

L2/05-081

CHILLING EFFECTS OF THE CHILLU

author: Mahesh T. Pai ; printed 3/14/2005

Encoding Problems in Malayalam.

The vexed problem of chillu encoding has been a serious road block in adoption of the Unicode standard for Malayalam computing. People have been very quick to dismiss the Chillu as mere half a consonant.

I am no linguist, and will, for the time being, define a chillu as a special form of five specific consonants in Malayalam, rendered in a particular way. (**Update**:- see below). Did I say five? Actually, there are seven. The Malayalam language has seen ever so many ``script reforms" at the behest of the printing industry, which are better called ``simplifications" and each reform has resulted in reduction of the number of characters used in the language. The two chillus no longer used were victims of such reforms. I will not blame such reforms because limitations of printing technology, and volume of a print run for an average printed book in the Malayalam Language simply did not justify the existence and costs of a large number (900+) of glyphs in the Malayalam Language. Leaving history to itself, we will now look at the five chillus in frequent use and one chillu whose glyph I have with me right now in the table below.

The first row is the correct unicode encoding for the chillu in the 2nd row. The 2nd row shows the character, and the proper chillu form will show only if you have a unicode encoded font with Malayalam characters installed. The third row shows the accurate rendering of the chillu. If you get the same character in column 2 as the image in column 3, chances are that you are running a Unicode compliant operating system.

Table 1 - The Chillus In Malayalam

Unicode Sequence	Malayalam character as text	Malayalam text as Image
u+0D25 + u+0D4D + u+200D		൩
u+0D30 + u+0D4D + u+200D		രി
u+0D32 + u+0D4D + u+200D		ൽ
u+0D33 + u+0D4D + u+200D		ൾ
u+0D23 + u+0D4D + u+200D		൩
u+0D15 + u+0D4D + u+200D	(Not frequently used)	൩
u+0D2F + u+0D4D + u+200D	(Not in use)	(The glyph is not available with me at the moment)

Note:- I am not sure if the sequence in column 2 is really (consonant) + U+0D4D +

[Navigation](#)
[Writings](#)
[Index](#)
[When standards hold up development](#)
[Photo Gallery](#)
[Snaps](#)
[RMS in Cochin](#)
[Downloads](#)
[Downloads](#)
[Miscellany](#)
[GPG Public Key](#)
[Feedback](#)
[Home Page](#)
[Home Page \(old\)](#)
[Search with Google](#)

U+200D. I am crossing my fingers and hoping that my keyboard has correctly input the U+0D4D + U+200D sequence accurately.

What is the big deal about chillus?

This apparently simple issue is holding up Malayalam software development and localisation efforts. The reason chillus are very fundamental to the Malayalam language, and are what the character `e' is to English. You really cannot write anything more than one (or may be, a few) sentence without a chillu in it. Chillus, or their absence can change meanings of the words too. For example, see the following two sentences. The only difference between them is that a zwj is absent in the second, with the result that the chillu is not formed. Note the change in meaning.

Table 2 -- Chillus and their absence - how words change meanings

The Original Malayalam	English Translation
ആ മനുഷ്യൻ കൊടുക്കുന്നു	That man is giving
ആ മനുഷ്യന് കൊടുക്കുന്നു	Giving to that man

If all rendering engines were very uniform in implementing the consonant + virama + ZWJ sequence, there should not have been any problem at all. But, that is not the case. For example, see [this](#) mail where a leader of one of the localisation teams explains how, and **why** he implements a particular standard, which is obviously wrong. The amount of efforts required to correct this is simply *phenomenal*, and happily, *correctable* at this stage.

The role played by a *wrong interpretation* ([another link](#)) of the Unicode Standard in adding to this confusion is certainly *not trivial*.

That said, here are some examples of chillu usage in Malayalam, complete with translations, so that non-Malayalam speakers can understand how important it is.

Chillu and cons + explicit virama - different meanings

Table 3

1.	വില്ലൻ പോയി	``The villain left''
2.	വില്ലന് പോയി	``A loss for the villain''
3.	അത് വാര്	``Dig it up''
4.	അത് വാർ	``That is a metal strip''
5.	മനുഷ്യന്റെ കാല്	``Man's leg''
6.	മനുഷ്യന്റെ കാൽ	``One fourth of a human being''
7.	കുഞ്ഞ് ആണ്	``It's a child!!''

