SCRIPT GRAMMAR FOR MALAYALAM LANGUAGE

Prepared by

Technology Development for Indian Languages (TDIL) Programme of DIT, GoI in association with

Centre for Development of Advanced Computing (C-DAC)

Table of Contents

0. INTRODUCTION

The term **script grammar** refers to the behaviour pattern of the writing system of a given language. Languages which have written representations do not use a haphazard manner of storing the information within the system, but use a coherent pattern which is similar to the linguistic grammar of a given language. With the help of specialists (not necessarily linguist) who work in the area of the written representation of the language, the manner in which the shapes of the characters of the language and the representation of the conjunct forms is provided. In other words the Script Grammar deals with the surface structure of the language and tries to provide the best possible "fit" for shapes and their representation. Since this is a highly subjective issue, the shapes provided here are recommendations at the best and conform to the perception of the mandating body/evaluators who consensually arrive at the "best possible fit" which is acceptable to a majority of users. An example will make the above clear. Although Marathi and Hindi share the same script Devanāgarī, not only do they not share the same character inventory but in addition the representation of certain characters is different. Thus the Marathi /la/ is different from the Hindi /la/ in so far as the placement of the stem is concerned Hindi ल

Marathi ल. This ensures that the Script Grammar conforms to the language in question and provides the character shapes acceptable to a given user community. It should be noted that this does not mean monotony. The Hindi and Marathi /la/ can have a variety of forms once the intrinsic structure of the character is determined.

Script Grammar is the term used to define:
- the writing system used to inscribe a given language
- the history of the script and language (wherever available)
- the syllabic structure of the writing system of the language
- the rule ordering of the characters within the syllable (akshar)
- description of the syllabic clusters
- collation order of the characters: lexical / dictionary sorting order

1.   OBJECTIVES OF SCRIPT GRAMMAR

The Objectives of the script grammar for each language can be divided into two major parts:

Societal:
- Provide a visual representation of shapes that are deemed to be in conformity with the perception of a given community

- Ensure thereby that this perception is safe-guarded

- Through wide-spread dissemination and creation of appropriate tools ensure that within the given linguistic community, all media tries to adopt the given shape.

Technical:
- Classify the language in terms of its ISO and also whether it belongs to the Abjad, Akshar (Alphasyllabary) class.

- Provide an inventory of the characters pertinent to the language and classify the same in terms of their taxonomy.

- As a corollary determine whether the inventory is in conformity to the Syllable formalism as stipulated in ISCII'91 and subsequently adopted by Unicode.

- Since Brahmi is written from left to right, and since certain characters do not follow the linear LtoR order, provide an inventory of displaced catenators i.e. characters such as Mātrās that concatenate to the Consonant

- Propose the best shape representation of the individual characters as well as of the ligatures used within a given script. As a corollary request the expert(s) to identify the largest possible strings of such ligatures.

- Finally provide the collation order pertinent to that script/language, which would be of great utility to high-end NLP as well as to CLDR's in the pertinent language. The collationorder for Marathi is different from Hindi although both languages share the same script. Thus in Marathi क्ष, ज्ञ are placed at the end of the consonant inventory i.e. after ह in the sort order. In Hindi क्ष is sorted along with क and ज्ञwith ज

2.   END USERS FOR SCRIPT GRAMMAR

The script-grammar specific to a given language can be used by a large number of users.

- Most importantly it can be used by font developers desirous of developing a font which is compliant with the perception of the characters and ligatures of a language by its user community.

- Certain features of the script grammar such as the shapes can also be used for testing OCR and OHWR. Similarly information regarding Ligatures as well as collation order can help in high-end NLP work such as detecting invalid combinations, correct implementation of syllable structure, prediction routines to name a few. Information regarding collation and character sets can be also used for CLDR.

- They allow the font designer to design a font which is in compliance with the norms and standards of that particular script. A major problem which will be dealt with in the template is one of ligatures. The final list of ligatures defined by the script grammar allows the font designer to write specific rules for such glyphs.

- It permits the software developer to design and implement the keyboard and the input mechanism which will meet the requirement of the particular linguistic community.

- The collation or sort order as described in a Script Grammar permits the software developer to write software functions/ routines for sorting data in all applications.

- Script Grammars are equally important for keyboard design, especially when supplemented by frequency data from a corpus.

As can be seen the script grammar has a wide range of use and can be of utility to font developers, Indian language developers and linguists in the area of computation.

3. SCOPE

This script grammar document contains following information about the language and the script used for writing the language.

1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1 & 639.3 standard.
2. Script used to inscribe the given language
3. The structure of the script used for writing the language
   - Rule ordering of the characters within the syllable formation is a language
   - Description of the syllabic clusters of the script
   - Collation order of the characters: lexical / dictionary sorting order
   - Compliance of the script with Unicode.

These will be treated within the relevant sections of the script grammar

4. TERMINOLOGY[1]

Abjad: A writing system in which each symbol always or usually stands for a consonant. The long vowels are indicated. However the short vowels are rarely marked and the reader needs to supply these. Example: Urdu written in Perso-Arabic Script is an example of this writing system.

Abugida: also called an alphasyllabary, is a segmental writing system in which consonant–vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel notation is obligatory but secondary[2]

Akshar: see **Abugida**

Allographs: Variants of the representation of a character. Thus ae and æ [U+00E6] in Latin alphabet are allographs.

Allo-Script: The term relates to languages which share a common script. Thus Devanāgarī is used to write 9 official languages. However these languages do not use the same set of characters. Thus Marathi uses the retroflex lla ळ [U+ 0933] which Hindi does not use. Flaps used in Hindi ड़ [U+095C] ढ़ [U+095D] are not used in Konkani. These sub-sets of scripts based on a single "matricial" script are termed as allo-scripts.

Alphabet: A set of letters used in writing a language. Example: The English Alphabet.

Aspirated consonant: A consonant which is pronounced with an extra puff of air coming out at the time of release of the oral obstruction. This has a sound of an extra "h".

Basic alphabet: The minimal set of letters which can be used for uniquely encoding every word of a language. The basic alphabet for English consists of only the upper-case letters A-Z

Catenators: Also termed as Concatenators are characters which are concatenated to another character. In the Brahmi script these are the Mātrās or Vowel modifiers which are adjoined to the consonant and add a vocalic value to the consonant.

Conjunct: The Indic scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms. This abbreviation takes place only in the context of a consonant cluster….Under normal circumstances, a consonant cluster is depicted with a conjunct glyph if such a glyph is available in the current font. In the absence of a conjunct glyph, the one or more dead

---

[1] As in the case of the BIS Document, in order to make the terminology accessible for all readers, examples have been chosen from English/Latin scripts, wherever possible. Some definitions have been excerpted from the BIS ISCII91 document and suitably modified where necessary.
[2] Wikipedia definition

consonants that form part of the cluster are depicted using half-form glyphs. In the absence of half-form glyphs, the dead consonants are depicted using the nominal consonant forms combined with visible virama signs.[3]

Consonant: A letter representing a speech sound in which the breath is at least partly obstructed,

Diacritic:A mark added to a letter which distinguishes it from the same letter without a mark, usually having a different phonetic value or stress.

Displaced Catenator: (see Catenator) Within the Brahmi script, the writing system is linear and moves from left to right. However in the case of some catenators this rules is not observed and the catenator (wholly or partially) is placed to the right of the consonant to which it relates. The short vowel I in Devanāgarī is an example of a displaced catenator.

Display composing: The process of organizing the basic shapes available in a font in order to display (or print) a word.

Display rendition: The process by which a string of characters is displayed (or printed). In this process several consecutive characters may combine with each other on the screen. The sequence of display of the characters may become different.

Eyebrow repha: (See **Eyelash ra)**

Eyelash ra:The eyelash ra is used in Konkani, Nepali and Marathi. It is treated as different from the र (repha) by certain linguists. While the former is treated as a flap, the latter is a continuant trill (*cf.*, Kalyan Kale and Anjali Soman. 1986). There are cases in Marathi of minimal pairs such as: आचार्यास"to the teacher" vs. आचाऱ्यास"to the cook" or दर्या /darya/ "ocean" vs. दऱ्या /darya/ "valleys".

