



PAN  
Localization

# 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](http://www.nu.edu.pk)

IDRC  CRDI

Canada

[www.idrc.ca](http://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

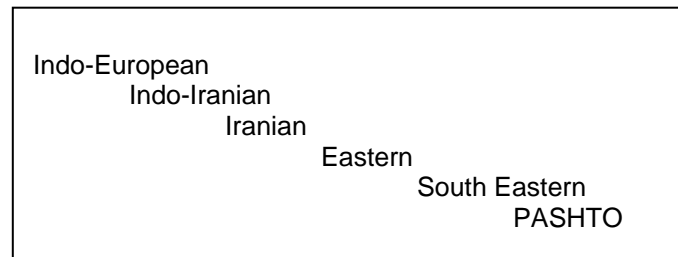
Printed by Walayatsons, Pakistan

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.*

# Pashto

Pashto is an Indo-Iranian language spoken by about 25 million people in Afghanistan, India, Iran, Pakistan, Tajikistan, the UAE and the UK. There are three main varieties of Pashto: Northern Pashto and Central Pashto (spoken mainly in Pakistan), and Southern Pashto (spoken mainly in Afghanistan). Pashto was made the national language of Afghanistan by royal decree in 1936. Today both Dari and Pashto are official languages in that region. Pashto is also a state language of Pakistan.



**Figure 1: Language Family Tree for Pashto [1]**

Pashto first appeared in writing during the 16th century in the form of an account of Sheikh Mali's conquest of Swat. It is written with modified Arabic script [2].

## Character Set and Encoding

Unicode Arabic script code block 0600-06FF is used for Pashto text encoding. Most of contemporary Pashto is supported. However, older version of script developed by Khushal Khan Khattak is still not supported. A few other ad hoc encodings have also been in use. However, most of the recent work is being done using Unicode.

## Fonts and Rendering

Pashto fonts are available in Open Type format (e.g. [3]).

### Microsoft Platform

Tahoma and Sans Serif fonts provide complete support for Pashto character set. These are built-in fonts shipped with XP. Some of the Pashto fonts have been developed by Pashto localization vendors. Figure 2 shows Microsoft Tahoma, Pashto Breshna and Nafees Pakistani Naskh on Microsoft platform.

افغانستان کې په یوه برید کې لږ تر لږه لس چنیایي کارکونکي

(a)

افغانستان کې په یوه برید کې لږ تر لږه لس چنیایي کارکونکي

(b)

افغانستان کې په یوه برید کې لږ تر لږه لس پنیایي  
(c)

**Figure 2: (a) Tahoma, (b) Nafees Pakistani Naskh and (c) Pashto Breshna [8] on MS Office 2003**

## Linux Platform

Linux supports rendering of four shaped fonts for Arabic script. Pashto written through these fonts can be adequately rendered on Linux platform. The following figures show rendering Pashto text in various applications running in Linux platform. However there are still some rendering issues with some characters. Open Type Pashto fonts cannot be used on KDE. The platform only supports True Type Pashto fonts.

دور ستیوکلونورو ستی ستی ستره لاسته  
(a)

دورستموکلونورو ستی ستی ستره لاسته  
(b)

د ازتر نیت د غچو د جالونو په  
(c)

**Figure 3: (a) Four-Shaped Pashto Font on Open Office, (b) Nafees Naskh in GNOME, and (c) Pashto TTF Naskh on KDE**

## Keyboard

No standard keyboard layout exists for Pashto language. There are keyboard layouts available based on Arabic or Persian layouts or phonetic layouts based on English QWERTY. These keyboard layouts have been developed by local vendors (e.g. [5, 6]). A recent study has been conducted by UNDP which recommends specifications [4]. Work is currently in progress to standardize the keyboard (and other localization related issues) through Afghanistan Localization Program [7]. The keyboard layouts in following figures are recommended by UNDP in the draft locale requirements for Afghanistan, and those developed by Khapla Pashto Software and Liwal Software.

