

Request for clarification on historical conjuncts involving Chillus

Cibu Johny cibucj@gmail.com
 Sunil V S vssun9@gmail.com
 Shiju Alex shijualexonline@gmail.com
 2018-Dec-26

1. Introduction

Malayalam chillu characters and their properties are described in section 12.9 *Malayalam* (page 505-507) of Unicode Standard 11.0.0. The subsection *Chillu Forms* mentions: “Occasionally, chillu forms may take vowels or be elements of conjuncts.” However, no examples are provided in there to indicate this behavior.

Meanwhile, under the subsection *Special Cases Involving rra*, the conjoining behavior of CHILLU N with VIRAMA and RRA, along with vowel signs, is shown using ന്റ conjunct (/nt̪a/; <CHILLU N, VIRAMA, RRA>). See table 12-38:

A very similar situation exists for the combination of ന് ചില്ലു-n and റ റ്ര-r. When used side by side, ന്ററ can be read either /nt̪a/ or /nt̪a/, while stacked ന്റ is always read /nt̪a/.

The sequence <0D7B, 0D31> is rendered as ന്റ, regardless of the reading of that text. The sequence <0D7B, 0D4D, 0D31> is rendered as ന്റ. In both cases, vowels signs can be used as appropriate, as shown in Table 12-38.

Table 12-38. Malayalam /nr/ and /nt/

ആന്ററ	0D06 0D7B 0D47 0D31 0D3E	/āntō/	a proper name
ആന്റ	0D06 0D7B 0D4D 0D31 0D47 0D3E		
എന്ററൾ	0D0E 0D7B 0D31 0D47 0D3E 0D7A	/engōl/	enroll

The conjuncts involving chillus, other than ന്റ (/nt̪a/; <CHILLU N, VIRAMA, RRA>), are not used in contemporary Malayalam. However, they are well attested in historical texts. As the conjoining behavior for those cases are not clearly mentioned in the standard text, input method developers find it hard to decide on the right sequences to generate those archaic chillu conjuncts. Similarly, the type designers need to incorporate right shaping logic for such sequences.

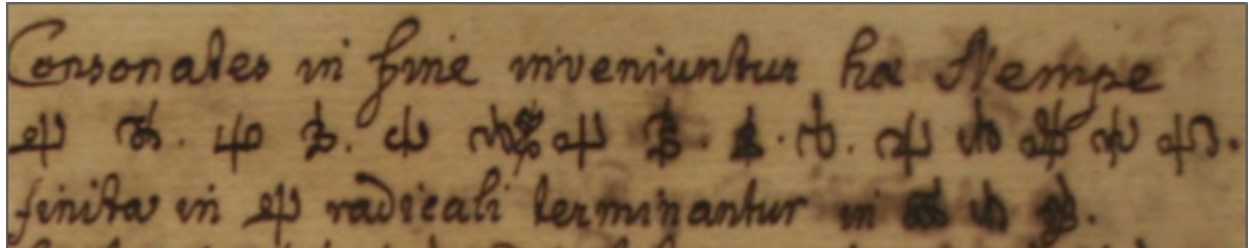
Recently, the LGR proposal (<https://www.icann.org/en/system/files/files/proposal-malayalam-lgr-25sep18-en.pdf> section 6.1: *In-script variants*) was found to be in conflict with Unicode standard due the confusion about the chillu conjuncts.

For some chillu conjuncts, we could potentially use either atomically encoded chillus or VERTICAL BAR VIRAMA. This is another source of confusion. For example, CHILLU form of N-NA can be thought of encoded either as <NA, VIRAMA, CHILLU N> or as <NA, VIRAMA, NA, VERTICAL BAR VIRAMA>. It could be argued that unless the base form is clearly an atomically encoded chillu, the VERTICAL BAR VIRAMA should be used. Further details of VERTICAL BAR VIRAMA and its attestations can be found in its accepted proposal L2/14-015R (<https://unicode.org/L2/L2014/14015r-vertical-virama.pdf>). This would imply that there is no need to encode any further chillus. Also, the already atomically encoded chillus should not be represented by ligating a base consonant with VERTICAL BAR VIRAMA.