Font: A set of symbols used for display or printing of a script in a particular style.

International numerals: The conventional 0 to 9 digits used in English for denoting numbers. these are also known as Indo-Arabic numerals (to differentiate them from the Roman numerals like IX for 9).

Latin alphabet: The alphabet used for writing the language of ancient Rome. Also known as the Roman alphabet. The alphabet is used today for writing English and European languages.

Letter: A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

---

[3]Unicode 6.0 Chapter 9.0 pp 6-7

Ligature: (see **Conjunct**)

Nasal consonant: A consonant pronounced with the breath passing through the nose. Example *m n*inEnglish.

Nasalized vowel: A vowel pronounced with the breath passing both through the nose and the mouth. In Indian scripts this is denoted by a Chandrabindu and gives the vowel/vowel modifier over which it placed a nasal value. Example: जाँच

Phonetic alphabet: An alphabet which has direct correspondence between letters and sounds Example: The International Phonetic Alphabet..

Pure consonant: A consonant which does not have any vowel implicitly associated with it.

Rafar: A special case of a ligature constituted by the adjunction of ra followed by a halanta to consonant. The resultant combination places the ra on top of the consonant to which it is adjoined. In case the consonant itself is adjoined to another consonant, the rafar is placed above the consonant e.g. र्+क र्क, र्+घ्+य र्घ्य

Rakar: A special case of a ligature constituted by the adjunction of a consonant followed by a halanta to ra. In a large number of Brahmi scripts the ra is adjoined to the stem of consonant to which it relates. In the case of consonants which have no stem such as the dental retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates.

Repha: (see **Rafar**)

Roman script: The script based on the ancient Roman alphabet, with the letters A-Z and additional diacritic marks. Used for writing a language which is not usually written in the Roman alphabet.

Script: A distinctive and complete set of characters used for the written form of one or more languages.

Script numerals: The 0 to 9 digits in a script, which have shapes distinct from their international counterparts.

Syllable: A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel or diphthong sound optionally surrounded by one or more consonants

Transliteration: Representation of words with the closest corresponding letters in an alphabet of a different language.

Vowel: A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction

Vowel sign: A graphic character associated with a letter, to indicate a vowel to be associated with that character (Mātrā in Hindi).

5.PHILOSOPHY AND UNDERLYING PRINCIPLES

The script grammar is based on the following principles:

1. The Grammar aims to depict the surface grammar of the written language: the manner in which characters as well as conjuncts are depicted

2. Where a given script admits many languages, it is pre-suppose that such languages will prescribe different representations for a given shape or conjunct according to the perception of the native users of that language

3. Corollary to the above the result is a script and allo-scripts i.e. a given script shared by many languages is not uniformly deployed across all the languages but is subject to variations and modulations.

4. The term Grammar is used here in a non-normative sense: what is prescribed is in the form of recommendations provided by experts who visualize the shape of the given script in their mother tongue in a specific manner. Subjective variations may occur[4]

5. The Grammar is limited to its synchronic use i.e. the manner in which a given language as of today admits a character set within the script used to write it. It is not diachronic or historical in nature and does not study the evolution of the given script across centuries.

---

[4] It is recommended that such variations be culled by placing the Grammars of different scripts in public review.

6. SCRIPT GRAMMARSTRUCTURE

The script grammar provided below has the following parts.

Part 6.1. deals with peripheral elements such as the ISO of the language, the writing system used: (Alphasyllabic) Abugida or Abjad.

Part 6.2. treats of the syllabic structure. It verifies whether the character set of the language complies with the ISCII syllabic structure and if not which cases are not compliant.

Part 6.3 is the script grammar proper and describes the character set as well as the conjunct shapes of the given script along with the collation order

12

6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR

These constitute the elements that are peripheral to the Script Grammar. The main parameters considered are the mnemonic and name of the language (needed for CLDR and also for language tags), the writing system used to inscribe the language and wherever possible a short history of the language.

6.1.1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1. & 639.3

Name of the Language: MALAYALAM
ISO Mnemonics: *mal*
This refers to a one line description of the language and its mnemonic representation as per the ISO. In the case of Malayalam, the above information is pertinent.

6.1.2. Identification of the writing system(s) used to inscribe the given language

Malayalam is written using the Malayalam script. It is an alphasyllabary with the akshar as its core.
This is a one line description of the script used to write the language. However in case the language uses more than one script, all the scripts in question are specified, provided these constitute the official language of the given state.

All scripts derived from Brahmi are Abugidas i.e. syllabary driven systems. The main features of Abugidas are as under:

- The consonant has an implicit vowel built-in which is normally the schwa.
- The inherent vowel can be modified by the addition of other vowels or muted by a diacritic termed as a Virama or Halanta
- Vowels can be handled as full vowels with a vocalic value

When two or more consonants join together they form ligatures eg. � **.**

Abugidas/Alphasyllabaries because of their syllabic structure require a special description which is the subject of the discussion in 6.2. below.

6.1.3. Amendments needed inUnicode for Malayalam language
The experts have not deemed fit to propose any amendments to Unicode for Malayalam

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

Malayalam language complies with the syllable (akshar) structure described above. It can admit up to 4 consonant clusters.

Alphasyllabaries are determined by the notion of the syllable or the Akshar. The compositional grammar of the syllable determines it well-formedness. This is through a series of formal constraints based on a Backus-Naur Formalism which is given below. The syllable (akshar), first defined in the ISCII document (1991), identifies the following character 'sub-sets' for the purposes of identifying the syllable (akshar). In what follows the syllable analysis will be restricted to Malayalam.

(C) Consonants

ക	ഖ	ഗ	ഘ	ങ
ച	ഛ	ജ	ഝ	ഞ
ട	ഠ	ഡ	ഢ	ണ
ത	ഥ	ദ	ധ	ന
പ	ഫ	ബ	ഭ	മ
യ	ര	ല	വ	
ശ	ഷ	സ	ഹ	
ള	ഴ	റ		

(V) Vowels

അ	ആ	ഇ	ഈ	ഉ	ഊ	ഋ	എ	ഏ
ഐ	ഒ	ഓ	ഔ					

(M) Mātrās or Vowel Modifiers

ാ	ി	ീ	ു	ൂ	ൃ	െ	േ	ൈ	ൊ	ോ	ൗ

(D) Diacritics

ം Anuswaram	For example, അംഗം.The Anuswaram is placed to the right of the Character to which it is adjoined. This falls under Nasal category.
ഃ Visarga	Visarga, denoted by two dots placed one above the other and placed to the right of the Character to which it is adjoined. For example,ദുഃഖം
ഽ : Avagraha/ Praslesham	Avagraha which is rarely used has been omitted by the experts.

(H):Chandrakkala (Halanta): $\overset{\circ}{\bigcirc}$

Chandrakkala (Halanta) used in most writing systems of the Indian subcontinent to signify the lack of an inherent vowel.

(N)[5] Nukta $\bigcirc$ -is Not used in Malayalam

Each of these sub-types has its restrictions in terms of what can precede or follow it within a syllable (akshar), as shown in the table below:

PRECEDED BY	SUBTYPE	FOLLOWED BY
-, H	**C**	M,D,H
-	**V**	D
C	**M**	D
C,V,M	**D**	-
C	**H**	C

C can be preceded by H or no subtype and followed by any one of the following: M,D,H
V can be preceded by no subtype and followed by D but not by another sub-type.
M can be preceded by C and followed by D.
D can be preceded by C, V, M and followed by no other subtype. It closes the syllable (akshar).
H can be preceded by C alone and followed only by C and no other sub-set.