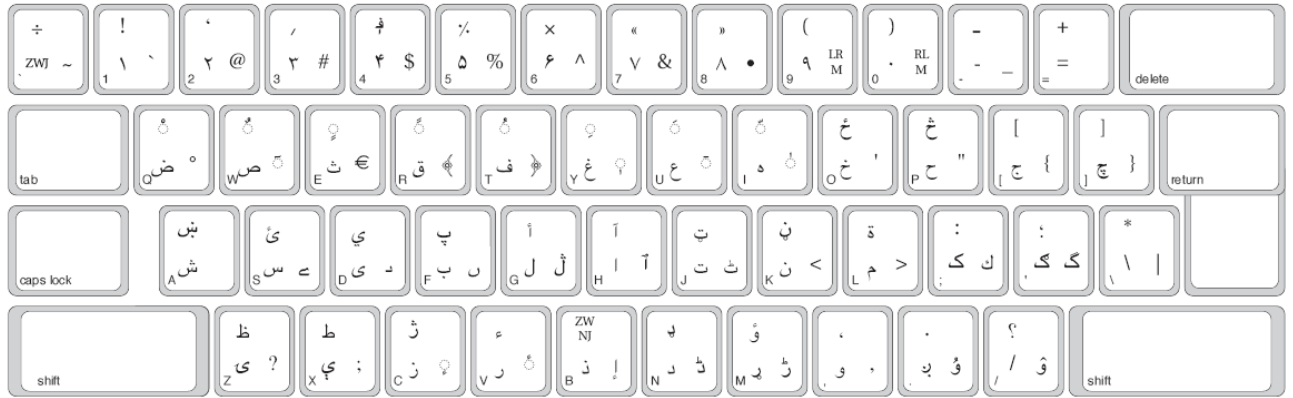


Figure 4: UNDP Recommended Pashto Keyboard (see [4] for clearer Image)



Figure 5: Pashto Keyboard by Khpala Software [5]



*Figure 6: Pashto Keyboard by Liwal Software [6]*

Microsoft or Linux platforms do not provide built-in keyboard layout for Pashto.

## Collation

UNDP has given recommendations on collation order of Pashto text [4]. However, it has not been standardized. Work is currently in progress through PAN Localization project to standardize Pashto sequence for Afghanistan [7].

Microsoft and Linux platforms do not support Pashto collation.

## Locale

A draft for Pashto locale requirements for Afghanistan has been developed by UNDP [8]. Incomplete versions of locale are available through IBM ICU and CLDR 1.3. The locale is also being standardized by Afghanistan (pa\_AF). No such work has been initiated in Pakistan (pa\_PK).

There is no locale support in Microsoft or Linux platforms. However, work is in progress to develop Microsoft Language Interface Pack for Pashto in Afghanistan [7].

## Interface Terminology Translation

No standard terminology exists for Pashto. Work is in progress through Afghanistan Localization program to develop this terminology [7].

## Status of Advanced Applications

Few language processing resources are available for Pashto. Some initial work has been done on Pashto lexica (e.g. [9]). Work has also been done on Arabic script based OCR by vendors, which can potentially be extended for Pashto. Pashto TTS and ASR systems are also being developed. Pashto-English machine translation systems are available through private vendors, e.g. [10, 11]. Most of this work is being done outside Afghanistan and Pakistan.

## References

- [1] [www.ethnologue.com](http://www.ethnologue.com)
- [2] [www.omniglot.com](http://www.omniglot.com) / [www.ethnologue.com](http://www.ethnologue.com)
- [3] <http://www.crulp.org/nafeesPakistaniNaskh.html>
- [4] <http://www.evertype.com/standards/af/af-locales.pdf>
- [5] <http://www.khpalapashtu.com/sitee/pashtusw/paskeyb.htm>
- [6] <http://www.liwal.com/windows/pashto/keyboard.htm>
- [7] <http://www.moc.gov.af/Projects/localization.asp>
- [8] <http://www.evertype.com/standards/af/>
- [9] <http://www.yorku.ca/twainweb/troberts/pashto/lexicon.html>
- [10] <http://linguistical.com/shop/index.cgi?ID=143217967&PID=IT492&code=13>
- [11] <http://www.aramedia.com/>