

Changes related to Malayalam in Unicode 5.1.0 from 5.0

Cibu C. Johny
cibu@google.com
2007-Apr-21
version 0.2

Effective of Unicode 5.1, representation of Malayalam Chillus, /nta/ conjunct and dot style Reph would change. The new representation could be summarized as:

- All the Chillus are atomically encoded
- The /nta/ conjunct is represented by <Chillu-NA, Virama, RRA>
- Dot reph is represented by <Chillu-RA, Virama, Consonant>

The tables below detail these changes with many examples.

Changes related to representation of Chillus:

Table 1A	Visual	Codepoint sequence in 5.0	Codepoint sequence in 5.1.0 (atomic encoding)	Sample Words
C1	ൺ	൩, ്, ZWJ NNA, VIRAMA, ZWJ	ൺ	ആൺ /aann/
C2	ൻ	ന, ്, ZWJ NA, VIRAMA, ZWJ	ൻ	അവൻ /avan/
C3	ർ	ര, ്, ZWJ RA, VIRAMA, ZWJ	ർ	അയർ /njaayarr/
C4	ൽ	ല, ്, ZWJ LA, VIRAMA, ZWJ	ൽ	ഉൽപ്പത്തി /ulpatti/
C5	ൾ	ള, ്, ZWJ LLA, VIRAMA, ZWJ	ൾ	അവൾ /avall/
C6	ൽ	Not defined	ൽ	ദ്രക്സാക്ഷി /drksaakssi/

Table 1B	Levels of rendering fallbacks for the code sequence <C1, VIRAMA, C2>					
	Unicode 5.0				Unicode 5.1	
	Microsoft Interpretation		Linux Interpretation			
	Visual	Example	Visual	Example	Visual	Example
F1	C1 + C2 Conjunct	ൻ	C1 + C2 Conjunct	ൻ	C1 + C2 Conjunct	ൻ
F2	C1, Virama, C2	ൻ്	Chillu-C1, C2	ൻ	C1, Virama, C2	ൻ്
F3			C1, Virama, C2	ൻ്		

Table indicating the changes specific to Chillu NA(ൻ) and RRA(റ) combinations:

Table 2	Visuals	Shortest Codepoint sequence in 5.0		Sequence in 5.1.0	When joiner is lost	Fallback	Sample Words
		Microsoft interpretation	Linux interpretation				
N1	ൻറ	ന, ്ര, റ	ന, ്ര, ZNWJ, റ	ന, ്ര, റ			ഉപ്പുമാവിൻറവ /uppumaavini-rava/
N2	ൻറ	ന, ്ര, ZWJ, ZWNJ, റ	ന, ്ര, ZWJ, റ	ൻ, റ			ഹെൻറി /henri/
N3	ൻറ	Not defined		ൻ, ്ര, ZWJ, റ	N4	N4	എൻറെ /ente/
N4	ൻറ	ന, ്ര, ZWJ, റ	ന, ്ര, റ	ൻ, ്ര, റ		N3	എൻറെ /ente/

Position of left part of the vowel sign, if used.

Following table indicates the changes related to dot-reph representation:

Table 3	#	Visual	Codepoint sequence in 5.0	Codepoint sequence in 5.1.0	Fallback	Sample Words
ര(RA) with യ(YA)	D1	യ്	Not defined	ര, ി, യ		No examples found yet.
	D2	ര്യ	ര, ി, ZWJ, യ	ര, യ		ചതുര്യഗം /chaturyugam/
	D3	ര്യ	ര, ി, യ	ര, ി, യ	D4	ഭാര്യ /bha:rya/
	D4	ര്യ	ര, ി, ZWNJ, യ	ര, ി, ZWNJ, യ		No examples found yet.
	D5	ര്യ	ര, ി, യ, ി, യ	ര, ി, യ, ി, യ		ഭാര്യ /bha:rya/
	D6	ര്യ	ര, ി, ZWJ, യ, ി, യ	ര, യ, ി, യ	ര്യ	No examples found yet.
ര(RA) with വ(VA)	D7	വ്	ര, ി, വ	ര, ി, വ		നിവാണം
	D8	രവ	ര, ി, ZWJ, വ	ര, വ		പാർവതി /bha:rya/
	D9	രവ	Not defined	ര, ി, വ	D10	വരവ /varvo/ (colloquial)
	D10	രവ	ര, ി, ZWNJ, വ	ര, ി, ZWNJ, വ		No examples found yet.
	D11	രവ	ര, ി, വ, ി, വ	ര, ി, വ, ി, വ		പാർവതി /bha:rya/
	D12	രവ	ര, ി, ZWJ, വ, ി, വ	ര, വ, ി, വ	രവ	സർവ്വം /sarvvam/
RA with other consonants	D13	ഷ്	ര, ി, ഷ	ര, ി, ഷ		വർഷം /varsham/