

Fonts

An Overview of Indic Fonts for T_EX

Anshuman Pandey

1 Introduction

Many scholars and students in the humanities have preferred T_EX over other “word processors” or document preparation systems because of the ease T_EX provides them in typesetting non-Roman scripts, the availability of T_EX fonts of interest to them, and the ability T_EX has in producing well-structured documents.

However, this is not the case amongst Indologists. The lack of Indic fonts for T_EX and the perceived difficulty of typesetting them have often turned Indologists away from using T_EX. Little do they realize that T_EX is *the* foremost tool for developing Indic language/script documents. With an increase over the past few years in the development and availability of Indic language and font packages, the introduction of other fonts and style packages, the flexibility of the L^AT_EX 2_ε system, and the creation of TUGIndia (which may revolutionize the typesetting of Indic scripts) there is now even more reason for Indologists to implement T_EX in their work.

There are roughly thirteen major Indic scripts (Tibetan is included in this list) which are used throughout South Asia to write the major languages and dialects of the region. As of this article all of these major scripts can be typeset with T_EX, the exception being Assamese (see Section 6).

Not only is it fascinating that the major scripts of South Asia can be typeset with T_EX, but the ease with which such a task can be accomplished is itself an amazing feat. Anyone who has ever tried writing a document with multiple non-Roman scripts and diacritic text in an environment other than T_EX understands the complexity of such a task. T_EX takes the user beyond such difficulties by facilitating the implementation of multiple scripts without the hassle of worrying about various fonts and their encodings, manual font switching, and other such hindrances to productivity caused by common “word processors”.

T_EX enables the incorporation of several non-Roman scripts within a single document through transliterated input of the scripts. Indic scripts are based on the phonetic template of the languages they represent, a template which is uniform in both

the Indo-Aryan and Dravidian language families of India. Such uniformity in phonetics is reflected in orthography, which in turn enables all scripts to be transliterated through a single scheme. This uniformity has subsequently been reflected in the transliteration schemes of the Indic language/script packages.

Most packages have their own transliteration scheme, but these schemes are essentially variations on a single scheme, differing merely in the coding of a few vowel, nasal, and retroflex letters. Most of these packages accept input in one of the two primary 7-bit transliteration schemes — ITRANS or Velthuis — or a derivative of one of them. There is also an 8-bit format called CS/CSX which a few of these packages support. CS/CSX is described in further detail in Section 3.

2 The Fonts and Packages

Figure 1 shows examples of the various fonts described in this article. Table 1 lists the sites from which all of the fonts and packages described in this article are available.

3 CS/CSX

CS/CSX (Classical Sanskrit/Classical Sanskrit extended) is the closest thing to an accepted standardization of 8-bit transliteration of Indic scripts. Adopted in 1990 at the 8th World Sanskrit Conference in Vienna, CS/CSX enables Indologists to exchange electronic data in a variety of platform-independent media.

CS/CSX is an encoding convention based on IBM Code Page 437. CS is a basic inventory of diacritic letters which are traditionally used to transliterate Sanskrit written in the Devanagari script. CSX is an extension of this basic inventory to include accented and other characters. Contrary to what the name indicates, the inventory of CS/CSX characters is not limited to Sanskrit, and may be used to transliterate other Indic languages.

Introductory information on CS/CSX is found in an article by Dominik Wujastyk titled *Standardization of Sanskrit for Electronic Data Transfer and Screen Representation* [1]. This document, as well as supporting screen fonts and drivers for DOS-based machines, is available from the INDOLOGY site as well as from CTAN.

