
Rendering Thirukkural in L^AT_EX – an experiment in creating large, bi-lingual documents

S. Parthasarathy
drpartha@gmail.com

kuralengtam3a.tex
Ver.: 20130716c

1 Preamble

This experiment demonstrates the creation of documents with both English and non-English text. We have chosen Tamil as the non-English language, and the “Thirukkural” as the demonstration text for display. Thirukkural¹ may perhaps be considered as the most fundamental and popular work amongst all literary works in Tamil. Tamil itself, is an ancient Dravidian language, very rich in literary and cultural content. Tamil is spoken in the southern part of India (Tamilnadu state), in parts of Sri Lanka, and in many south-Asian countries (e.g. Malaysia).

This is NOT a document ² about Thirukkural. It is a document to explain the use of L^AT_EX in creating bilingual documents. We explore some tricky aspects of L^AT_EX, and the mechanisms of introducing a non-English text (Tamil), in a predominantly English document. Only a few of the 1330 couplets (5 out of 1330 kurals, randomly chosen) are given in this document, since our aim is just to demonstrate a concept in L^AT_EX programming. Beginners, wishing to learn L^AT_EX programming will find this document, as well as a whole lot of similar demonstration documents, on the web, at [3].

¹More details at : <http://en.wikipedia.org/wiki/Tirukkural>

²You can get the L^AT_EX source of this document from the author.

They can use this model, for creating bilingual documents in other languages.

This work is based on an earlier work, which involved the creation of an English-Chinese glossary [1]. Unfortunately, the \LaTeX source of this original work is not available now. It is also very similar to an example given in the “ \LaTeX companion” book (Chapter on supertab – making multipage tabulars) [4].

Handling non-English characters in \LaTeX is not as simple as it appears to be. As we can see below, the process of font encoding of non-English characters is still not frozen, and its rendering in a \LaTeX text is equally difficult. All this gets more complex when we try to combine English text and a non-English text in the same \LaTeX document.

2 About Thirukkural ³

Thirukkural (aka the Kural, Tirukural, Tirukkural) is a Tamil classic, consisting of 1330 couplets or kurals (anglicised plural form of kural). In this document, we use the term “Kural” (with a capital K), to denote the entire work i.e. Thirukkural. We use the term “kural” (with a lower case k), to denote a single couplet. The 1330 couplets are divided into 3 sections and 133 chapters. Each chapter contains 10 couplets. A couplet consists of seven cirs, with four cirs on the first line and three on the second. A cir is a single word, or a combination of more than one Tamil word. For example, Thirukkural is a cir formed by combining the two words Thiru and Kural, i.e. Thiru + Kural = Thirukkural. It is highly venerated by the Tamilians of South India. It was authored by Thiruvalluvar, and is considered to be the first work to focus on ethics, in Dravidian Literature. It does not promote or advocate any specific religion. Each chapter has a specific subject ranging from “ploughing a piece of land” to “ruling a country” to “domestic bliss”. The Kural is so profound in meaning and impact, that many learned scholars have written their own interpretations of the kural [13], [14], and [15].

The Kural has also been translated into many languages. We have used the English translation by P S Sundaram [2] (and have added our own comments). Like always, the translation is not very perfect. It does not capture the nuances of the words in the original. This is particularly normal in trans-

³Most of this material is borrowed from wikipedia

lating poetic works like this, where the original is in a cryptic form.

It is claimed that the Thirukkural, was authored by Thiruvalluvar, a sage-poet, a weaver by profession. But this claim is contested by many (compare with similar controversies about Shakespeare). There is a recent claim by Kanyakumari Historical and Cultural Research Centre (KHCRC) that Thiruvalluvar was a king, who ruled Valluvanadu, in the hilly tracts of Kanyakumari District of Tamil Nadu, India.

His birth and origins is also a subject of speculation. By all estimates, Thiruvalluvar was born at least a quarter of a century before Christ (no one knows the exact date). One legend associates him to Madurai, the ancient capital of the Pandya rulers who vigorously promoted Tamil literature. According to another legend, he was born and lived in Mylapore, a part of present day Chennai city, and traveled to Madurai to submit his work, the Thirukural, for approval of the king (Pandian) and his college of poets. His wife, “Vasuki”, is considered to be an epitome of an ideal wife.