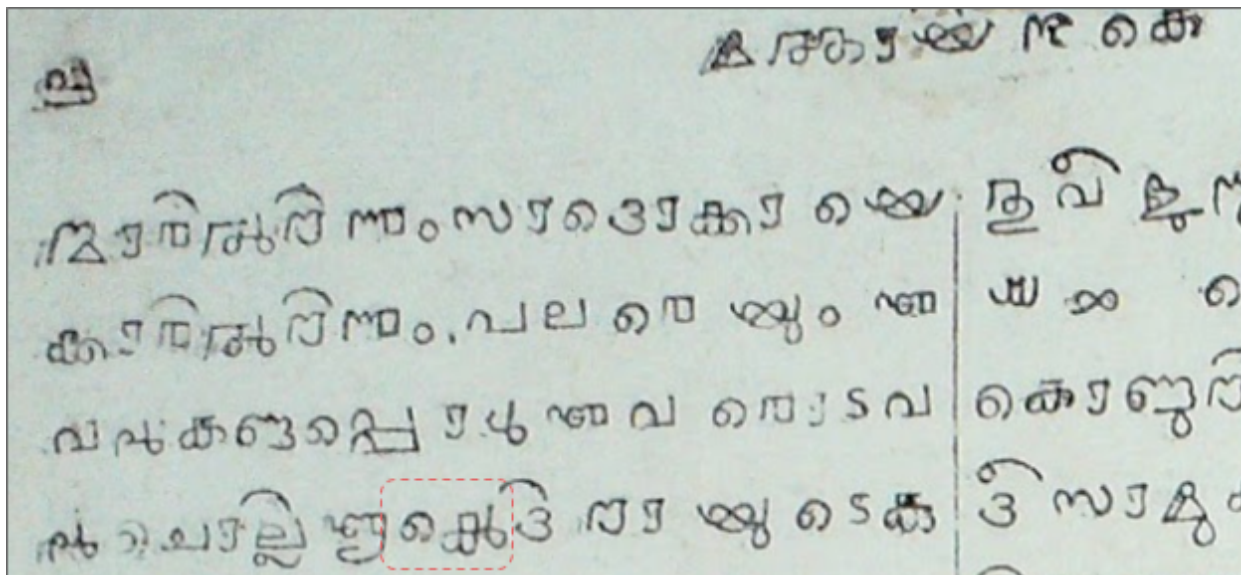
To avoid such ambiguities, the conjoining and the shaping behavior of chillu letters need to be clearly articulated in the standard under the section *Chillu Forms*. Otherwise, this could result in double encoding in public data.

2. Attestations

Listed below are the attestations of various archaic chillu conjuncts found in various sources ranging from year 1681 to 1918. The /ṇṭa/ conjunct is not specifically listed, as its contemporary usage is well documented in the standard.



Attestation 1: Grammatica Grandonica : the Sanskrit Grammar of Johann Ernst Hanxleden S.J; Page 60; year 1681-1732
https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/6251/file/hanxleden_grammatica.pdf
 Chillu forms of various consonants without atomically encoded chillus.



Attestation 2: Rampan Bible; Currier Press; Bombay; Page 13; year 1811
<https://archive.org/details/1811RampanBibleMalayalam/page/n13>
 The conjunct visual: <<SIGN E, CHILLU K>>
 Corresponding character sequence is: CHILLU K, SIGN E

Note: The word in the context is a transliteration of *Echidna* /ekh'-id-nah/, the Greek mythological she-viper. In modern Malayalam, it can be transliterated as എകിട്ന. So, it seems, in the above attestation, the independent vowel *E* has been erroneously replaced with vowel sign *E*. Also, the vowel sign *I* seems to be shared between CHILLU K and DA. Alternatively it could be thought of as ആക്ടൊ, using the preceding ആ as well. Here also, the vowel sign *I* seems misplaced. In either case, this spelling could be showing a peculiar Syriac pronunciation of *Echidna*. The name *Echidna* is the same in Greek and Syriac and Syriac was the liturgical language of the native church in that period. Until further attestations of <<SIGN E, CHILLU K>> is uncovered, this usage may be treated as a printing mistake.

