language Computing in Asia
2005

Sarmad Hussain
Nadir Durrani
Sana Gul

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences

www.nu.edu.pk

www.idrc.ca

Published by

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences
Lahore, Pakistan

Copyrights © International Development Research Center, Canada

Printed by Walayatsons, Pakistan

ISBN: 969-8961-00-3

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada, administered through the Centre for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences (NUCES), Pakistan.

Dzongkha

Dzongkha is a Sino-Tibetan language related to Tibetan. It has 0.13 million first-language speakers [1] and approximately 0.5 million total speakers [3] in Bhutan. Dzongkha is the native language of eight western districts of Bhutan (Thimphu, Paro, Punakha, Wangdue, Phodrang, Gasa, Ha, Dhakana, and Chukha) and also recognized as the national and official language of the country. Dzongkha speakers also reside in India (specifically West Bengal) and Nepal [2].

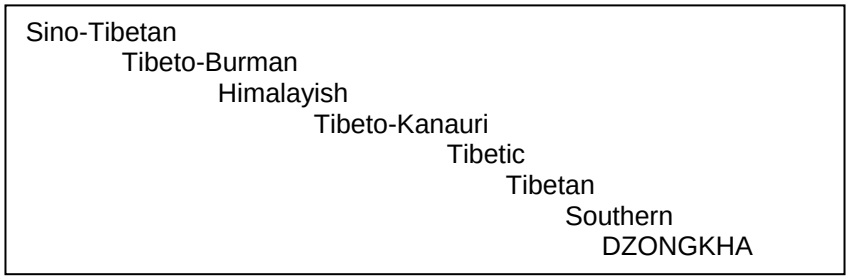


Figure 1: Language Family Tree of Dzongkha [1]

Dzongkha is written in Tibetan script, which was modeled on Devanagari script [3].

Character Set and Encoding

Dzongkha character set was not standardized prior to the release of Unicode 4.0. Unicode block 0F00-0FFF is the standard character set encoding used for Tibetan script in computers [4], which is also accepted as national as well as international standard for encoding Dzongkha text [5].

Fonts and Rendering

Tibetan script is complex and not possible to implement using True Type fonts. Open Type fonts have been developed for Tibetan script for Tibetan and Dzongkha by different people and organizations. These fonts include fonts developed by Department of IT and Dzongkha Development Authority in Bhutan, e.g. Tsuyig, Joyig, Tashi, xTashi, Uchen, Wangdi, [5, 6, 7].

Microsoft Platform

Until recently Dzongkha was not supported on the Microsoft platform. However, the latest version of Uniscribe (USP10.dll, version 1.453.3665.0) supports layout tables for Tibetan script. This version is shipped with Office 2003 Service Pack 1. Inclusion of Tibetan in the latest version of Uniscribe has facilitated typing and web-browsing in Tibetan script for Dzongkha.

Microsoft does not ship fonts for Tibetan script. However, third-party Tibetan script fonts can be used on Microsoft platform for Dzongkha text input and display [6, 8]. Rendering results of some of these fonts are shown in Figure 2 below.

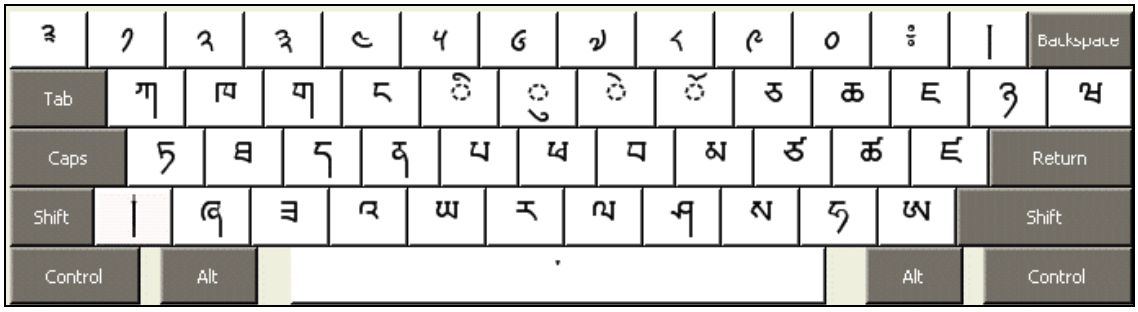
Figure 2: Unicode Dzongkha Fonts on Microsoft Platform [6]