3 Building this document

The first step was to identify and clearly define the expectations and the objectives of the document.

- The kurals will be displayed in a tabular form
- Each kural will be quoted in Tamil, and in English, in a single row, one row per kural.
- The kural will also be explained briefly, in the same language
- The explanation in English may be longer, or shorter, than the explanation in Tamil.
- Maintaining this document should be easy. Addition of the kurals, and modifications of the explanations should be done in a modular fashion, without messing up the table.
- The structure and layout should be maintained across page boundaries (there are 1330 kurals to be quoted and explained). The entries for all kurals should have the same “look-and-feel”. We should be able to

handle 1330 kurals, in two languages. With an estimated 2-3 kurals per page, we are likely to end up with a document of about 450 pages !

- In the Tamil version, the visual structure of the kural i.e. 4 words in the first line, and three words in the second line, should be maintained.
- It should be possible to create cross-references between the explanations of the kurals.

Considering the fact that we are dealing with a huge document consisting of discrete chunks of information (kurals), we adopted a very modular design for the document. The document was made out of these components:

- A set of \LaTeX files, one for each kural in Tamil
- A set of \LaTeX files, one for each kural, in English
- An envelope document incorporating the above

This structure ⁴ makes it easy to modify the text for any kural, without disturbing the rest of the document (and creating errors).

The next challenge was to find a suitable font encoding/decoding scheme (font transcoding), for entering non-English text. There are two approaches for storing/displaying non-English, say Tamil text :

1. Use a phonetics based transliteration scheme. Each letter of the non-English text, is replaced by its phonetic equivalent as a string of English characters. A non-English text would be displayed using the English alphabet, and can be “read” using a phonetic based decoding. Thus the kural #0001 would be written as:

agara mudala ezhutellAm Adi
Bagavan mudatrE ulagu

This does not make much sense to someone who does not know Tamil. It is also illegible to someone who does not know English. One must know both English and Tamil, and the transliteration scheme used, to read this document correctly. Understanding the text is a different matter altogether.

⁴The entire package, consisting of all the above files, is available as a single tarball, from the author.

2. Use a Tamil font transcoding scheme, to store and display the Tamil text. This approach requires one to know Tamil, to read the text. It is also unbelievably complex, as explained in the next paragraph.

Font transcoding itself has two complementary components:

- Storage (encoding) :: Given a non-English character (font) generate a unique set of Unicode characters to store it.
- Display (decoding) :: Given a Unicode character code, generate the original non-English character (glyph) and display it.

We quote V Venkataraman (venkat@tamillinux.org), the author of “Tamil Linux HOWTO” [5]

QUOTE

It can seem like anarchy. There are an unknown number of fonts, each encoded with their own tables, driven by arbitrary keyboard layouts and outputs. In my opinion, Tamil can seriously compete with any other language for maximum number of font tables. Added to this commotion are the dynamic fonts for the web pages, that enable anyone to get away with a non-standard font as long as his pages are viewable.

Adding to all these is the official Indian Standard Code for Information Interchange (ISCII), the Government of India sponsored "unifying" scheme to bring all Indian fonts under the Devanagari umbrella. Anyone familiar with the way the characters are written in Tamil and in Devanagari script will understand the lack of any rationale in this approach.

Needless to say, this is serving to only add to the confusion. A good analysis of this and the unicode for Tamil is once again written by Sivaraj and can be found at : <http://www.tamil.net/people/sivaraj/tamil-unicode.html> . For those not familiar with the Tamil script, a good introduction written by Sivaraj is at <http://www.tamil.net/people/sivaraj/write-tamil.html>.