Various fonts and packages have been developed which enable T_EX to typeset documents encoded in the CS/CSX convention. These are enumerated below:

cp437csx The file `cp437csx.def` is an input encoding definition file for L^AT_EX 2_ε which enables

क ख ग घ ङ	क ख ग घ ङ	क ख ग घ ङ
Devnag	Devnac	Devnag Pen
क ख ग घ ङ	ক খ গ ঘ ঙ	ક ખ ગ ઘ ઙ
Sanskrit	ItxBengali	ItxGujarati
ਕ ਖ ਗ ਘ ਙ	ਰ ਖ ਗ ਘ ਙ	க ங ச ஞ
Punjabi	Gurmukhi	Washington Tamil
క ఖ గ ఘ ఙ	ക ഖ ഗ ഘ ങ	ක ඛ ග ඝ ඞ
Telugu	Malayalam	Sinhala
କ ଖ ଗ ଘ ଙ	କ ଖ ଗ ଘ ଙ	ಕ ಖ ಗ ಘ ಙ
Konark	Cuttack	AAI Kannada
ཀ ཁ ག ང ཅ	ا ب پ ت ث	+ ٲ ٳ ٴ ٵ
GTibetian	Naskh	Washington Brahmi

Figure 1: Example of Indic Fonts

CS/CSX encoded documents to be typeset in \LaTeX without need for conversion. The file is available from CTAN as part of the `csx` package.

`csxtimes` John Smith has made available the font metrics and virtual fonts of the commonly-used PostScript fonts re-encoded with the CSX encoding. The use of these fonts enable CS/CSX documents to be typeset directly by \TeX without the need for any conversions. This is facilitated through the `csx.def` file which provides the CS/CSX input encoding definitions for standard \LaTeX and for the standard \TeX fonts.

`cscharter` Dominik Wujastyk produced an extension and re-encoding of the Bitstream Charter font according to the CS/CSX convention called ‘CS Charter’. Users should note that ‘CS Charter’ supports *only* the characters of the Classical Sanskrit encoding and does not support the Extended encoding. This font is available from

the INDOLOGY site and is also bundled with the ITRANS package.

`csutopia` Dominik Wujastyk produced an extension and re-encoding of the Adobe Utopia font according to the CS/CSX convention called ‘CS Utopia’. Users should note that ‘CS Utopia’ supports *only* the characters of the Classical Sanskrit encoding and does not support the Extended encoding. This font is available from the INDOLOGY site and is also bundled with the ITRANS package.

`wnri` Thomas Ridgeway developed a package called ‘Washington Roman Indic’ which contained a family of fonts based on Computer Modern and which were encoded with the CS/CSX and other supplementary conventions. I recently revised the package for use with $\LaTeX 2\epsilon$ and added a style file and input encoding definition file which does away with the need for the fonts.

Ridgeway also developed screen fonts and drivers for `wnri` for DOS-based machines. The updated package is available from CTAN.

4 `babel`

Recently Jun Takashima introduced two Indic language modules to the `babel` fold. These packages enable support for Romanized Sanskrit and for Kannada in both the original and Roman scripts. Please refer to Section 5.5 for a description of Takashima's Kannada package.

`skthyph` This module provides the hyphenation patterns for Romanized Sanskrit. As of this article these files are not distributed with the current version of `babel` but will be included in the next release. This module is presently available only from the developer's FTP site.

5 `ITRANS`

The `ITRANS` package developed by Avinash Chopde is the primary component of an on-going project to make the typesetting of all Indic scripts possible by means of a single tool. As of this article, `ITRANS` supports the Bengali, Devanagari, Gujarati, Gurmukhi, Kannada, Tamil, and Telugu scripts. It also supports the 'CS Utopia' diacritic Roman font.

In addition to the default \TeX output, `ITRANS` can produce direct HTML and PostScript output from the input file. `ITRANS` versions for both DOS and UNIX systems are available from the developer's website.

5.1 Bengali

`arogaon` The *AroSgaon* package was developed by Muhammad Masroor Ali as an extension to the 'SonarGaon' HP Laserjet softfont designed by Anisur Rahman. This package contains an `ITRANS` module which provides glyphs for certain characters not available in the original 'SonarGaon' font. Although this package was designed as a supplement to the Bengali support of `ITRANS`, it may be used as an independent package with $\LaTeX 2_{\epsilon}$. The 'SonarGaon' font is not bundled with `arogaon` or `ITRANS`, and must be obtained separately.

