

Proposed Changes for Section 9.9 of Unicode Standard

Santhosh Thottingal santhosh.thottingal@gmail.com

Malayalam Conjuncts example

Table 9-26 of lists ര + റ് + പ → റ്പ. This is wrong, since the small dot on top of the പ has a separate code point - U+0D4E MALAYALAM LETTER DOT REPH.

The sequence U+0D4E + U+0D2A forms the /rpa/ glyph shown in the table. ര + റ് + പ forms ര്പ and does not qualify to name as conjunct since it does not form any ligature.

Table 9-26. Malayalam Conjuncts

ക	+	റ്	+	പ്പ	→	കപ്പ	(kpa)
ക	+	റ്	+	ക	→	കക	(kka)
ജ	+	റ്	+	ഞ	→	ജഞ	(jña)
ട	+	റ്	+	ട	→	ട്ട	(tta)
പ	+	റ്	+	പ	→	പ്പ	(ppa)
ച	+	റ്	+	ച	→	ച്ഛ	(ccha)
ബ	+	റ്	+	ബ	→	ബ്ബ	(bba)
ന	+	റ്	+	യ	→	ന്യ	(nya)
പ	+	റ്	+	ര	→	പ്ര	(pra)
ര	+	റ്	+	പ	→	ർപ	(rpa)
ശ	+	റ്	+	വ	→	ശ്വ	(śva)

Malayalam Chandrakkala

Table 9-27 lists 3 examples for *Chandrakkala*. The first one is *Samvruthorakaram* example. പാല് Written using Chandrakkala(0D4D) is the example for Chandrakkala usage. Samvruthorakaram often used without the explicit /ഉ/ vowel sign, also known as *pseudosamvruthokaram*(Refer L2/05-210 R Chitrajakumar, N Gangadharan) is written just with Chandrakkala. See <http://ml.wikipedia.org/wiki/പാല്> and http://en.wikipedia.org/wiki/Malayalam_alphabet#Chandrakkala

പാല് as given in the table is a not-widely used writing form where the /ഉ/ vowel sign is explicitly given.

Table 9-27. Candrakala Examples

പാല്	/paalə/ milk	0D2A, 0D3E, 0D32, 0D41, 0D4D
എന്നാ	/ənnaa/ on which day?	0D0E, 0D4D, 0D28, 0D4D, 0D28, 0D3E
ഐശീല്	/aishiiləm/ than ice	0D10, 0D36, 0D40, 0D32, 0D4D, 0D02

The second example, എന്നാ is also wrong since Chandrakkala cannot be applied to a Vowel letter. And there is no Samvruthorakaram form for Vowels. It is written as എന്ന in contemporary Malayalam usage. See [https://www.google.co.in/#hl=ml&q="എന്ന"](https://www.google.co.in/#hl=ml&q=)

The last one in the table ഐശീല് is an invalid example. Ice is written as ഐസ് in Malayalam, /aishiilem/ does not mean 'than ice'. 'Than ice' can be 'ഐസിനെക്കാൾ'. See

[https://www.google.co.in/#hl=ml&q="](https://www.google.co.in/#hl=ml&q=)ഐസിനെക്കാൾ"

The /tta/ and /rra/ sequences

Table 9-29 shows the example of /rr/ and /tt/ usage. One obvious mistake in this table is spelling mistake in 3rd and 4th rows. /maattoli/ is മാറ്റോലി and not മാറ്റേലി /maatteli/. Moreover the given sequence at row 4 corresponds to: മ + ഞ + റ + റ് + ഞ + റ + ഞ + ല + ി, is wrong, it does not use the correct /o/ vowel sign- ഞ 0D4A.

This section introduces lot of confusion to the /rr/ /tta/ writing.

റ്റ is written as റ + റ് + റ == 0D31 0D4D 0D31. Which can produce a stacked form in Old orthography and horizontal form ററ in new orthography.

ററ is written as റ + റ == 0D31 0D31. Both modern and old orthography produce a horizontal layout of two റ letters.

According to the table, /paatta/ can be written use two different unicode sequences.

0D2A 0D3E 0D31 0D31,
0D2A 0D3E 0D31 **0D4D** 0D31.

The example used from English to explain this behavior is *cathode* and *cathouse*. The th in those words does not differ by code points.

