Malayalam Orthographic Reforms:
Impact on Language and Popular Culture

Kavya Manohar  Swanthantra Malayalam Computing
sakhi.kavya@gmail.com

Santhosh Thottingal  Swanthantra Malayalam Computing
santhosh.thottingal@gmail.com

June 8, 2018

Abstract

Malayalam is a language spoken in India, predominantly in the state of Kerala with about 38 million native speakers. The Malayalam script evolved from Brahmi through Grantha alphabet and Vattezhuthu writing systems. The script orthography has acquired its uniqueness with its complex shaped ligatures formed by the combination of consonants and vowel sign forms. The number of unique graphemes in this system exceeds 1200. The orthographic styles were constantly evolving. In 1971 there was a Governmental intervention in the orthography, to reduce its complexity to address the difficulties in typesetting and printing. This paper is an attempt to explore the impact of this orthographic reforms on various aspects of script usage including popular culture, media, textbooks, graffiti and handwriting. We will also analyse the impact of Unicode and the advancement in digital typography on the orthographic diversity of Malayalam script.

Keywords: Malayalam Script, Orthography Reforms, Unicode, Graphemes, Media and Communication, Digital Typography

1 Introduction

With 38 million native speakers Malayalam is the official language of Kerala, in southern India. Malayalam used to be written in Vattezhuthu, the earliest known sample dating back to 830 AD. The modern Malayalam script evolved from Grantha

*https://smc.org.in
alphabet which was a script for Sanskrit. Both Vattezhuthu and Grantha have their roots in the Brahmi script\textsuperscript{1}. Unicode has encoded 18 vowels and 37 consonants as of today, some of them are archaic. The Figure 1 illustrates the Malayalam code block as per Unicode 10.0\textsuperscript{2}.

![Figure 1: Unicode 10.0 Malayalam Code block](https://en.wikipedia.org/wiki/Malayalam_script)

Malayalam script is abugida, or alphasyllabary. That is, consonant–vowel sequences are written as a unit: each unit is based on a consonant or conjunct letter, and vowel notation is secondary. Vowel notations that modify the consonants is in the form of diacritics or vowel signs. Vowels have independent existence only at word beginnings. This is the common characteristic of Brahmic family of scripts from South and Southeast Asia.

![Figure 2: Few samples of graphemes in Malayalam](https://en.wikipedia.org/wiki/Malayalam_script)

The script has acquired its uniqueness with its complex shaped ligatures formed by the consonants and conjuncts with signed vowel forms. Conjuncts are formed

\textsuperscript{1}Malayalam Script in English Wikipedia: https://en.wikipedia.org/wiki/Malayalam_script

\textsuperscript{2}Malayalam Unicode block: https://unicode.org/charts/PDF/U0D00.pdf
by a sequence of two more consonants. The conjunct grapheme usually has a shape
smoothly blend from the constituent consonants. Figure 2 illustrates some samples.

2 Nature of Malayalam Script

Malayalam script as known today has 18 vowels, 39 consonants apart from various
numerals, measuring units etc. Some are archaic in nature. The general nature of the
script since its early days of evolution can be consolidated as below [1].

1. Vowels and consonants are the basic building blocks of Malayalam script.

2. Vowels have stand alone existence in their pure form. Vowels in Malayalam are
   അ (a), ഇ (i), എ (e) etc. See Figure 3.

3. Vowels also appear as signed form modifying a consonant sound. Vowel signs
   have no existence without a consonant. Vowel signs in Malayalam are ി (i), ി (iː), ി (e), ി (o) etc. See Figure 3.

4. Consonants in Malayalam always have the inherent vowel /a/, also known as
   schwa present in them. Consonants in Malayalam are ക (ka), ത (ta), ഗ (ga), ദ (da), ധ (dʱa) etc.

5. Any other vowels sound, other than /a/ associated with a consonant is written
   as a signed form of the consonant. The vowel signs can appear left, right or both
   with respect to a consonant. Some signs modify the shape of base grapheme.
   Examples for consonants with vowel signs: ೖ (ki), ೑ (gu), ೓ (dʱe), ೔ (d̪o).
   See Figure 4.

Figure 3: The vowels in Malayalam. There are independent vowels and dependent
vowel sign forms.
Figure 4: The consonant (ka) is shown with all possible associated vowel signs.