In the case of Malayalam however, a set of consonants termed as Chillu characters need special treatment in the BNF. As per the experts there are 5 chillu characters in Malayalam. These are as under:

ണ്	ന്	ര്	ല്	ള്

Their theoretical distribution is as under:

PRECEDED BY	SUBTYPE	FOLLOWED BY
C, V	Chillu	C, V
C,V	Chillu	C,V

6.2.1.Syllable (akshar) Types

---

[5] The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving 5 other consonants loaned from Urdu क़,ख़,ग़,ज़,फ़

The formalism defines the syllable (akshar) in terms of both what can constitute a syllable (akshar) and what cannot. A valid syllable (akshar) as per this definition can be of only two types:

1. A vowel syllable (akshar): a full vowel.
2. A consonant syllable (akshar): a full consonant (having a weak vowel or a mātrā)

The three other subsets viz. Mātrās, Diacritics, Chandrakkala (Halanta) cannot constitute a syllable (akshar) by themselves or in combination among themselves.

6.2.1.1. The Vowel syllable (akshar) is of the following types:

6.2.1.1.1. A pure vowel all by itself: അ /a/ ,ആ/ā/ etc.
6.2.1.1.2. A vowel followed by a modifier i.e. either a nasal marker or a visarga:
ഇം /im /, അഃ /aH/
6.2.1.2. The Consonant syllable (akshar) can be of the following types:
6.2.1.2.1. A full consonant (with or without Nukta) i.e. with the inherent vowel : ക /ka/
6.2.1.2.2. A consonant followed by a mātrā i.e. the inherent vowel being substituted by another vowel: കി /ki:/
6.2.1.2.3. A consonant followed by a modifier: കം/kam/,കഃ/kaH/
6.2.1.2.4. A consonant followed by a mātrā and a modifier: കിം/kim/, ദുഃ/duH/.
6.2.1.2.5. A consonant cluster i.e. a dead or half consonant (Consonant+Chandrakkala (Halanta)) followed by a full consonant followed optionally by a mātrā, a modifier or a combination of both. These result in a ligature or what is often termed as *yuktakshara.*
        ത്സ/tsa/ ത്സം/tsam/,ക്കഃ /kkaH/ ത്സും /tsum/, ത്സു /tsu/.
The above permutations and combinations result in 7 major syllable (akshar) types. Of these the last type introduces the problem of the number of consonant clusters. ISCII (91, p.23) provides for up to four consonant clusters as the "worst case" i.e. the largest possible string.
This means that theoretically the following forms can be postulated:
    1.   Vowel Set: With the Vowel as the node.
            V   VD
    2.   Consonant set: With the Consonant as the node (an implicit or modified vowel is pre-implied).

Node	Mātrā	Modifier	Mātrā+Modifier
C	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD
CHCHCHC	CHCHCHCM	CHCHCHCD	CHCHCHCM

A total number of 16 theoretical syllables is therefore possible. It will be seen that the written syllable (akshar) is not very different in structure from the phonetic syllable and

that the movement from the written to the spoken levels is made feasible by application of certain rules.

This formal structure of the syllable (akshar) explained above is common to all Brahmi based scripts (with a few variations). It will form the basis of an exhaustive description of the characters as well as their ligatural representations.

6.3 SCRIPT GRAMMAR PROPER

This section lays down in detail the different parameters of the Script Grammar for Malayalam. These are:

6.3.1. The Character Set of Malayalam.
6.3.2. The Consonant mātrā combinations of Malayalam.
6.3.3. The Ligature Set of Malayalam.
6.3.4. Collocation Order of Malayalam

6.3.1. The Character Set of Malayalam.

This section provides detailed information about the characters in the language and the list of the same and also more importantly shows the manner in which the character is to be written. Each subsection comprises therefore two parts: the basic character set and the shape each character should have, as mandated by the experts who have designed the script grammar of Malayalam.

This comprises the following:
6.3.1.1. The Consonant Set
6.3.1.2. The Vowel Set
6.3.1.3. The Mātrā Set
6.3.1.4. Displaced Catenators
6.3.1.5. Shape of the combination of ra (rakar,repha)
6.3.1.6. The Set of Diacritics
6.3.1.7. Numerals
6.3.1.8. Punctuation marks
6.3.1.9. Other symbols
Each of these will be analysed in detail:

6.3.1.1. The Consonant Set

The Consonant set of Malayalam comprises the following characters:

Basic Consonant inventory arranged as per their Vargas.

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	ക	ഖ	ഗ	ഘ	ങ
Palatal	ച	ഛ	ജ	ഝ	ഞ
Retroflex	ട	ഠ	ഡ	ഢ	ണ
Dental	ത	ഥ	ദ	ധ	ന
B-labial	പ	ഫ	ബ	ഭ	മ

Other consonants

യ	ര	ല	വ	ശ
ഷ	സ	ഹ	ള	ഴ
റ				

Malayalam admits the following 5 Chillu[6] Characters.

ണ്	ന്	ര്	ല്	ള്

The exact shapes as desired by the experts are provided in the table below:

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	ക	ഖ	ഗ	ഘ	ങ
Palatal	ച	ഛ	ജ	ഝ	ഞ
Retroflex	ട	ഠ	ഡ	ഢ	ണ
Dental	ത	ഥ	ദ	ധ	ന
B-labial	പ	ഫ	ബ	ഭ	മ

Other consonants

യ	ര	ല	വ	ശ
ഷ	സ	ഹ	ള	ഴ
റ				

---

[6] These have been incorporated in Unicode 6.0. but are included here as per Expert's recommendation so that the Script grammar is truly representative of Malayalam

Chillu Characters.

ൺ	ൻ	ർ	ൽ	ൾ

6.3.1.2. The Vowel Set

The Vowel set of Malayalam is as under:

അ	MALAYALAM LETTER A
ആ	MALAYALAM LETTER AA
ഇ	MALAYALAM LETTER I
ഈ	MALAYALAM LETTER II
ഉ	MALAYALAM LETTER U
ഊ	MALAYALAM LETTER UU
ഋ	MALAYALAM LETTER VOCALIC R
എ	MALAYALAM LETTER E
ഏ	MALAYALAMLETTER EE
ഐ	MALAYALAM LETTER AI
ഒ	MALAYALAM LETTER O
ഓ	MALAYALAM LETTER OO
ഔ	MALAYALAM LETTER AU

As per expert recommendations the character set should be written as under:

അ ആ ഇ ഈ ഉ ഊ ഋ എ ഏ ഐ ഒ ഓ ഔ

6.3.1.3. The Mātrā Set

The Mātrā (Vowel Modifier Set) of Malayalam is as under:

Mātrās Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
1. Malayalam sign AA	ാ	ആ	ക്+ ആ=കാ
2. Malayalam sign I	ി	ഇ	ക്+ ഇ= കി
3 Malayalam sign II	ീ	ഈ	ക്+ഈ= കീ
4. Malayalam sign U	ു	ഉ	ക്+ഉ= കു

19

5. Malayalam sign UU	◌ൂ	ഊ	ക്+ ഊ= കൂ
6. Malayalam sign vocalic R	◌ൃ	ഋ	ക്+ഋ= കൃ
7. Malayalam sign E	െ◌	എ	ക്+എ= കെ
8. Malayalam sign EE	േ◌	ഏ	ക്+ഏ= കേ
9. Malayalam sign AI	ൈ◌	ഐ	ക്+ഐ=കൈ
10. Malayalam sign O	െ◌ാ	ഒ	ക്+ഒ= കൊ
11. Malayalam sign OO	േ◌ാ	ഓ	ക്+ഓ= കോ
12. Malayalam sign AU	◌ൗ	ഔ	ക്+ഔ= കൗ

As per expert recommendations the character set should be written as under:

◌ാ ◌ി ◌ീ ◌ു ◌ൂ ◌ൃ െ◌ േ◌ ൈ◌ െ◌ാ േ◌ാ ◌ൗ

6.3.1.4. Displaced Catenators

Malayalam only admits Two part vowel modifiers i.e. one part of the modifier is placed to the left and the other to the right of the consonant to which these modifiers are adjoined.

