



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

IDRC  CRDI

Canada

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.

Nepali

Nepali is an Indo Aryan language spoken by about 17 million people in Nepal, Bhutan and some parts of India, and is the national and official language of Nepal [1]. Figure 2 shows the family tree for Nepali language.

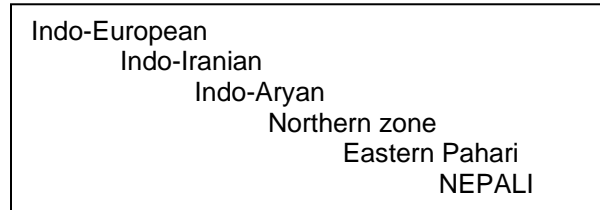


Figure 1: Language Family Tree of Nepali [1]

Bhujimol script was earlier used to write Nepali language. Now Nepali is written using Devanagari script [2,3].

Character Set and Encoding

Nepali uses the internationally standardized Devanagari block 0900-097F of Unicode. This standard is gaining popularity in Nepal. Unicode has adopted encoding characteristics from ISCII standard. However, there are still vendor specific encodings being used. Two of the other commonly used encoding schemes, Sabdatara and Anapurna are given in Figure 2 [4].

Another character set encoding standard was developed by the Nepali Fonts Standardization Committee in 1998 [4], but is not frequently used.

Sabdatara										Annapurna											
	0	1	2	3	4	5	6	7	8	9		0	1	2	3	4	5	6	7	8	9
30				ज	घ	झ	ञ	ठ			30				!	"	#	\$	%	&	'
40	ढ	ण	ड		.	(र	०	१	40	()	*	+	,	-	.	/	०	१
50	२	३	४	५	६	७	८	९	रु	स	50	२	३	४	५	६	७	८	९	:	;
60	?	.	क्ष	रु	ह	द्व	य	त्र	म	भ	60	<	=	>	?	@	अ	अ	इ	ई	उ
70	°	न	उ	क	०	८	ी	:	ल	इ	70	ऊ	ऋ	ए	ऐ	ओ	औ	क	व	कु	क
80	ए	त	र	क	त	र	ह	६	ह	६	80	ऋ	क्ष	क्ष	ख	रु	ग	र	ज	व	घ
90	श				ट)		व	द	अ	90	८	[\]	^	_	ड	ड	ड	ड
100	म	भ	ा	न	ज	०	व	प	ि	०	100	ड	उ	ड	ऋ	८	३	छ	ज	८	ज
110	ल	य	उ	त्र	च	क	त	ग	ख	घ	110	भ	क्त	भ	क्ष	ज	८	ट	ह	ड	ठ
120	ह	थ	श		,		०				120	डु	घ	ड	{		}	-		ह	ण
130	५	८	ड	ह	-	८	६	३	३	ड	130	प	□	८	...	८	त्र	थ	६	द	ह
140						६		५			140	ड	प	८	ड	थ	'	'	"	"	६
150											150	-	-	न	८	८	प	८	स	फ	५
160											160		क्र	व	८	८	भ	५	म	८	स
170											170	य	८	-	८	८	र	०	८	रु	रु
180											180	ल	ल	॥	.	स	व	८	व	५	५
190											190	५	श	५	क्ष	श्र	क्ष	ष	८	६	६
200											200	स	रु	स	स	ह	ह	स	ह	ह	ह
210											210	ह	ह			अ	५	८	.	.	.
220											220
230											230	.	६	ि	रु	ी	ी	०	०	५	५
240			य	६							240	५	ह	५	ह	८	८	८	८	५	५
250											250	.	.	ी	५	५	५	५	५	५	५

Figure 2: Popularly Used Nepali Encodings [4]

Fonts and Rendering

Microsoft provides support for Devanagari script in Arial and Mangal fonts, which are shipped with Windows and Office. These fonts can be used for Nepali. Many other Nepali Unicode fonts have also developed by other groups. Some of these fonts are Gauri, Himali, Fontasy Himali, Kanchan, Kantipur, MtEverest, Nepali, Kalimati, and Kanjiwari [5, 6, 7]. Figure 3 shows results of rendering Nepali text using Devanagari fonts on MS platform. Devanagari script is also supported on Linux, through work done by India Linux [8] and Madan Puraskar Pustakalaya (through PAN Localization Nepali component) [7,9].

Sample	Font
हिन्दी, मराठी, नेपाली	Akshar Unicode (akshar.ttf)
हिन्दी, मराठी, नेपाली	Arial Unicode MS (arialuni.ttf)
हिन्दी, मराठी, नेपाली	CDAC-GISTSurekh (CDACSRNT.TTF and CDACSRBT.TTF [bold])
हिन्दी, मराठी, नेपाली	CDAC-GISTYogesh (CDACYGNT.TTF [normal], CDACYGBT.TTF [bold], and CDACYGIT.TTF [italic])

Figure 3: Devanagari Fonts for Nepali [6]

Keyboard

Keyboard layout for Nepali has not been standardized yet. Commonly used Nepali keyboard layouts, Remington and phonetic, are shown in Figure 4 [4].

