



4.2.2 Malayalam Script Details

A. Explanations for Revised Malayalam Code Chart

1. Encoding principles

The Malayalam writing system is mostly syllabic. The predominant orthographic unit is a vowel ending syllable with the canonical structure (C)V. The obligatory V represents a short or long vowel. The optional C represents one or more consonants. Except in a few instances the system follows the principles of phonology and mostly corresponds to the pronunciation. The system involves the following distinct character types:

- (i) Signs representing a single consonant followed by the inherent short vowel /a/. e.g. ക (0D 15) represents the consonant /k/ followed by /a/.
- (ii) Seven signs representing pure consonants (i.e. without any inherent /a/). Five (അ ഊ ഓ ഐ ഏ) of these are referred to as *chillu*, one *anusvaaram* (ഓ) (0D 02) and the remaining one *visargam* (ഔ) (0D 03).
- (iii) Signs representing certain consonants in consonant clusters:
eg: signs for /g/ and /m/ in ഗമ ((0D 17) +(0D 4D) +(0D 2E)); those for /ñ/ and /c/ in ഞച്ച ((0D 1E) +(0D 4D) +(0D 1A)) etc.
- (iv) Signs representing independent vowels. e.g. അ(0D 05)/a/, ഇ(0D 07)/i/, ഉ(0D 09)/u/ etc.
- (v) Signs representing dependent vowels. e.g. ഐ (0D 3F)/i/, ഊ(0D 41) /u/ etc.

Principles of the Malayalam Script

Appearance of the characters in the Malayalam script is affected by the following factors: (1) Ordering of the concerned character with respect to other characters (2) The font employed and (3) The application or system environment. These variables can cause the appearance of the Malayalam characters to differ from their nominal glyphs (used in Unicode charts).

Consonant Letters

Each consonant letter represents a single consonant sound followed by the inherent vowel /a/ thereby making an orthographic syllable.

Consonant letters may also be rendered as **half forms** which go into the constitution of consonant conjuncts. Only those half forms which represent the final member of a consonant conjunct has an inherent /a/.

4. Independent Vowel Letters

Independent vowels in Malayalam are signs that stand on their own. These are used to write syllables, which start with a vowel.

5. Dependent vowel signs

These occur only in combination with a base consisting of a sign for a single consonant or a consonant cluster. When the vowel quality of the syllable is different from that of the inherent /a/, it is represented by the respective dependent vowel sign. Explicit appearance of a dependent vowel in a syllable overrides the inherent vowel of the consonant/consonant cluster. eg. കാ കി കീ കു കൂ ക്യ കെ കേ കൈ കൊ കോ കൗ കൗ

6. Adding dependent vowel signs to consonant sign bases

In the traditional writing system dependent non-spacing signs for the short and long varieties of /u/ were attached to the respective consonant sign base. But the present day Malayalam writing system uses spacing for all the dependent vowel signs. These belong to the following three types:

- (i) There is a set of eight signs which follow the base representing a consonant or a consonant cluster. Among them ഐ (0D 3E), ഐ (0D 3F), ഐ (0D 40), ഊ (0D 41) and ഊ (0D 42) respectively correspond and are phonetically equivalent to the following independent vowel signs: അ (0D 06), ഇ (0D 07), ഊ (0D 08), ഉ (0D 09) and ഉ (0D 0A) whereas ഊ (0D 43) and ഊ (0D 57) respectively correspond and have the same phonetic value as the independent sign for the vocalic R റ (0D 0B) and that for the diphthong /au/ ഓ (0D 14). The eighth dependent sign, *chandrakkala* ഏ (0D 4D) representing the central vowel or /u/ with spread lips does not have any corresponding independent vowel since none



is required by the code of Malayalam orthography as the concerned vowel never occurs utterance initially.

- (ii) Three signs which precede the consonant base: െ (OD 46), േ (OD 47), and ൈ (OD 48). These signs respectively correspond and are phonetically equivalent to the independent vowel signs: എ (OD 0E), ഐ (OD 0F) and ഐ (OD 10).
- (iii) Two signs which have the first glyph component preceding the consonant base and the second following it: ൊ (OD 4A), ോ (OD 4B). These are equivalent to the independent vowels ഒ (OD 12) and ഓ (OD 13) respectively.