8.	കുഞ്ഞ് ആൺ	``It is a male child!!"
9.	കോട്ട് അടിച്ചു	``Hit a jackpot"
10.	കോൾ നിലം	A kind of swampy land

Chillu and cons + explicit virama - similar meaning, different contexts

Table 4

1.	ആൾ	Used mostly as a singular suffix (as the ``man" in ``Chairman"
2.	ആള്	ditto.
3.	ആല്	A family of trees. Used in singular, it can refer to a single tree.
4.	പേരാൽ	A variety of the above.
5.	ആര്	``Who?"
6.	ആർ	A river/rivulet, not to be confused with this -- ആര് which means the number six.
7.	പാൽ	Milk
8.	പാല്	Milk
7.	മാൻ	Deer
8.	മാന്	(Suffix, usually to translation of English names ending in -man)

Valid chillu, invalid cons + explicit virama

Table 5

കാൾ	Transliteration of the English ``Karl/Carl"
ഞാൻ	The egotistic ``I"
കുൺ	Mushroom
വാർന്	``Lost" -- typically referring to loss / leakage of a liquid.
കൽ	Prefix for ``stone"

Invalid chillu, valid cons + explicit virama

Table 6

താൻ	Low, low lying, etc.
-----	----------------------

I would like to point out that above tables certainly do not constitute an exhaustive list. When it comes to actual prose, changes in meaning brought about by (absence of) chillus can be quite dramatic. If you have not noticed, I have used only words with one or two characters to illustrate the differences.

My objective in giving above is to illustrate the fact that Malayalam requires three forms of rendering of the Chillu Forming Consonants (CFC). First form requires the CFC is rendered like any other consonant, as the stand alone consonant, with the inherent vowel sound. The second form is without the inherent vowel sound. This form is to rendered as the consonant + virama, sometimes referred as the 'explicit virama' form. The third form of rendering is as the chillu character.

On Page 249 of the current Unicode standards, (accessible [here](#), (pdf link) usage of all these three forms of the CFC is given. The relevant portions are at the bottom of the page.

Chillus - a Grammatically correct definition

After putting up the initial version of this page, I came across G. Sankara Kurup's book, (Bhasha Deepika), (1955, Government of Travancore). This book is published by the Then Government of Travancore, which forms the Southern half of the present day Kerala as a text book for use in High School classes.

G. Sankara Kurup is a popular poet and a well known literary figure, and this book can be considered to be a very -- **the** -- authoritative source.

This is what Sankara Kurup has to say:-
(Rough translation of selected parts)

Varna are indivisible sounds (*dhwani*s). Some *varnas* can be spoken without conjoining with other *varnas*, and their pronunciation can be elongated as much as one's breath allows. Such *varnas*, sixteen in number, are called *swaras*. Other *varnas* are not legible without help from *swaras*. These are called *vyanjanas*.

(At p 63).

Swaras and *vyanjanas* which conjoining with *swaras* are *aksharas*. Mere *vyanjanas* are not *aksharas*; they are mere *varnas*. ... our language has no *varnamala* - only *aksharamala*.

(At p 64).

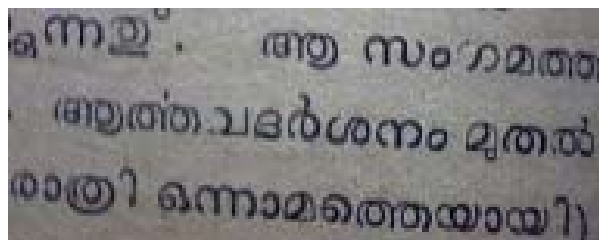
Some *vyanjanas* may be spoken without help of *swaras*. These are not *aksharas*, only *varnnas*. The character (specifies five characters) in the words, (gives some words) are chillus.

(At p 68).

The repha and the RR chillu - a possible problem

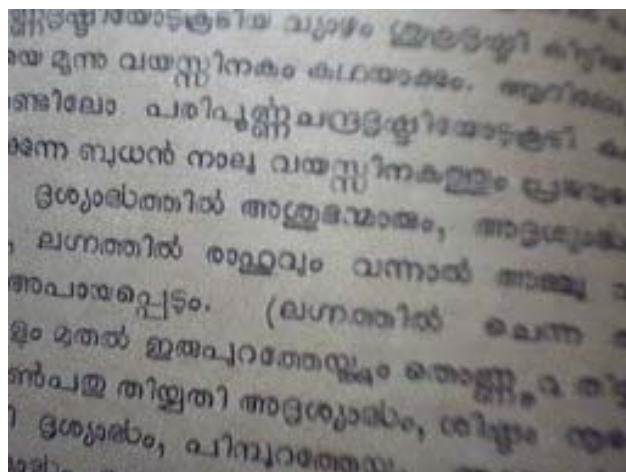
One confusion persisting in the standards formulating bodies is that a repha and the chillu form of RR chillu (the one found on row 2 of table 1, above) are same. I wish to point out a possibility of the repha RR and chillu RR are different. At least, there may be situations when user wants to force a chillu RR when the text is rendered using a font with the repha.

It is true that the RR chillu and the repha have similar pronunciation. For ``technical'' reasons connected to typography, people might have been using the chillu where a repha should be used, but certainly not vice - versa. For example, here are a two screen shots from the book, *Jatakadesam*, with Malayalam commentary by Puthen Pisharath Narayana Pisharody. This book is printed at Kunnamkulamm, in Malayalam era 1111, which corresponds to the year 1935 AD.



The above image shows use of both the repha and chillu in the *same word*.

Here is another image, from the same book, showing consistent use of rephas. The rephas are on the 3rd, 5th and last two lines.



I certainly do not wish to canvass the case that repha and the RR chillu are not different. That is a task I will leave to qualified linguists. My issue is only this - do the unicode standards have a mechanism for differentiating a case where the user wants an explicit repha, rather than a chillu RR? And how will they be distinguished while sorting and searching?