	~	ॐ	इ	घ	ढ	छ	ट	ठ	ड	ढ	ण	ˆ	-	. +	ˆ
	ज	१	२	३	४	५	६	७	८	९	०	ˆ	-	, =	ˆ
tab	त्त Q	ध W	ध E	च R	त् T	थ Y	ग् U	क्ष I	ड O	ए P	{	^ }	enter		
	त्र	ध	भ	च	त	थ	ग	ष	य	उ	[ˆ]			
cap lk	ठ A	क S	म D	फ F	ग G	ज H	ठ J	क K	ी L	र :	“				
	ब	क	म	ा	न	ज	व	प	ि	स ;	’				
shift	श् Z	ह X	ऋ C	व V	ब B	न N	: M	ड <	श्र >	रू ?	shift				
	श	ह	अ	ख	द	ल	ँ	, ,	। .	र /					
space															

(a)

ज ~ ः	!	@	#	रु Rs. ४	% ५	^ ६	ऊं ७	* ८	(९) ०	- -	+ =	 \	
tab	त्र Q क्ष	व W व	ए E ँ	र R र	थ T त	य Y य	उ U उ	इ I ि	ओ O ो	फ P प	{ रु [}	ड]	enter
cap lock	अ A ा	श S स	ध D द	् F ँ	घ G ग	ह H ह	ञ J ज	ख K क	ल L ल	:	:	:	:	:
shift	ऋ Z ृ	ऌ X ष	ऍ C च	व V व	भ B भ	ण N न	म M म	< ,	> .	? /	:	:	:	shift

space

(b)

Figure 4: (a) Remington, and (b) Phonetic Keyboard Layouts for Nepali

These keyboard layouts are also available for Linux platform [7], and may also be created on Microsoft platform through its MSKLC tool [10]. Microsoft has also released Nepali LIP, through which a Nepali on-screen Keyboard is provided.

Collation

Two different collation sequences are followed in Nepal. One which treats three conjoined characters a sequence of original characters and other which treats them as new characters. The former is taught in schools and used in phone books, and the latter has been used by Royal Nepal Academy to print its dictionary, Brihat Shabdakosh. Recently, after debate through Nepali language in IT committee, former was standardized nationally [11].

Microsoft Platform

Microsoft does not support Nepali sorting. Krama, a sorting utility developed by Madan Puraskar Pustakalya sorts Nepali strings. Sorting through this utility can be customized through various sort options provides in the utility [7].

Linux Platform

Support for Nepali collation has also been developed for Linux platform.

Locale

Nepali locale (ne_NP) has not been standardized, however some work on it has started [12]. Microsoft has released its Nepali LIP, which has Nepali locale data. Nepali Linux by MPP also has Nepali locale defined.

Interface Terminology Translation

Nepali glossary has been translated and standardized by Nepali language in IT committee. It has also been implemented to develop Nepali Linux through PAN Localization project.

Microsoft Platform

This translation has been used to produce Microsoft LIP for Nepali.

Linux Platform

On the Linux platform 84.85% of GNOME 2.12 has been done by the Nepali team [13] through PAN Localization project [9]. A linux distribution is available including localized Open Office, Nepali GNOME desktop and Mozilla browser.

Status of Advanced Applications

Madan Puraskar Pustakalya is currently developing support for advanced Nepali applications. Currently the MPP team has developed an encoding conversion utility Rupanter that converts non-Unicode Nepali text in to the Nepali Unicode text. Non-Unicode text might be in ad hoc True Type font encodings for Preeti, Kantipur etc. MPP has also developed a prototype version of Nepali spell checker, a Nepali 800 word dictionary, and a Nepali thesaurus of about 800 words. Work is also underway on English-Nepali Machine Translation project [7].

References

- [1] <http://www.ethnologue.com>
- [2] <http://www.omniglot.com>
- [3] http://en.wikipedia.org/wiki/Nepali_language
- [4] "Nepali Font Standards." <http://www.cicc.or.jp/english/hyoujyunka/mlit3/7-7-2.pdf>, 1998.
- [5] <http://www.nepalhomepage.com/reference/fonts/>
- [6] <http://salrc.uchicago.edu/resources/fonts/devanagarifonts.html>
- [7] <http://www.mpp.org.np/>
- [8] <http://www.lndLinux.org>
- [9] <http://www.PANL10n.net>
- [10] http://www.mpp.org.np/detail_guide/winxp.htm
- [11] Tuladhar, A. "Report on Activities of Standardization of Nepali In Computers." <http://www.unlimit.com/nepali/reports/malaysia.doc>
- [12] http://www.nepalinux.org/ldf/ne_NP
- [13] <http://l10n-status.GNOME.org/GNOME-2.12/index.html>