

Chillus, Samvrithòkaram and Chandrakkala
– A Problem Which is Not

Mahesh T. Pai

November 17, 2005

Contents

1	Introduction	3
1.1	Background and History	3
1.2	About the Malayalam Script	4
2	Definitions	5
2.1	The Chillu	5
2.2	The Samvrithòkaram	6
3	The Problem	9
3.1	Existing Situation	10
3.2	Repha and R chillu	10
3.3	What Happens After the Virama?	10
3.3.1	Manmohan's Problem	10
3.3.2	Problem of the Begging Vessel - the <i>NPA</i> Conjunct	12
3.4	Whose Chillu is it Anyway?	12
4	Some Possible Solutions	13
4.1	The ZWJ Way	13
4.2	Pros and cons - the ZWJ Way	13
4.3	The other ZWJ Way	14
4.4	Encoding the Chillu	14
4.5	Encoding the Samvrithokaram	14
5	Comments on the Rachana Document	15
5.1	The UTC is not Reforming the Language	15
5.2	Pseudo <i>Samvrithokaram</i> - a Myth?	16
5.3	The Ghost Which is Not	16
5.4	Sorting Weights for the Chillu	17

6	This Document	18
6.1	About the Author	18
6.2	Using this Document	18
6.3	Changelog	18
A	Extracts From Sankara Pillai's <i>Bhasha Deepika</i>	20
B	Summary of Script Reform	25
C	Extracts From a Dictionary	29

Chapter 1

Introduction

1.1. Background and History

The Unicode Consortium had resolved to accept the Chillu characters into the Malayalam range. This decision was later rescinded, apparently on the basis of a paper submitted by the Rachana Akshara Vedi¹.

I had previously submitted to the Unicode Consortium, a paper titled “Chilling Effects of the Chillu”². I had following in mind while writing that paper:-

1. Pure consonants are consonants without an inherent vowel, indicated by a chandrakkala - hence referred to as the explicit virama form.
2. Chillus. are, *linguistically*, pure consonants with a different glyph.
3. *Linguistically*, it is assumed that an explicit virama form of chillu forming consonant (CFC) is equivalent to a chillu form of the consonant.
4. Over a period of time and at least in current usage, the explicit virama form of the chillu forming consonant has acquired a different usage from the chillu form.

My argument in the “Chilling effects” paper centred around examples like:-

Malayalam	Translation
കോൾ നിലം	swampy land
കോള അടിച്ചു	Hit a jackpot
ആ മനുഷ്യൻ കൊടുക്കുന്നു	That man is giving
ആ മനുഷ്യന് കൊടുക്കുന്നു	Giving to that man

Table 1.1: Chillus and their consonants may have different meanings

There are a few more such examples in Table 3 of the “Chilling effects” document.

While I did not explicitly ask for encoding of chillus, I did say this:-

“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

¹ Available as “Chandrakkala. Samvruthokaram. Chillaksharam” by R. Chitrajakumar and N. Gangadharan from <http://www.unicode.org/rick/05210-malayalam.pdf>

² Now Available from http://in.geocities.com/paivakil/writings/chill_effect.html

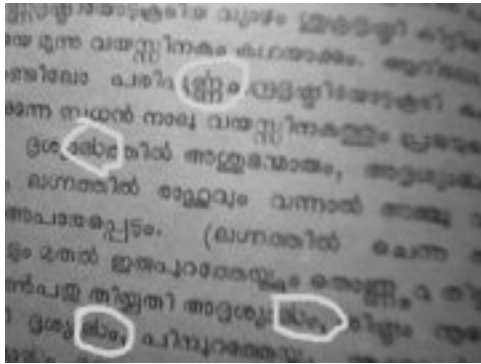


Figure 1.1: Consistent use of repha

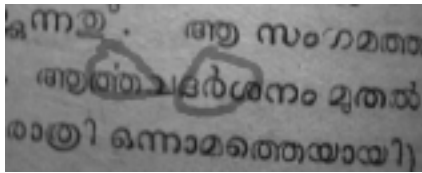


Figure 1.2: Interchangeable use of repha and RR chillu

consonant, as the stand alone consonant, with the inherent vowel sound. The second form is without the inherent vowel sound. This form is to be rendered as the consonant + virama, sometimes referred to as the 'explicit virama' form. The third form of rendering is as the chillu character."

I did not appreciate at that time that there is a fourth form of rendition for the rr chillu – as the *repha* – which is shown as a short vertical line above the next consonant/conjunct. The repha form and chillu form or RR appear to be used interchangeably. See figure 1.1 and 1.2, which use the repha and chillus interchangeably³.

1.2. About the Malayalam Script

Malayalam language has a fairly refined script which is capable of having an exact and unique visual reproduction for every sound in speech. This refinement was severely debilitated by the script reform, whose sole aim was to reduce the number of glyphs required in printing and typing.

The general consensus appears to be that Malayalam script derived from the similar *Vattezuthu* and *kolahezuthu*, both of which were used (apparently) before advent of paper, and the script as we see it now, *Aryaezuthu*, which became prevalent with large scale assimilation of Sanskrit words⁴.

The "script reform", as we use in the context of Malayalam language was solely at the instance of printers, the object apparently being reduction of costs in the printing process⁵.

Going by the text in Figure B.1, though the Malayalam script has only fifty three characters, but there are more than 500 conjuncts, signs and symbols in regular usage⁶.

³ Unfortunately, I have not documented from which book I took these two snaps. Both are from books on astrology; but I do not have the exact title right now.

⁴ See footnote on page 3 of Frohnmeyer's book, and also the image at B.1

⁵ Gopinatha Pillai's book lists Sri Kandathil Varghese Mappilai as spearheading the movement for script reform. He is the original founder of the Malayala Manorama, a news paper still in publication and having the largest circulation. See <http://www.nationmaster.com/encyclopedia/Malayalam-journalism>

⁶ See last four lines of the image.

Chapter 2

Definitions

2.1. The Chillu

G. Sankara Kurup¹ gives the following definition for chillus:-

(Rough translation of selected parts)

Varnas are indivisible sounds (*dhwanis*). 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*, (lists sixteen vowels), are called *swaras*. Other *varnas* are not legible without help from *swaras*. These are called *vyanjanas*².

This draws the distinction between what English calls *vowels* and *consonants*. *Varnas* are either *swaras* or *vyanjanas*. *Swaras* can be pronounced by themselves, (vowels); remaining *varnas* are called *vyanjanas* (consonants).

Swaras and *vyanjanas* which conjoin with *swaras* are *aksharas*. Mere *vyanjanas* are not *aksharas*; they are mere *varnas*. ... our language has no *varnamala* - only *aksharamala*³.

Some *vyanjanas* may be spoken without help of *swaras*. These are not *aksharas*, only *varnnas*. The character (specifies five chillu characters) in the words, (gives some example words) are chillus⁴.

In other words, those *vyanjanas* (consonants) which can be spoken without help from *swaras* (vowels) are not *aksharas* but are mere *varnas* - not deserving the status of alphabets.

The same opinion has been made by A. R. Raja Raja Varma⁵. A rough translation of the first paragraph in the image made available by Cibu would be "Vyanjanas without a samvritha occurring at end of a word can be called chillus. They also have special glyphs (he uses the term *lipi*) too." The relevant part is extracted in figure 2.1.

In plain speak, the underlined part in the extract from Raja Raja Varma means that **the Malayalam script contained some glyphs which were neither conjuncts nor characters**. Understanding this concept forms key to unravelling the mystery of the chillus.

Another author echoing same opinion is L. J. Frohnmeyer⁶.

¹ Bhasha Deepika, (1955, Government of Travancore). This book is published by the erstwhile 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

പദാന്തത്തിൽ സംവൃതംകൂടാതെ നില്ക്കാവുന്ന ഈ വ്യഞ്ജനങ്ങൾക്ക് 'ചില്ലുകൾ' എന്നു പേർ ചെയ്തിരിക്കുന്നു; അവയ്ക്കു പ്രത്യേകം ലിപികളും എഴുതപ്പെട്ടിട്ടുണ്ട്:

Figure 2.1: A. R. Raja Raja Varma's statemnt on chillus.