6. The removal of inherent /a/ in a consonant is marked in the script by a special character virama. Example for consonant with virama (◌’): ക (k). Virama after a consonant not only removes inherent /a/ but also indicate no vowel sound follows it.

\[ ക (k) + \text{virama} \rightarrow ക’ (k) \]

7. A conjunct is formed by a sequence of consonants connected using virama. Examples:

\[ ക (k) + \text{virama} + ത (t̪) \rightarrow ക്ത (kt̪) \]

\[ ഗ (ɡ) + \text{virama} + ദ (d̪) \rightarrow ഫ (ɡd̪) \]

A conjunct can again connect to another consonant using virama and form a longer conjunct as in:

\[ ഫ (ɡd̪) + \text{virama} + ധ (d̪ʱ) \rightarrow ഫ്സ (ɡd̪d̪ʱ) \]

8. Every conjunct can be modified by a vowel sign forming a new ligature.

\[ ഫ (ɡd̪) + \text{virama} + ഉ (u) \rightarrow ഫു (ɡd̪u) \]

9. Some consonants involved in the formation of conjuncts have signed forms. Examples:

\[ ക (k) + \text{virama} + ര (ɾ) \rightarrow കɾ (kɾ) \]

\[ ക (k) + \text{virama} + ല (l) \rightarrow കള (kla) \]

\[ ക (k) + \text{virama} + യ (j) \rightarrow കja (kja) \]

\[ ര (ɾ) + \text{virama} + ക (k) \rightarrow രക (ɾka) : This special sign is named dot reph. \]

10. Certain consonants have a unique grapheme representation in their pure form named chillu. ക (k), ല (l), ണ (ɳ), ന (n̪), ള (l), ര (r)

The above nature of the script makes the number of unique graphemes to exceed 1200[2]. Attempts of script reformation which happened during the later half of 20th century was aimed at simplifying the script to bring down the number of graphemes. Detailed discussion on script reformation will follow in a later section.
3 Script in Early Printing Era

The shapes of conjuncts, relative positioning of signs and their sizes have changed over time to match the needs of writing methods.

The first ever book in Malayalam script was printed in Rome, in 1772. Printing in Malayalam started natively during 1820s\[3\]. When printing technology started getting popular there was a requirement to cast movable types in huge numbers. Even though there were less than a hundred basic characters, the orthographic style demanded separate types for conjuncts, and their signed vowel forms. Apart from vowels, some consonants too have signed notations, further increasing the number of types needed in the foundry.

Figure 5: Samkshepavedartham - 1772
Figure 6: Samples of print documents. Complex graphemes formed by consonant sequences and signed conjuncts can be seen.
The first printed book in Malayalam using movable types, സംെ•പെവദാർം (Samkshepavedartham) in 1772 had more than thousand unique types[3]. The manual labor on typesetting and lay outing were high for the same reason. Figure 5 shows pages from the catechism book Samkshepavedartham. The script is mostly rectangular. The types were made and printing was done in Rome.

The first native type casting and printing was done by Benjamin Bailey, an Anglican missionary in 1829[3]. His contributions as a typographer made the curvy style of the Malyalam orthography popular[4]. Figure 6a, shows pages from The New Testament printed using the types designed by Benjamin Bailey, imprinted in 1829[3]. The script continued to evolve by separating some vowel sign types (◌ി, ◌ീ) from the consonant or conjunct grapheme. Still the richness of conjuncts and their signed forms were largely retained. It can be seen from Figure 6b.

4 Script Reformation

Typewriters became very popular around 1960s in Kerala. These typewriters followed the same design of English typewriters with the keys re-purposed for Malayalam. Obviously the keys were not enough to support all complex ligatures. The end result was Malayalam with all ligatures split up. It was a painful experience for reading and did not do any justice to the beauty of script as we can see from Figure 7.

To solve this either the typewriter or language had to be redesigned. There were demands from newspaper and publishing industries to reduce the script complexity so that Malayalam is better suited for typewriters and printing. Based on this, in 1967 Kerala government appointed a committee to study script reformation. The committee submitted their report and in 1971 Kerala government published an order to reduce the complexity of the script[5].

The order was to discard the usage of complex conjuncts and to detach the vowel notations from the consonants and conjuncts. Being a forced intervention, this was a major event to be marked in the history of orthographic evolution.