The present scenario requires use of (probably - because usage is not clear) a ``U+0D30 + U+0D4d + U+200D'' sequence to encode the chillu form. How will the user force non-formation of a repha, if chillu and repha are semantically equivalent? Use of ``U+0D30 + U+0D4D + U+200D + U+200D'' for express repha form? Or should he use ``U+0D30 + U+0D4D + U+200D + U+200C'' (zero width non joiner) for the explicit chillu form?

On the other hand, could we not avoid the above confusion by simply assigning a

separate code point for the chillu, and require a zwj for the explicit repha?

Update - repha and the rr chillu

Sankara Kurup (see above), at page 86, describe behaviour of consonants appearing after the rr chillu in compound words. He says that in sometimes, consonants appearing after the rr chillu are to be doubled. eg:- + = . In this section he interchangeably uses both the rr chillu and the vertical (repha) mark above the consonant succeeding the rr chillu. Obviously, both the repha and chillu should be derived from the same character.

Sorting rules and separate code point for chillus

I have illustrated above that the CFC requires to be rendered in three different ways, and meanings in all the three situations are/may be different, depending on the word in which the CFC appears. This requires a bit of thinking over, and I might update this section later on. Some quick thoughts though.

Suppose a sequence, (u+0NXY u+0NXZ) is replaced by a new codepoint, u+0NXA, is there some thing wrong or difficult with having a search algorithm looking for u+0NXY to return both u+0NXY and u+0NXA?

If the different words with the CFC were mere variants (like the word 'bar' 'barred' 'barring' in English), one certainly could have said that encoding the chillu form will impact searching. But, like the words at rows 8 and 9 of Table 3 above, chillu form and explicit repha form of the CFC have entirely different meaning. and the CFC with the inherent vowel is the transliteration for cola (as in Coca Cola). Cola, jackpot, and swampy lands are quite dissimilar. So, we certainly have a hard nut here. But I have not said the last word on this yet.

Conclusion

First, I need to emphasise that I am not a trained linguist, software professional or a grammarian. It just happens that I know more than one language, and also contribute a bit towards software localisation efforts, mostly in the form of translation. And the fact is, I have not really done any serious translation for a loong time, for the simple reason that the project I am a part of, interprets the Unicode standards differently from what I believe the Unicode standards intends them to mean.

Apart from the above examples, [here](#) is another, well articulated view, which I should admit is not entirely acceptable for me, at least in so far as the issues relating to the chillus are concerned. My issue is about rendering of the explicit virama form of a chillu-forming consonant. *None of the non-chillu consonants require a zero width non joiner to form an explicit virama.* Why should the chillus be different?

I have already illustrated the problem of a case where the rr chillu should **not** form a repha in the preceding section.

I am aware that there are various factors to be taken into consideration while encoding a character. Ease of constructing a rendering engine may or may not be one of them. One thing is clear -- for a long time, localisation efforts based on the

Unicode standard in Malayalam is sort of frozen - on the Chillu issue. Whatever efforts in progress have interpreted the Unicode standards differently, in their own way.

A simple clarification on usage, sorting algorithms, and guidelines on the characters to form particular glyphs certainly *might* solve the problem. Advantages of a multi-code point chillu, with having to jump through the complexities of a rendering engine and font encoding issues may be compared to the benefit which may be gained from lesser loads on rendering engines and memory usage(!!!) on (at least) the smaller, embedded devices.

And I have serious doubts about the validity that encoding chillus separately will affect sorting order. First, Unicode is not concerned with sorting order at all. Secondly, at least on POSIX compliant systems, (especially GNU/Linux), sorting is controlled by the LC_COLLATE environmental variable, and it is possible to allocate equal weights to multiple characters or codepoints or combinations of codepoints. In fact, two users on the same system sorting the same file may get different results, if the sorting order specified by their respective LC_COLLATE variables are different. So, sorting is a non-issue, so far as Unicode standards are concerned.

I do not see any impact on existing data, since we are adding a new character.

I shall leave the issue of actually encoding the chillus to be sorted out by experts. I am a simple user, a non-technical one that, and though born and brought up in Kerala, my mother tongue is Konkani - it is not Malayalam. But Malayalam was my second language, and I have done courses on translation of English to Malayalam and vice-versa. I use Malayalam on a daily basis, while talking to my friends, clients, preparing legal documents for them (Land records are in Malayalam). So, in the event that I am wrong somewhere, do feel free to correct me. My contact email address is else where on this page.

Changelog

1. Originally published on 5th September, 2004 to avoid sending a long mail to the Unicode's Indic mailing list..
2. Slight modifications, additions, and updates based on information from Sankara Kurup's book, made on 9th September, 2004. Added the section on sorting. Also added this changelog.
3. 9th September, 2004 - noted that Sankra Kurup was misspelled in a couple of places.

Copyright © Mahesh T. Pai
Verbatim copying and redistribution allowed.

Last modified: Thu Sep 9
23:25:46 IST 2004