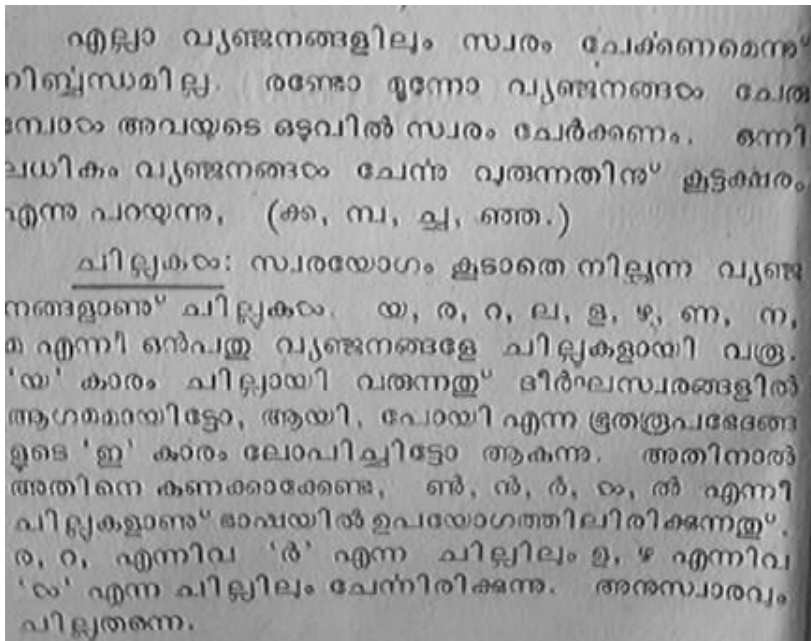


Figure 2.2: Gopinatha Pillai's definition of Chillu

Figure 2.2 shows the definition of chillus given by Gopinatha Pillai⁷. For him, nine consonants can form chillus, of which the chillu of “ya” can be ignored because it occurs in only conjunction with other consonants, or is absorbed by other vowel signs like “i”. He also states that both ra, and Rra form the rr chillu; and the LL chillu is formed by both LLa and zha. The last sentence states that “anuswaram too is a chillu”.

2.2. The Samvrithòkaram

Samvruthòkaram, is more a rule of pronouncing the u vowel; than a spelling rule. Going by the tex tbooks, there are two ways to pronounce the “u” sound in Malayalam. The first is “*vivritha*” or open; and the second is “*samvritha*” or reserved. In writing these are distinguished by the presence of the chandrakkala (virama) mark at the end of th glyph which is to be pronounced “*samvritha*”.

Citing Gundert, Gopinatha Pillai says that *samvrithokaram* is the “*ara ukaram*” – the half u⁸. He points out that there is a stream of thought which says that *samvrithokaram* is actually a half “a” and not half “u”; and that school of thought writes such consonants without the *ukar*; and uses only the *chandrakkala*. Also see Figure 2.3.

literary figure, and this book can be considered to be a very – the – authoritative source. Extracts from this book are in Appendix A

² *Ibid*, see A.1. Words in italics are transliterations of the original Malayalam word.

³ *id.*, At p 64 – see A.2

⁴ *id.*, At p 68 – see A.4

⁵ See <http://photos1.blogger.com/blogger/1048/152/1600/anuswaaram.jpg> cited by Cibu at <http://varamozhi.blogspot.com/2005/06/unicode-chillu-desired-homogeneity.html>

⁶ Available from <http://www.unicode.org/~emuller/iwg/sources/pl14713.f7.1979/p04.png>. See § 5 on that page. Frohnmeyer also seems to conclude that the anuswara is the chillu form of ma.

⁷ Gopinatha Pillai, *Malayala Vyakaranavum Rachanayum* (Malayalam Grammar and Writing), 4th ed., (1983), H & C Stores, Kunnamkulam. The quoted part appears to be a verbatim “lift” from Raja Raja Varma’s work cited by Cibu, *Supra*.

⁸ *Ibid.*, p. 13.

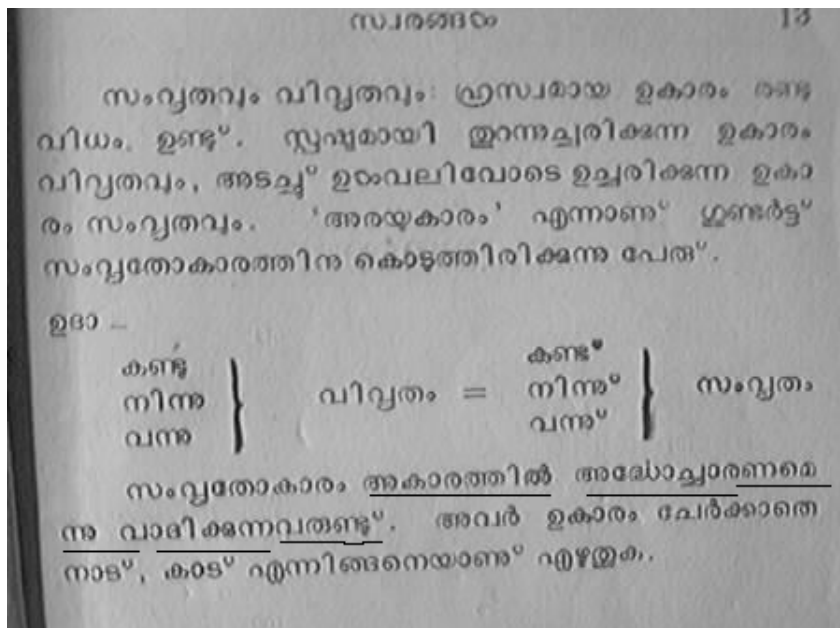


Figure 2.3: Defining Samvrithokaram (From Gopinatha Pillai's book

Frohnmeier⁹ states that the u + chandrakkala indicates half form of the u occurring at the end of a vowel.

Sankara Kurup's definition of samvrithokaram is given in Figures 2.4 and 2.5.

He asserts that when an ukaram is to be pronounced in its samvrithokaram form, the writer should put the chandrakkala above the vowel with the ukar. He neither specifies when the samvrithokaram should be used; nor does he give any rules to distinguish between the two forms of ukars.

At any rate, none of the above authors indicate that the samvrithokaram has anything to do with pure form of a consonant; or a chillu forming consonant, for that matter.

Samvrithokaram **is a variant of the ukar** – the vowel sign represented by unicode character U+0D41 – and for encoding purposes, is to be represented as:-

consonant + U+0D41 + U+0D4D

If people do not use an ukar before the chandrakkala, it is because **they do not believe in the existence of an samvrithokaram** - something explained in the last sentence of Figure 2.3. It is a mistake to state that it is the script reform which resulted in the abolition (or disuse) of the samvrithokaram by use of a chandrakkala, with the user having to infer a samvrithokaram when a word / consonant ended with an explicit virama.

⁹ http://www.unicode.org/~emuller/iwg/sources/pl14713.f7_1979/p08.png

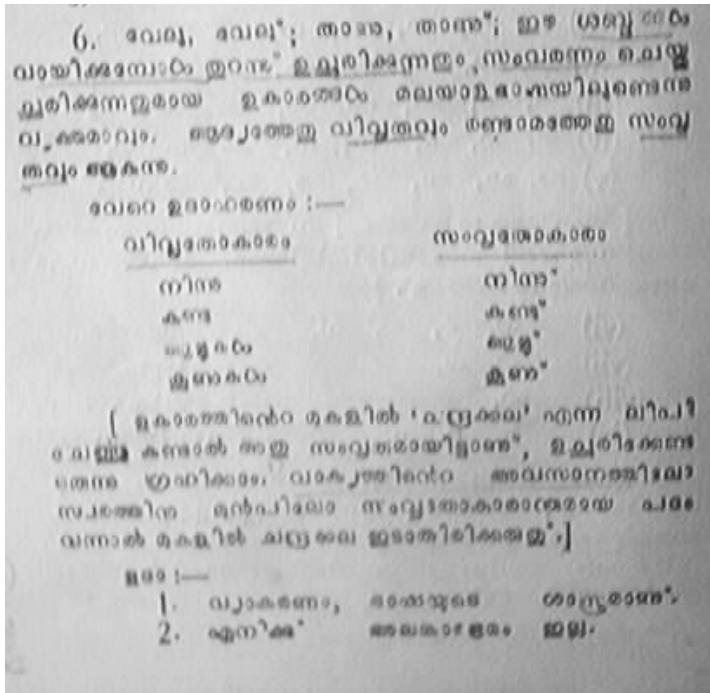


Figure 2.4: Defining Samvrithokaram - by G. Sankara Pillai

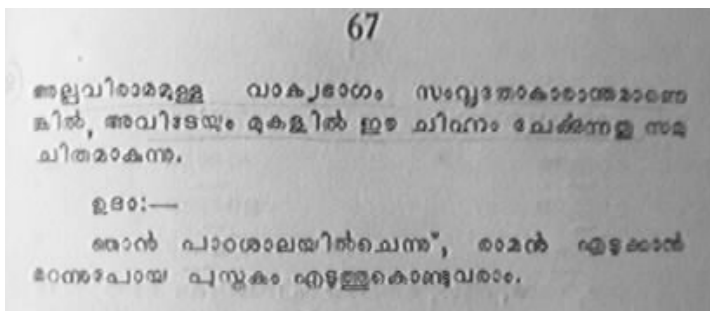


Figure 2.5: Defining Samvrithokaram (Continuation of Figure 2.4)