Let us ignore the anarchy for a moment and get a picture of the frequently used font encodings. There are two main contenders and luckily they will converge soon. The first and most popular one is the Tamil Standard Code for Information Interchange (TSCII), developed by volunteers throughout the world, and the other, TAmil Monolingual (TAM), and TAmil Bilingual (TAB)

encodings, were proposed by the Tamil Nadu Government. Once again, TAM is of limited use in an OS environment and we can safely ignore that. Almost all Linux efforts are in TSCII (Console, KDE, GNOME localizations).

UNQUOTE

The next step, which involves handling such encoded fonts in \LaTeX , is still a matter of great turbulence. The first major step in this direction was “omega” [6]. Omega is an extension of the \TeX typesetting system that uses the Basic Multilingual Plane of Unicode. It was authored by John Plaice and Yannis Haralambous. Although the project seemed very promising from the beginning, the development has been slow and the functionality rather unstable. A separate project, known as Aleph, and led by Giuseppe Bilotta, was started with the goal of stabilizing the code and extending it with e- \TeX functionality [7]. Aleph alone is not being developed any more, but most of its functionality has been integrated into another project called Lua \TeX . Lua \TeX uses Lua as an integrated lightweight programming language. Lua \TeX , it is claimed, will be a successor of both Aleph (which itself superseded Omega) and pdf \TeX . Thus, we seem to be drifting away progressively from traditional \LaTeX . Each such innovation adds one more layer of complexity and uncertainty in the use of \LaTeX for non-English scripts.

A recent alternative, and relatively more robust approach consists of extending the basic \LaTeX to support Unicode characters. This is the approach of Xetex and XeLa \TeX . Xe \TeX [9] is a \TeX typesetting engine using Unicode and supporting modern font technologies such as OpenType or Apple Advanced Typography (AAT). It natively supports Unicode. The input file is assumed to be in UTF-8 encoding by default. Xe \TeX integrates well with \LaTeX and pdf \TeX

3.1 The approach adopted

There are several possible alternatives and variants, each of which has its own set of dependencies and restrictions. We used the following combination of tools for the various steps encountered in this experiment :

- Use yudit [11], a Unicode aware text editor, to create files of Tamil text. The output is a Unicode encoded file of Tamil text. In [10] we can get a list, and details, of text editors which support Unicode encoded characters.

- Use the inbuilt phonetic keymap for mapping the US/English keyboard to Tamil characters, for entering Tamil text.
- Use the above Unicode encoded text file as input to the L^AT_EX document.
- Use xelatex [9] to compile the L^AT_EX document and create a pdf file

3.2 A quick tour of Thirukkural

We now give a bilingual tour of Thirukkural. We choose 5 out of the 1330 kurals. The sequential number of each kural is given at the end of the corresponding kural.

அகர முதல எழுத்தெல்லாம் ஆதி
பகவன் முதற்றே உலகு (0001)

**A begins the alphabet
And GOD, primordial, the
world.**

‘A’ is the first letter in the Tamil alphabet, and in many other languages. Its variants (e.g. aleph, alpha) are also the first letters in many non-Indian languages. ‘A’ is treated by Valluvar, as a metaphor for God.

This very first kural is proof of the secular and neutral nature of the Kural. Without naming any specific god, or religion, Valluvar begins his magnum opus with an invocation.

There is a belief that Adi and Bhagavan indicate respectively, the names of Valluvar’s mother and father.

நில்லாத வற்றை நிலையின என்றுணரும்
புல்லறி வான்மை கடை (0332)

**Great wealth, like a crowd at
a concert,
Gathers and melts.**

This is a very subtle way of confirming the transitory nature of wealth.

சொல்லுதல் யார்க்கும் எளிய அரியவாம்
சொல்லிய வண்ணம் செயல் (0664)

**It is easy for anyone to talk,
But hard to act thereon.**

This is a confirmation of the English saying "better said than done".

அரம்போலும் கூர்மைய ரேனும் மரம்போல்வர்
மக்கட்பண் பில்லா தவர் (0996)

**The world goes on because
of good men –
else it will turn to dust.**

Valluvar drives home the point of having good men in this world, to compensate for the damage done by evil men.

