Proposal for a Malayalam Script Root Zone Label Generation Ruleset (LGR)

*LGR Version:* 3.0

*Date:* 2018-05-01

*Document version:* 0.01 pre alpha (production)

*Authors:* Neo-Brahmi Generation Panel [NBGP]

### 1. General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Malayalam LGR in the XML format and the rationale behind the design decisions taken. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

Proposed-LGR-mlym20180424.xml

Labels for testing can be found in the accompanying text document:

Labels-mlym-version-20180424.txt

### 2. Script for which the LGR is proposed

ISO 15924 Code: Mlym

ISO 15924 Key N°: 347

ISO 15924 English Name: Malayalam

Latin transliteration of native script name: malayāḷaṁ

Native name of the script: മലയാളം

Maximal Starting Repertoire (MSR) version: MSR-3

### 3. Background on Script and Principal Languages Using It

Malayalam is a Dravidian language with about 38 million speakers spoken mainly in the south west of India, particularly in Kerala, the Lakshadweep Islands and neighboring states, and also in Bahrain, Fiji, Israel, Malaysia, Qatar, Singapore, UAE and the UK.

Malayalam was first written with the Vatteluttu alphabet (വട്ടെഴുത്ത് Vaṭṭeḻuttŭ), which means 'round writing' and developed from the Brahmi script. The oldest known written text in Malayalam is known as the Vazhappalli or Vazhappally inscription, is in the Vatteluttu alphabet and dates from about 830 AD.

A version of the Grantha alphabet originally used in the Chola kingdom was brought to the southwest of India in the 8th or 9th century and was adapted to write the Malayalam and Tulu languages. By the early 13th century it is thought that a systemised Malayalam alphabet had emerged. Some changes were made to the alphabet over the following centuries, and by the middle of the 19th century the Malayalam alphabet had attained its current form.

As a result of the difficulties of printing Malayalam, a simplified or reformed version of the script was introduced during the 1970s and 1980s. The main change involved writing consonants and diacritics separately rather than as complex characters. These changes are not applied consistently so the modern script is often a mixture of traditional and simplified letters.

#### **Notable features**

* Type of writing system: syllabic alphabet in which all consonants have an inherent vowel. Diacritics, which can appear above, below, before or after a consonant, are used to change the inherent vowel.
* When they appear at the beginning of a syllable, vowels are written as independent letters.
* Chillaksharam is another feature of Malayalam. A chillu is a pure consonant without the use of a virama. It kills the inherent vowel of a consonant.
* When certain consonants occur together, special conjunct symbols are used which combine the essential parts of each letter.
* Direction of writing: left to right in horizontal lines.
* Used to write: Malayalam (മലയാളം), a Dravidian language spoken mainly in the south west of India.

Ref:<https://www.omniglot.c/writing/malayalam.htm>

#### **3.1 The Evolution of Malayalam Script:**

Malayalam was first written in the Vatteluttu alphabet, an ancient script of Tamil. However, the modern Malayalam script evolved from the Grantha alphabet, which was originally used to write Sanskrit. Both Vatteluttu and Grantha evolved from the Brahmi script, but independently.

##### **Vatteluttu alphabet**

Vatteluttu (Malayalam: വട്ടെഴുത്ത്, *Vaṭṭeḻuttŭ*, “round writing”) is a script that had evolved from Tamil-Brahmi and was once used extensively in the southern part of present-day Tamil Nadu and in Kerala.

Malayalam was first written in Vatteluttu. The Vazhappally inscription issued by Rajashekhara Varman is the earliest example, dating from about 830 CE. In the Tamil country, the modern Tamil script had supplanted Vatteluttu by the 15th century, but in the Malabar region, Vatteluttu remained in general use up to the 17th century, or the 18th century. A variant form of this script, Kolezhuthu, was used until about the 19th century mainly in the Kochi area and in the Malabar area. Another variant form, Malayanma, was used in the south of Thiruvananthapuram.