Chapter 3

The Problem

I have already pointed above that the CFC requires to be rendered in *three* different ways (four for the rr chillu's repha form). These are:-

1. As a consonant with the inherent A vowel - which has to be rendered and encoded like any other consonant.
2. The explicit virama form - where the CFC is followed by a virama (chandakkala) - indicating that the CFC is without an inherent vowel; which again has to be encoded and rendered in same way as any other consonant.
3. The chillu form, which is *apparently* linguistically equivalent of the above.
4. The special case of the repha form for the RR chillu, which has to be encoded such that it turns up in collation and searches without any changes in the search algorithm.

This can be illustrated in the following table:-

Chillu	Base Character	Explicit virama	Chillu	Repha
rr/RR	ര	ര്	രി	See figure 1.1, above
L	ല	ല്	ലി	(Not Applicable)
LL/zha	ള	ള്	ളി	(Not Applicable)
N	ന	ന്	നി	(Not Applicable)
NN	ണ	ണ്	ണി	(Not Applicable)
KK	ക	ക്	കി	(Not Applicable)

Table 3.1: There is a need to render chillus forming consonants in up to 4 different ways

What is required is a uniform sequence and rules to represent and properly render the character / conjunct / glyph (whatever people wish to call it) accurately in all the above four situations.

The solution also has to ensure that there are no difficulties in text manipulation – that is, while spell checking, searching and collation.

3.1. Existing Situation

The Unicode solution was to use the CFC + virama + zwj combination to render chillus¹.

But, this was being interpreted by different people in different ways.

Microsoft, (and consequently, the Open-Type Specification) for example, adopted the view that chillus were mere halant form of certain consonants².

Contrary interpretations too abound³.

3.2. Repha and R chillu

It has been already pointed out that both the r chillu and repha are used interchangeably. Figure 1.1 on page 4, above shows a consistent use of the *repha* while figure 1.2 shows inconsistent usage of the same mark.

If *cons + virama + zwj* is to represent a chillu, and if chillu r is to be shown as both repha and r chillu, how will a rendering engine choose between the repha and r chillu, when encountered with the R + virama + zwj sequence?

Cibu discusses this in more detail on his page⁴.

3.3. What Happens After the Virama?

3.3.1. Manmohan's Problem

Not only *Manmohan*, but also *Velmurugan*, *Mohandas* have it.

But, Man Mohan, Vel Murugan and Mohan Das will not face that problem.

The text in italics in the first line above are liable to form conjuncts, and the first consonant (in bold italics) in those words are to be rendered as chillus. The words themselves are names of persons, and are compound words. When written with a space, as in the second line, the second part is to be interpreted as a second name / surname. The normal user expects that the explicit virama (the chillus, in these specific examples) appear as such when the words are spelt singly. In the above examples, the words in italics should not form conjuncts and therefore render separately, since they are pronounced as distinct syllables. But the same sequence of vowelless consonant + consonant are common in other words, and the user expects the second case to form conjunct⁵.

While the Malayalam language has only 53 basic characters, approximately 68 non-numeral characters have been encoded as against at least 123 glyphs required even under the reformed script⁶. The traditional rule for forming conjuncts, whether they be the “dead” consonant, doubled (strong) consonant, half/partial, or complex consonants has been this:-

C1 + virama + C2

¹ See page 291 of Unicode's standards - <http://www.unicode.org/versions/Unicode4.0.0/ch09.pdf>

² See <http://www.microsoft.com/typography/otfntdev/indicot/features.htm>. However, I am told that when it comes to actual implementation, Malayalam fonts released by Microsoft use the CFC + chandrakkala + ZWJ sequence.

³ See for example, Vinod's views at <http://lists.sarovar.org/pipermail/smc-discuss/2004-May/000483.html> and the interpretation by Sajith V. K. at <http://lists.sarovar.org/pipermail/smc-discuss/2004-May/000486.html>

⁴ <http://varamozhi.blogspot.com/2005/07/unicode-malayalam-eyelash-repha.html>

⁵ One of the major drawbacks of script reform was that the user was left guessing as to what exactly the author / printer intended in such cases – that is, the user had to deduct, from the context in which a character sequence appears, whether the sequence was a single syllable or not.

⁶ See page 40 of April 2002 issue of ViswaBharati, the journal of TDIL, available from <http://tdil.mit.gov.in/news.htm>

where C1 and C2 could possibly be same consonant.

Practically, the sequence “consonant + virama + a consonant” acts as a directive to substitute the code sequence for a conjunct glyph and/or ligate, if the subsequent consonant is capable of such ligation.

Sometimes, especially under the traditional script, more than two consonants could form conjuncts. In such situations, when same character sequence is represented under the new script, the user has to infer from the context whether the sequence is to be pronounced as a single syllable, or whether the virama represents that the characters separated by the virama are to be pronounced separately⁷.

Sometimes, the sequence would be consonant1 + virama + consonant2 + vowel modifier, where the vowel modifier is to apply to the second consonant. Under the present rule, the vowel modifier will apply to the whole sequence of consonant1 + virama + consonant2 and split vowel sign will appear to both sides of the whole sequence, and signs to appear on the left of consonants may appear on the left of consonant1 rather than on the left of consonant2.

But, there are situations where the author and reader of a document expects that a consonant immediately following a consonant with explicit virama *does not* ligate or form conjunct.

Consider the following word, which is the name of a panchayat in Trichur district.

Wilvattom

U+0D35 + U+0D3F + chillu ll + U+0D35 + U+0D14 + U+0D4D + U+0D14

The right way to pronounce this word is

Wil/vattom

where the “/” represents the syllable break; hence (chillu ll + U+0D35) sequence should display the U+0D35 va character separately.

But, occurrence of the virama in the chillu sequence will cause the va to ligate, and the resultant rendition will be pronounced as:-

Wilva/ttom

Another example of this can be found in the third and fourth lines on the table at the bottom of the page Cibu gives as an example ⁸ page from A. R. Raja Raja Varma’s book.

Obviously, there is an encoding issue here and the question is how to identify situations where the post-base ligature of va should be formed (like in “vilwa”) and it should not be formed (like in “ayurveda”)⁹.

Also see this¹⁰ on Cibu’s page.

This issue is related with the chillu question because, under the present solution, the chillu is a code sequence with the last code being the chandrakkala. Naturally, the rendering engine will, if a corresponding glyph is available, ligate / conjoin the succeeding consonant with the chillu, even in situations where such ligating is not desired.

Non-Joiners could be used to solve the issue; but that is certain to cause confusion in minds of people. The common expects that chillus stand alone, and not to form conjunct with the next conjunct.

At the same time, where a non-chillu forming consonant is expected to display an explicit virama, (and not conjoin with the subsequent consonant), the user will naturally include a non joiner. One of the solutions to this problem would be encode a chillu without a joiner character.

⁷This issue is similar to what is sought to be dealt with in <http://www.unicode.org/review/pr-37.pdf>. Thanks to Cibu C. J. for pointing this out. However, the issue here is about how to *prevent* ligating.

