



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.

Telugu

Telugu is a Dravidian language spoken in southern Indian states, and is the official language of Andhra Pradesh. It is considered to be the the second most widely spoken language in India after Hindi. There are 75 million first language and about 5 million second language speakers of Telugu [1]. Figure 1 shows the language tree for Telugu.

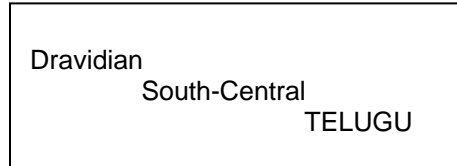


Figure 1: Language Family Tree of Telugu [1]

Telugu script derives from Brahmi script, and is similar to Kannada script. The earliest known inscriptions in the Telugu language date back from the 6th century AD. Telugu has been written in an old style. However, the writing system was modernized in second half of 20th century to agree more with the spoken language [2].

Character Set and Encoding

ISCII includes encoding for Telugu characters, with code page identifier 57006. Telugu script has also been incorporated in Unicode from 0C00-0C7F [3].

There are also other ad hoc encodings which are in use. Encoding converters between ISCII and these Telugu ad hoc encodings have also been developed [17].

Fonts and Rendering

Many Open Type fonts are available, which can be used for rendering Telugu script, e.g. Akshar Unicode, Code 2000, Pothana 2000 and Vemana 2000 [4].

Microsoft Platform

Microsoft supports Telugu rendering, and provides font support for this script, e.g. in Arial Unicode MS font. Figure 2 shows results of rendering some of these fonts on Microsoft platform [4]. Microsoft also provides a guide to develop Open Type fonts for Telugu [8].

Linux Platform

Telugu Open Type fonts are not rendered on Linux platform in regular distributions. However, rendering support has been developed in Pango for GNOME and Firefox, and Qt engine for KDE. Upgraded versions are available at [5].

Sample	Font
తెలుగు	Akshar Unicode (akshar.ttf)
తెలుగు	Code2000 (code2000.ttf)
తెలుగు	Gautami (gautami.ttf)
తెలుగు	Pothana2000 (Pothana2000.ttf)
తెలుగు	Vemana2000 (Vemana.ttf)

Figure 2: Telugu Unicode fonts [4]

Keyboard

Three different keyboard layouts are used by Telugu, Inscript, RTS and WX, of which first two are very popular. RTS and WX are phonetic based [5]. Inscript is standardized through Government of India [6, 7] and is shown in Figure 3.

~	2	1	@	#	4	5	6	7	8	9	0	-	+ ఋ	BS
TAB	Q ట	W ట	E ట	R ట	T ట	Y ట	U ట	I ట	O ట	P ట	{ ట }	ట	\	
CONTROL	A ట	S ట	D ట	F ట	G ట	H ట	J ట	K ట	L ట	: ట	" ట	' ట	RETURN	
SHIFT	Z ట	X ట	C ట	V ట	B ట	N ట	M ట	< ట	> ట	? ట	/ ట	SHIFT		

INSCRIPT OVERLAY FOR TELUGU

Notes :
 - The Macro-keys in the top-row generate :
 (Rakar) (= ట
 ట = ట ట ట = ట ట ట = ట ట ట = ట ట

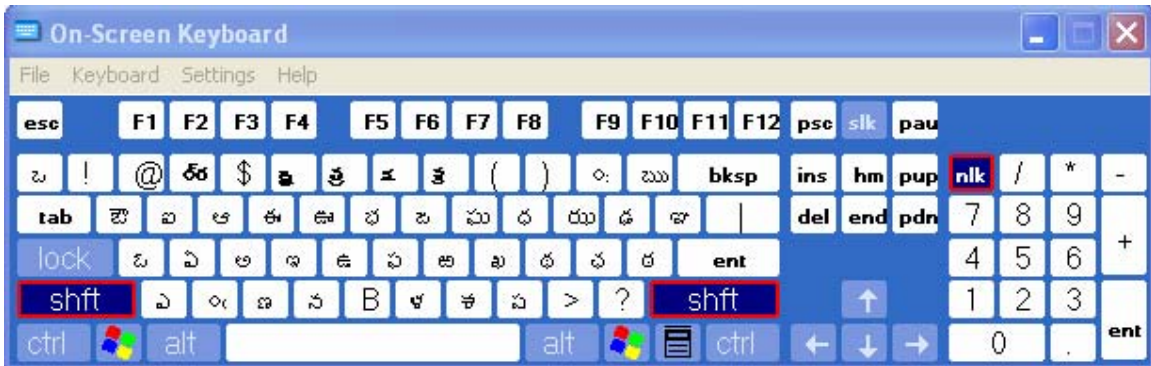
Figure 3: Inscript Telugu Keyboard Layout [7]

Microsoft Platform

Microsoft provides a Telugu keyboard, enabling support in all Microsoft applications. Figure 4 shows the keyboard layout provided by Microsoft.



(a)



(b)

Figure 4: Microsoft On-Screen Keyboard for Telugu: (a) Normal, and (b) Shift State

Linux Platform

All keyboard layouts mentioned above (RTS, Inscript and WX) are available on Linux platform. The support is not built-in and has to be installed [5].