Diximus usque modo de litteris simplicibus, atque compositis; nunc finales perpendendae sunt, ac de iis debemus singillatim loqui. Hae igitur sunt octo, nempe

æt, ɔ, ɔ, ɔ, æt, ɔ, æt, ɔ.

el, er, el, iz, en, en, en, am:

Attestation 3: Alphabetum Grandonico Malabaricum; Page 69; year 1772

<http://books.google.com/books?id=qmETAAAAQAAJ&pg=PA69>

The conjunct visual is: <<N-NA ligature with chillu tail>>

Corresponding character sequence is: NA, VIRAMA, NA, VERTICAL BAR VIRAMA

தேவ-நாகாபலகம்-உய்ணவரன்குறையினு
 ன-சொட்டி-சொண்டி-விழை-விண்ணை-சொட்டி-சொண்டி-
 உடிக-உய்யோலம்-ப.க. உய்யோடு-ம. வயட்-வெள்ளை-வெ
 ளையாறு-புலம்-பருமிக்காது-நாகாபலகம்-மீளாது-வாயிடும்-ஆ

Attestation 4: Malayālabhāsā vyākaranam; H. Gundert; Tellicherry; Page 12; year 1851

<http://idb.ub.uni-tuebingen.de/pendigi/CiXIV35#p=20>

The conjunct visuals are:

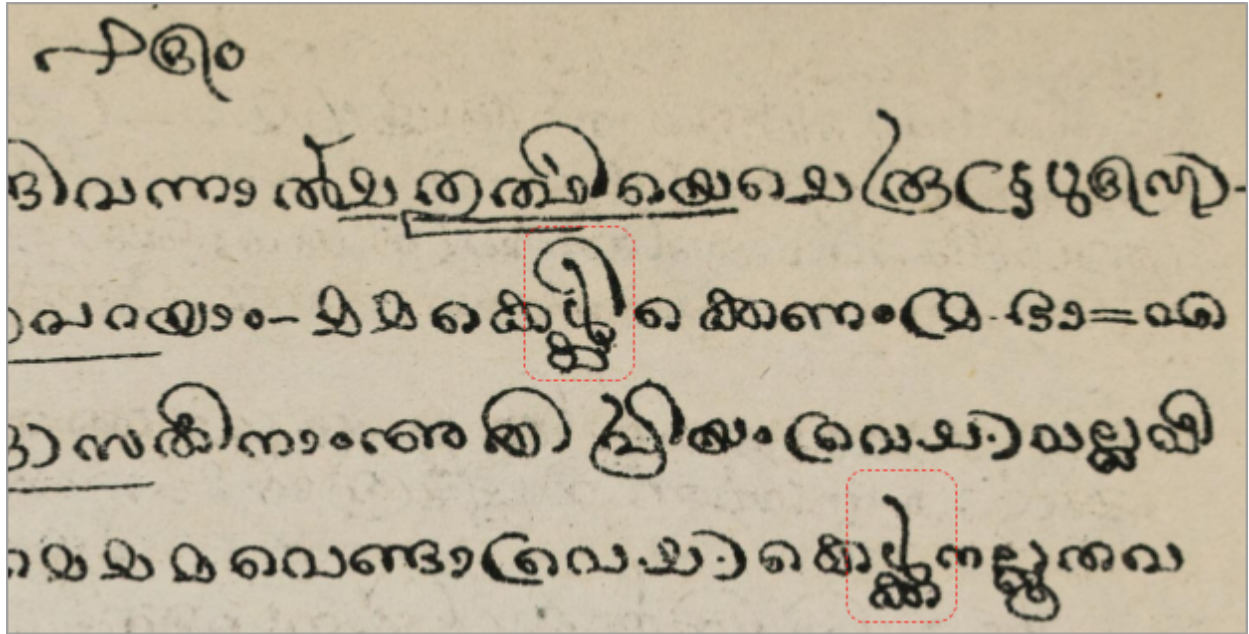
<<CHILLU LL base, subscript ligature N-T-U>>