⁸ See the image <http://photos1.blogger.com/blogger/1048/152/1600/anuswaaram.jpg> on Cibu’s <http://varamozhi.blogspot.com/2005/06/unicode-chillu-desired-homogeneity.html> page.

⁹ Parvam and Parvati are two other words which immediately come to mind, where the chillu is immediately followed by another consonant; and first sequence requires a post-base va ligature; while the second does not

¹⁰ <http://varamozhi.blogspot.com/2005/06/unicode-chillu-challenge.html> — especially the portion on man.vikshObham meaning “explosion of mind” and manvikshObham meaning “fury of a lady”. Other possible candidates for this difficult situation are words “ayur/veda” and “soft/ware”.

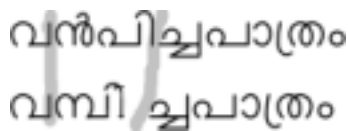


Figure 3.1: How will a spell checker distinguish between the two lines?

3.3.2. Problem of the Begging Vessel - the *NPA* Conjunct

Consider the two compound words in figure 3.1. Only difference between the two is in the encoding of the text between the two horizontal lines. The first word has the sequence, u+0D28 + u+0D4D + zwj + u+0D2A; and means “the big begging vessel”.

The second word has the sequence u+0D28 + u+0D4D + u+0D2A and means “the big vessel”.

Since the zwj is to be ignored, how will a grammar checker (when it actually becomes functional for Malayalam), distinguish between the two? The only difference between the two words is, going by the current Unicode prescription, (and ought to be, according to Rachana), the presence of a zwj after the virama which follows n.

I trust that it will be noticed that this section deals with sorting and searching in a situation where multiple visual representations of what linguists like the Rachana team insist are equivalent are involved; and the different visual representation has a different, dissimilar meaning and sorting value.

The previous section dealt with the rendering issues in similar situations.

To put it in general terms, there are sequences where a “chillu-character” sequence exists; and so may the sequences “conjunct of CFC+character” exist. While rendering engines *may* distinguish between the two by presence of the zwj; the zw[n]j are characters to be ignored.

3.4. Whose Chillu is it Anyway?

At least in case of certain chillus, there is evidence that more than one basic consonant forms the chillu.

This problems exists in case of the RR, ll, and LL chillus. The RR chillu is said to be formed by both U+0D30 and U+0D31, the ll chillu by U+0D32 and U+0D24 and the LL chillu by U+0D33 and U+0D34 respectively¹¹.

This dispute creates the following problems.

1. Data entry - should the second consonant be allowed to form a chillu?
2. Sorting - should the chillu be treated as half form of first or second consonant?

In my opinion, all statements are almost like “chillu (*CH*) is **usually** formed by consonant (*C*); but some authors are of view that consonant (*A*) too forms a chillu”.

This is not a critical issue, and the data entry problem would vanish if the chillu gets a separate code point; and the solution to collation issue would be to define the chillu as the vowel-less form of consonant (*C*) and ignore the chillu form of (*A*).

¹¹ Also see figure C.2 in appendix C. On a side note, I have often noticed that Tamilians pronounce u+0D34 as u+0D33. I have always attributed this to the inability to pronounce u+0D34. But I have no proof whatever for Tamilians verbal skill.

Chapter 4

Some Possible Solutions

These solutions are not intended to be exhaustive.

4.1. The ZWJ Way

One way was to clarify the exist text of the UTS and specify that explicit virama form was to be formed by

consonant + U+0D4D

and the chillu was to be formed by

consonant + U+0D4D + ZWJ

4.2. Pros and cons - the ZWJ Way

Cons

1. Problems with the r chillu, discribed by Cibu¹
2. Inability to force [non]formation of the *repha* in case of the r chillu.
3. Fails to take into consideration the linguistic differences between chillus and explicit virama forms of chillu forming consonants (see table 1.1).
4. The Zero Width Joiner and Zero Width Non joiners are used in non-indic scripts only as font directives. Each application will have to add Support for different linguistic functions of the ZW[N]J for each Indic script.
5. Inability to force (non) formation of a conjunct with succeding consonant².

Pros

1. This scheme is light on searching and collation algorithms etc.
2. Linguistic equivalence maintained in encoding also.

¹ see <http://varamozhi.blogspot.com/2005/06/unicode-chillu-rachana-document.html>

² See the issues relating post-base form of the U+0D35 by Cibu in <http://varamozhi.blogspot.com/2005/06/unicode-chillu-challenge.html>

4.3. The other ZWJ Way

This would involve encoding the chillu as:-

consonant + ZWJ + U+0D4D

The ZWJ “joins” the virama with the CFC; the succeeding consonant is still free to form a conjunct with the CFC, something which can be explicitly prevented by presence of a ZWNJ.

The user will still be confounded with having to insert a zwnj when he does not want the CFC to form conjunct with the succeeding consonant (because a conjunct will be caused by the presence of a virama at the end of the chillu).

The advantage of this solution is that the lexical equivalence between the explicit virama form and chillu form of the CFC is maintained, because the ZWJ is ignored for lexical ordering.

4.4. Encoding the Chillu

1. Simplicity in sequencing
2. Can force the repha using the simpler U+0DRX + zwj where U+0DRX is the code for the RR chilu.
3. Considers the linguistic differences in usage
4. takes away excess load from the virama/chandrakkala
5. No chandrakkala involved - forming ligature / conjunct with succeeding consonant would require yet another zwj / chandrakkala

If this method is adopted, we will have to specify, for sake of uniformity, that when a conjunct is expected to form, like the (cons1 + virama + cons2) sequence, usage of the the encoded chillu as cons1 is invalid. This because all other conjuncts are expected to form with the sequence consonant + virama + consonant; but the encoded chillu already assumes the virama in it.

4.5. Encoding the Samvrithokaram

The vowel U, when conjoined with any consonant, (the *ukàra* might be pronounced freely - “vivriṭha”, or reserved - “samvritha”. The former is the general pronounced, and is encoded as (consonant + U+0D41). Since “samvritha” is a quality of the *ukàra*, the chandrakkala, which is used to represent the samvrithokara, should be appended to the ukara; and not to the preceding consonant.

Hence, the proper sequence should be (consonant + U+0D41 + U+0D4D).

While I did see an opinion on the mailing list that this could be (consonant + U+0D4D + U+0D41), lexically, we will end up with the same result as the ordinary ukara; since the ordinary ukara is pure consonant (ie., cons + U+0D4D) conjoined with U+0D41.

The samvrithokara is sure to throw up issues in search and collation, we are really helpless; because the two methods of writing the samvrithokara are because one stream of thought believes that the samvrithokara is in fact the half form of akara, rather than ukara. The situation is akin to the “photo” – “foto” dichotomy in the English language.

Chapter 5

Comments on the Rachana Document

5.1. The UTC is not Reforming the Language

The Rachana team has a serious problem understanding the role of the Unicode Consortium. They seem to suspect that the Unicode Consortium will confer the status of characters to the chillus. The authors' confusion about the difference between the term “character” in the linguistic sense, and “encoding” in the standards sense is borne out by sentences like:-

In Malayalam, it is the sort order which provides the rationale for distinctions in an encoding. This is because, it is the sort order which assigns a value (or weight). Only when two characters or sequences differ in value (or weight) at the primary level, is there a need to differentiate them at the encoding level. Actual pronunciation of the character does not imply a difference in sorting value.

(at page 2, first paragraph)

and

“Giving specific code points to the chillu divorces the chillu characters from their base (or underlying characters), which is not logical.”

and

“Divorcing the chillus from their underlying characters also causes more trouble, due to the extra rules that will be required in the *already complicated implementations of spell checkers and grammar checkers, not to mention collation rules.*¹”

(page 2, section 4 of the document)

Obviously, there is little understanding of the Unicode standards, since codepoints have nothing to do with either sorting or spell checking. In fact only contribution of the Rachana document is confusion. In saying:-

This pseudo-*samvrithokaram* also paved the way for all the problems of encoding chillaksharam. *The real problem is not in distinguishing chillaksharam from samvrithokaram, rather it is in the distinction between samvrithokaram and chandrakkala.*

This obviously equates what are dissimilar. There is a confusion between two separate issues - the difficulty of encoding chillus and the totally unrelated question of *samvrithokaram*. They have successfully identified ghosts where there are none.