7. Vowel omission sign

A crescent mark ൿ (OD 4D) called *chandrakkala*, placed on the right top of the base sign of the respective consonant or consonant cluster, indicates that the quality of the inherent vowel /a/ is to be omitted from the value of the orthographic syllable, thereby making the remaining consonant 'dead'. Note that in utterance final position as has already been stated (6) the same sign represents the central vowel, pronounced with spread lips. As no dead consonant can occur as final utterance in Malayalam, this dual value associated with the same sign leads to no ambiguity.

8. Consonant conjuncts

Malayalam has a large number of consonant conjunct forms, which serve as orthographic abbreviations of two or more adjacent letter forms. A consonant cluster is depicted with a conjunct glyph if available in the current font(s). In the absence of a conjunct glyph, the conjunct is depicted with the nominal Consonant forms with *chandrakkala* in between.

The following types of consonant conjunct formations are present.

- (i) Signs for 'strong' (or the so called 'duplicate') consonants:
 - (a) Glyphs involving duplication of the signs, repetition of the basic sign being effected at its bottom or to its right: ള്, ഴ്, ഺ്, ഻്, ി, ഴ്, ഴ്, ഴ്/ഠ

- (b) Subscripting a half form to the consonant base: ഴ, ഴ
- (c) Duplicating the basic sign to its right: ഴ, ഴ, ഴ
- (d) Post-posing a half form to the consonant base: ജ്, ഴ, ന്, ക്
- (e) Pre-posing a half form to the consonant base: ഴ
- (f) Subscripting a special sign to the consonant base: ഴ, ഴ, ഴ, ഴ, ഴ
- (g) Subscripting the sign of an aspirated consonant to that of an un-aspirated one: ഴ

- (ii) Signs for complex consonant conjuncts:

They are of the following types:

- (a) Those consisting of a consonant sign followed by a spacing half form or a special sign which represent the final 'live' consonant. ക്, ക്
- (b) Those consisting of a consonant sign representing a dead consonant and a subscripted dependent consonant sign which represents the final live consonant.
- (c) Those consisting of a consonant sign and a special sign; they represent the dead consonant and the final live consonant respectively: ക്, ഴ, ക്, ഴ, ക്

The special sign representing either റ or റ depending on the context is separated from the consonant letter and appears to be pre-posed to (ക്, ഗ്) it thereby resulting in the reversal of the sequence of elements in pronunciation.

Ligatures for complex consonant conjuncts involve the following types of combination:

- (a) The signs of the concerned consonants are ligated in one stroke: ക്, ഗ്
- (b) The consonant sign and the pre-posed partial form represents respectively the live and dead consonants: ക്, ഗ്, ഴ, ജ്, ന്, ഴ



(c) The consonant sign and the post-posed partial form represent respectively the dead and live consonants: ൠ, ൡ, ൢ

There are a few instances of misfit between the visual sign and the pronunciation:

ക, ഡ, ക്ഷ

Here the phonetic value of the ligature differs from that of the combination of the usual value of the component signs.

9. Rendering

Rules for Rendering

These rules describe the mapping between the Unicode characters and the Glyphs in a Malayalam font. It also describes the combining and ordering of those Glyphs.

It is important to note that in a font that is capable of rendering Malayalam, the set of glyphs is greater than the number of fonts in the Malayalam Unicode characters

Notations

The following notation applies to the description that follows.

C_n	Nominal glyph form of a consonant C as it appears in the code charts.
C_l	A live consonant, depicted identically to C_n
C_d	Glyph depicting the dead consonant form of the consonant C .
L_n	Nominal glyph form of a conjunct ligature consisting of two or more component consonants
V_{vs}	Glyph depicting the dependent vowel sign form of a vowel V
CH_n	The nominal glyph form of the vowel omission sign chandrakkala.
YAKAR	Dependent sign for the consonant YA
RAKAR	Dependent sign for the consonant RA or RRA
LAKAR	Dependent sign for the Consonant LA
VAKAR	Dependent sign for the Consonant VA

Dead Consonant Rule

When a consonant C_n precedes a *chandrakkala* CH_n it is considered to be a dead consonant C_d . A consonant C_n that does not precede a *chandrakkala* is considered to be a live consonant C_l .

$$KA_n + CH_n = KA_d$$

$$ക + ൠ = ക്$$

Rule for YAKAR: YAKAR is formed when Malayalam letter YA (0D2F) is at the end of a consonant / consonant cluster. The special sign (൧) is post posed to the consonant in such cases.

$$KA_d + YA_l \quad KA_d + YAKAR \quad \text{displayed output}$$

$$ക് + യ \quad ക + ധ \quad ക്ക$$

This rule does not apply to a strong consonant of ഐ. Note that when ഐ is added to ഐ, ഐ is formed (ഐ + ഐ = ഐ).