Collation

Telugu locale is defined but not standardized. Basic support is available on Microsoft and Linux platforms e.g. see [10].

Locale

Telugu locale (te_IN) is partially defined and is standardized through posting at IBM ICU and CLDR 1. Further information is available at [5].

Microsoft Platform

Microsoft provides a complete LIP for Telugu. Figure 5 shows a snapshot of the locale settings for Telugu.

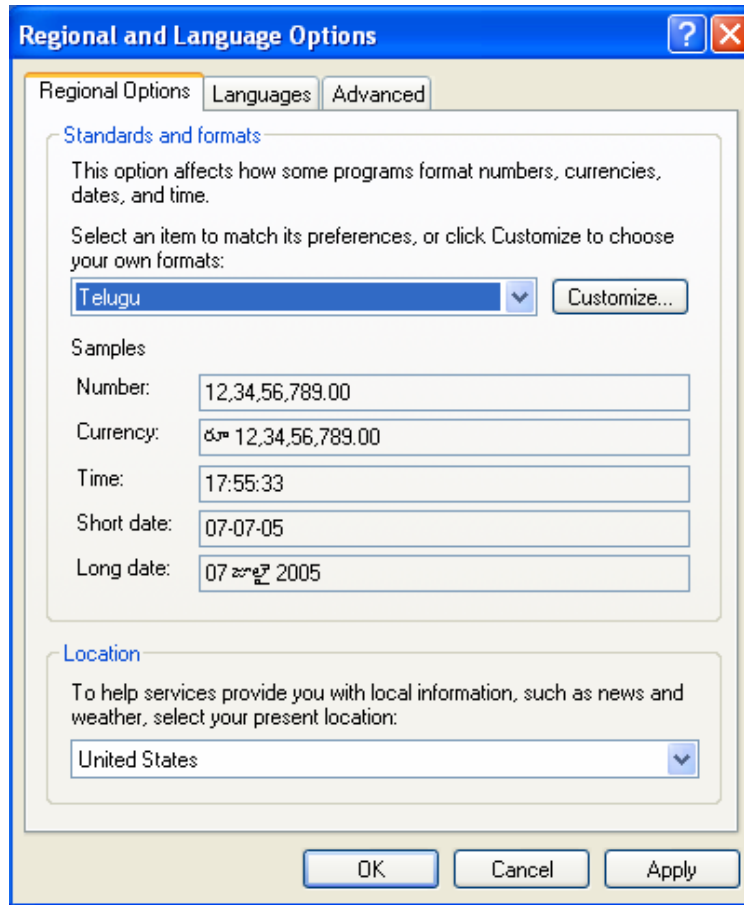


Figure 5: Telugu Locale on Windows XP

Linux Platform

Support for Telugu locale is available in Glib C. The Telugu locale for India has also been defined officially in Red Hat Linux [9]. It is also supported on other open source platforms [10].

Interface Terminology Translation

Interface terminology is not standardized, but is available on many platforms.

Microsoft Platform

Microsoft supports Telugu LIP, which contains Telugu interface for its applications [11].

Linux Platform

A “Native Language Project” has been initiated for localizing Open Office version 2.0. Its team has just started therefore no output from this group has been committed [12]. The India Linux project also includes a language team for localizing Linux in Telugu [14]. However no output of the team has been published. GNOME translation has started for Telugu and about 85% translations of glossary have been accomplished [15]. This team is working independently and has not registered at the www.GNOME.org for translating GNOME 2.8 GUI messages to local languages [15]. A Telugu localization team has also registered on the Mozilla Localization Project (MLP). This team is working on localizing Mozilla version 1.2.1. No outputs of this project are posted [13].

Status of Advanced Applications

Telugu support has been added into the Unicode text editor Yudit [10]. Work has also been done for Telugu Lexicon, open source spell checker Aspell [10] and prototypical work on optical character recognition system. Significant work on Telugu text-to-speech, lexicon and Telugu-Hindi machine translation is also under progress, with working systems already released [16, 17].

References

- [1] http://www.ethnologue.com/show_language.asp?code=tel
- [2] <http://www.omniglot.com/writing/telugu.htm>
- [3] <http://www.unicode.org/charts/PDF/U0C00.pdf>
- [4] <http://salrc.uchicago.edu/resources/fonts/telugufonts.html>
- [5] <http://telugu.sarovar.org/wiki/>
- [6] <http://tdil.mit.gov.in/keyoverlay.htm>
- [7] <http://tdil.mit.gov.in/isciichart.pdf>
- [8] <http://www.microsoft.com/typography/OpenType%20Dev/telugu/intro.mspix>
- [9] http://www.indlinux.org/downloads/locale/Locales/te_IN
- [10] <http://telugu.sarovar.org/wiki/index.php/Status>
- [11] <http://www.bhashaindia.com/downloadsV2/Category.aspx?ID=8>
- [12] <http://te.openoffice.org>
- [13] http://www.mozilla.org/projects/l10n/mlp_status.html
- [14] <http://indlinux.org/lang/>
- [15] <http://l10n-status.GNOME.org/GNOME-2.8/index.html>
- [16] <http://lrc.iiit.net/showfile.php?filename=research/>
- [17] <http://lrc.iiit.net/showfile.php?filename=downloads/>