CATENATOR	POSITION	EXAMPLE
െ◌ാ	Consonant in the middle	കൊ
േ◌ാ	Consonant in the middle	കോ

6.3.1.5. Shape of the combination of ra "ര" or " റ"

In Malayalam the combination of Consonant+ Chandrakkala+ ര or റ is shown as the

adjoining of a " ്ര"
ക്ര - ക ◌് ര
പ്ര - പ ◌് ര
ക്റ - ക ◌് റ
പ്റ - പ ◌് റ
The experts have recommended as under:

> Although " ്ര" can be allograph of both ര and റ, depending upon the context, a unique representation should be adopted in keyboard driver and font implementations. For compatibility reasons, ര is preferred at present. However there should be a detailed discussion in this matter and a final decision should be taken to avoid future lexicography problems.

In addition to the above, the experts have also added the following comments on allographs:

1. **ALLOGRAPHS OF വ**

   ക്വ - ക ് വ

   ത്വ - ത ് വ

2. **ALLOGRAPHS OF യ**

   ക്യ - ക ് യ

   പ്യ - പ ് യ

3. **ALLOGRAPHS OF ല or ള**

   ക്ല - ക ് ല

   പ്ല - പ ് ല

   ക്ള - ക ് ള

   പ്ള - പ ് ള

Th expert recommendation is as under:

> *Although* 𝌀[7]*can be an allograph of both* ല *and* ള*, depending upon the context, a unique representation should be adopted in keyboard driver and font implementations. For compatibility reasons,* ല *is preferred at present. However there should be a detailed discussion in this matter and a final decision should be taken to avoid future lexicography problems.*

6.3.1.6. Diacritics

ം Anuswaram

ഃ Visargam

Avagraha/ Praslesham which is rarely used has been omitted.

6.3.1.7. Numerals

The experts have deemed that English Numerals (Latino-Arabic set: 0,1,2,3,4,5,6,7,8,9) are alone used.
The Malayalam numeral set given below is rarely used.

---

[7] The character can exist only as conjunct shape and hence has been inserted as an image.

ഠ ൧ ൨ ൩ ൪ ൫ ൬ ൭ ൮ ൯ ൰ ൱ ൲

A sample specimen of a calendar using Malayalam numerals is provided below by the experts:

6.3.1.8. Punctuation Markers

Malayalam uses punctuation markers from the Latin set. such as . , ; : " ' ( ) [ ] etc.
A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
1.	Full Stop or Period	.
2.	Comma	,
3.	Question Mark	?
4.	Exclamation Mark	!
5.	Apostrophe	,
6.	Semi Colon	;
7.	Colon	:
8.	Hyphen	-
9.	Dash	--
10.	Ellipsis mark	...
11.	Oblique	/
12.	Double quotation mark	""
13.	Single quotation mark	"
14.	Cross	XXX
15.	As Above	- - " - -
16.	Round Brackets	( )
17.	Square Brackets	[]
18.	Curly Brackets	{ }

6.3.1.9 Other Symbols

These are religious, currency markers etc. included in Unicode:

൹-Malayalam Date Mark

൰-Malayalam number Ten

൱- Malayalam number one hundred

൲-Malayalam number one thousand

൳-Malayalam Fraction one quarter

൴- Malayalam Fraction one half

൵- Malayalam Fraction three quarters

₹: Rupee Sign as mandated by Government of India.

6.3.2. Consonant Mātrā Combinations.

These refer to the shapes generated when a Mātrā is adjoined to the Consonant. The layout of these is in the shape of a matrix where the first horizontal row refers to the active consonant and the first vertical column refers to the vowel-modifier.

Due to constraints of space and also clarity, for each class a series of 3 tables are provided.

Table 1:	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ
Table 2:	ട	ഠ	ഡ	ഢ	ണ	ത	ഥ	ദ	ധ	ന
Table 3:	പ	ഫ	ബ	ഭ	മ	യ	ര	ല	വ	ശ
	ഷ	സ	ഹ	ള	ഴ	റ				

Wherever there is an X it implies that the combination does not exist. For the font developer this is an indication that for this particular combination which is not possible in the language but needs to be accommodated in the font table, a simple linear combination be provided.

e.g. Although the combination of ഖ+Chandrakkala+ഖ is theoretically not possible it needs to be handled at the font level in the anticipation that a user could type this combination. The font would show the following: ഖ്ഖ

The classes are as under:

6.3.2.1. refers to a simple concatenation of Consonant and Mātrā combinations.

6.3.2.2. refers to a concatenation of Consonant and Mātrā + Anuswaram marker combinations. Other diacritics such as avagraha and visarga have been avoided, since these are linear in nature, are adjoined to the combination and do not in any way modify the structure of the shapes.

6.3.2.1 Consonant and Mātrā combinations.

This set refers to a simple concatenation of Consonant and Mātrā.

Consonant and Mātrā combinationsSet 1

	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ
ാ	കാ	ഖാ	ഗാ	ഘാ	ങാ	ചാ	ഛാ	ജാ	ഝാ	ഞാ
ി	കി	ഖി	ഗി	ഘി	ങി	ചി	ഛി	ജി	ഝി	ഞി
ീ	കീ	ഖീ	ഗീ	ഘീ	ങീ	ചീ	ഛീ	ജീ	ഝീ	ഞീ
ു	കു	ഖു	ഗു	ഘു	ങു	ചു	ഛു	ജു	ഝു	ഞു
ൂ	കൂ	ഖൂ	ഗൂ	ഘൂ	ങൂ	ചൂ	ഛൂ	ജൂ	ഝൂ	ഞൂ
ൃ	കൃ	ഖൃ	ഗൃ	ഘൃ	ങൃ	ചൃ	ഛൃ	ജൃ	ഝൃ	ഞൃ
െ	കെ	ഖെ	ഗെ	ഘെ	ങെ	ചെ	ഛെ	ജെ	ഝെ	ഞെ
േ	കേ	ഖേ	ഗേ	ഘേ	ങേ	ചേ	ഛേ	ജേ	ഝേ	ഞേ
ൈ	കൈ	ഖൈ	ഗൈ	ഘൈ	ങൈ	ചൈ	ഛൈ	ജൈ	ഝൈ	ഞൈ
ൊ	കൊ	ഖൊ	ഗൊ	ഘൊ	ങൊ	ചൊ	ഛൊ	ജൊ	ഝൊ	ഞൊ
ോ	കോ	ഖോ	ഗോ	ഘോ	ങോ	ചോ	ഛോ	ജോ	ഝോ	ഞോ
ൌ	കൗ	ഖൗ	ഗൗ	ഘൗ	ങൗ	ചൗ	ഛൗ	ജൗ	ഝൗ	ഞൗ

Consonant and Mātrā combinations Set 2

This set is in continuation of set 1 which shows consonant and Matra combinations.

	ട	ഠ	ഡ	ഢ	ണ	ത	ഥ	ദ	ധ	ന
ാ	ടാ	ഠാ	ഡാ	ഢാ	ണാ	താ	ഥാ	ദാ	ധാ	നാ
ി	ടി	ഠി	ഡി	ഢി	ണി	തി	ഥി	ദി	ധി	നി
ീ	ടീ	ഠീ	ഡീ	ഢീ	ണീ	തീ	ഥീ	ദീ	ധീ	നീ
ു	ടു	ഠു	ഡു	ഢു	ണു	തു	ഥു	ദു	ധു	നു
ൂ	ടൂ	ഠൂ	ഡൂ	ഢൂ	ണൂ	തൂ	ഥൂ	ദൂ	ധൂ	നൂ
ൃ	ടൃ	ഠൃ	ഡൃ	ഢൃ	ണൃ	തൃ	ഥൃ	ദൃ	ധൃ	നൃ
െ	ടെ	ഠെ	ഡെ	ഢെ	ണെ	തെ	ഥെ	ദെ	ധെ	നെ
േ	ടേ	ഠേ	ഡേ	ഢേ	ണേ	തേ	ഥേ	ദേ	ധേ	നേ
ൈ	ടൈ	ഠൈ	ഡൈ	ഢൈ	ണൈ	തൈ	ഥൈ	ദൈ	ധൈ	നൈ
ൊ	ടൊ	ഠൊ	ഡൊ	ഢൊ	ണൊ	തൊ	ഥൊ	ദൊ	ധൊ	നൊ
ോ	ടോ	ഠോ	ഡോ	ഢോ	ണോ	തോ	ഥോ	ദോ	ധോ	നോ
ൌ	ടൗ	ഠൗ	ഡൗ	ഢൗ	ണൗ	തൗ	ഥൗ	ദൗ	ധൗ	നൗ

Consonant and Mātrā combinations Set 3

This set is in continuation of set 2 which shows consonant and Matra combinations.