##### **Grantha, Tigalari and Malayalam scripts**

According to Arthur Coke Burnell, one form of the Grantha alphabet, originally used in the Chola dynasty, was imported into the southwest coast of India in the 8th or 9th century, which was then modified in course of time in this secluded area, where communication with the east coast was very limited. It later evolved into Tigalari-Malayalam script was used by the Malayali, Havyaka Brahmins and Tulu Brahmin people, but was originally only applied to write Sanskrit. This script split into two scripts: Tigalari and Malayalam. While Malayalam script was extended and modified to write vernacular language Malayalam, the Tigalari was written for Sanskrit only. In Malabar, this writing system was termed Arya-eluttu (ആര്യ എഴുത്ത്, *Ārya eḻuttŭ*), meaning “Arya writing” (Sanskrit is Indo-Aryan language while Malayalam is a Dravidian language).

Vatteluttu was in general use, but was not suitable for literature where many Sanskrit words were used. Like Tamil-Brahmi, it was originally used to write Tamil, and as such, did not have letters for voiced or aspirated consonants used in Sanskrit but not used in Tamil. For this reason, Vatteluttu and the Grantha alphabet were sometimes mixed, as in the Manipravalam. One of the oldest examples of the Manipravalam literature, *Vaishikatantram* (വൈശികതന്ത്രം, *Vaiśikatantram*), dates back to the 12th century, where the earliest form of the Malayalam script was used, which seems to have been systematized to some extent by the first half of the 13th century.

Thunchaththu Ezhuthachan, a poet from around the 17th century, used Arya-eluttu to write his Malayalam poems based on Classical Sanskrit literature. For a few letters missing in Arya-eluttu (*ḷa*, *ḻa*, *ṟa*), he used Vatteluttu. His works became unprecedentedly popular to the point that the Malayali people eventually started to call him the father of the Malayalam language, which also popularized Arya-eluttu as a script to write Malayalam. However, Grantha did not have distinctions between *e* and *ē*, and between *o* and *ō*, as it was as an alphabet to write a Sanskrit language. The Malayalam script as it is today was modified in the middle of the 19th century when Hermann Gundert invented the new vowel signs to distinguish them.

By the 19th century, old scripts like Kolezhuthu had been supplanted by Arya-eluttu – that is the current Malayalam script. Nowadays, it is widely used in the press of the Malayali population in Kerala.

Malayalam and Tigalari are sister scripts descended from the Grantha alphabet. Both share similar glyphic and orthographic characteristics.

##### **Orthography reform**

In 1971, the Government of Kerala reformed the orthography of Malayalam by a government order to the education department. The objective was to reduce the labour in the process of print and typewriting technology of that time, by reducing the number of glyphs required. In 1967, the government appointed a committee headed by Sooranad Kunjan Pillai, who was the editor of the Malayalam Lexicon project. It reduced number of glyphs required for Malayalam printing from around 1000 to around 250. Above committee's recommendations were further modified by another committee in 1969.

None of the major newspapers implemented it completely. But every newspaper took its own subset from the proposal. The reformed script came into effect on 15 April 1971 (the Kerala New Year), by a government order released on 23 March 1971.

##### **Languages using the Malayalam script**

The script is also used to write several minority languages such as Paniya, Betta Kurumba, and Ravula (all at EGIDS 5). The Malayalam language itself was historically written in several different scripts

[ZWJ/ZWNJ] To be completed

* Usage
* its impact not having them in the domain name system]

#### **3.2 Overall Development Process and Methodology**

Neo-Brahmi Generation Panel (NBGP) has been formed by a members having experience in linguistics and computational linguistics. Under the Neo-Brahmi Generation Panel, there are nine scripts belonging to separate Unicode blocks. Each of these scripts will be assigned a separate LGR; however Neo-Brahmi GP will ensure that the fundamental philosophy behind building those LGRs are all in sync with all other Brahmi derived scripts.

NBGP considered all the languages with EGIDS scale 1 to 4 and found that Malayalam is the one out of two languages using Malayalam script (viz Malayalam & Sanskrit), to be placed among the 22 scheduled languages of India.