Rule for RAKAR: RAKAR is formed when Malayalam letter RA (0D30) or RRA (0D31) is at the end of a consonant / consonant cluster, a special sign '൲' is preposed to the consonant.

$$KA_d + RRA_l \quad KAl + RAKAR \quad \text{displayed output}$$

$$ക് + റ \quad ക + ൴ \quad ക്ക$$

$$GA_d + RA_l \quad GAl + RAKAR \quad \text{displayed output}$$

$$ഗ് + റ \quad ഗ + ൴ \quad ഗ്ക$$

This rule does not apply to a strong consonant of റ. Note that when റ is added to റ, റ (tta) is formed.

$$റ + ൴ + റ = റ്റ$$

Rule for LAKAR: LAKAR is formed when the Malayalam letter LA (0D32) is at the end of a consonant / consonant cluster. A special sign ഓ is put at the bottom of the consonant / consonant cluster.

$$KA_d + LA_l \quad KA_d + LAKAR \quad \text{displayed output}$$

$$ക് + ല \quad ക + ഓ \quad ക്ക$$

Rule for VAKAR: VAKAR is formed when the Malayalam letter VA (0D35) is at the end of a consonant cluster. The special sign '൵' is post posed to the consonant / consonant cluster.

$$KA_d + VA_l \quad KA_d + VAKAR \quad \text{displayed output}$$

$$ക് + വ \quad ക + ൶ \quad ക്ക$$



(Rule does not apply to a strong/duplicate consonant of വ. Note that when വ is added to വ, വു is formed)

വ + ് + വ = വു

Ligature Rules

Certain rules governing formation of ligatures in Malayalam is given below. The precise application of these rules depends on the availability of glyphs in the current font(s) being used to display the text.

If a dead consonant immediately precedes another dead consonant or a live consonant, then the first dead consonant may join the subsequent element to form a two part conjunct ligature form.

The different types of consonant conjunct formations are given in section 8.

Rule for strong or 'duplicate' consonant.

Conjunct ligatures formed by repetition of basic sign (slightly smaller in size) being affected at its bottom.

ഗ് + ഗ = ഗ്ഗ
 ഡ് + ഡ = ഡ്ഡ്
 ണ് + ണ = ണ്ണ
 പ് + പ = പ്പ
 ശ് + ശ = ശ്ശ
 സ് + സ = സ്സ
 റ് + റ = റ്റ

When the same combination റ + ് + റ is added to dead സ then the resulting complex conjunct glyph will be as shown below.

സ + ് + റ + ് + റ = സ്ട്ര

(a) Conjuncts formed by subscripting a half/partial form to the consonant base.

ട + ് + ട = റ്ട
 ദ + ് + ദ = റ്ദ

(b) Duplicating the basic sign to its right.

ത + ് + ത = ത്ത
 മ + ് + മ = മ്മ
 ഉ + ് + ഉ = ഉു

(c) Post posing a half form/partial to the consonant base.

ജ + ് + ജ = ജ്ജ
 ണ + ് + ണ = ണ്ണ
 ന + ് + ന = ന്ന
 ക + ് + ക = ക്ക

(d) Pre posing a half form to the consonant base.

ങ + ് + ണ = ണ്ങ

(e) Subscripting a special sign to the consonant base.

വ + ് + വ = വു
 ബ + ് + ബ = ബു
 ച + ് + ച = ചു
 യ + ് + യ = യു
 ല + ് + ല = ലു

(f) Subscripting the sign of an aspirated consonant to that of an unaspirated one.

ച + ് + ഹ = ചുഹ

Complex Consonant Conjuncts

Ligature formations for complex consonant are given below.

(g) The signs of concerned consonants are legated in one stroke.

ക + ് + ത = ക്ത
 ഗ + ് + മ = ഗ്മ

(h) The consonant sign and the pre-posed partial form represents respectively the live and dead consonants.