	പ	ഫ	ബ	ഭ	മ	യ	ര
○ാ	പാ	ഫാ	ബാ	ഭാ	മാ	യാ	രാ
○ി	പി	ഫി	ബി	ഭി	മി	യി	രി
○ീ	പീ	ഫീ	ബീ	ഭീ	മീ	യീ	രീ
○ു	പു	ഫു	ബു	ഭു	മു	യു	രു
○ൂ	പൂ	ഫൂ	ബൂ	ഭൂ	മൂ	യൂ	രൂ
○ൃ	പൃ	ഫൃ	ബൃ	ഭൃ	മൃ	യൃ	രൃ
െ○	പെ	ഫെ	ബെ	ഭെ	മെ	യെ	രെ
േ○	പേ	ഫേ	ബേ	ഭേ	മേ	യേ	രേ
ൈ○	പൈ	ഫൈ	ബൈ	ഭൈ	മൈ	യൈ	രൈ
ൊ○	പൊ	ഫൊ	ബൊ	ഭൊ	മൊ	യൊ	രൊ
ോ○	പോ	ഫോ	ബോ	ഭോ	മോ	യോ	രോ
○ൗ	പൗ	ഫൗ	ബൗ	ഭൗ	മൗ	യൗ	രൗ

Consonant and Mātrā combinations Set 4
This set is in continuation of set 3 which shows consonant and Matra combinations.

	റ	ല	വ	ശ	ഷ	സ	ഹ
○ാ	റാ	ലാ	വാ	ശാ	ഷാ	സാ	ഹാ
○ി	റി	ലി	വി	ശി	ഷി	സി	ഹി
○ീ	റീ	ലീ	വീ	ശീ	ഷീ	സീ	ഹീ
○ു	റു	ലു	വു	ശു	ഷു	സു	ഹു
○ൂ	റൂ	ലൂ	വൂ	ശൂ	ഷൂ	സൂ	ഹൂ
○ൃ	റൃ	ലൃ	വൃ	ശൃ	ഷൃ	സൃ	ഹൃ
െ○	റെ	ലെ	വെ	ശെ	ഷെ	സെ	ഹെ
േ○	റേ	ലേ	വേ	ശേ	ഷേ	സേ	ഹേ
ൈ○	റൈ	ലൈ	വൈ	ശൈ	ഷൈ	സൈ	ഹൈ
ൊ○	റൊ	ലൊ	വൊ	ശൊ	ഷൊ	സൊ	ഹൊ
ോ○	റോ	ലോ	വോ	ശോ	ഷോ	സോ	ഹോ
○ൗ	റൗ	ലൗ	വൗ	ശൗ	ഷൗ	സൗ	ഹൗ

Consonant and Mātrā combinations Set 5

This set is in continuation of set 4 which shows consonant and Matra combinations with retroflex consonants ള and ഴ

	ള	ഴ
○ാ	ളാ	ഴാ
○ി	ളി	ഴി
○ീ	ളീ	ഴീ
○ു	ളു	ഴു
○ൂ	ളൂ	ഴൂ
○ൃ	ളൃ	ഴൃ
െ○	െള	െഴ
േ○	േള	േഴ
ൈ○	ൈള	ൈഴ
ൊ○	ൊള	ൊഴ
ോ○	ോള	ോഴ
○ൗ	ളൗ	ഴൗ

6.3.2.2 Consonant and Mātrā +Anuswaram

This set refers to a Consonant and Mātrā + Anuswaramcombinations.

Consonant and Mātrā +Anuswaram combinations -Set 1

	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ
○ാ○ം	കാം	ഖാം	ഗാം	ഘാം	ങാം	ചാം	ഛാം	ജാം	ഝാം	ഞാം
○ി○ം	കിം	ഖിം	ഗിം	ഘിം	ങിം	ചിം	ഛിം	ജിം	ഝിം	ഞിം
○ീ○ം	കീം	ഖീം	ഗീം	ഘീം	ങീം	ചീം	ഛീം	ജീം	ഝീം	ഞീം
○ു○ം	കും	ഖും	ഗും	ഘും	ങും	ചും	ഛും	ജും	ഝും	ഞും
○ൂ○ം	കൂം	ഖൂം	ഗൂം	ഘൂം	ങൂം	ചൂം	ഛൂം	ജൂം	ഝൂം	ഞൂം
○ൃ○ം	കൃം	ഖൃം	ഗൃം	ഘൃം	ങൃം	ചൃം	ഛൃം	ജൃം	ഝൃം	ഞൃം
െ○○ം	കെം	ഖെം	ഗെം	ഘെം	ങെം	ചെം	ഛെം	ജെം	ഝെം	ഞെം
േ○○ം	കേം	ഖേം	ഗേം	ഘേം	ങേം	ചേം	ഛേം	ജേം	ഝേം	ഞേം
ൈ○○ം	കൈം	ഖൈം	ഗൈം	ഘൈം	ങൈം	ചൈം	ഛൈം	ജൈം	ഝൈം	ഞൈം
ൊ○○ം	കൊം	ഖൊം	ഗൊം	ഘൊം	ങൊം	ചൊം	ഛൊം	ജൊം	ഝൊം	ഞൊം
ോ○○ം	കോം	ഖോ	ഗോ	ഘോ	ങോ	ചോ	ഛോ	ജോ	ഝോ	ഞോ
○ൗ○ം	കൗം	ഖൗം	ഗൗം	ഘൗം	ങൗം	ചൗം	ഛൗം	ജൗം	ഝൗം	ഞൗം

Consonant and Mātrā +Anuswaram combinations - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Anuswaram

	ട	ഠ	ഡ	ഢ	ണ	ത	ഥ	ദ	ധ	ന
◌ാ◌ം	ടാം	ഠാം	ഡാം	ഢാം	ണാം	താം	ഥാം	ദാം	ധാം	നാം
◌ി◌ം	ടിം	ഠിം	ഡിം	ഢിം	ണിം	തിം	ഥിം	ദിം	ധിം	നിം
◌ീ◌ം	ടീം	ഠീം	ഡീം	ഢീം	ണീം	തീം	ഥീം	ദീം	ധീം	നീം
◌ു◌ം	ടും	ഠും	ഡും	ഢും	ണും	തും	ഥും	ദും	ധും	നും
◌ൂ◌ം	ടൂം	ഠൂം	ഡൂം	ഢൂം	ണൂം	തൂം	ഥൂം	ദൂം	ധൂം	നൂം
◌ൃ◌ം	ടൃം	ഠൃം	ഡൃം	ഢൃം	ണൃം	തൃം	ഥൃം	ദൃം	ധൃം	നൃം
െ◌◌ം	ടെം	ഠെം	ഡെം	ഢെം	ണെം	തെം	ഥെം	ദെം	ധെം	നെം
േ◌◌ം	ടേം	ഠേം	ഡേം	ഢേം	ണേം	തേം	ഥേം	ദേം	ധേം	നേം
ൈ◌◌ം	ടൈം	ഠൈം	ഡൈം	ഢൈം	ണൈം	തൈം	ഥൈം	ദൈം	ധൈം	നൈം
ൊ◌◌ം	ടൊം	ഠൊം	ഡൊം	ഢൊം	ണൊം	തൊം	ഥൊം	ദൊം	ധൊം	നൊം
ോ◌◌ം	ടോം	ഠോം	ഡോം	ഢോം	ണോം	തോം	ഥോം	ദോം	ധോം	നോം
◌ൌ◌ം	ടൌം	ഠൌം	ഡൌം	ഢൌം	ണൌം	തൌം	ഥൌം	ദൌം	ധൌം	നൌം

Consonant and Mātrā +Anuswaram combinations - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Anuswaram