Table 9-29. Malayalam /rr/ and /tt/

പാറ്റ	0D2A 0D3E 0D31 0D31	/paatta/	cockroach
പാറ്റ	0D2A 0D3E 0D31 0D4D 0D31		
മാറ്റോലി	0D2E 0D3E 0D31 0D46 0D31 0D3E 0D32 0D3F	/maattoli/	echo
മാറ്റേലി	0D2E 0D3E 0D31 0D4D 0D31 0D46 0D3E 0D32 0D3F		
ബാറ്ററി	0D2C 0D3E 0D31 0D31 0D31 0D3F	/baattari/	battery
ബാറ്ററി	0D2C 0D3E 0D31 0D4D 0D31 0D31 0D3F		
സൂറ്റർ	0D38 0D42 0D31 0D31 0D31 0D4D	/suuratt/	(name of a place)
സൂറ്റർ	0D38 0D42 0D31 0D31 0D4D 0D31 0D4D		
ടെമ്പററി	0D1F 0D46 0D02 0D2A 0D31 0D31 0D3F	/temparari/	temporary (English loan word)
ലെക്ചററോട്	0D32 0D46 0D15 0D4D 0D1A 0D31 0D31 0D4B 0D1F 0D4D	/lekcararoot/	to the lecturer

The sequence 0D2A 0D3E 0D31 0D31 is /paarara/ and there is no reason to read it /paatta/
 The sequence in 3rd row 0D2E 0D3E 0D31 0D46 0D31 0D3E 0D32 0D3F is /maarareli/ and
 not /maattoli/ while the 4th row is correct.

Row 1 conflicts with Row 9, where 0D31 0D31 is given as /tt/ and in row 9 as /rara/. In the
 same way it conflicts with row 10.

I propose a new table and it is given below.

Code point Sequence	Read as	Old Orthography	New Orthography	Meaning	Comment
0D2A 0D3E 0D31 0D4D 0D31	/paatta/	പാറ്റ	പാററ	Cocroach	
0D2A 0D3E 0D31 0D31	/paarara/	പാററ	പാററ	No meaning	Invalid word
0D2E 0D3E 0D31 0D4D 0D31 0D4A 0D32 0D3F	/maattoli/	മാറ്റൊലി	മാരൊലി	Echo	
0D2E 0D3E 0D31 0D31 0D4A 0D32 0D3F	/maararoli/	മാരൊലി	മാരൊലി	No meaning	Invalid word
0D2C 0D3E 0D31 0D4D 0D31 0D31 0D3F	/baattari/	ബാറ്ററി	ബാററി	Battery	
0D2C 0D3E 0D31 0D31 0D31 0D3F	/baarari/	ബാററി	ബാററി	No meaning	Invalid word
0D38 0D42 0D31 0D31 0D4D 0D31 0D3F	/suuratt/	സൂരട്ട്	സൂററ്	Name of a place	
0D38 0D42 0D31 0D31 0D31 0D3F	/suurarar	സൂററ്	സൂററ്	No meaning	
0D32 0D46 0D15 0D4D 0D1A 0D31 0D31 0D4B 0D1F 0D4D	/lekcararoot/	ലെക്ചററോട്	ലെക്ചററോട്	To the lecturer	

Looking at the original or proposed table, one can understand that the new orthography

does not differentiate between /rra/ and /tta/ visually. But it should be also note that there is no rule or convention in new orthograhya that /tta/ should not be stacked. Many new orthography fonts follow stacked form of /tta/ to avoid this ambiguity.

The /nta/ sequence

The /nta/ sequences given in Table 9-30 also require corrections as explained below. The /nta/ is a commonly used letter and it existed before Unicode 5.1 when 0D7B introduced.

First of all the code points used is wrong, 0D06 0D7B 0D47 0D31 0D3E corresponds to അ + ന് + ഞ + ഞ . What is intended is അ + ന് + ഞ + ഞ .

Just like new orthography does not distinguish between ഞ and ഞ, It has same visual form for അ, ഞ. But the underlying code point sequence is same. - അ + ഞ + ഞ => 0D28+ 0D4D+ 0D31.

Table 9-30. Malayalam /nr/ and /nt/

അന്തോ	0D06 0D7B 0D47 0D31 0D3E	/aantoo/	(proper name)
അന്തോ	0D06 0D7B 0D4D 0D31 0D47 0D3E		
എൻറോൾ	0D0E 0D7B 0D31 0D47 0D3E 0D7A	/enrool/	enroll (English word)

The അ + ഞ + ഞ rendering of ഞ with new and old orthograhya is given in the below table.

Code point Sequence	Read as	Old Orthography	New Orthography	Meaning	Comment
0D06 0D28 0D4D 0D31 0D4B	/aantoo/	അന്തോ	അന്തോ	A name- Anto	
0D0E 0D7B 0D31 0D4B 0D7A	/enrool/	എൻറോൾ	എൻറോൾ	enroll	

1. It is logically incorrect to have two code point sequences for the same word as given in First two rows of table 9-30.
2. ഞ was never written using a chillu code point(atomic or non atomic-zwj based).
3. Chillu (0D7B in this case) being dead consonant, a virama will not attach to it for another ligature. Ie the sequence 0D7B + 9D4D is wrong.
4. There is no reason to redefine the sequence of an existing widely used ഞ, when it is widely present in digital space, people write their names using that, “my” is എന്റെ in Malayalam, and there are too many people with name Anto written using അ + ഞ + ഞ. Eg: http://ml.wikipedia.org/wiki/അന്തോ_അന്തോ.
5. Similar to the /rra/-/tta/ case, Modern orthography also uses stacked form of nta, to avoid ambiguity with side by side rendering.