ഹ + ് + മ = ഹ്മ
 ഹ + ് + ന = ഹ്ന
 ന + ് + മ = ന്മ
 ണ + ് + മ = ണ്മ
 ണ + ് + ട = ണ്ട
 ജ + ് + ണ = ജ്ഞ
 ണ + ് + ച = ണച്ച
 ന + ് + ധ = ന്ധ
 ന + ് + ദ = ന്ദ

(i) The consonant sign and the post posed partial form represent respectively the dead and the live consonants.

ത + ് + ഫ = തഫ
 സ + ് + ഫ = സഫ
 ക + ് + ഷ = കഷ



The commonly used conjuncts and their formation sequence is given below.

Conjunct glyph	Component characters	Unicode representation
ക	ക + ് + ക	0D 15 + 0D 4D + 0D 15
ച	ച + ് + ച	0D 1A + 0D 4D + 0D 1A
ട	ട + ് + ട	0D 1F + 0D 4D + 0D 1F
ത	ത + ് + ത	0D 24 + 0D 4D + 0D 24
പ	പ + ് + പ	0D 2A + 0D 4D + 0D 2A
ഗ	ഗ + ് + ഗ	0D 17 + 0D 4D + 0D 17
ജ	ജ + ് + ജ	0D 1C + 0D 4D + 0D 1C
ഡ	ഡ + ് + ഡ	0D 21 + 0D 4D + 0D 21
ദ	ദ + ് + ദ	0D 26 + 0D 4D + 0D 26
ബ	ബ + ് + ബ	0D 2C + 0D 4D + 0D 2C
ഭ	ഭ + ് + ഭ	0D 19 + 0D 4D + 0D 19
ണ	ണ + ് + ണ	0D 1E + 0D 4D + 0D 1E
ന	ന + ് + ന	0D 23 + 0D 4D + 0D 23
മ	മ + ് + മ	0D 28 + 0D 4D + 0D 28
യ	യ + ് + യ	0D 2E + 0D 4D + 0D 2E
ര	ര + ് + ര	0D 2F + 0D 4D + 0D 2F
ല	ല + ് + ല	0D 32 + 0D 4D + 0D 32
വ	വ + ് + വ	0D 35 + 0D 4D + 0D 35
ശ	ശ + ് + ശ	0D 36 + 0D 4D + 0D 36
സ	സ + ് + സ	0D 38 + 0D 4D + 0D 38
ള	ള + ് + ള	0D 33 + 0D 4D + 0D 33
റ	റ + ് + റ	0D 31 + 0D 4D + 0D 31
ക്ത	ക + ് + ത	0D 15 + 0D 4D + 0D 24
ക്ഷ	ക + ് + ഷ	0D 15 + 0D 4D + 0D 37
ന്ത	ന + ് + ത	0D 23 + 0D 4D + 0D 28
ന്മ	ന + ് + മ	0D 23 + 0D 4D + 0D 2E
ങ്ക	ക + ് + ങ	0D 15 + 0D 4D + 0D 19
ച	ച + ് + ഛ	0D 1A + 0D 4D + 0D 1B
ജ	ജ + ് + ഞ	0D 1C + 0D 4D + 0D 1E
ബ	ബ + ് + പ	0D 2C + 0D 4D + 0D 1A
ഭ	ഭ + ് + ട	0D 19 + 0D 4D + 0D 1F
ണ	ണ + ് + മ	0D 23 + 0D 4D + 0D 2E
ത	ത + ് + ഡ	0D 24 + 0D 4D + 0D 25
മ	മ + ് + മ	0D 28 + 0D 4D + 0D 2E
ഭ	ഭ + ് + ട	0D 19 + 0D 4D + 0D 2D
സ	സ + ് + സ	0D 38 + 0D 4D + 0D 38
ദ	ദ + ് + ഡ	0D 26 + 0D 4D + 0D 27

ന	ന + ് + ത	0D 28 + 0D 4D + 0D 24
ന	ന + ് + ഡ	0D 28 + 0D 4D + 0D 25
ന	ന + ് + ട	0D 28 + 0D 4D + 0D 26
ന	ന + ് + ഡ	0D 28 + 0D 4D + 0D 27
മ	മ + ് + പ	0D 2E + 0D 4D + 0D 2A
മ	മ + ് + മ	0D 28 + 0D 4D + 0D 2E
ന	ന + ് + റ	0D 28 + 0D 4D + 0D 31
ബ	ബ + ് + ട	0D 2C + 0D 4D + 0D 26
ബ	ബ + ് + ഡ	0D 2C + 0D 4D + 0D 27
ശ	ശ + ് + ച	0D 36 + 0D 4D + 0D 1A
ഷ	ഷ + ് + ട	0D 37 + 0D 4D + 0D 1F
സ	സ + ് + ഡ	0D 38 + 0D 4D + 0D 25
ഹ	ഹ + ് + ന	0D 39 + 0D 4D + 0D 28
ഹ	ഹ + ് + മ	0D 39 + 0D 4D + 0D 2E

Rule for alternate representation of conjuncts

Conjuncts can also be shown with the *chandrakala* in between the constituent consonants, without formatting the conjunct. This can be achieved by typing *chandrakala* twice.