	പ	ഫ	ബ	ഭ	മ	യ	ര
◌ാ◌ം	പാം	ഫാം	ബാം	ഭാം	മാം	യാം	രാം
◌ി◌ം	പിം	ഫിം	ബിം	ഭിം	മിം	യിം	രിം
◌ീ◌ം	പീം	ഫീം	ബീം	ഭീം	മീം	യീം	രീം
◌ു◌ം	പും	ഫും	ബും	ഭും	മും	യും	രും
◌ൂ◌ം	പൂം	ഫൂം	ബൂം	ഭൂം	മൂം	യൂം	രൂം
◌ൃ◌ം	പൃം	ഫൃം	ബൃം	ഭൃം	മൃം	യൃം	രൃം
െ◌◌ം	പെം	ഫെം	ബെം	ഭെം	മെം	യെം	രെം
േ◌◌ം	പേം	ഫേം	ബേം	ഭേം	മേം	യേം	രേം
ൈ◌◌ം	പൈം	ഫൈം	ബൈം	ഭൈം	മൈം	യൈം	രൈം
ൊ◌◌ം	പൊം	ഫൊം	ബൊം	ഭൊം	മൊം	യൊം	രൊം
ോ◌◌ം	പോം	ഫോം	ബോം	ഭോം	മോം	യോം	രോം
◌ൌ◌ം	പൌം	ഫൌം	ബൌം	ഭൌം	മൌം	യൌം	രൌം

27

Consonant and Mātrā +Anuswaram combinations - Set 4

This set is in continuation of set 3 above which shows combinations of Consonant and Mātrā + Anuswaram

	ര	ല	വ	ശ	ഷ	സ	ഹ
○ാ○ം	രാം	ലാം	വാം	ശാം	ഷാം	സാം	ഹാം
○ി○ം	രിം	ലിം	വിം	ശിം	ഷിം	സിം	ഹിം
○ീ○ം	രീം	ലീം	വീം	ശീം	ഷീം	സീം	ഹീം
○ു○ം	രും	ലും	വും	ശും	ഷും	സും	ഹും
○ൂ○ം	രൂം	ലൂം	വൂം	ശൂം	ഷൂം	സൂം	ഹൂം
○ൃ○ം	രൃം	ലൃം	വൃം	ശൃം	ഷൃം	സൃം	ഹൃം
െ○○ം	രെം	ലെം	വെം	ശെം	ഷെം	സെം	ഹെം
േ○○ം	രേം	ലേം	വേം	ശേം	ഷേം	സേം	ഹേം
ൈ○○ം	രൈം	ലൈം	വൈം	ശൈം	ഷൈം	സൈം	ഹൈം
ൊ○○ം	രൊം	ലൊം	വൊം	ശൊം	ഷൊം	സൊം	ഹൊം
ോ○○ം	രോം	ലോം	വോം	ശോം	ഷോം	സോം	ഹോം
○ൌ○ം	രൌം	ലൌം	വൌം	ശൌം	ഷൌം	സൌം	ഹൌം

Consonant and Mātrā +Anuswaram combinations - Set 5

This set is in continuation of set 4 above which shows combinations of retroflex Consonants and Mātrā + Anuswaram

	ള	ഴ
○ാ○ം	ളാം	ഴാം
○ി○ം	ളിം	ഴിം
○ീ○ം	ളീം	ഴീം
○ു○ം	ളും	ഴും
○ൂ○ം	ളൂം	ഴൂം
○ൃ○ം	ളൃം	ഴൃം
െ○○ം	ളെം	ഴെം
േ○○ം	ളേം	ഴേം
ൈ○○ം	ളൈം	ഴൈം
ൊ○○ം	ളൊം	ഴൊം
ോ○○ം	ളോം	ഴോം
○ൌ○ം	ളൌം	ഴൌം

6.3.3. The Ligature Set of Malayalam.

Malayalam has a large set of ligatural forms. These are combinations of Consoanant+Chandrakkala (Halanta)+Consonant (CHC) or CHCHC or even rarer CHCHCHC. The CHC combinations which are the most frequent are arranged in the shape of a matrix: the abscissa or horizontal axis refers to the Consonant which constitutes the ligature and the ordinate or vertical axis shows the consonant which forms the ligature and which is followed by a Chandrakkala (Halanta).

As in 6.3.2. the ligature sets are divided into the following

6.3.3.1 CHC (in a matrix)

6.3.3.2 CHCHC

6.3.3.3.CHCHCHC

6.3.3.1. CHC (combination of two consonants)

These ligatures are presented as in the earlier case of Consonant+Mātrā combinations in four sets. A lot of "slots" have an X marked, showing that the experts have deemed that such a ligature is not possible in the language. However in these cases, the font developer is to assume that the ligature is linear in nature.

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ക" and "ക"is ligature "ക്ക".

CHC(combination of two consonants) - Set 1										
	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ
ക	ക്ക	X	X	X	X	X	X	X	X	X
ഖ	X	X	X	X	X	X	X	X	X	X
ഗ	X	X	ഗ്ഗ	X	X	X	X	X	X	X
ഘ	X	X	X	X	X	X	X	X	X	X
ങ	ങ്ക	X	X	X	ങ്ങ	X	X	X	X	X
ച	X	X	X	X	X	ച്ച	ച്ഛ	X	X	X
ഛ	X	X	X	X	X	X	X	X	X	X
ജ	X	X	X	X	X	X	X	ജ്ജ	X	ജ്ഞ
ഝ	X	X	X	X	X	X	X	X	X	X
ഞ	X	X	X	X	X	ഞ്ച	X	ഞ്ജ	X	ഞ്ഞ
ട	X	X	X	X	X	X	X	X	X	X
ഠ	X	X	X	X	X	X	X	X	X	X
ഡ	X	X	X	X	X	X	X	X	X	X
ഢ	X	X	X	X	X	X	X	X	X	X
ണ	X	X	X	X	X	X	X	X	X	X
ത	X	X	X	X	X	X	X	X	X	X

ധ	X	X	X	X	X	X	X	X	X	X
ഭ	X	X	X	X	X	X	X	X	X	X
ധ	X	X	X	X	X	X	X	X	X	X
ന	X	X	X	X	X	X	X	X	X	X
പ	X	X	X	X	X	X	X	X	X	X
ഫ	X	X	X	X	X	X	X	X	X	X
ബ	X	X	X	X	X	X	X	X	X	X
ഭ	X	X	X	X	X	X	X	X	X	X
മ	X	X	X	X	X	X	X	X	X	X
യ	യ്യ	X	X	X	X	X	X	X	X	X
ര	X	X	X	X	X	X	X	X	X	X
റ	X	X	X	X	X	X	X	X	X	X
ല	ല്ല	X	X	X	X	X	X	X	X	X
വ	X	X	X	X	X	X	X	X	X	X
ശ	X	X	X	X	X	ശ്ല	X	X	X	X
ഷ	ഷ്ട	X	X	X	X	X	X	X	X	X
സ	സ്ത	X	X	X	X	X	X	X	X	X
ഹ	X	X	X	X	X	X	X	X	X	X
ള	X	X	X	X	X	X	X	X	X	X
ഴ	ഴ്ക	X	X	X	X	ഴ്ച	X	X	X	X

30

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ക" and "ട"is ligature "ക്ട".