### 4. Languages considered

Below is the tabular representation of the languages that have been considered for the Malayalam LGR. Languages belonging to the EGIDS scale 1 to 4 have been considered.

|  |  |  |  |
| --- | --- | --- | --- |
| **EGIDS Scale 1** | **EGIDS Scale 2** | **EGIDS Scale 3** | **EGIDS Scale 4** |
|  | Malayalam |  |  |

Main languages considered under Malayalam LGR

### 5. Repertoire

The Malayalam grapheme cluster is based on the Malayalam phonological system, with the following basic phonological template:

#### **5.1 Unicode Code Points Inclusion**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Unicode Code Point** | **Character** | **Character Name** | **Unicode General Category (gc)** | **Indic Syllabic Category** | **Reference** |
| 1 | 0D02 | ം | MALAYALAM SIGN ANUSVARA | MC | Anusvara | Starting Repertoire - MSR-3 |
| 2 | 0D03 | ഃ | MALAYALAM SIGN VISARGA | MC | Visarga | Starting Repertoire - MSR-3 |
| 3 | 0D05 | അ | MALAYALAM LETTER A | Lo | Vowel | Starting Repertoire - MSR-3 |
| 4 | 0D06 | ആ | MALAYALAM LETTER AA | Lo | Vowel | Starting Repertoire - MSR-3 |
| 5 | 0D07 | ഇ | MALAYALAM LETTER I | Lo | Vowel | Starting Repertoire - MSR-3 |
| 6 | 0D08 | ഈ | MALAYALAM LETTER II | Lo | Vowel | Starting Repertoire - MSR-3 |
| 7 | 0D09 | ഉ | MALAYALAM LETTER U | Lo | Vowel | Starting Repertoire - MSR-3 |
| 8 | 0D0A | ഊ | MALAYALAM LETTER UU | Lo | Vowel | Starting Repertoire - MSR-3 |
| 9 | 0D0B | ഋ | MALAYALAM LETTER VOCALIC R | Lo | Vowel | Starting Repertoire - MSR-3 |
| 10 | 0D0E | എ | MALAYALAM LETTER E | Lo | Vowel | Starting Repertoire - MSR-3 |
| 11 | 0D0F | ഏ | MALAYALAM LETTER EE | Lo | Vowel | Starting Repertoire - MSR-3 |
| 12 | 0D10 | ഐ | MALAYALAM LETTER AI | Lo | Vowel | Starting Repertoire - MSR-3 |
| 13 | 0D12 | ഒ | MALAYALAM LETTER O | Lo | Vowel | Starting Repertoire - MSR-3 |
| 14 | 0D13 | ഓ | MALAYALAM LETTER OO | Lo | Vowel | Starting Repertoire - MSR-3 |
| 15 | 0D14 | ഔ | MALAYALAM LETTER AU | Lo | Vowel | Starting Repertoire - MSR-3 |
| 16 | 0D15 | ക | MALAYALAM LETTER KA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 17 | 0D16 | ഖ | MALAYALAM LETTER KHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 18 | 0D17 | ഗ | MALAYALAM LETTER GA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 19 | 0D18 | ഘ | MALAYALAM LETTER GHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 20 | 0D19 | ങ | MALAYALAM LETTER NGA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 21 | 0D1A | ച | MALAYALAM LETTER CA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 22 | 0D1B | ഛ | MALAYALAM LETTER CHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 23 | 0D1C | ജ | MALAYALAM LETTER JA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 24 | 0D1D | ഝ | MALAYALAM LETTER JHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 25 | 0D1E | ഞ | MALAYALAM LETTER NYA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 26 | 0D1F | ട | MALAYALAM LETTER TTA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 27 | 0D20 | ഠ | MALAYALAM LETTER TTHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 28 | 0D21 | ഡ | MALAYALAM LETTER DDA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 29 | 0D22 | ഢ | MALAYALAM LETTER DDHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 30 | 0D23 | ണ | MALAYALAM LETTER NNA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 31 | 0D24 | ത | MALAYALAM LETTER TA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 32 | 0D25 | ഥ | MALAYALAM LETTER