`itrans` `ITRANS` provides support for the Bengali script through the 'ItxBengali' PostScript Type 1 font developed by Shrikrishna Patil.

5.2 Devanagari

`devnag` The *Devanagari for \TeX* package developed by Frans Velthuis was the original package for Devanagari. This package uses the 'Devnag' font also developed by Velthuis which con-

tains the characters required to typeset Sanskrit, Hindi, Marathi, and any other languages which use the Devanagari script. The font 'Devnag Pen' developed by Thomas Ridgeway is a variation on 'Devnag' which resembles Devanagari written with an ordinary pen and is bundled with `devnag`.

Dominik Wujastyk, John Smith, myself, and a few others have recently upgraded `devnag` for use with $\LaTeX 2_{\epsilon}$. The package is now NFSS-compliant.

`sanskrit` The *Sanskrit for $\LaTeX 2_{\epsilon}$* package developed by Charles Wikner is an extensive package which enables the typesetting of Devanagari text with Vedic accents and other special characters not supported by the `devnag` package. Numerous options may be set in regard to transliteration, alternate characters, inter-character spacing, and other preferences. Only support for the Sanskrit language is available. The font 'Sanskrit', also developed by Wikner, is bundled with the package. It is a rather complete font in that it contains many complex ligatures and variants which enable excellent typesetting of Devanagari. This package is available from CTAN.

`itrans` Four fonts provide Devanagari support in `ITRANS`: the 'Devnag' and 'Devanagari Pen' fonts described above and two more called 'Devnac' and 'Xdvng'. The 'Devnac' font is a PostScript Type 3 font developed by Avinash Chopde for the `ITRANS` package. 'Devnac' was developed to enable users unfamiliar with \TeX to still produce texts in Devanagari through the "dumb textual interface" mode of `ITRANS`. 'Xdvng' is a PostScript Type 1 font, rendered by Sandeep Sibal from Velthuis's 'Devnag' METAFONT, which enables users to produce direct HTML output of Devanagari text in addition to the standard \TeX and PostScript output.

`jtex` Developed by John Smith, *Jaisalmer \TeX* is a Perl preprocessor which enables Ken Bryant's 'Jaisalmer' font to be used with \TeX . The font is not freely available and must be purchased from Bryant for a nominal fee. More information about this package is available from the URL given in Table 1.

5.3 Gujarati

`itrans` Currently the only package available for Gujarati is `ITRANS`, which uses the 'ItxGujarati'

PostScript Type 1 font developed by Shrikrishna Patil. The ‘ItxGujarati’ font is bundled with ITRANS.

5.4 Gurmukhi

gurmukhi Developed by Amarjit Singh, the *Gurmukhi for T_EX* package enables support for the Gurmukhi script in Plain T_EX. The package includes a preprocessor and the ‘Gurmukhi’ METAFONT also developed by Singh. This package has not been updated since the initial release of **gurmukhi** in October 1995. This package is available from CTAN.

itrans ITRANS supports Gurmukhi through the PostScript font ‘Punjabi’ developed by Hardip Singh Pannu. The font metric file module used by ITRANS for Gurmukhi was developed by me.

5.5 Kannada

kannadatex The *KannadaT_EX* package developed by Jun Takashima provides Kannada support for L^AT_EX and **babel**. The package includes the METAFONT source for the ‘AA Institute Kannada’ font, a preprocessor, and a hyphenation pattern for the Kannada language. The Kannada hyphenation patterns will be included in the next release of **babel**. The package is available from the developer’s FTP site.

kantex The *KanT_EX* package developed by G. S. Jagadeesh and Venkatesh Gopinath enables the Kannada script to be typeset with T_EX. The ‘Kannada’ METAFONT is bundled with **kantex**. This package is available from the developers’ website.

itrans To typeset Kannada ITRANS uses a modified module of the **kantex** package. The ‘Kannada’ METAFONT is bundled with ITRANS.