CHC(combination of two consonants) – Set 2									
	ട	ഠ	ഡ	ഢ	ണ	ത	ഥ	ദ	ധ	ന
ക	ക്ട	Χ	Χ	Χ	Χ	ക്ത	Χ	Χ	Χ	Χ
ഖ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഗ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	ഗ്ന
ഘ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	ഘ്ന
ങ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ച	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഛ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ജ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഝ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഞ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ട	ട്ട	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഠ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഡ	Χ	Χ	Χ	ഡ്ഢ	Χ	Χ	Χ	Χ	Χ	Χ
ഢ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ണ	ണ്ട	Χ	ണ്ഡ	Χ	ണ്ണ	Χ	Χ	Χ	Χ	Χ
ത	Χ	Χ	Χ	Χ	Χ	ത്ത	ത്ഥ	Χ	Χ	ത്ന
ഥ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ദ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	ദ്ദ	ദ്ധ	Χ
ധ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ന	Χ	Χ	Χ	Χ	Χ	ന്ത	ന്ഥ	ന്ദ	ന്ധ	ന്ന
പ	Χ	Χ	Χ	Χ	Χ	പ്ത	Χ	Χ	Χ	പ്ന
ഫ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ബ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ
ഭ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ	Χ

മ	X	X	X	X	X	X	X	X	X	X
ത	X	X	X	X	X	ത്ത	X	X	X	X
ഭ	X	X	X	X	X	X	X	X	X	X
ന	X	X	X	X	X	X	X	X	X	X
ല	X	X	X	X	X	X	X	X	X	X
വ	X	X	X	X	X	X	X	X	X	X
ശ	X	X	X	X	X	X	X	X	X	X
ഷ	ഷ്ട	ഷ്ഠ	X	X	ഷ്ണ	X	X	X	X	ഷ്ണ
സ	X	X	X	X	X	സ്ത	X	X	X	സ്ന
ഹ	X	X	X	X	X	X	X	X	X	ഹ്ന
ള	X	X	X	X	X	X	X	X	X	X
ഴ	X	X	X	X	X	ഴ്ത	X	X	X	X

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ക" and "യ"is ligature "ക്യ".

CHC(combination of two consonants) – Set 3										
	പ	ഫ	ബ	ഭ	മ	യ	ര	റ	ല	വ
ക	X	X	X	X	X	ക്യ	ക്ര	ക്ര	ക്ല	ക്വ
ഖ	X	X	X	X	X	ഖ്യ	ഖ്ര	ഖ്ര	ഖ്ല	ഖ്വ
ഗ	X	X	X	X	ഗ്മ	ഗ്യ	ഗ്ര	ഗ്ര	ഗ്ല	ഗ്വ
ഘ	X	X	X	X	X	ഘ്യ	ഘ്ര	ഘ്ര	ഘ്ല	ഘ്വ
ങ	X	X	X	X	X	X	X	X	X	X
ച	X	X	X	X	X	ച്യ	ച്ര	ച്ര	ച്ല	ച്വ
ഛ	X	X	X	X	X	ഛ്യ	ഛ്ര	ഛ്ര	ഛ്ല	ഛ്വ
ജ	X	X	X	X	X	ജ്യ	ജ്ര	ജ്ര	ജ്ല	ജ്വ
ഝ	X	X	X	X	X	ഝ്യ	ഝ്ര	ഝ്ര	X	ഝ്വ
ഞ	X	X	X	X	X	X	X	X	X	ഞ്വ
ട	X	X	X	X	X	ട്യ	ട്ര	ട്ര	X	ട്വ
ഠ	X	X	X	X	X	ഠ്യ	X	X	X	ഠ്വ
ഡ	X	X	X	X	X	ഡ്യ	ഡ്ര	ഡ്ര	X	ഡ്വ
ഢ	X	X	X	X	X	ഢ്യ	ഢ്ര	ഢ്ര	X	ഢ്വ
ണ	X	X	X	X	ണ്മ	ണ്യ	ണ്ര	ണ്ര	X	ണ്വ
ത	X	X	X	ത്ഭ	ത്മ	ത്യ	ത്ര	ത്ര	ത്ല	ത്വ
ഥ	X	X	X	X	X	ഥ്യ	ഥ്ര	ഥ്ര	X	ഥ്വ
ദ	X	X	X	X	X	ദ്യ	ദ്ര	ദ്ര	X	ദ്വ
ധ	X	X	X	X	X	ധ്യ	ധ്ര	ധ്ര	X	ധ്വ
ന	X	X	X	X	ന്മ	ന്യ	X	X	X	ന്വ
പ	പ്ഫ	X	X	X	X	പ്യ	പ്ര	പ്ര	പ്ല	പ്വ
ഫ	X	X	X	X	X	ഫ്യ	ഫ്ര	ഫ്ര	ഫ്ല	ഫ്വ
ബ	X	X	ബ്ബ	X	X	ബ്യ	ബ്ര	ബ്ര	ബ്ല	ബ്വ

ഭ	X	X	X	X	X	ഭ്യ	ഭ്ര	ഭ്ര	X	ഭ്യ
മ	മ്പ	X	X	X	മ്മ	മ്യ	മ്ര	മ്ര	മ്ല	മ്യ
യ	യ്പ	X	X	X	യ്മ	യ്യ	X	X	X	യ്യ
ര	X	X	X	X	X	ര്യ	X	X	X	ര്യ
റ	X	X	X	X	X	റ്യ	X	റ്റ	X	X
ല	ല്പ	X	X	X	X	ല്യ	X	X	ല്ല	ല്യ
വ	X	X	X	X	X	വ്യ	വ്ര	വ്ര	വ്ല	വ്യ
ശ	X	X	X	X	ശ്മ	ശ്യ	ശ്ര	ശ്ര	ശ്ല	ശ്യ
ഷ	X	X	X	X	ഷ്മ	ഷ്യ	ഷ്ര	ഷ്ര	X	ഷ്യ
സ	സ്പ	X	X	X	സ്മ	സ്യ	സ്ര	സ്ര	സ്ല	സ്യ
ഹ	X	X	X	X	ഹ്മ	ഹ്യ	ഹ്ര	ഹ്ര	ഹ്ല	ഹ്യ
ള	X	X	X	X	X	ള്യ	X	X	X	ള്യ
ഴ	X	X	X	X	X	ഴ്യ	X	X	X	ഴ്യ

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ക" and "ഷ"is ligature "ക്ഷ".

CHC(combination of two consonants) – Set 4					
	ശ	ഷ	സ	ഹ	ള	ഴ
ക	X	ക്ഷ	X	X	ക്ള	X
ഖ	X	X	X	X	ഖ്ള	X
ഗ	X	X	X	X	ഗ്ള	X
ഘ	X	X	X	X	ഘ്ള	X
ങ	X	X	X	X	X	X
ച	X	X	X	X	ച്ള	X
ഛ	X	X	X	X	ഛ്ള	X
ജ	X	X	X	X	ജ്ള	X
ഝ	X	X	X	X	X	X
ഞ	X	X	X	X	X	X
ട	X	X	X	X	X	X
ഠ	X	X	X	X	X	X
ഡ	X	X	X	X	X	X
ഢ	X	X	X	X	X	X
ണ	X	X	X	X	X	X
ത	X	X	ത്സ	X	ത്ള	X
ഥ	X	X	X	X	X	X
ദ	X	X	X	X	X	X
ധ	X	X	X	X	X	X
ന	X	X	X	X	X	X
പ	X	X	X	X	പ്ള	X
ഫ	X	X	X	X	ഫ്ള	X
ബ	X	X	X	X	ബ്ള	X
ഭ	X	X	X	X	X	X
മ	X	X	X	X	മ്ള	X
യ	X	X	X	X	X	X
ര	X	X	X	X	X	X

ന	X	X	X	X	X	X
ല	X	X	X	X	X	X
വ	X	X	X	X	വ്വ	X
ശ	ശ്ര	X	X	X	ശ്ര	X
ഷ	X	X	X	X	X	X
സ	X	X	സ്ന	X	സ്ത	X
ഹ	X	X	X	X	X	X
ള	X	X	X	X	ള്ള	X
ഴ	X	X	ഴ്ന	X	X	X

In what follows the Chillu combinations with Consonants are presented:

Chillu + Consonant Clusters Set1										
	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ
ൻ	X	X	X	X	X	X	X	X	X	X
ൽ	X	X	X	X	X	X	X	X	X	X
ർ	X	X	X	X	X	X	X	X	X	X
ൺ	X	X	X	X	X	X	X	X	X	X
ൾ	X	X	X	X	X	X	X	X	X	X