THA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 33 | 0D26 | ദ | MALAYALAM LETTER DA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 34 | 0D27 | ധ | MALAYALAM LETTER DHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 35 | 0D28 | ന | MALAYALAM LETTER NA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 36 | 0D2A | പ | MALAYALAM LETTER PA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 37 | 0D2B | ഫ | MALAYALAM LETTER PHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 38 | 0D2C | ബ | MALAYALAM LETTER BA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 39 | 0D2D | ഭ | MALAYALAM LETTER BHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 40 | 0D2E | മ | MALAYALAM LETTER MA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 41 | 0D2F | യ | MALAYALAM LETTER YA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 42 | 0D30 | ര | MALAYALAM LETTER RA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 43 | 0D31 | റ | MALAYALAM LETTER RRA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 44 | 0D32 | ല | MALAYALAM LETTER LA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 45 | 0D33 | ള | MALAYALAM LETTER LLA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 46 | 0D34 | ഴ | MALAYALAM LETTER LLLA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 47 | 0D35 | വ | MALAYALAM LETTER VA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 48 | 0D36 | ശ | MALAYALAM LETTER SHA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 49 | 0D37 | ഷ | MALAYALAM LETTER SSA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 50 | 0D38 | സ | MALAYALAM LETTER SA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 51 | 0D39 | ഹ | MALAYALAM LETTER HA | Lo | Consonant | Starting Repertoire - MSR-3 |
| 52 | 0D3E | ാ | MALAYALAM VOWEL SIGN AA | MC | Matra | Starting Repertoire - MSR-3 |
| 53 | 0D3F | ി | MALAYALAM VOWEL SIGN I | MC | Matra | Starting Repertoire - MSR-3 |
| 54 | 0D40 | ീ | MALAYALAM VOWEL SIGN II | MC | Matra | Starting Repertoire - MSR-3 |
| 55 | 0D41 | ു | MALAYALAM VOWEL SIGN U | MC | Matra | Starting Repertoire - MSR-3 |
| 56 | 0D42 | ൂ | MALAYALAM VOWEL SIGN UU | MC | Matra | Starting Repertoire - MSR-3 |
| 57 | 0D43 | ൃ | MALAYALAM VOWEL SIGN VOCALIC R | MC | Matra | Starting Repertoire - MSR-3 |
| 58 | 0D46 | െ | MALAYALAM VOWEL SIGN E | MC | Matra | Starting Repertoire - MSR-3 |
| 59 | 0D47 | േ | MALAYALAM VOWEL SIGN EE | MC | Matra | Starting Repertoire - MSR-3 |
| 60 | 0D48 | ൈ | MALAYALAM VOWEL SIGN AI | MC | Matra | Starting Repertoire - MSR-3 |
| 61 | 0D4A | ൊ | MALAYALAM VOWEL SIGN O | MC | Matra | Starting Repertoire - MSR-3 |
| 62 | 0D4B | ോ | MALAYALAM VOWEL SIGN OO | MC | Matra | Starting Repertoire - MSR-3 |
| 63 | 0D4D | ് | MALAYALAM SIGN VIRAMA | MN | Chandrakkala / Virama | Starting Repertoire - MSR-3 |
| 64 | 0D57 | ൗ | MALAYALAM AU LENGTH MARK | MC | Various Signs | Starting Repertoire - MSR-3 |
| 65 | 0D7A | ൺ | MALAYALAM LETTER CHILLU NN | LO | Chillu Letters | Starting Repertoire - MSR-3 |
| 66 | 0D7B | ൻ | MALAYALAM LETTER CHILLU N | LO | Chillu Letters | Starting Repertoire - MSR-3 |
| 67 | 0D7C | ർ | MALAYALAM LETTER CHILLU RR | LO | Chillu Letters | Starting Repertoire - MSR-3 |
| 68 | 0D7D | ൽ | MALAYALAM LETTER CHILLU L | LO | Chillu Letters | Starting Repertoire - MSR-3 |
| 69 | 0D7E | ൾ | MALAYALAM LETTER CHILLU LL | LO | Chillu Letters | Starting Repertoire - MSR-3 |
| 70. | 0D7F | ൿ | MALAYALAM LETTER CHILLU K | LO | Chillu Letters | Starting Repertoire - MSR-3 |