¹ (Emphasis added). The italicised part of the penultimate quote (the part about spell and grammar checkers) shows that the authors of the Rachana document do not have any grasp of the technical aspects in implementing the Unicode standards.

കുട	forest
കുട	Japanese quail
കുട	(meaningless)

Figure 5.1: The first word can take *samvrithokaram* form

5.2. Pseudo *Samvrithokaram* - a Myth?

The authors seem to be labouring under a severe misconception about what they describe as the “pseudo *samvrithokaram*” – defined as “usage of chandrakkala as a substitute for the *samvrithokaram* in the Typewriter script”. This definition misunderstands the concept of *samvrithokaram*. I hope that concept of *samvrithokaram* is already clear from the discussion in previous chapters (especially, Chapter 2).

I suspect that Rachana Akshara VEDI put up their document about a non-issue because some of the examples I gave in my “Chilling Effects” document would be written by some users using *samvrithokaram*. But the fact remains that at least since 100 years, there are two opinions about existence of (or the need for) the *samvrithokaram*. Obviously, the dispute of whether or not to use *samvrithokaram* is a matter of style and / or usage, and never is, nor can be dealt with, at the encoding level.

The Rachana team fails to note that **all** consonants — not merely the CFCs – may be pronounced with the *samvrithokaram* form. See figure 5.1. In figure 5.1, the first word, with the explicit virama means forest. This word may also be spelt in the *samvrithokaram* form, which is not shown. When the last character is spelt with the inherent a vowel, the word means “Japanese quail”.

Rachana also assumes that the practise of using the chandrakkala as a “substitute for the *samvrithokaram*, sneaked into the language via the Typewriter script”. The fact is that (as far as I know) the typewriter allows the user to input the *consonant + u + chandrakkala* sequence; and any user could easily type in a *samvrithokaram* on a typewriter. However, at page 3 of their document, they dismiss the distinction between *avaN* (chillu form) and *avan* (explicit virama form) as a creature of the typewriter script.

While I have not actually read the report of Script reform committee, a summary of its recommendations is available². Probably, the committee did recommend discontinuation of the use of the *samvrithokaram*. Frohnmeyer’s text noted earlier explicitly makes it clear that the practise in Travancore is to put a chandrakkala at the end of a vowel with u. Thus, there is a divergent practise even amongst users of Malayalam regarding the [non] use of the *samvrithokaram*. Rachana’s comments on *samvrithokaram*, seem to suggest that only way for writing the explicit virama form of all characters is the *samvrithokaram* form. If accepted, the argument would result in all users being forced to adopt the *samvrithokaram* form; which is not desirable.

Finally, how has Rachana’s erudition on *samvrithokaram* thrown any light on the issue of how to distinguish between the explicit virama form of a CFC (CFC + U+0D4D) and the chillu?

5.3. The Ghost Which is Not

While the Rachana document fails to make out any relationship between the chillus and *samvrithokaram*, starting from page 4, they complain, with well laid out illustrations, that what are not equals, are, well, not equals.

All arguments (including the sorting issues) in the document centre around the *samvrithokaram* issue; nothing about chillus.

The argument being that all cases of the sequence *consonant + U+0D4D* (that is, consonant + virama) should be considered equivalent to the sequence *consonant + U+0D41 + U+0D4D* (consonant + ukar + virama). This distinction is not a result of simplification of scripts; it is a result of existence of two schools of writing, one of which considers that *samvrithokaram* is half form of “a”, while the other school considers it to be the half form of the ukara vowel. an issue already discussed and explained.

² Malayala Vyakaranavum Rachanayum, by Prof. Vattaparambil Gopinatha Pillai, 4th ed., H & C Stores, Kunnankulam – extracts in appendix B

At any rate, the issues of sorting and spell checking the explicit virama form when the user intends a *samvrithokaram* form is an issue to be considered similar to “photo - foto” dichotomy in English.

5.4. Sorting Weights for the Chillu

This again is a non-issue, since it is simply a question of defining that sequence **CFC + chandrakkala** is equal to the (newly encoded) chillus.

Chapter 6

This Document

6.1. About the Author

The author, Mahesh T. Pai, is a lawyer by profession. While his mother tongue is Konkani, he is more fluent in English and Malayalam, and is also familiar with the Devanagari script. His interest in language issues stem from an intense desire to have the people of Kerala use the computer in their own language.

I wish to point out that I am supremely unqualified to make a linguistic analysis of Malayalam. For one, I am neither a software developer; nor a linguist. I am a lawyer by profession; and non-native speaker of Malayalam to boot. My mother tongue is Konkani, and I speak Konkani at home. But I was born and brought up in Kerala, I speak more Malayalam than Konkani. I studied Malayalam in school and college; and have to use Malayalam for my professional work. Most title deeds are in Malayalam. The executive branch of the government prefers to conduct correspondence in Malayalam. Language of the judiciary in Kerala is English, I need to communicate with my clients in Malayalam; and even in courts, witnesses are examined in Malayalam; and several judges, especially in the lower rungs of the judiciary are more comfortable when arguments are advanced in Malayalam. As such, I need to use the language in my day to day life; and know how the “man on the street” uses it.

6.2. Using this Document

This document is released for the explicit purpose of placing information before the [Unicode Consortium](#). The Unicode consortium is free to republish all or any part of this document without any restrictions.

Any other person or entity is free to redistribute unmodified versions of this document without any restrictions or royalty provided this notice is included.

You are free to quote from this document without any restrictions.

6.3. Changelog

Original version published on September 12, 2005

November 17, 2005:–

1. Converted all images to grayscale and scaled down the size - this saves plenty of bytes in the pdf.
2. Fixed typos and grammar errors in several places.
3. Included notes to table [3.1](#) about the RR and LL chillu.

4. Slight modifications to section 3.3 – thanks to Cibu for inputs on this and several other issues.
5. Modifications in Chapter 5