5.6 Malayalam

malayalam The *Malayalam-T_EX* package was developed by Jeroen Hellingman. The **malayalam** package enables text in to be typeset in both the traditional and reformed Malayalam scripts. The package includes a preprocessor and fonts in the regular, slanted, bold, and calligraphic typefaces. This package also supports the Devanagari and Tamil scripts through the ‘Devnag’ and ‘Washington Tamil’ fonts. A modified version of the Velthuis scheme is used for transliterated input. This package is available from CTAN.

5.7 Oriya

oriyatex The *Oriya-T_EX* package is being developed by Jeroen Hellingman. The **oriyatex** package currently provides the two fonts ‘Cuttack’ and ‘Konark’ designed by Hellingman. The first is a regular face while the second is a calligraphic variation of the former. The preprocessor is still being developed. *Oriya-T_EX* uses a modified version of the Velthuis scheme for transliterated input. The beta-version of this package is available from CTAN.

5.8 Perso-Arabic

In South Asia the Perso-Arabic script is used predominantly to write the Urdu, Sindhi, and Kashmiri languages. Each language has distinct forms for certain letters, but the character shapes are generally identical.

arabtex *ArabT_EX* was developed by Klaus Lagally and functions through a system of style files, eliminating the use for a preprocessor. *ArabT_EX* supports the typesetting of almost all languages whose orthography is based on the Perso-Arabic system. As concerns South Asian orthography *ArabT_EX* currently supports Urdu and Sindhi; the extensions for Kashmiri are being developed.

Currently only the font ‘Naskh’ is supported. ‘Naskh’ is a METAFONT designed by Lagally and is based on the *naskhī* style of Arabic calligraphy. A font based on the *nast‘alīq* style, used predominantly for Urdu, has not yet been developed.

A variety of both 7- and 8-bit input encoding schemes are supported, yet the *ArabT_EX* encoding itself (based on the transliteration scheme of the *Zeitschrift der Deutschen Morgenländischen Gesellschaft*) is the only scheme which fully accommodates the extended Perso-Arabic script used by Indic languages.

This package is available from Lagally’s FTP site as well as from CTAN.

5.9 Sinhalese

sinhala The *Sinhalese T_EX* package has two different versions. The first and original package was developed by Yannis Haralambous. The second version is a modification of the first by Prasad Dharmasena to accommodate a second transliteration scheme called ‘Samanala’. Both versions require the use of the *Indica* preprocessor, bundled with the package. The second version includes a DOS executable of the *Indica*

program. The original package is available from the INDOLOGY site and from CTAN; and the second version from Dharmasena's FTP site.

5.10 Tamil

tamilize *Tamilize* is a preprocessor developed by Thomas Ridgeway for the 'Washington Tamil' METAFONT. This font was designed at the former *Humanities and Arts Computing Center*¹ at the University of Washington for a Tamil-English dictionary project. It is no longer supported by the University of Washington. The **tamilize** package is available from CTAN.

itrans The interface provided by ITRANS for typesetting Tamil makes it easier to use than the *Tamilize* program. ITRANS makes use of the *Washington Tamil* font as well.

5.11 Telugu

telugutex The *TeluguTeX* package was developed Lakshmi V. S. Mukkavilli. It uses the 'Telugu' METAFONT also developed by Mukkavilli. The package is available from CTAN.

itrans ITRANS supports Telugu through a modified module of the **telugutex** package. ITRANS uses the 'Telugu' METAFONT.

5.12 Tibetan

The three Tibetan packages for \TeX are essentially revisions and enhancements of the original package. These three packages are called **sparkes**, **sirlin**, and **steiner** after their developers and are all available from CTAN.

sparkes *Tibetan L^AT_EX* was the original Tibetan package written by Jeff Sparkes. The package includes a preprocessor and the 'Tibetan' font.

sirlin Sam Sirlin fixed minor bugs in the **sparkes** package and provided an improved preprocessor. This package uses a METAFONT developed by Sirlin called 'GTibetan' and requires the **sparkes** package.

steiner *TEXTib* or *Tibetan Transcript Translator*, developed by Beat Steiner, introduces a major overhaul of the first two packages. Steiner created an improved preprocessor enabling support for different input schemes, better handling of ligatures, and more logical typesetting of Sanskrit in the Tibetan script. The **steiner** package uses the 'GTibetan' font and requires the **sirlin** package.