Chillu + Consonant Clusters Set2										
	ട	ഠ	ഡ	ഢ	ണ	ത	ഥ	ദ	ധ	ന
ൻ	X	X	X	X	X	X	X	X	X	X
ൽ	X	X	X	X	X	X	X	X	X	X
ർ	X	X	X	X	X	X	X	X	X	X
ൺ	X	X	X	X	X	X	X	X	X	X
ൾ	X	X	X	X	X	X	X	X	X	X

Chillu + Consonant Clusters Set3										
	പ	ഫ	ബ	ഭ	മ	യ	ര	റ	ല	വ
ൻ	X	X	X	X	X	X	X	ൻറ	X	X
ൽ	X	X	X	X	X	X	X	X	X	X
ർ	X	X	X	X	X	X	X	X	X	X
ൺ	X	X	X	X	X	X	X	X	X	X
ൾ	X	X	X	X	X	X	X	X	X	X

Chillu + Consonant Clusters Set4						
	ശ	ഷ	സ	ഹ	ള	ഴ
ൻ	X	X	X	X	X	X
ൺ	X	X	X	X	X	X

37

ർ	X	X	X	X	X	X
ൽ	X	X	X	X	X	X
ൾ	X	X	X	X	X	X

Consonant + Chillu set 1					
	ൻ	ൽ	ർ	ൺ	ൾ
ക	X	X	X	X	X
ഖ	X	X	X	X	X
ഗ	X	X	X	X	X
ഘ	X	X	X	X	X
ങ	X	X	X	X	X
ച	X	X	X	X	X
ഛ	X	X	X	X	X
ജ	X	X	X	X	X
ഝ	X	X	X	X	X
ഞ	X	X	X	X	X
ട	X	X	X	X	X
ഠ	X	X	X	X	X
ഡ	X	X	X	X	X
ഢ	X	X	X	X	X
ണ	X	X	X	X	X
ത	X	X	X	X	X
ഥ	X	X	X	X	X
ദ	X	X	X	X	X
ധ	X	X	X	X	X
ന	X	X	X	X	X
പ	X	X	X	X	X
ഫ	X	X	X	X	X
ബ	X	X	X	X	X
ഭ	X	X	X	X	X
മ	X	X	X	X	X
യ	X	X	X	X	X
ര	X	X	X	X	X
റ	X	X	X	X	X
ല	X	X	X	X	X
വ	X	X	X	X	X

ശ	X	X	X	X	X
ഷ	X	X	X	X	X
സ	X	X	X	X	X
ഹ	X	X	X	X	X
ള	X	X	X	X	X
ഴ	X	X	X	X	X

6.3.3.2 CHCHC (combination of three consonants)

These are not as frequent as the CHC combinations. Only the major are listed below.

ഷ്+ട്+ര=ഷ്ട്ര	ണ് +ഡ്+യ=ണ്ഡ്യ
സ് +ത് +ര = സ്ത്ര	ന് +ധ് +ര= ന്ധ്ര
ദ് +ധ് +യ= ദ്ധ്യ	ത് +ര് +യ= ത്ര്യ
ന് +ത് +യ= ന്ത്യ	സ് +ത് +യ=സ്ത്യ
ന് +ധ് +യ=ന്ധ്യ	ക് +ത് +ര=ക്ത്ര
ക് +ത് +യ=ക്ത്യ	ത്+ സ് +ന=ത്സ്ന

6.3.3.3.CHCHCHC (Combination of four Consonants)

This is rare and is found only in the following:

ന് +ത് +ര്+യ=ന്ത്ര്യ
ഗ് +ദ് +ധ് +യ =ഗ്ദ്ധ്യ
ത്+ സ് +ന്+യ= ത്സ്ന്യ

6.3.4 The Collation Order of Malayalam.

The collation order refers to the order in which the characters in a given language are sorted. Sorting order is approved by govt. of kerala in G.O.(MS)No:24/2001/ITD/22.11.01.It is published in the extra ordinary Gazette dated 18.12.2001. Report of the committee on Malayalam character Encoding and Keyboard layout standardization – See chapter 5 reproduced below:

The following order of characters is recommended for the lexicographic ordering of Malayalam words.

അ	ആ	ഇ	ഈ	്‌ (സംവൃത ഉകാരം)
ഉ	ഊ	ഋ	എ	ഏ
ഐ	ഒ	ഓ	ഔ	
ക	ഖ	ഗ	ഘ	ങ
ച	ഛ	ജ	ഝ	ഞ
ട	ഠ	ഡ	ഢ	(ൺ ണ)
ത	ഥ	ദ	ധ	(ൻ ന)
പ	ഫ	ബ	ഭ	(ം മ)
യ	ര	(ർ റ)	(ൽ ല)	വ
ശ	ഷ	സ	(ഃ ഹ)	(ൾ ള) ഴ

The text above is reproduced as a soft copy below:

അ	ആ	ഇ	ഈ	്‌(സംവൃത ഉകാരം)
ഉ	ഊ	ഋ	എ	ഏ
ഐ	ഒ	ഓ	ഔ	
ക	ഖ	ഗ	ഘ	ങ
ച	ഛ	ജ	ഝ	ഞ
ട	ഠ	ഡ	ഢ	(ൺണ)
ത	ഥ	ദ	ധ	(ൻന)
പ	ഫ	ബ	ഭ	(ംമ)
യ	ര	(ർറ)	(ൽല)	വ
ശ	ഷ	സ	(ഃഹ)	(ൾള) ഴ

7. REFERENCES

1. http://www.unicode.org
2. ISCII'91

8. ANNEXURES

Annexure 1: Names of experts who have contributed to the script grammar

Annexure :2 Report of the committee on Malayalam character Encoding and Keyboard layout standardization

Annexure 3: Unicode Table of MALAYALAM

[1] The Unicode chart provided is for version 5.1 since the Script Grammar was prepared at that time. No considerable change in the script grammar can be seen in the updated versions of Unicode, with the possible addition of the Rupee Sign U+02B9

Additional vowels for Sanskrit

0D60	ഌ	MALAYALAM LETTER VOCALIC RR
0D61	ൡ	MALAYALAM LETTER VOCALIC LL

Dependent vowels

0D62	ി	MALAYALAM VOWEL SIGN VOCALIC L
0D63	ൃ	MALAYALAM VOWEL SIGN VOCALIC LL

Reserved

For viram punctuation, use the generic Indic 0964 and 0965.

0D64	▨	\<reserved\>
		→ 0964 । devanagari danda
0D65	▨	\<reserved\>
		→ 0965 ॥ devanagari double danda

Digits

0D66	൦	MALAYALAM DIGIT ZERO
0D67	൧	MALAYALAM DIGIT ONE
0D68	൨	MALAYALAM DIGIT TWO
0D69	൩	MALAYALAM DIGIT THREE
0D6A	൪	MALAYALAM DIGIT FOUR
0D6B	൫	MALAYALAM DIGIT FIVE
0D6C	൬	MALAYALAM DIGIT SIX
0D6D	൭	MALAYALAM DIGIT SEVEN
0D6E	൮	MALAYALAM DIGIT EIGHT
0D6F	൯	MALAYALAM DIGIT NINE

Malayalam numerics

0D70	൰	MALAYALAM NUMBER TEN
0D71	൱	MALAYALAM NUMBER ONE HUNDRED
0D72	൲	MALAYALAM NUMBER ONE THOUSAND

Fractions

0D73	൳	MALAYALAM FRACTION ONE QUARTER
0D74	൴	MALAYALAM FRACTION ONE HALF
0D75	൵	MALAYALAM FRACTION THREE QUARTERS

Date mark

0D79	൹	MALAYALAM DATE MARK

Chillu letters

0D7A	ൺ	MALAYALAM LETTER CHILLU NN
0D7B	ൻ	MALAYALAM LETTER CHILLU N
0D7C	ർ	MALAYALAM LETTER CHILLU RR
		• historically derived from the full letter ra
		• also used for chillu r
0D7D	ൽ	MALAYALAM LETTER CHILLU L
		• historically derived from the full letter ta
		• used for chillu t and chillu d
0D7E	ൾ	MALAYALAM LETTER CHILLU LL
0D7F	ൿ	MALAYALAM LETTER CHILLU K

47