Glyph	Alternate representation	Formation sequence
ക	ക്ക	ക + ് + ് + ക
ത	ക്ത	ക + ് + ് + ത
ത	ക്ത	ത + ് + ് + ത
പ	പ്പ	പ + ് + ് + പ
സ	സ്മ	സ + ് + ് + ഡ

Rule :

YAKAR, RAKAR, VAKAR, and LAKAR combinations wouldn't come after chillu characters.

ന + റ = ന്റ not ന്റ

ത + വ = ത്വ not ത്വ

ശ + യ = ശ്യ not ശ്യ

ബ + ല = ബ്ല not ബ്ല

However the Malayalam letter റ 'RRA' (0D 31) when combined with chillu 'ന്റ' (0D 5C) will give rise to the soft conjunct ന്റ (nda).

If 'ന്റ' and 'റ' are to be separately displayed as in name ഹെന്ററി (Henry) then 'ന്റ' and 'റ' has to be separated by '്'.



Memory Representation and Rendering Order

The order for storage of plain text in Malayalam follows the phonetic order. A CV syllable with a dependent vowel is always encoded as a consonant letter C followed by a vowel sign V in the memory representation if the dependent vowel is one to depicted at the right side of the consonant . This order is employed by the ISCII standard and corresponds with both the phonetic and keying order of the textual data.

Rendering Order

Character Order

$$KA_n + AA_{vs} = KAA_n$$

$$ക + ഓ = കോ$$

Since some of the dependent Malayalam vowel signs must be depicted to the left side of the consonant letter, and some others on both sides, the software that renders the Malayalam script must be able to reorder elements in mapping from the logical (character) store to the presentational (glyph) rendering.

Dependent vowels on the left side of the consonant sign

When the dependent vowel signs E_{vs} (െ), EE_{vs} (േ) and AI_{vs} (ഈ) are used to override the inherent vowel of a syllable it is always written to the extreme left of the orthographic syllable. If the orthographic syllable contains a consonant cluster then this vowel sign is always depicted to the left of that cluster.

$$KA_n + E_{vs} = KE_n \text{ (Displayed output)}$$

$$ക + െ = കെ$$

$$KA_n + AI_{vs} = KAI_n$$

$$ക + ൈ = കൈ$$

Two part vowel

Two of the dependent vowel signs in Malayalam consist of two discontinuous elements as in other cases of discontinuous elements, two sequences of Unicode values can be used to express the component elements.

$$O_{vs} = E_{vs} + AA_{vs}$$

$$ഌ (0D 4A) = െ (0D 46) + ഓ (0D 3E)$$

$$OO_{vs} = EE_{vs} + AA_{vs}$$

$$ഔ (0D 4B) = േ (0D 47) + ഓ (0D 3E)$$

Rule :

When the dependent vowel O (ഌ (0D 4A) = െ (0D 46) + ഓ (0D 3E)) and OO (ഔ (0D 4B) = േ (0D 47) + ഓ (0D 3E)) are used to override the inherent vowel of the syllable, it pieces on both sides of the syllable. The first sign (െ) corresponding to E (0D 46) and EE (0D 47) is placed on the left side of the syllable and the last sign (ഓ) (corresponding to 0D 3E) is placed at the right side of the syllable.

$$KA_n + O_{vs} = KO_n \text{ (Displayed Output)}$$

$$ക + ഌ = കോ$$

$$KA_n + OO_{vs} = KOO_n \text{ (Displayed Output)}$$

$$ക + ഔ = കോ$$

(Courtesy : Shri R. Ravindra Kumar
ER & DCI, Vellayambalam,
Thiruvananthapuram 695 033
Tel. 0471-723333 Fax: 0471-723456
E-mail : ravi@erdcitvm.org)