¹ The *Humanities and Arts Computing Center* was replaced by the *Center for Advanced Research in the Arts and Humanities*.

6 What's Next?

1. \TeX support is currently being developed for the following Indic scripts.

Assamese A \LaTeX package for Assamese is currently being developed by Jugal Kalita. Information about the font design and package is available from the URL given in Table 1.

Brahmi The Brahmi script is the ancestor from which all scripts mentioned in this article are derived. This script was employed by, and perhaps even developed under, the Mauryan king Aśoka to have his edicts inscribed during the third century BCE. I am designing a METAFONT of the Brahmi script called 'Washington Brahmi'. The style is an approximation of the early Mauryan Brahmi style and based upon the characters found on the inscriptions at Gīrnār. The font is not yet complete, however a brief description and examples of the font are available from my website.

2. ISO/TC46/SC2/WG12 (the Working Group for the Transliteration of Indic Scripts) is progressing toward a standardized 7- and 8-bit scheme. Perhaps all Indic \TeX packages will support this standard.
3. Unicode? Will there be a need for these packages once Unicode is firmly established?

References

- [1] Wujastyk, Dominik. *Standardization of Romanized Sanskrit for Electronic Data Transfer and Screen Representation* [results of a session held at the 8th World Sanskrit Conference, Vienna, 1990], in *Sesame Bulletin* 4(1), 1991, pp. 27-29. Also available as a PostScript document from CTAN/fonts/csx/csx-doc.ps.

◇ Anshuman Pandey
University of Washington
Department of Asian Languages
and Literature
225 Gowen Hall, Box 353521
Seattle, WA 98195
apandey@u.washington.edu
<http://weber.u.washington.edu/~apandey/>

arabtex	ftp://ftp.informatik.uni-stuttgart.de/pub/arabtex/ CTAN/language/arabtex/
arosgaon	CTAN/language/bengali/arosgn/
assamese	http://www.acsu.buffalo.edu/~talukdar/assam/language/assamlang.html
brahmi	http://weber.u.washington.edu/~apandey/texts/
csx	CTAN/fonts/csx/
csxtimes	ftp://bombay.oriental.cam.ac.uk/pub/john/software/fonts/
devnag	CTAN/language/devanagari/
gurmukhi	CTAN/language/gurmukhi/
itrans	http://www.aczone.com/itrans/
jtex	ftp://bombay.oriental.cam.ac.uk/pub/john/software/jtex/
kannadateX	ftp://ftp.aa.tufs.ac.jp/pub/tool/TeX/languages/kannada/
kantex	http://langmuir.eecs.berkeley.edu/~venkates/
malayalam	CTAN/language/malayalam/
oriyatex	CTAN/language/oriya/
sanskrit	CTAN/language/sanskrit/
sinhala	ftp://ftp.mq.edu.au/home/vsaparam/sinhala_tex/ CTAN/language/sinhala/
sirlin	CTAN/language/tibetan/sirlin/
skthypH	ftp://ftp.aa.tufs.ac.jp/pub/tool/TeX/languages/sanskrit/
sonargaon	http://www.winsite.com/info/pc/win3/fonts/sgaon.zip
sparkes	CTAN/language/tibetan/original/
steiner	CTAN/language/tibetan/steiner/
tamilize	CTAN/language/tamil/tamilize/
telugutex	CTAN/language/telugu/
wnri	CTAN/fonts/wnri/
'bombay'	ftp://bombay.oriental.cam.ac.uk/pub/john/
INDOLOGY	http://www.ucl.ac.uk/~ucgadkw/indology.html
CTAN	ftp://ftp.tex.ac.uk/tex-archive/ ftp://ftp.dante.de/tex-archive/ ftp://ftp.tug.org/tex-archive/

Table 1: Package Sites