#### **5.2 Unicode Code Point Exclusion**

#### 

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Unicode Code Point** | **Character** | **Character Name** | **Unicode General Category (gc)** | **Indic Syllabic Category** | **Reference** |
| 1. | 0D0C | ഌ | MALAYALAM LETTER VOCALIC L | Lo | Vowel | Starting Repertoire - MSR-3 |
| 2. | 0D44 | ൄ | MALAYALAM VOWEL SIGN VOCALIC RR | MC | Matra | Starting Repertoire - MSR-3 |
| 3. | 0D29 | ഩ | MALAYALAM LETTER NNNA | Lo | Consonant | Starting Repertoire - MSR-3 |

#### **5.3 Vowels**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **അ**  **A** | **ആ**  **AA** | **ഇ**  **I** | **ഈ**  **II** | **ഉ**  **U** | **ഊ**  **UU** | **ഋ**  **R** |
| **ഌ**  **LL** | **എ**  **E** | **ഏ**  **EE** | **ഐ**  **AI** | **ഒ**  **O** | **ഓ**  **OO** | **ഔ**  **AU** |

#### **5.4** **Vowel diacritics with ക (0D15)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ക    KA | കാ    KAA | കി  KI | കീ  KII | കു    KU | കൂ    KUU | |  | | --- | | കൃ | | KR | |
| കെ    KE | കേ  KEE | കൈ  KAI | കൊ    KO | കോ  KOO | കൗ    KAU | കം |

#### **5.5** **Consonants**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ക  KA | ഖ  KHA | ഗ  GA | ഘ  GHA | ങ  NGA |
| ച  CA | ഛ  CHA | ജ  JA | ഝ  JHA | ഞ  NYA |
| ട  TTA | ഠ  TTHA | ഡ  DDA | ഢ  DDHA | ണ  NNA |
| ത  TA | ഥ  THA | ദ  DA | ധ  DHA | ന  NA |
| പ  PA | ഫ  PHA | ബ  BA | ഭ  BHA | മ  MA |
| യ  YA | ര  RA | റ  RRA | ല  LA | ള  LLA |
| ഴ  LLLA | വ  VA | ശ  SHA | ഷ  SSA | സ  SA |
| ഹ  HA |  |  |  |  |

#### **5.6** **Chillu letters (chillaksharam)**

Pure consonants, without any vowel

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ൺ  0D7A  NN | ൻ  0D7B  N | ർ  0D7C  RR | ൽ  0D7D  L | ൾ  0D7E  LL | ൿ  0D7F  K |

#### **5.7** **A selection of conjunct consonants**

A Consonant combined with another Conjunct or Consonant using Virama. Conjuncts with more than 4 consonants are rare. ഗ്ദ്ധ്ര്യ is formed by 5 consonants.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | kka | ṅka | ṅṅa | cca | ñca | ñña | ṭṭa | ṇṭa | ṇṇa | tta | nta | nna | ppa | mpa | mma |
| NLF | ക്‌ക | ങ്‌ക | ങ്‌ങ | ച്‌ച | ഞ്‌ച | ഞ്‌ഞ | ട്‌ട | ണ്‌ട | ണ്‌ണ | ത്‌ത | ന്‌ത | ന്‌ന | പ്‌പ | മ്‌പ | മ്‌മ |
| LF | ക്ക | ങ്ക | ങ്ങ | ച്ച | ഞ്ച | ഞ്ഞ | ട്ട | ണ്ട | ണ്ണ | ത്ത | ന്ത | ന്ന | പ്പ | മ്പ | മ്മ |

NLF - Non-ligated form has a visible virama (chandrakkala)

LF- Ligated form in which consonants are conjoined fully or partially (this how fonts render them).

#### **5.8** **Consonant diacritics with യ (0D2F), ര (0D30), ല (0D32), വ (0D35)**

|  |  |  |  |
| --- | --- | --- | --- |
| Consonant + യ (0D2F) | Consonant + ര (0D30) | Consonant + ല (0D32) | Consonant + വ (0D35) |
| ക്യ | ക്ര | ക്ല | ക്വ |
| പ്യ | പ്ര | പ്ല | പ്വ |

Consonant used above are ക (0D15) and പ (0D2A). Other consonants can be combined in similar fashion.

#### **5.9 Samvruthokarams**

All consonant can have consonant + ു (0D41) + ◌് (0D4D) and forming samvruthokaram form of that consonant.

Examples: തു് , കു് , പു് , രു് .