Appendix A

Extracts From Sankara Pillai's *Bhasha Deepika*

The following images are extracted from Sankara Pillais *Bhasha Deepika*

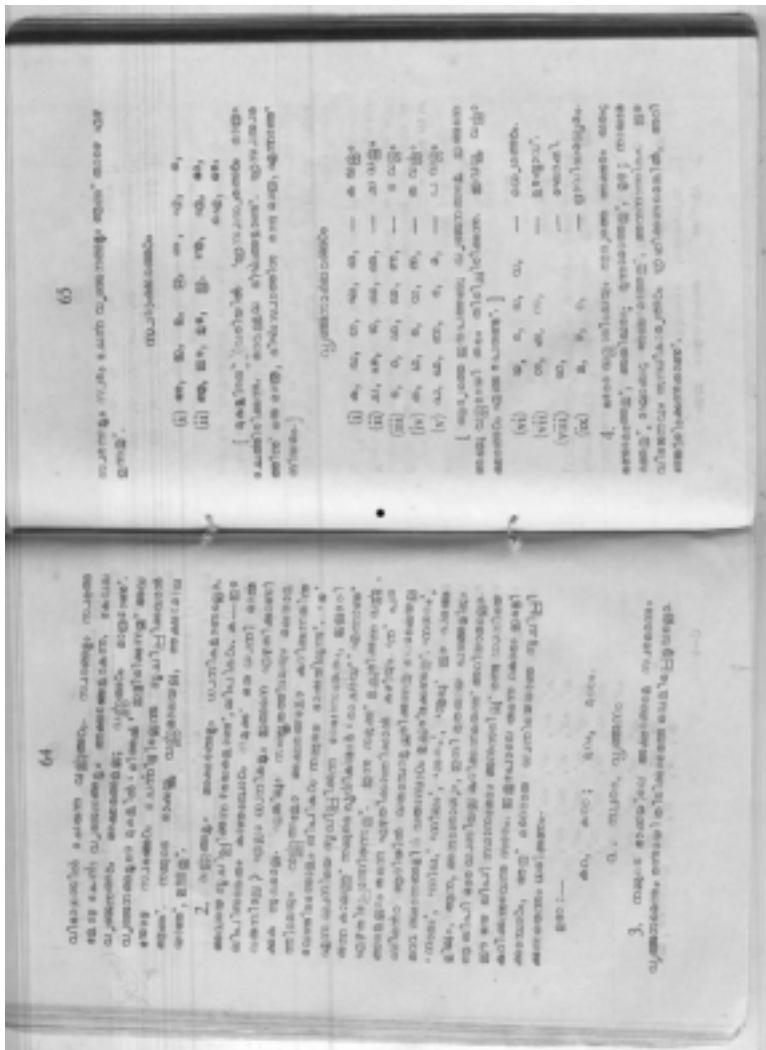


Figure A.2: Pages 64 and 65 of Sankara Pillai's book

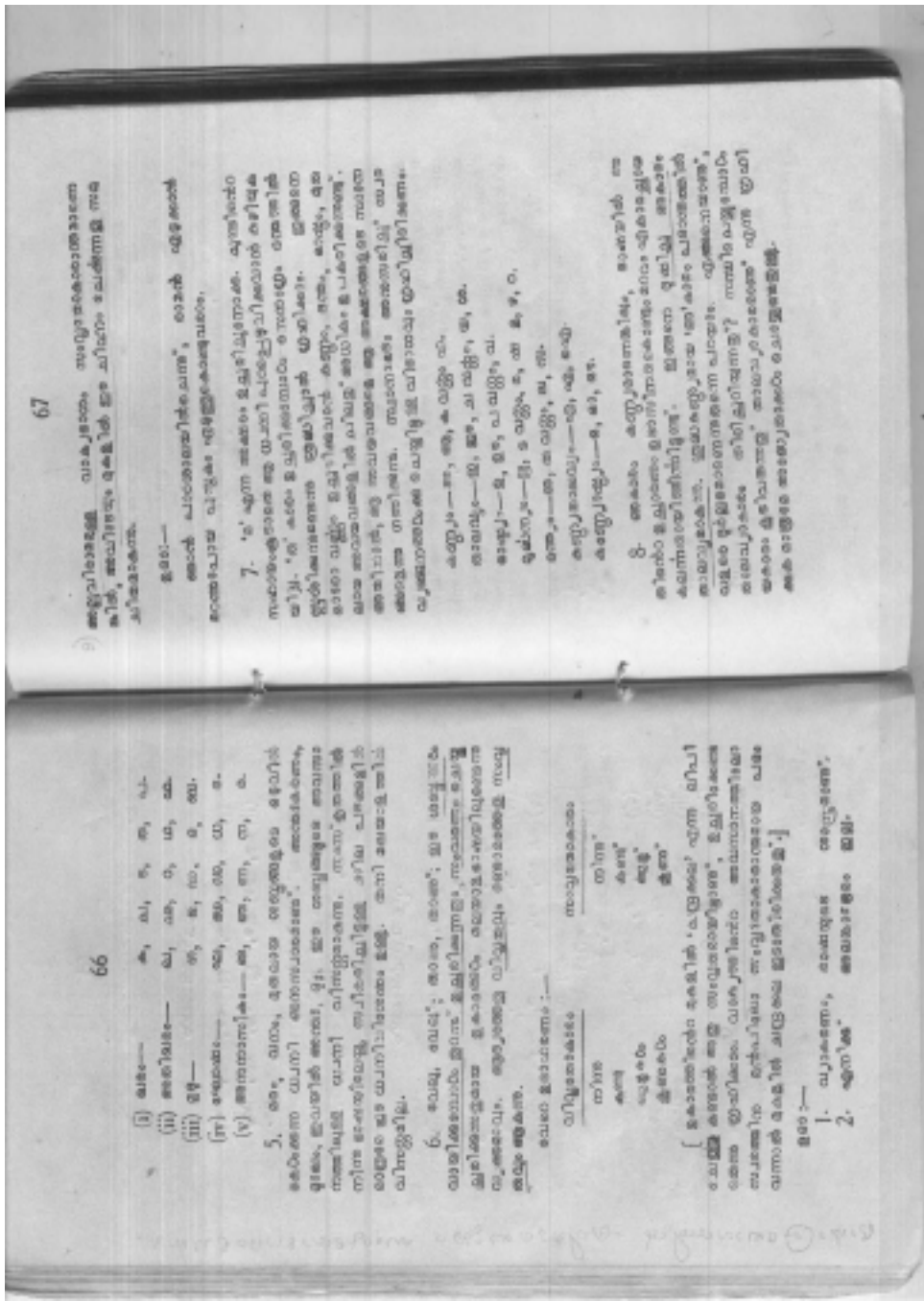


Figure A.3: Pages 66 and 67 of Sankara Pillai's book

അവവ്യക്തം:-

കറസ്സു

ഉരസ്സു

കിളസ്സു

ഇതുകൾ:-

കറക്കു

ഉരക്കു

കിളക്കു

അവ, അമ്മ, ആൾ, ഉദയം പദങ്ങളുടെ അന്ത്യസ്വരമായ അകാരം അവവ്യക്തമാണെന്നു്, ഒരു സ്വരം പിന്നാലെ വരുമ്പോൾ സന്ധിയിൽ വരുന്ന അകാരം വ്യക്തമാകും.

അ + ഇൽ = അവയിൽ.

അ + ഉൾ = അയ്ക്കൽ.

ആൾ + ഇല്ല = ആൾക്കിടയിൽ.

9. ത, ങ, ണ, മ, ന, യ, ള, ൾ ഇവ ശിമിലവ്യഞ്ജനങ്ങളാകുന്നു. പദങ്ങളിൽ ഞ്ഞിയിൽ വരുമ്പോൾ ഏകദേശമായ കലർന്ന അകാരമാണു് നാം ഉച്ചരിക്കുന്നതു്. തം, ങം, ണം, മം, നം, യം, ൾ എന്നിവയെക്കൊണ്ടു് ഏഴുതരത്തിലും, നാം വായിക്കുന്നതു് തം, ങം, ണം, മം, നം, യം, ൾ എന്നിവയാണു്. അകാരം ഉച്ചരിച്ചു അവവ്യക്തമായിപ്പോയതു കൊണ്ടാണു്, ഉച്ചരിക്കുന്നതിൽ ഈ കാര്യം വന്നതു് എന്നു് അറിയണം.

10. സ്വരസഹായം കൂടാതെ ഉച്ചരിക്കാവുന്ന വ്യഞ്ജനങ്ങളാണു്. അവ അക്ഷരങ്ങളല്ല; വർണ്ണങ്ങളാണു്. വൻ, കൻ, വാൾ, മെൽ, മുതൽ, ഈ ശബ്ദങ്ങളിലെ, ന്, ണ്, ൾ, ക്, ൾ, ൾ, ഇവ ചിലപ്പോൾ ഉപയോഗിക്കുന്നു.

അദ്ധ്യായം ൨ ന്

1. വർണ്ണവും അക്ഷരവും തമ്മിലുള്ള അന്തരം സ്പഷ്ടമാക്കണം വെളിപ്പെടുത്തുക.

2. സ്വരം, വ്യഞ്ജനം, ഇവയ്ക്കിടയിൽ ഉപയോഗിക്കേണ്ടതെന്തു്.

3. സംവൃത്തംകൊണ്ടു് എന്താണതു്? എന്തെ അതു് എഴുതിക്കാണിക്കാം.

Figure A.4: Pages 68 of Sankara Pillai's book

Appendix B

Summary of Script Reform

See the following pictures for Gopinatha Pillai's summary of the script reform.

മലയാള ലിപിയും ലിപി പരിഷ്കരണവും

ഏ. ഡി. പതിനേഴാം നൂറ്റാണ്ടോടുകൂടി ഇന്നു നാം ഉപയോഗിക്കുന്ന മലയാള ലിപി രൂപം പ്രാപിച്ചുവെന്നാണ് പല പണ്ഡിതന്മാരുടേയും അഭിപ്രായം. പഴയ കാലത്തു മലയാളമെഴുതാൻ വട്ടെഴുത്തും കോലെഴുത്തുമാണുപയോഗിച്ചിരുന്നത്. "കല്ലിൽവെട്ടി (ഉളികൊണ്ടു കൊത്തി) എഴുതിയിരുന്നതുകൊണ്ടാണ് ആ ലിപിക്കു വെട്ടെഴുത്തു എന്ന പേർ സിദ്ധിച്ചതെന്നും, ആ പേർ കാലാന്തരത്തിൽ വട്ടെഴുത്തു എന്നായി പരിണമിച്ചു എന്നും പറയുന്നു. കോലുകൊണ്ട് (കോലുപോലുള്ള നാരായം കൊണ്ട്) എഴുതിയിരുന്നതാണ് കോലെഴുത്തു" വട്ടെഴുത്തിൽ നിന്നാണ് കോലെഴുത്തിന്റെ ഉല്പത്തി. വട്ടെഴുത്തിനും കോലെഴുത്തിനും ലിപികൾക്കും തമ്മിൽ വലിയ വ്യത്യാസമൊന്നുമില്ല. "വട്ടെഴുത്തിൽ നിന്നു തന്നെ ഉണ്ടായതും അതിൽ നിന്നു ഈഷദ് വ്യത്യാസങ്ങൾ മാത്രം ഉള്ളതുമായ മറ്റൊരു വകഭേദമാണ് മലയാള ലിപികൾ."