The script variant of Malayalam that came into existence after the reformation order will henceforth be referred as reformed orthography. Reformed orthograhy consists only of a smaller set of graphemes than in the exhaustive set described in section 2.
Figure 7: A typewriter and a sample of Malayalam document prepared using typewriter. No complex graphemes are used. Consonant sequences remain separated with virama (◌்) in between.

Figure 9 shows the front page of government order proposing the new orthography style. The proposal aimed at reducing the grapheme usage in Malayalam by 75%. The major proposals in the order are following: [5]

- Detach the signs of vowels ു (u), ു (uː) and ു (ri) from the base grapheme. 
  ു → ു, ു → ു, ു → ു.

- Detach the consonant sign of ്, that is ് from the base grapheme
  ് → ്.
Figure 8: Examples of complex graphemes in Malayalam script

Figure 9: The Government Order on Malayalam Script Reform in 1971

GOVERNMENT OF KERALA

Abstract

MALAYALAM SCRIPT-ADOPTION OF NEW SCRIPT FOR USE-ORDERS ISSUED

EDUCATION ‘P’ DEPARTMENT

G. O. (P) 3771/Edn.

Dated, Trivandrum, 23rd March 1971.

Read: G.O. (P) 329/68/Edn dated 11-7-1968

ORDER

The question of reducing the unwieldy number of alphabets and signs in Malayalam which consume much time and labour in the process of printing and typewriting, has been under consideration of Government for some time. In 1967 Government appointed a Committee with Shri Soormad F. N. Kunjan pillai, Editor, Malayalam Lexicon as convener to advise them on the question of reformation of Malayalam script. The committee in its report has made recommendations to reduce 75% of the total number of existing characters in printing and typewriting. The reformed Malayalam script recommended by the above Committee was revised with slight modifications by another committee appointed in 1969 to expedite the adoption of the new script for use. The recommendations of the above two committees in the matter of reformation of the Malayalam script are in brief as follows:

i. 

ii.

Figure 9: The Government Order on Malayalam Script Reform in 1971
<table>
<thead>
<tr>
<th>No</th>
<th>Characters</th>
<th>IPA</th>
<th>Traditional Orthography</th>
<th>Reformed Orthography</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>്, ു</td>
<td>ka:</td>
<td>്</td>
<td>്</td>
</tr>
<tr>
<td>2</td>
<td>്, ു</td>
<td>ḍe</td>
<td>്</td>
<td>്</td>
</tr>
<tr>
<td>3</td>
<td>്, ു</td>
<td>j̪a</td>
<td>ു</td>
<td>ു</td>
</tr>
<tr>
<td>4</td>
<td>്, ു</td>
<td>k̪a</td>
<td>ു</td>
<td>ു</td>
</tr>
<tr>
<td>5</td>
<td>്, ു</td>
<td>ku</td>
<td>ു</td>
<td>ു</td>
</tr>
<tr>
<td>6</td>
<td>്, ു</td>
<td>gu</td>
<td>ു</td>
<td>ു</td>
</tr>
<tr>
<td>7</td>
<td>്, ു, ു</td>
<td>g̪a</td>
<td>ു</td>
<td>ു</td>
</tr>
<tr>
<td>8</td>
<td>്, ു, ു</td>
<td>g̪u</td>
<td>ു</td>
<td>ു</td>
</tr>
<tr>
<td>9</td>
<td>്, ു, ു</td>
<td>Şt̪a</td>
<td>Ş̪</td>
<td>Ş̪</td>
</tr>
<tr>
<td>10</td>
<td>്, ു</td>
<td>kra</td>
<td>Ş̪</td>
<td>Ş̪</td>
</tr>
<tr>
<td>11</td>
<td>്, ു, ു</td>
<td>st̪ra</td>
<td>Ş̪</td>
<td>Ş̪</td>
</tr>
<tr>
<td>12</td>
<td>്, ു, ു</td>
<td>k̪r̪a</td>
<td>Ş̪</td>
<td>Ş̪</td>
</tr>
</tbody>
</table>

Table 1: Illustrating traditional and reformed orthography differences

- Discard the usage of ' in the consonant sequence in the form of dot reph sign. Instead use the alternate form of ു. ് ു ു → ് ു ു ു.

- Discard the use of rare conjuncts by splitting them down into constituent consonant sequence separated by the virama sign. Those retained are: ്, ു, ു, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ, ʃ. Others are split down as: ു → ʃ̪.