ஊடுதல் காமத்திற் கின்பம் அதற்கின்பம்
கூடி முயங்கப் பெறின் (1330)

**The body held back is love's
joy,
and the joy of that joy embrace
forthcoming.**

This very last kural epitomises the joy of separation from one's lover and the greater joy of getting back again. The Kural begins with A the first letter of the alphabet. It ends with "en" the last letter of the Tamil alphabet.

4 Closing remarks

This document is the first step in a much larger and more ambitious project. Creating a large bilingual document in L^AT_EX has its own difficulties.

This is a L^AT_EX document, created under Linux, using Kile. The pdf version was created by compiling the source document using XeLaTeX. You can get the L^AT_EX source of this document from drpartha@gmail.com. Please mention the Reference Code, and Version code, given at the top of this document.

If you found this article useful, please send a note to drpartha@gmail.com. As always, suggestions and constructive comments are always welcome.

This document is released under a Creative Commons By Attribution - Non Commercial - ShareAlike 3.0 Unported License. See[12]

This document is an attempt to understand how to handle the difficulties involved in typesetting bilingual documents. It can serve as a demonstration for those who would like to create similar documents (in other languages). Or, it can serve as a template for those who would like to create a better (or different) commentary on Thirukkural.

5 Acknowledgments

This project was not easy. I would never have been able to complete this work, without the help of a lot of people. In particular, I would like to thank the following people who were my gurus at one time or another during this project :

- **T Shrinivasan** (tshrinivasan@gmail.com)
- **Aravinda VK** (hallimanearavind@gmail.com)
- **Duvvuri Venugopal** (venugopal_duvvuri@rediffmail.com)
- **Ravi Teja R**(raviteja.r@research.iiit.ac.in)

I apologise, if I have missed out anybody.

I use a kural to express my gratitude :

நன்றி மறப்பது நன்று அன்று நன்றல்லது
அன்றே மறப்பது நன்று (0108)

To forget a good turn is not good, and good it is
To forget at once what is not good

6 About the author

Parthasarathy teaches discrete mathematics, and preaches LaTeX and Linux, to undergraduate students of Computer Science, in India. He is based in Secunderabad, India.

His web site : <http://algolog.tripod.com/nupartha.htm> will give more details about him.

References

- [1] UNU/IIST *PRaCoSy Project Bilingual terminology (English-Chinese)*, Doc. No. jdh/terms/03, June 1994
- [2] Sundaram P.S., *Tiruvalluvar – the Kural*, Pub.: Penguin Books, 1990.
- [3] Parthasarathy S, *Teach yourself L^AT_EX* ,
URL: <http://www.profpartha.webs.com/teachlatex.htm>.
- [4] Goossens M, Mittelbach F, Samarin A, *The L^AT_EX companion*, Pub. Addison Wesley Longman Inc.
- [5] Venkataraman V, *Tamil Linux HOWTO*
URL: <http://www.linux.org/docs/ldp/howto/Tamil-Linux-HOWTO/index.html>
- [6] *Omega – an extension to TeX*
URL : <http://www.ctan.org/tex-archive/systems/omega/>
- [7] *The Aleph system*
URL : <http://www.ctan.org/tex-archive/systems/aleph/>
- [8] *Omega-Tamil*
URL : <http://www.ctan.org/tex-archive/macros/omega/latex/contrib/tamil-omega/>
- [9] *XeTeX*
URL: <http://www.tug.org/xetex/>
- [10] *Unicode and Multilingual Editors and Word Processors for Unix and Linux*
URL : http://www.alanwood.net/unicode/utilities_editors_unix.html

- [11] *yudit* :: *Unicode text editor*
URL : <http://www.yudit.org/>
- [12] Creative Commons By Attribution - NonCommercial - ShareAlike 3.0 Unported License. http://creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US
- [13] *Commentaries on Thirukkural*
URL : <http://algolog.tripod.com/kuralcomment.htm>
- [14] மு. கருணாநிதி, திருக்குறள் - கலைஞர் உரை, திருமகள் நிலையம், சென்னை
- [15] பரிமேலழகர், திருக்குறள் - பரிமேலழகர் உரை, கங்கை புத்தக நிலையம், சென்னை

* * *