### 6. Variants

#### **6.1 In-script variants**

|  |  |  |  |
| --- | --- | --- | --- |
| Variant Set | Combination | Codepoint Combination | Variant |
|
| 1. | ൻ + റ | 0D7B + 0D31 | ൻറ |
| ന്‌ + റ | 0D28 +0D4D +0D31 | ന്റ |
| ൻ + ് + റ | 0D7B + 0D4D + 0D31 |  |
| 2. | ള് + ള | 0D33 + 0D4D + 0D33 | ള്ള |
| ള + ള | 0D33 + 0D33 | ളള |

#### **6.2 Cross-script variants**

Some characters of Malayalam look somewhat similar to some characters in Tamil, Kannada, Telugu and Sinhala. The table below lists them.

##### **6.2.1 Cross-script variants for Tamil and Malayalam**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variant Set | Tamil | | Malayalam | |
| CP | Glyph | CP | Glyph |
| 1 | 0B9C | ஜ | 0D1C | ജ |
| 2. | 0BB5 | வ | 0D16 | ഖ |
| 3. | 0BAE | ம | 0D25 | ഥ |
| 4. | 0BBF | ி | 0D3F | ി |
| 5. | 0BC6 | ெ | 0D46 | െ |
| 6. | 0BC7 | ே் | 0D47 | േ |

##### 

##### **6.2.2 Cross-script variants for Oriya and Malayalam**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variant Set | Oriya | | Malayalam | |
| CP | Glyph | CP | Glyph |
| 1. | 0B02 | ଂ | 0D02 | ം |
| 2. | 0B03 | ଃ | 0D03 | ഃ |
| 3. | 0B20 | ଠ | 0D20 | ഠ |

### 7. Whole Label Evaluation Rules (WLE)

This section provides the WLEs that are required by all the languages mentioned in section 4 when written in Malayalam Script. The rules have been drafted in such a way that they can be easily translated into the LGR specifications.

Below are the symbols used in the WLE rules, for each of the "Indic Syllabic Category" as mentioned in the table provided in Code point repertoire

#### **7.1 Variables or definitions:**

**V**  → Vowel

**M**  → Matra (Vowel Sign)

**C** → Consonant

**CH**  → Chillu

**H**  → Chandrakkala/Halant/Virama ( ◌് 0D4D)

**B**  → Anusvara ( ം 0D02)

**X**  → Visarga (ഃ 0D03)

#### **7.2 Rules for forming aksharam or a cluster of one unit:**

* Rule 1: H must be preceded by C or the chillu ൻ (0D7B ) or the M ു (0D41) (samvruthokaram)
* Rule 2: M must be preceded by C
* Rule 3: B must be preceded by C, V or M
* Rule 4: X must be preceded by C, V or M
* Rule 5: CH cannot be preceded by B, X or H
* Rule 6: Label does not begin with CH

### 8. Contributors

Neo-Brahmi Generation Panel (NBGP)

Prasad P.K.

Santhosh Thottingal

Veena Solomon ([veena.ycet@gmail.com](mailto:veena.ycet@gmail.com))

### 9. References

[101] <http://unicode.org/reports/tr31/#Layout_and_Format_Control_Characters>

[102] Report on Malayalam Unicode Issues (2012) prepared by Santhosh Thottingal (also part of NEGP) and submitted to Unicode via Wikimedia Foundation. It discusses both chillu and nta issues

[103] <http://thottingal.in/documents/ReportonMalayalamUnicodeIssues.pdf>

[104] A report by Roozbeh Pournader and Cibu Johny of Google on Unicode on usage statistics of Old and New Malayalam Chillus.

[105] <http://www.unicode.org/L2/L2013/13036-chillus-uptake.pdf>

[106] [https://ml.wikipedia.org/wiki/മലയാളം\_അക്ഷരമാല](https://ml.wikipedia.org/wiki/)

[107] [https://en.wikipedia.org/wiki/Malayalam\_script](https://ml.wikipedia.org/wiki/)

[108] [https://si.wikipedia.org/wiki/ෆ](https://si.wikipedia.org/wiki/%E0%B7%86)