സംസ്കൃതാക്ഷരമാല മലയാളത്തിൽ സ്വീകരിച്ചതോടെ ഗ്രന്ഥാക്ഷരം എന്ന പേരിലറിയപ്പെടുന്ന ലിപി മലയാളത്തിൽ നടപ്പിലായി. ഈ ഗ്രന്ഥലിപിയുടെ രൂപാന്തരമാണ് 'ആര്യ' എഴുത്തു എന്നു കൂടി പേരുള്ള നമ്മുടെ മലയാള ലിപിയെന്ന് ശ്രീ. എസ്. ശങ്കര അയ്യപ്പം, ടി. ഗോപിനാഥറാവുവും അഭിപ്രായപ്പെടുന്നു. മലയാളത്തിൽ അൻപത്തിമൂന്നക്ഷരങ്ങളുണ്ട്. കൂട്ടക്ഷരങ്ങൾ വള്ളി പുള്ളികൾ മുതലായവ ഉൾപ്പെടെ അഞ്ഞൂറോളം അക്ഷരങ്ങളും ചിഹ്നങ്ങളും ഭാഷയിൽ നടപ്പുണ്ട്.

Figure B.1: Recommendations of Script Reform Committee – part 1

അച്ചടിയുടെ ആരംഭത്തോടെ മലയാളത്തിലെ ലിപി പരിഷ്കരണത്തിനുപേണ്ടി പലതും ശ്രമിച്ചിട്ടുണ്ട്. ഇതിന് ആദ്യമായി നേതൃത്വം നല്കിയത് ശ്രീ. കണ്ടത്തിൽ വഴിസ് മാപ്പിളയാണ്. തുടന്ന് പലരും പരിഷ്കരണശ്രമങ്ങളുമായി മുന്നോട്ട് വന്നു. ഇവയെല്ലാം കാലക്രമേണയായ ലിപി പരിഷ്കരണം സ്മൃതയാർഹമായ ഒന്നാണ്.

1968-ൽ ശ്രീ. ശുരനാട്ട് കണ്ണൻപിള്ള കണ്വീനായി രൂപീകരിച്ച ലിപി പരിഷ്കരണക്കമ്മിറ്റിയുടെ ശുപാർശസംഗ്രഹം താഴെ ചേർക്കുന്നു.

1. ഉ, ഊ, ഋ എന്നീ മൂന്ന് സ്വരങ്ങൾക്ക് വ്യഞ്ജനങ്ങളോട് ചേരുമ്പോൾ മറ്റു സ്വരങ്ങളുടെ കാര്യത്തിലെ നപോലെ പ്രത്യേകം ചിഹ്നങ്ങൾ ഏർപ്പെടുത്തുക. 'ഉ' എന്നതിന് 'ചുനിപ്പം' 'ഊ' എന്നതിന് 'ചുറിക്കെട്ടു' ചുനിപ്പം വ്യഞ്ജനം ശബ്ദോടുകൂടി ചേർക്കുന്നതിനുള്ള ക്രമം കാണിക്കാൻ കൊള്ളത്തുരേഖയുടെ അറ്റത്തു് ചേർക്കുന്നു—ക, ക്ക, ച, ച്ച. പ്രകാരത്തിന് ഇപ്പോഴുള്ള ചിഹ്നം തന്നെ വേർതിരിച്ചു പ്രത്യേകം ചേർക്കാം—കു, തു, ടി ഇത്യാദി.

2. മുൻമ്പിൽ രേഖം (ർ) ചേർന്ന കൂട്ടക്ഷരങ്ങൾക്ക് നിലവിലുള്ള രണ്ടുതരം ലിപികളിൽ ഒരു തരത്തിലുള്ളത് (തലയിൽകത്തുള്ള) മുഴുവനും ഉപേക്ഷിക്കാം. അതരം കൂട്ടക്ഷരങ്ങളുടെ മുൻമ്പിൽ 'ർ' ചേർത്തഴുതാം—'ക്ക' എന്നത് 'ർക്ക' എന്ന പോലെ.

3. കൂട്ടക്ഷരങ്ങളിൽ 18 എണ്ണം തന്നെ മലയാള പദങ്ങളിൽ അധികമായി വരുന്ന ക, ക്ക, ങ്ങ, ച്ച, ഞ്ഞ, ട്ട, ണ്ണ, ത്ത, ന്ന, പ്പ, വ്വ, മ്മ, യ്യ, ല്ല, വ്വ എന്നിവ ഇപ്പോഴത്തെപ്പോലെതന്നെ നിലനിർത്താം. മറ്റുള്ള സംയുക്താക്ഷരങ്ങളെ (അവ ഏറിയകൂറും സംയുക്തത്തിൽനിന്നുള്ളവയാണ്) ചന്ദ്രശേഖരയിട്ട് വേർതിരിക്കുന്നത് ഇപ്പോൾത്തന്നെ നടപ്പിലുള്ളതാണ്. ആ രീതി സർവ്വസാധാരണമാക്കാം.

Figure B.2: Recommendations of Script Reform Committee –part 2

4. യ്, ര്, ല്, വ് എന്ന മധ്യമങ്ങൾ പിന്നിൽ ചേർന്ന് വരുന്ന സംയുക്താക്ഷരങ്ങളിൽ യ്, വ് ചേർന്നവ ഇപ്പോൾത്തന്നെ പിരിച്ചെഴുതുന്നു. യ്, വ് എന്ന അംശങ്ങൾക്കു വേർതിരിച്ചു ചിഹ്നങ്ങൾ കൊടുക്കുന്നു. ഈ രീതി (ക്യ, ക്വ) എന്നു സർവ്വസാധാരണമാക്കാം. പിന്നിൽ ര് (റ), ല് (ള) ചേർന്ന കൃ, ക്ല എന്നുപോലുള്ള ലിപികളും വേർതിരിച്ചു എഴുതാവുന്നതാണ്. ര് എന്ന അംശത്തിന് ഇപ്പോഴുള്ള ചിഹ്നം വലതുവശത്തു തൊടാത്തതരത്തിൽ ആക്കാം. ല (ള) ഇപ്പോഴത്തെ ഇപ്പോലെയോ ചരുകലയിട്ട് പിരിച്ചോ കാണിക്കാം.

5. ചില്ലുകൾ (ൺ, ന്, റ്, ത്, ൾ) ഇപ്പോഴത്തെ ഇപ്പോലെ നിലനിർത്താം.

6. ഏഴത്തിൽ ഇപ്പോഴത്തെ ലിപി സമ്പ്രദായം കൂടുതലാകട്ടെ.

7. ഈ നിർദ്ദേശങ്ങൾ നടപ്പിൽ വരുത്തിയാൽ ഇപ്പോഴുള്ള ആഞ്ഞുറിലധികം ലിപികളിൽ മുക്കാൽ പങ്കും കറവു ചെയ്യാവുന്നതാണ്. പുതിയ ലിപി 1971 ഏപ്രിൽ 15 മുതൽ നടപ്പിൽ വന്നു.

സ്വരങ്ങളുടേയും ചില വ്യഞ്ജനങ്ങളുടേയും ചിഹ്നം

(പുതിയ അച്ചടി മാതൃകയിൽ)

ഇ=ി	ഈ=ീ	ഉ=ു
ഊ=ു	ഘ=ഘ	റ=ർ

Figure B.3: Recommendations of Script Reform Committee – part3

Appendix C

Extracts From a Dictionary

These images are taken from the Assissi Malayalam Hindi English Dictionary, by P. Madhavan Pillai, 1981, Assissi Printing and Publishing House, Changanacherry.