<<CHILLU NN base, subscript NA, SIGN AA>>

Corresponding character sequences are:

CHILLU LL, VIRAMA, NA, VIRAMA, TA, SIGN U

CHILLU NN, VIRAMA, NA, SIGN AA



Attestation 5: Malayāḷabhāṣā vyākaraṇam; H. Gundert; Tellicherry; Page 150; year 1851

<http://idb.ub.uni-tuebingen.de/pendigi/CiXIV35#p=158>

The conjunct visuals are:

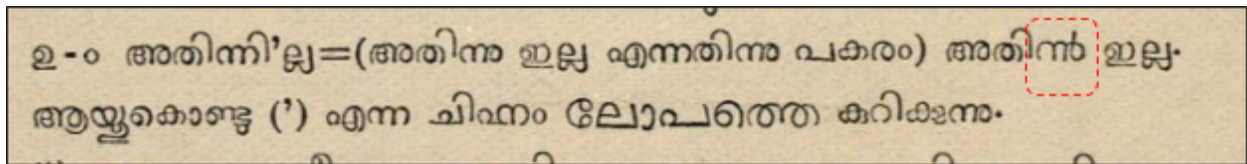
<<CHILLU LL base, subscript ligature P-PA, SIGN I>>

<<CHILLU LL base, subscript ligature K-KA>>

Corresponding character sequences are:

CHILLU LL, VIRAMA, PA, VIRAMA, PA, SIGN I

CHILLU LL, VIRAMA, KA, VIRAMA, KA

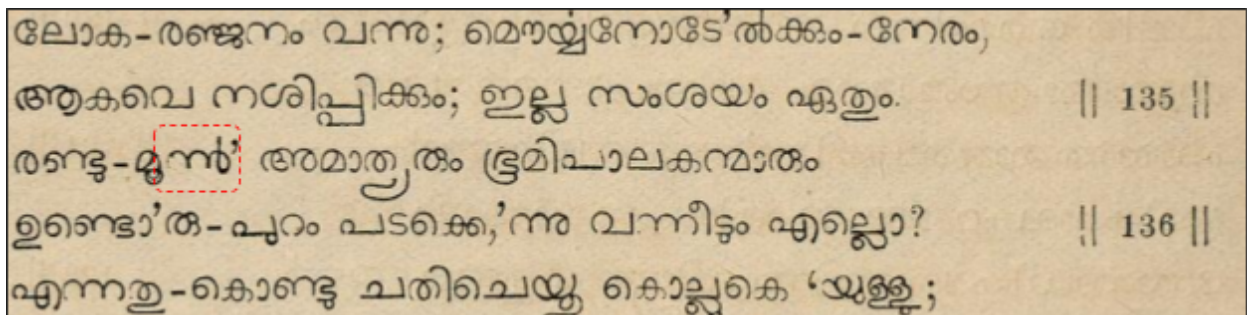


Attestation 6: Cāṇakyaśūtram allenkil mudrārākṣasam; Page X1V; year 1868

<http://idb.ub.uni-tuebingen.de/pendigi/CiXIV139#p=18>

The conjunct visual is: <<N-NA ligature with chillu tail>>

Corresponding character sequence is: NA, VIRAMA, NA, VERTICAL BAR VIRAMA



Attestation 7: Cāṇakyaśūtram allenkil mudrārākṣasam; Page 10; year 1868

<http://idb.ub.uni-tuebingen.de/pendigi/CiXIV139#p=30>

The conjunct visual is: <<N-NA ligature with chillu tail>>

Corresponding character sequence is: NA, VIRAMA, NA, VERTICAL BAR VIRAMA

— ൨൧ —

ഉണ്ടായിരിക്കുമ്പോൾ ആ പെരുമാളെ കൊന്നുകൊ
 ്ലാൻ പ്രയാസം തന്നെ എന്നു കണ്ടു, ആ ഭൂതങ്ങളെ
 അകറ്റേണ്ടതിന്നു ഒരു ചതിപ്രയോഗം ചെയ്യേണം
 എന്നു നിശ്ചയിച്ചു. ഒരു ഭട്ടത്തിരി ഞാൻ ചെന്നു