- The signed form of consonants are to be separated from the base grapheme as in ു, ു, ു).

- The signed below base modifiers of ് (ш) may be retained as such ʃ or split using virama sign as ʃ̪.

As per the government order the reformed orthography would retain only 90 unique graphemes.

Table 1 compares the graphemes formed by sequence of basic characters in traditional and reformed orthography. As can be seen from the first four rows, the detached sign forms in traditional orthography are retained as such in the reformed one. Also some commonly used conjuncts are retained as such. The difference between two orthography variants becomes spectacular in the forthcoming rows. Complex ligatures formed by sequence of consonants gets split up by placing virama sign in be-
tween. Joined signed forms in traditional orthography get detached in the reformed variant.

It is important to note that the reformation order introduce the detached form of vowel signs for ഉ and ഊ as ഓ and ഒ respectively. In the exhaustive set of traditional orthography u and uː had very diverse sign forms \[6\]. Their usage is evident from rows 5, 6, 8 and 12 in Table 1.

4.1 Adoption of Reformed Orthography

The print media switched to the reformed orthography to varying extends. The official prints of the government almost completely switched to the reformed style. Some publishers retained the graphemes for signed form of consonants but detached the signed vowel forms. Publishers adopted a set of conjuncts as per their choice and split down the others using virama sign.

Students started to learn reformed orthographic style from the textbooks. But they continue to watch and learn the usage of traditional complex orthography widely seen in wall graffiti, poster designs and handwriting. Figure 10a and 10b illustrates the co-existence of both orthography - One in textbook and other in a movie poster.

5 Script in Digital Era

The digitization of printing by early 1990s was yet another remarkable event. The pre-Unicode digital fonts in Malayalam were Malayalam letters mapped to the ASCII character space. Such fonts retained only a limited repertoire of conjuncts, because ASCII had the limitation of 256 code points. Also the signed notations of vowels and consonants were detached from the base grapheme. Digital fonts before the Unicode era embraced the reformed orthography more closely. The publishing industry largely depended on these fonts for decades.

At the same time, writing Malayalam in non-digital, non-printing contexts continue to use traditional orthography. Wall paintings, artistic lettering used in magazines, movie titles continued using it as illustrated above.

*Rachana Aksharavedi*, an organization formed 1998 was successful in bringing back the exhaustive traditional orthography set with the help of then existing technology.
At that time, Malayalam was not encoded in Unicode. Rachana Aksharavedi developed a font named Rachana with about 1200 glyphs. Since there is no Unicode or opentype technology, it was a set of 6 fonts, each covering about 200 glyphs mapped to ASCII codepoints. A special editor known as Rachana Editor was required to automatically switch between these fonts and display Malayalam with data being English. This brave attempt was widely appreciated. A couple of years later, in 2001 Unicode encoding for Malayalam happened.

5.1 Unicode and Advanced Digital Typography

Unicode did not differentiate between the traditional and reformed orthography. Orthography style was left to the typography and fonts layers. The ISO 639-1 also does not differentiate between the traditional and reformed orthography[7]. As per ISO 639-1 the language code for Malayalam is ml. That means, the digital representation using unicode code points remains same, the readers see that data using a font following traditional or reformed orthography as per their choice. Because of this abstraction, the reformation had no impact on the processing of textual data.
With the advent of Unicode based digital typography, complex conjunct formations and their rendering were no longer an impossibility. With only the basic graphemes encoded in Unicode, any long sequence of consonants and signs could be mapped to a single conjunct grapheme in signed or unsigned form. Complex rendering rules of the script can easily be handled by modern rendering engines. With these technical advancements, fonts which could very well support the traditional orthographic scheme of the Malayalam script emerged.

The Rachana font was ported to Unicode. Parallel to that, more unicode fonts emerged, notably AnjaliOldLipi. In 2006, Swathanthra Malayalam Computing (SMC), a free software developer community became active in Malayalam computing. Along with various language processing tools and technology improvements, SMC released a dozen of Malayalam fonts. With one exception, all fonts followed traditional orthography and embraced the opentype technology.

