

Omega

Typesetting Malayalam using Ω

Alex A.J.

Abstract

This paper explains the installation and usage of a package for typesetting the Malayalam language using the Ω system. This package supports both the Traditional (Old Lipi) and Reformed (New Lipi) Malayalam scripts and provides two font families. The Ω TPs and macros are explained in detail.

Introduction

Typesetting Malayalam with \TeX was first implemented by Mr. Jeroen Hellingman, who developed a package for plain \TeX . In this package, Malayalam text is written in an ASCII transliteration scheme, which in turn was converted into proper \TeX source using a preprocessor written in C. He also developed a primitive `METAFONT` for both the Traditional and Reformed Malayalam scripts.

A new and improved package for \LaTeX was written by Alex A.J. in 2003. Two professional quality fonts were provided along with proper hyphenation and many other improvements. However, it still used the original preprocessor and transliteration scheme from Hellingman's package.

Using a preprocessor has several disadvantages. First of all, there are two source files, a `.mm` file (the \LaTeX source file with transliterated Malayalam text) and a `.tex` file which is produced by the preprocessor. There is no control over hyphenation; the preprocessor simply put discretionary hyphens after every character. It was necessary to use an external editor to delete unwanted hyphens.

The Ω system, developed by John Plaice and Yannis Haralambous, has proven to be a useful solution for Indic language typesetting. It supports the Unicode standard and accepts UTF-8 encoded text as input files. Thus the use of a preprocessor is entirely avoided and Malayalam text can be directly processed by Ω .

Installation and usage

The package is available as a tarball and can be installed using the installation script provided with it. It can also be downloaded from CTAN. The package contains a number of Ω TPs, two font families in Type 1 format, and additional files supporting other commercial fonts etc.

To typeset Malayalam text, you simply use the command `\mal` inside a group, and enter your text using a UTF-8 enabled editor, such as Yudit (<http://yudit.org>).

Ω TPs in detail

This is a technical description of the Ω TPs included in this package. It assumes some knowledge of Ω and Malayalam script.

mal-uni01.otp In this short Ω TP, a ‘soft hyphen’ (`@"OD4F`) is added to every Malayalam syllable, using lines like the following:

```
{consonant}{depA} => \1 \2 @"OD4F ;
{consonant}{depi} => \1 \2 @"OD4F ;
{consonant}{depI} => \1 \2 @"OD4F ;
```

...

The current Unicode standard does not provide slots for Malayalam ‘Cillu letters’. The general practice among developers is to use ZWJ and ZWNJ to differentiate between Cillu letters and normal virama forms. The government of Kerala has proposed assigning code positions 0D7A, 0D7B, 0D7C, 0D7D and 0D7E to the five Cillu letters. In this Ω TP, the ZWJ forms are mapped to the above locations using the following lines:

```
@"OD23 {virama}{zwj} => @"OD7A @"OD4F ;
@"OD28 {virama}{zwj} => @"OD7B @"OD4F ;
...
@"OD33 {virama}{zwj} => @"OD7E @"OD4F ;
```

mal-uni02.otp In this (also short) Ω TP, unwanted hyphen characters are removed in several places: before *anuswara*, *visarga*, Cillu letters, and others. Hyphens at the end of words are also removed.

The final stage In this stage, the Unicode locations are mapped to character positions in the actual fonts. Four Ω TPs are provided with the package enabling the use of many font families. Two of them, **mal-uni2keli.otp** and **mal-uni2rch.otp** provides typesetting in the Malayalam Reformed Script (പുതിയ ലിപി) using the Keli (കേളി) and Rachana (രചന) font families. **mal-uni2oldrch.otp** provides Traditional Malayalam script (പഴയ ലിപി) from the Rachana family. It is explained below in detail. The final Ω TP, **mal-uni2ism.otp**, supports over 35 Malayalam font families from the Indic language software ‘CDAC ISM Publisher’.

mal-uni2rch.otp First of all, the ligature expressions (കൂട്ടക്ഷരങ്ങൾ) are identified and mapped to the corresponding glyph in the font. For example, the line:

```
{ka}{virama}{ka} => "\<183>" ;
```

identifies the sequence ‘ക + ̣ + ക’ and produces the ligature ‘ക്ക’, which is glyph 183 in the Rachana font.

One thing to note is that ligatures coming along with dependent vowels in which a character is placed to the left of the ligature (in the above example: കക്ക, കേക്ക, കൈക്ക, കോക്ക, and കോക്ക) must have entries before the plain ligature (ക്ക).

Similarly, all ligatures that are present in the font file have corresponding lines in this Ω TP.

Next, the dependent vowels are mapped to appropriate locations in the font with the following:

```
{depA} => "n" ;
{depi} => "o" ;
{depI} => "p" ;
{depu} => "q" ;
{depU} => "r" ;
{depr} => "s" ;
...
{consonant}{depe} => <= "t" \1 ;
{consonant}{depE} => <= "u" \1 ;
...
```

The dependent form of ‘ഓ’ gets special treatment similar to the above (for example, ക + ̣ + ഓ = ക്കോ).

Finally, the independent vowels and consonants are mapped to glyphs in the font with lines like the following:

```
{a} => "A" @"OD4F ;
{A} => "B" @"OD4F ;
{i} => "C" @"OD4F ;
{I} => "\<164>" @"OD4F ;
...
{ka} => "I" ;
{kha} => "J" ;
{ga} => "K" ;
{gha} => "L" ;
...
```

The fake Unicode character OD4F, which is used as a soft hyphen, is replaced with a \TeX discretionary hyphen.

```
@"OD4F => "\- " ;
```

Producing UTF-8 Indic \TeX files

One main objective of developing this package was that the user must be able to see Malayalam text as the source file is prepared.

Although OpenOffice supports Malayalam and produces decent UTF-8 text, the character display is very primitive; most of the dependent vowels are shown on the wrong side of the characters they are associated with. The same is true of the standard editor Vim.

A decent solution is the Yudit editor developed by Gaspar Sinai. It uses OpenType fonts for displaying text and comes with many transliteration schemes for almost all Indic languages. Writing a new transliteration scheme is also very easy. This package includes support for using Yudit to prepare Malayalam \TeX source files using a phonetic transliteration scheme.

Fonts

Two font families are included in this package in Type 1 format. The first one is Keli (കേളി) whose character set includes Malayalam Reformed script (New Lipi). The second one, Rachana (രചന) contains more than 900 glyphs in six font files. This font enables typesetting in the Traditional Malayalam script (Old Lipi).

Conclusion

Ω seems to be capable of handling all the complexities involved in typesetting Indic languages. Having UTF-8 support makes it all the easier for developers to write Ω TP’s for any given font. Another advantage is that any UTF-8 compatible editor can be used for creating source files.

We hope this article will help developers to create support for other Indic languages.

This work was supported in part by the \TeX Development Fund (<http://tug.org/tc/devfund>).

◊ Alex A.J.
Lilly Dale
Mukkolakkal, Nedumangad P.O.
Thiruvananthapuram-695541
India
indicTeX (at) gmail.com
<http://sarovar.org/projects/malayalam>