Attestation 8: Keralolpatti; H. Gundert; Page 25; year 1868
https://archive.org/details/1868_Keralolpathi_Hermann_Gundert/page/n25
 The conjunct visual is: <<CHILLU LL base, subscript ligature VA, SIGN AA>>
 Corresponding character sequence is: CHILLU LL, ZWJ, VIRAMA, VA, SIGN AA

42. ടകാരം സവണ്ണങ്ങൾക്കും ഷകാരത്തിന്നും പകരം
 (ശബ്ദം-ചഞ്ച; ഫലകം-ഇടക; ഖണ്ഡം-കഞ്ച; മേഘം-മേടം; പൂർവ്വം-പുരാടം;
 ഗോഷം-ഗോട്ടം) ഷഡ്ഭാഗം രാട്ട മുതലായവറ്റിൽ ഉകാരം അധികം
 നടപ്പു. (ജൾ-രാൾ). പിന്നെ മലയാള ടകാരം പലതും നെളുകാരങ്ങ
 ഉൽ നിന്നുജനിച്ചവ (ഇരുട്ട-രു; കാട്ടുക-രു; കേട്ട-കേ രു)

Attestation 9: Malayāḷa bhāṣavyākaraṇam; H Gundert; Page 10; year 1868
<http://idb.ub.uni-tuebingen.de/opensig/CiXIV68a#p=22>
 All conjunct visuals are: <<CHILLU LL base, subscript ligature T-T-U>>
 Corresponding character sequence is: CHILLU LL, VIRAMA, TA, VIRAMA, TA, SIGN U

97 ആൾമാരി—ആഴുക

1. ആൾമാരി ālmāri Port. = അൾമയി Almeida.
 y ആൾമാർ ālvār & ആഴവാർ T. M. C. Te.
 2 ("the rulers" or "the humble") Title of the
 b 12 Vishnu apostles. see ആഴവാർ.

Attestation 10: A Malayalam English Dictionary; H. Gundert; Mangalore; Page 97; year 1872

<https://archive.org/details/amalayalamanden00gundgoog/page/n132>

The conjunct visuals are:

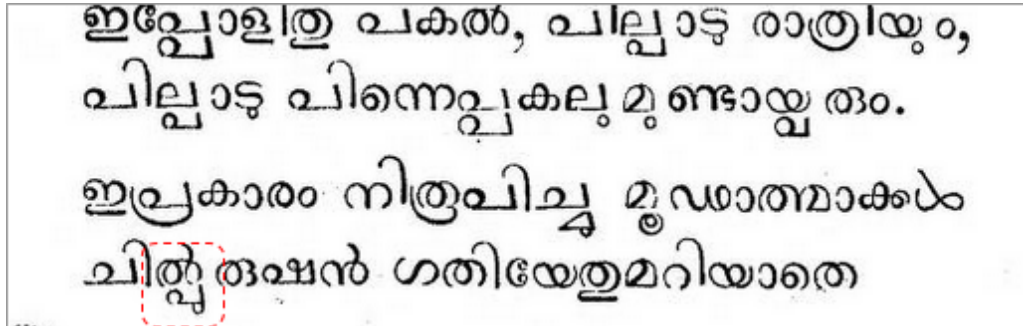
<<CHILLU LL base, subscript MA, SIGN AA>>

<<CHILLU LL base, subscript VA, SIGN AA>>

Corresponding character sequences are:

CHILLU LL, VIRAMA, MA, SIGN AA

CHILLU LL, ZWJ, VIRAMA, VA, SIGN AA

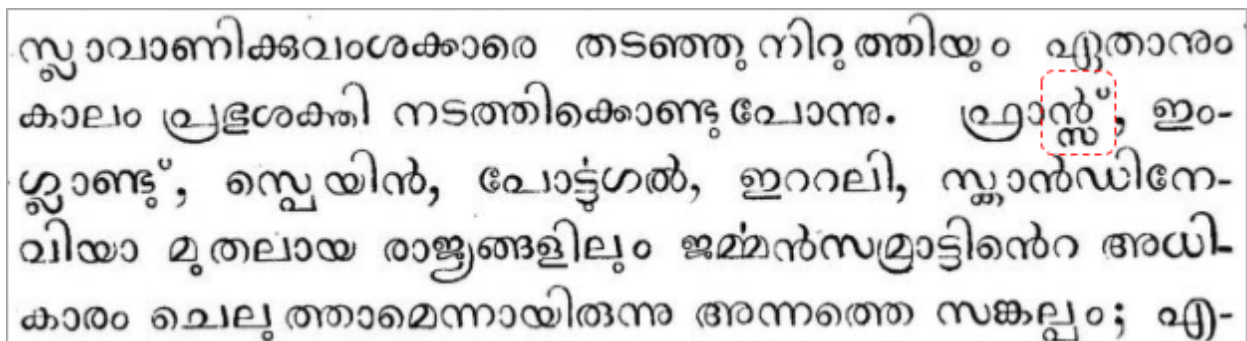


Attestation 11: Malayalam Fifth Reader; Basel Mission Press; Mangalore; Page 177; year 1918

https://archive.org/stream/MalayalamFifthReader1918Images/Malayalam_fifth_reader_1918#page/n177

The conjunct visual is: <<CHILLU L base, subscript ligature P-U>>

Corresponding character sequence is: CHILLU L, VIRAMA, PA, SIGN U



Attestation 12: Malayalam Fifth Reader; Basel Mission Press; Mangalore; Page 195; year 1918

https://archive.org/stream/MalayalamFifthReader1918Images/Malayalam_fifth_reader_1918#page/n195

The conjunct visual is: <<CHILLU N base, subscript SA, explicit VIRAMA>>

Corresponding character sequence is: CHILLU N, VIRAMA, SA, VIRAMA

3. Nature of the conjuncts with chillu bases

So far, following chillus are found to be participating in some form of conjuncts:

CHILLU N,

CHILLU NN,

CHILLU L,

CHILLU LL,

There are no attestations found regarding the conjuncts using the remaining chillus:

CHILLU RR,

CHILLU K

CHILLU LLL,

CHILLU M,

CHILLU Y

The attested conjuncts follow one of these patterns of codepoint sequences:
 CHILLU, LEFT PART VOWEL SIGN
 CHILLU, VIRAMA, CONSONANT, (ZWJ? VIRAMA, CONSONANT)* (VOWEL SIGN)*

There are no attestation found for chillu letters with right part vowel sign or with virama alone.

That said, it is possible that, conjuncts that are not attested today can turn up, when new archives are examined. So it is prudent to consider chillus like consonants for textual shaping and to make it future proof.

4. Editorial Update Requests



Following editorial updates are proposed to clarify the standard sequences to be used for encoding historical Chillu usages, mentioned in the above sections.

Request 1

In order to cover the cases found in the above attestations, please add following description and table under the subheading *Chillu Forms* - the section 12.9 (page 505) of Unicode Standard 11.0.0.:

The *chillu* letters behave like consonants, forming conjuncts with the help of a virama following it. They also accept a following vowel sign, similar to a consonant.

Table 12-38. Examples of historic Malayalam grapheme conjuncts involving *chillu*




	0D7A 0D4D 0D28	CHILLU NN, VIRAMA, NA	/nna/
	0D7B 0D4D 0D33 0D46	CHILLU N, VIRAMA, RRA, SIGN E	/nte/
	0D7B 0D4D 0D38 0D4D	CHILLU N, VIRAMA, SA, VIRAMA	/nsu/
	0D28 0D4D 0D7B	NA, VIRAMA, CHILLU N	/nn/
	0D7D 0D4D 0D2A 0D41	CHILLU L, VIRAMA, PA, SIGN U	/lpu/
	0D7E 200D 0D4D 0D35 0D3E	CHILLU LL, ZWJ, VIRAMA, VA, SIGN AA	/lvā/
	0D7E 0D4D 0D2A 0D4D 0D2A 0D3F	CHILLU LL, VIRAMA, PA, VIRAMA, PA, SIGN I	/lppi/

As in case of any consonant conjunct, *chillu* conjuncts encoded with virama should not be laid out horizontally, as if, virama is omitted. For example, it is incorrect to layout <CHILLU N, VIRAMA, RRA> as ന്നറ. This conjunct has the standard straightforward encoding as <CHILLU N, RRA>.