Linux Platform

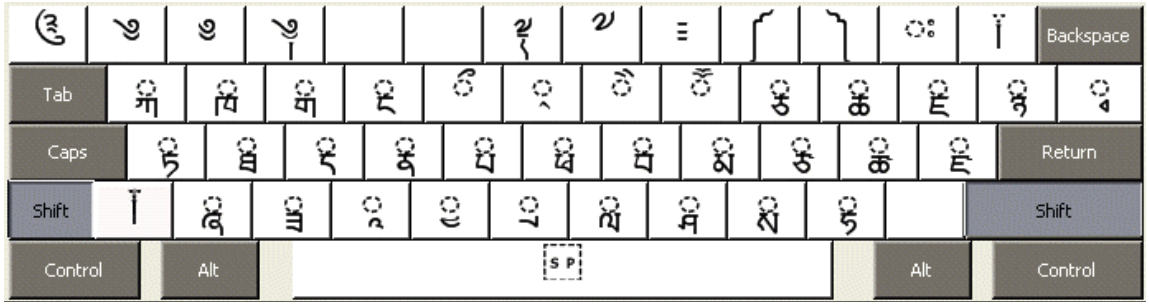
Dzongkha Development Authority, Department of Information Technology and Sherubste College are working together to enable Dzongkha computing on Linux operating system [9, 10] through the PAN Localization project [12]. Up till now the project has developed support for inclusion and rendering of Dzongkha Open Type fonts in X-Windows, Red Hat Linux, Fedora Core2 and Open Office 2.0 (though some technical challenges are still faced). Dzongkha Open Type fonts developed have also been successfully rendered through Pango in GNOME [11].

Keyboard

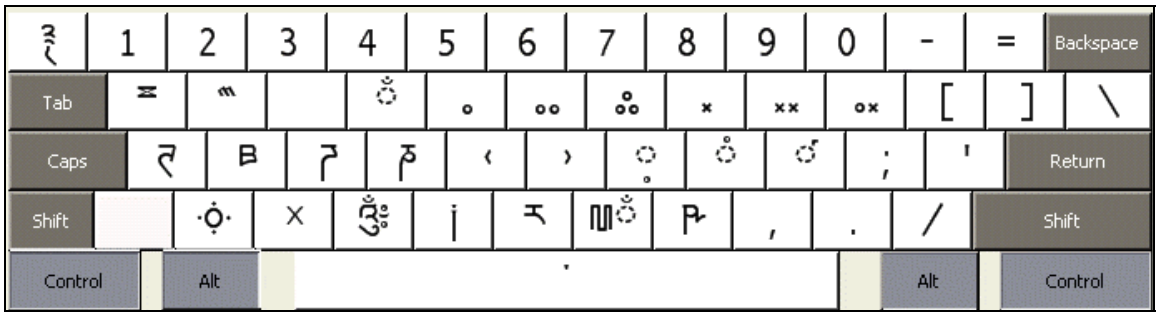
The Royal Government of Bhutan has nationally standardized a keyboard layout for Dzongkha. This has been designed by the Dzongkha Development Authority following consultation with all the major Dzongkha users and the Department of Information Technology. Figure 3 shows the standard keyboard layout [5].



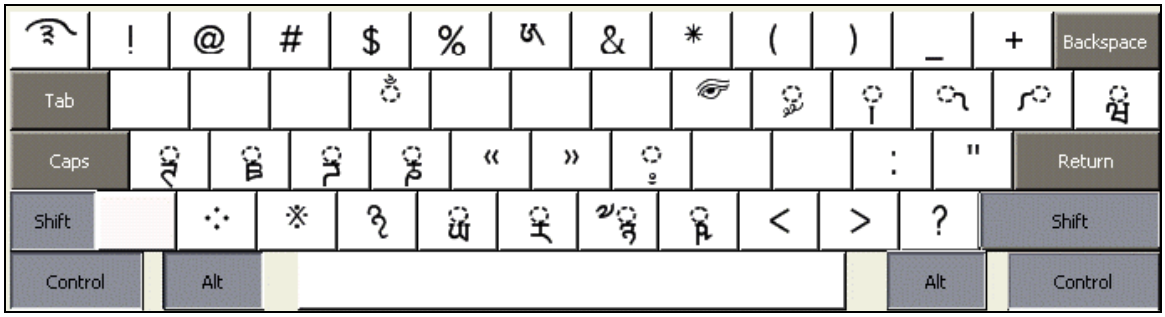
(a)



(b)



(c)



(d)

Figure 3: Standardized Dzongkha Keyboard in (a) Normal, (b) Shift, (c) Alt+Ctrl and (d) Alt-Ctrl-Shift States [5]

Microsoft Platform

Based on layout designed jointly by the Dzongkha Development Authority (DDA) and Department of Information Technology (DoIT), Royal Government of Bhutan, Tibetan and Himalayan Digital Library (THDL) project has created keyboards for Dzongkha using MSKLC for Microsoft platform. It can be used to input Dzongkha or Tibetan Unicode text [13, 15].

Linux Platform

Keyboard support for Dzongkha on Linux platform has also been developed using the standard and is being distributed with the Linux distribution being developed by Department of IT through PAN Localization project [12].

Collation

Collation rules are being finalized through collaboration of Dzongkha Development Authority (DDA) and Department of IT in Bhutan. They are based on the dictionary published by DDA.

Microsoft Platform

Microsoft lists Dzongkha in the Sort Options in its latest releases, but the sort is based on DUCET. Thus, Dzongkha sort is not realized. Figure 4 shows the Sort Options on MS Office 2003.