This book uses the reformed script; But, I may hasten to add that I am not particularly happy with this dictionary, which is why I have given the examples of apparently inconsistent sorting.

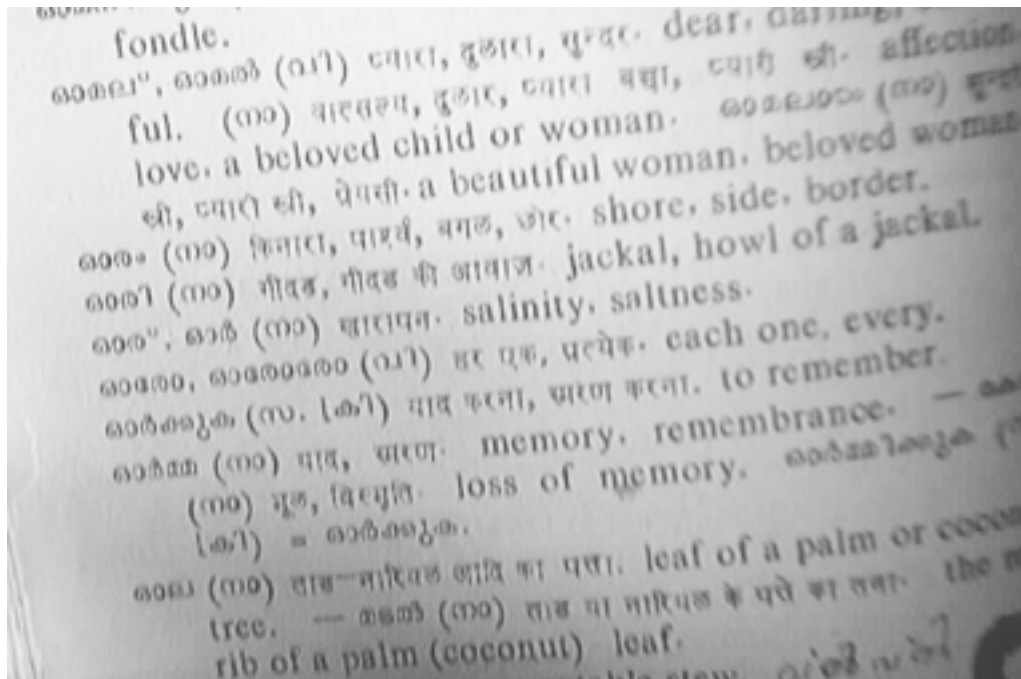


Figure C.3: Explicit virama form of rr is same as its chillu in the entry for “salinity”.

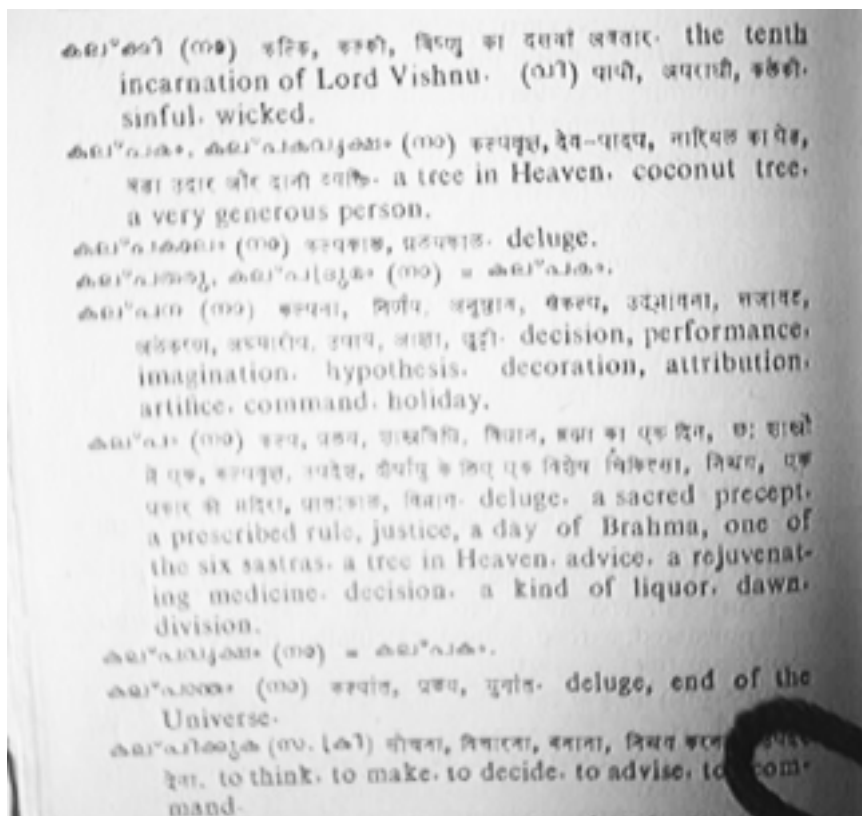


Figure C.4: Where is the ll chillu?? Only explicit virama form of U+0D32 is seen here.

കരർത്തികേയൽ (നം) കാർതികേയ, സ്കന്ദ. Lord Subrahmanya.
 കരർപ്പണം (നം) കാർഷ്യ, കൃഷ്ണതാ, കേജുടി; നവീന. niggardliness,
 miserliness, poverty.
 കരർമ്മുക (നം) കർമ്മുക, മനുഷ്യ, വീട, നീമ, ഇന്ദ്രവസു. a bow, bamboo.
 the neem tree, rain-bow.
 കരർത്ഥത (നം) കാർയ്കൃതത. efficiency.
 കരർത്ഥശി (നം) കാർയ്ദശി. secretary.
 കരർത്ഥപരിപാടി (നം) കാർയ്കര. programme.
 കരർത്ഥ (നം) കാർയ്, വാട, കാര, നാടക കീ അർത്ഥകൃതിയോ മേ പുക, ത്വർത്തി, ഉത്പതി,
 ഉദ്ദേശ്യ, പരിണാമ. deed, matter, work, the denouement of
 a drama, health, origin, motive, result.
 കരർത്ഥശി (നം) കാർയ്കർത, കാർയ്കാര, അമിർത. a manager, an agent.
 കരർത്ഥശി (നം) കാർയ്കർത. office.
 കരർത്ഥശി (നം) കാർയ്കർത, അമിർത. one with dark-complexion. Sri Krishna.
 കരർത്ഥശി (നം) കാർയ്കർത. agricultural.
 കരർത്ഥ, കരർത്ഥ (നം) വേ, വാട, വാട, വീട, ചതുർത്ഥ, സഹാര, ലം, വാലി.
 leg, a quarter or one-fourth. prop, pillar, a sprig of
 corn.

Figure C.5: Note the sorting order.

കേവ്കേവ് (നം) കേവ്കേവ്. canopy.
 കേവ്കേവ്കേവ് (നം) കേവ്കേവ്, അവിവേ. suzerainty.
 കേവ്കേവ് (നം) കേവ്കേവ്, മീന. high position, salvation.
 കേവ്കേവ് (നം) കേവ്കേവ്. as above.
 കേവ്കേവ് (നം) കേവ്, കേവ്കേവ്. a roof, an upper room.
 കേവ്കേവ് (നം) കേവ്കേവ് = കേവ്കേവ്.
 കേവ്കേവ് (നം) കേവ്കേവ്, അകാ. the upper portion, the
 sky.
 കേവ്കേവ് (നം) കേവ്കേവ്. an upper garment.
 കേവ്കേവ് (നം) കേവ്കേവ്. address.
 കേവ്കേവ് (നം) കേവ്കേവ്, മേവ്കേവ്, മേവ്കേവ്. chief priest.
 കേവ്കേവ് (നം) കേവ്കേവ്, മേവ്കേവ്, മേവ്കേവ്, മേവ്കേവ്, മേവ്കേവ്, മേവ്കേവ്.
 to dwell, to agree, to desire, to love, to become
 familiar.
 കേവ്കേവ് (നം) കേവ്കേവ്, table.

Figure C.6: Note the sorting order. Chillu rr, rr with vowel signs, again chillu rr. The last entry says chillu ll is same as its explicit virama form.