[109] <https://wiki.smc.org.in/CDAC-IDN-Critique>

[110] <https://smc.org.in/fonts/>

[111] <https://www.omniglot.com/writing/malayalam.htm>

[112] Maximal Starting Repertoire - MSR-3, Annotated Repertoire Tables, Non-CJK, Pages 66-68 <https://www.icann.org/en/system/files/files/msr-3-non-cjk-28mar18-en.pdf>

[113] <http://thottingal.in/blog/2017/05/27/a-formal-grammar-for-malayalam-syllables/>

[114] <http://tdil-dc.in/tdildcMain/articles/282860Malayalam%20Script%20Grammar.pdf>

[115] The Unicode Standard, Version 10.0., Chapter 12 “South and Central Asia I: Official Scripts of India”, <https://www.unicode.org/versions/Unicode10.0.0/ch12.pdf>

[116] [– see response document for additional suggested reference]

### 10. Appendix

#### **10.1 Excluded in-script variants based on pure visual similarity**

As these formations are not valid as per akshar formation rules, these cases are not proposed as variants.

|  |  |  |  |
| --- | --- | --- | --- |
| 1. | ഈ | 0D08 | ഈ |
| ഇ + ൗ | 0D07 + 0D57 | ഇൗ |
| 2. | ഊ | 0D0A | ഊ |
| ഉ + ൗ | 0D09 + 0D57 | ഉൗ |
| 3. | ഔ | 0D14 | ഔ |
| ഒ + ൗ | 0D12 + 0D57 | ഒൗ |
| 6. | ഓ | 0D13 | ഓ |
| ഒ + ാ | 0D12 + 0D3E | ഒാ |
| 4. | ഐ | 0D10 | ഐ |
| എ + െ | 0D0E + 0D46 | എെ |

|  |  |  |
| --- | --- | --- |
| Code Point 1 + Glyph 1 | Code Point 2 + Glyph 2 | Type (allocatable or blocked) |
| ൊ  (0D4A) | െ (0D46) + ാ (0D3E) | Although, Unicode defines this canonical decomposition, the Standard recommends not to use the sequence. (See[115], p. 501). Therefore, it is not advisable to use them in IDN labels; they are blocked here by akshar formation rule. |
| ോ  0D4B | േ (0D47) + ാ (0D3E) |
| ൗ  0D57 | െ (0D46) + ൗ (0D57) |

##### 

#### **10.2 Excluded cross-script variants based on pure visual similarity**

The code-points below are visually confusing only in smaller fonts and can be excluded from consideration as variants.

|  |  |
| --- | --- |
| **Tamil** | **Malayalam** |
| ஸ (0BB8) | സ (0D38) |

At Sri Lanka F2F, it was decided to exclude the code points below from variant list as these do not look alike, due to round - square structural differences.

|  |  |
| --- | --- |
| **Kannada** | **Malayalam** |
| ಲ (0CB2) | ല (0D32) |

|  |  |
| --- | --- |
| **Telugu** | **Malayalam** |
| ల (0C32) | ല (0D32) |

#### **10.3 Excluded cross-script variants based on the insufficient number of code point variants**

Most of the code points below would qualify as cross-script code point variants but there are not enough of them to form a variant labels, therefore these cases can be excluded. (If only combining marks are variants for a given script, no label can be formed without using at least one non-variant code point). In the case of Sinhala, the relevant base character is distinct.

##### **10.3.1 Kannada and Malayalam**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variant Set | Kannada | | Malayalam | |
| CP | Glyph | CP | Glyph |
| 1 | 0C82 | ಂ | 0D02 | ം |
| 2. | 0C83 | ಃ | 0D03 | ഃ |

##### 

##### **10.3.2 Telugu and Malayalam**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variant Set | Telugu | | Malayalam | |
| CP | Glyph | CP | Glyph |
| 1. | 0C02 | ం | 0D02 | ം |
| 2. | 0C03 | ః | 0D03 | ഃ |

##### **6.2.3 Sinhala and Malayalam**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variant Set | Sinhala | | Malayalam | | Note |
| CP | Glyph | CP | Glyph |
| 1. | 0D82 | ം | 0D02 | ം | Not enough code points |
| 2. | 0D83 | ඃ | 0D03 | ഃ |
| 3. | 0DC6 | ෆ | 0D28 | ന | Distinguishable |

##### 