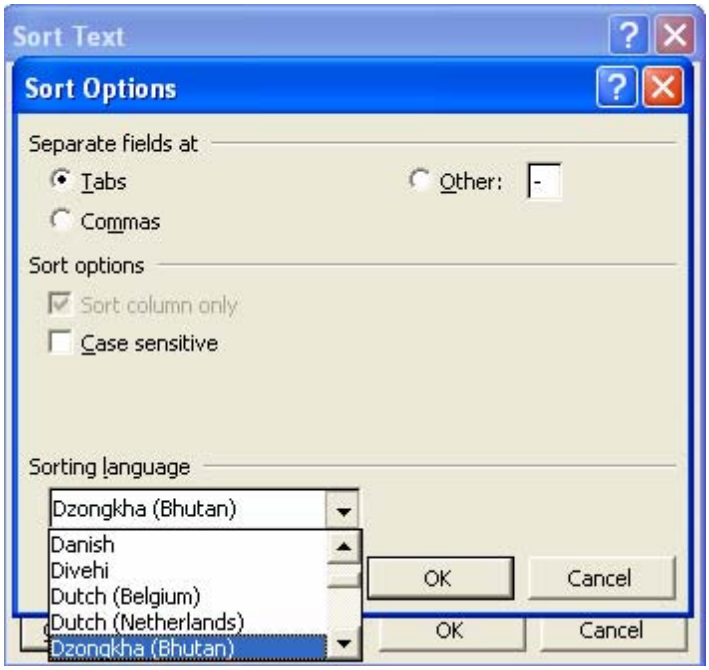


Figure 4: Sort Options for Dzongkha on Microsoft Platform

Linux Platform

Collation rules developed by DDA have been implemented on Linux platform. They are supported in Dzongkha version of Open Office 2.0 developed by Department of IT of Government of Bhutan, through PAN Localization project [12].

Locale

A nationally standardized locale definition for Dzongkha for Bhutan (dz_BT) has been compiled by Dzongkha Development Authority (DDA) in consultation with major computer vendors and language experts. Dzongkha (also known as Bhutani) is now included as a distinct language/culture within ISO 639, with the language codes "dz" and "dzo" [16]. Dzongkha locale definition has been included in CLDR 1.3 in 2004. The locale definitions described include date, time, calendar formats, name of days, months and numbers, etc.

Microsoft Windows XP does not include locale definition for Dzongkha.

Through the research efforts of Bhutan team of PAN Localization project, Dzongkha locale is now supported on Linux platform [11]. Locale for GNU C library has been created and implemented in

Linux operating system. Support for locale and collation rules has also been added to Open Office 2.0.

Interface Terminology Translation

Microsoft Platform

Dzongkha version of Microsoft is currently not available. However development of a localized version of Windows in Dzongkha is on the short term agenda of Microsoft [14].

Linux Platform

Glossary translation of KDE for Dzongkha has been initialized but there is still no significant progress [17]. However, GNOME desktop is complete and work is underway for complete translation of Open Office through PAN Localization project [18].

Status of Advanced Applications

There is little progress on the development of advanced applications in Dzongkha. Most of the work under progress is on localization of Linux, specifically GNOME and Open Office platforms.

References

[1] http://www.ethnologue.com/show_language.asp?code=dzo
[2] http://en.wikipedia.org/wiki/Dzongkha
[3] http://www.omniglot.com/writing/tibetan.htm
[4] http://www.unicode.org/charts/PDF/U0F00.pdf
[5] “Technology Standards and Resources for Computing in Dzongkha.” Department of IT, Royal Govt. of Bhutan. http://www.dit.gov.bt/guidelines/dzongkhastandard.pdf, 2004.
[6] http://salrc.uchicago.edu/resources/fonts/tibetanfonts.html
[7] http://www.dit.gov.bt/downloads/dzongkhafonts.zip
[8] Fynn, C. and Garson, T. “Tibetan fonts." http://iris.lib.virginia.edu/tibet/xml/show.php?xml=/tools/tibfonts.xml
[9] http://www.iosn.net/country/bhutan/news/dzongkha-on-linux
[10] http://sourceforge.net/projects/dzongkha/
[11] http://dzongkha.sourceforge.net/
[12] www.PANL10n.net
[13] http://iris.lib.virginia.edu/tibet/tools/dzkeyboard.html
[14] http://archives.cnn.com/2002/BUSINESS/asia/08/06/bhutan.windows/
[15] http://iris.lib.virginia.edu/tibet/tools/dzkeylayout.html
[16] “Oficial Nacional Standard of Dzongkha-Bhutan Locale.” Dzongkha Development Authority. http://www.dit.gov.bt/guidelines/locale_culture.pdf, 2004.
[17] http://i18n.KDE.org/teams/index.php?a=i&t=dz
[18] http://l10n.openoffice.org/languages.html