<table>
<thead>
<tr>
<th>No.</th>
<th>Font</th>
<th>Orthography style</th>
<th>Vendor</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Rachana</td>
<td>Traditional</td>
<td>SMC</td>
<td>Available in Ubuntu and other Linux distros</td>
</tr>
<tr>
<td>2.</td>
<td>Meera</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>3.</td>
<td>Manjari</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>4.</td>
<td>AnjaliOldLipi</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>5.</td>
<td>Chilanka</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>6.</td>
<td>Dyuthi</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>7.</td>
<td>Keraleeyam</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>8.</td>
<td>Uroob</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>9.</td>
<td>Manjari</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>10.</td>
<td>Karumbi</td>
<td>Traditional</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>11.</td>
<td>RaghuMalayalam</td>
<td>Reformed</td>
<td>SMC</td>
<td>”</td>
</tr>
<tr>
<td>12.</td>
<td>NotoSans Malayalam</td>
<td>Reformed</td>
<td>Google</td>
<td>Default in Android OS</td>
</tr>
<tr>
<td>13.</td>
<td>Nirmala</td>
<td>Reformed</td>
<td>Microsoft</td>
<td>Default in Windows OS</td>
</tr>
<tr>
<td>14.</td>
<td>Sangam</td>
<td>Reformed</td>
<td>Apple</td>
<td>Default in Apple products</td>
</tr>
</tbody>
</table>

Table 2: List of popular unicode fonts in Malayalam
GNU/Linux systems came with these traditional orthography fonts by default. Schools and government institutions were using GNU/Linux systems because of Kerala government policy to use Free Software. The userbase of traditional orthography started to expand among digital Malayalam users. The IT education curriculum in schools also widely used these fonts.

Table 2 lists the Malayalam unicode fonts available by default in various operating systems. Availability of good quality traditional orthography fonts accelerated the usage of traditional orthography in digital space.

![Malayalam Textbook - 2011. ASCII based new orthography in print](image)

The typesetting tools and software adapted to ASCII based fonts were the default in the publishing industry since 1990s. Even after the encoding of Malayalam in Unicode in 2001, the printing and publishing industry continued their practices. The Figure 11, shows the usage of ASCII based reformed orthography in school textbooks printed in 2011. This was largely due to lack of unicode and complex script rendering support in major typesetting systems like Adobe Indesign. But these typesetting systems started supporting complex scripts and now we are seeing a highly accelerated adoption of Unicode and traditional orthography in print.
6 Contemporary Script Usage

Currently everyone learns to read and write reformed orthography as part of school curriculum, but practically uses and are accustomed to the traditional orthographic style in everyday life. Non-digital media including wall writings, graffiti, bill-boards and handwriting were always sticking to the traditional orthographic set.

![Image of Samakalika Malayalam, a popular illustrated weekly announcing their return to traditional orthography. 2017 October](image)

Now that the technology has matured enough to support the traditional orthography, it is becoming more and more popular in the digital domain as well. There are newspapers and portals that switched to traditional orthography. Popular illus-
treated weeklies and science magazines switched to traditional orthography style in print. See Figures 12 and 13.

Figure 13: Eureka, a famous science magazine for children's changed to traditional orthography printing in November 2017.

Kerala government is actively promoting Unicode usage in the official documents. Government orders are now mostly in Meera font, a traditional orthography font by SMC. Identity cards used for voting and public distribution system (See Figure 14) etc also uses traditional orthography.

The traditional orthography fonts by SMC are widely used in web content. See an example in Figure 15.
Figure 14: 2017 Identity card for public distribution system (Ration system)

Figure 15: Internet meme example in Malayalam. Uses traditional orthography font. 2017 November

Figures 16, 17 illustrate the adoption of traditional orthography in contemporary movie posters.
Figure 16: A Malayalam movie poster from 2017. Uses a mix of reformed and traditional orthography for title.

Figure 17: A Malayalam movie poster from 2016. Uses traditional orthography for title.

7 Conclusion

Technology played a crucial role in defining the orthography of Malayalam from printing to digital age. When technology, such as typewrites, had limitations, the Malayalam script went through a difficult reformation. But it flourished again with the help of digital technology. A single human generation is witnessing Malayalam’s
transition from traditional orthography to reformed orthography and then again to traditional orthography. The fact has been explained through numerous examples in this paper.

Reformed orthography is sometimes referred as modern and traditional as old. But as traditional script is getting more popular in contemporary usage, calling it old may not be right. So we consciously avoided those phrases in this paper.

References