Chillu forms other than the atomically encoded *chillus* (U+0D7A .. U+0D7F, U+0D54 ..

U+0D56), should be produced using VERTICAL BAR VIRAMA. The VERTICAL BAR VIRAMA in a cluster can choose to ligate or not, with a preceding consonant or consonant ligature. However, already encoded nine *chillu* characters should not be formed by ligating VERTICAL BAR VIRAMA with a base consonant.

Table 12-39. Examples of *chillu* forms with VERTICAL BAR VIRAMA

	0D28 0D4D 0D28 0D3B	NA, VIRAMA, NA, VERTICAL BAR VIRAMA	/nn/
	0D28 0D3B	CA, VERTICAL BAR VIRAMA	/c/
	0D26 0D3B	DA, VERTICAL BAR VIRAMA	/d/

Request 2

In the section 12.9 (page 505) of Unicode Standard 11.0.0., the subheading *Chillu Forms* needs to be updated as additional three chillus are already encoded. So the text should start as:

Chillu Forms. The ~~six~~nine characters, U+0D7A..U+0D7F, U+0D54..U+0D56, encode dead consonants (those without an inherent vowel) known as *chillu* or *cillakṣaram*. In Malayalam language text, chillu forms never start a word. Occasionally, *chillu* forms may take vowels or be elements of conjuncts and textually behave as regular consonants. The *chillu* forms *nn*, *-n*, *-rr*, *-l*, and *-ll* are quite common; *chillu-k* is relatively rare in contemporary usage; and *chillu-m*, *-y*, *-lll* are only found in historical texts.

Request 3

In the section 12.9 (page 501) of Unicode Standard 11.0.0., the subheading *Historic Characters* needs to be updated as many more historic characters have been encoded in recent versions of the standard:

Historic Characters. The ~~four~~ characters, *avagraha*, *vocalic rr*, *vocalic rr sign*, *vocalic l*, *vocalic l sign*, *vocalic ll*, *vocalic ll sign*, *combining anusvara above*, and *chandrabindu*, are only used to write Sanskrit words in the Malayalam script. The *vocalic l sign* and *avagraha* is are the most common of the ~~four~~ these. The *avagraha* is also called *praślēṣam*. The *vocalic l sign* is also commonly used in Sanskrit words.

Two specific forms of viramas are found in historical materials. The U+0D3B MALAYALAM SIGN VERTICAL BAR VIRAMA was used to indicate a pure consonant when transliterating Sanskrit words in 17th century manuscripts, while the U+0D3C MALAYALAM SIGN CIRCULAR VIRAMA was employed to indicate a pure consonant in native Malayalam texts from the turn of 20th century.

The U+0D29 MALAYALAM LETTER NNNA is the counterpart of U+0BA9 TAMIL LETTER NNNA. This along with U+0D3A MALAYALAM LETTER TTTA it was used to represent Malayalam

phonetic reading.

The archaic vowel *ii* was in use until 20th century. However, historical chillu letters *-m*, *-y*, *-lll* are rare in historical texts itself.

Malayalam numerics U+0D70..U+0D72 represents ten, one hundred and one thousand. The fractions from $\frac{1}{160}$ to $\frac{3}{16}$ are specified in the range U+0D58..U+0D5E, U+0D73..U+0D78. They are found in historical texts only. This includes the measurement sign *para* also.

Request 4

In the Unicode chart for Malayalam range (<https://unicode.org/charts/PDF/U0D00.pdf>), the chillus U+0D54..U+0D56 should be categorized as *Additional historical chillu letters* to differentiate them from the chillus U+0D7A..U+0D7E in common use today:

Additional historic chillu letters

0D54	ച	MALAYALAM LETTER CHILLU M
0D55	യ	MALAYALAM LETTER CHILLU Y
0D56	ല	MALAYALAM LETTER CHILLU LLL

