

Visuals and corresponding storage representations of the edge cases in Malayalam

Cibu C. Johny
cibu@google.com
2007-02-22
version 1.0

Purpose of this document is to clarify the storage representation of specific Malayalam visuals for which the representation is not apparent from the Indic model. These representations use the upcoming atomic Chillu encoding.

There is a list of tables in the following pages. These tables enumerate the various visuals possible for Malayalam NA and RRA combinations along with corresponding storage representation. The pattern of storage representation is roughly restricted to <NA, RRA, Vowel>. Dot-Reph cases are also covered.

These visuals are based on the assumption that, the font has all the possible conjuncts. If the font lacks a specific conjunct, the fallback visual is also specified. Following are the list of tables:

1. NA, RRA combinations without any vowel sign
2. NA, RRA combinations with a vowel sign that has only left part
3. NA, RRA combinations with a vowel sign that has only right part.
4. NA, RRA combinations with a vowel sign that has both left and right parts.
5. Dot Reph combinations with YA, VA and other consonants
6. U-sign variants and Samvruthokaram
7. VA and RA-sign forms
8. Common conjunct that are often misrepresented in currently available fonts

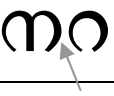
Table 1	Visuals without vowel sign	Proposed Storage Code	When joiner is lost	Fallback	Sample Words	Notes
N1		ന, ്, റ			ഉപ്പുമാവിൻ റവാ /uppumaavini- rrava/	Valid visual; however, not typically used. Seen only at word boundaries when space is lost.
N2						Not a valid visual for the given storage representation pattern
N3		ന, റ			നറു /narru/	
N4						Not a valid visual for the given storage representation pattern
N5		ൻ, റ			ഹെൻറി /henri/	
N6		ൻ, ്, ZWJ, റ	N7	N7	എൻറെ /ente/	Same reading as N7. The words that do not use left part vowel sign for this conjunct may have confusable pair as N5.
N7		ൻ, ്, റ		N6	എൻ്റെ /ente/	Same reading as N6. This form is preferred over N6.

Table 2	Visuals with left only vowel sign	Proposed Storage representation	Fallback	Sample Words	Notes
L1	പ്റെ	ന, ്, റ, െ	None	പൊന്നിൻ- റെക്കോഡ് /ponnini-rrekko:di/	
L2					Placeholder for N2
L3	മ്പറെ	ന, റ, െ	None	ഭാവനരേഡി /bha:vana-rreddi/	
L4					Placeholder for N2
L5	മ്പ്റെ	മ്പ, റ, െ	None	മുൻറെക്കോഡ് /mun'rrekko:di/	
L6	െമ്പറ	മ്പ, ്, ZWJ, റ, െ	None	എൻെറ /ente/	Same reading as L7
L7	െമ്പ്റ	മ്പ, ്, റ, െ	L6	എൻെ /ente/	
മ്പ = Chillu-NA ് = Virama െ = vowel sign E					

Table 3	Visuals with right only vowel sign	Storage Representation	Fallback	Sample Words	Notes
R1	പ്റു	പ, ്,റ, ു	None	പൊന്നിൻറുപ്പിക /ponnini-rruppika/	
R2					Place holder corresponding to N2
R3	നറു	ന, റ, ു	None	നറു /narru/	
R4					Place holder corresponding to N4
R5	തറു	ത, റ, ു	None	ടിൻറാക്ക് /tin-rack/	
R6	തറു	ത, ്, ZWJ, റ, ു	None	പിൻറു /pintu/	Same reading as R7
R7	തൂ	ത, ്, റ, ു	R6	പിൻറു /pintu/	
ത = Chillu-NA ് = Virama ു = vowel sign U					

Table 4	Visuals with two part vowel sign	Storage Representation	Fallback	Sample Words	Notes
S1	ൻറൊ	ന, ്, റ, െ	None	പൊന്നിൻറൊക്കം	
S2					Place holder for N2
S3	നറൊ	ന, റ, െ	None	തിനറൊട്ടി /tinarrodi/	
S4					Place holder for N4
S5	ൻറൊ	ന, റ, െ	None	തേൻറൊട്ടി /te:n'rrodi/	
S6	ൻറൊ	ന, ്, ZWJ, റ, െ	None	എൻറൊപ്പം /entoppam/	Same reading as R7
S7	ൻറൊ	ന, ്, റ, െ	S6	എൻറൊപ്പം /entoppam/	
ൻ = Chillu-NA ് = Virama െ = vowel sign O					

Table 5	Visual: Dot Reph	Storage Code	Fallback	Sample Words
D1	യ്	ര, ്, യ	None	No examples found yet.
D2	രയ	ര, യ	None	ചതുരയുഗം /chaturyugam/
D3	ര്യ	ര, ്, യ	ര്യ	ഭാര്യ /bha:rya/
D4	ർയ്യ	ര, ്, യ, ്, യ	None	ഭാര്യ /bha:rya/
D5	രയ്യ	ര, യ, ്, യ	രയ്യ	No examples found yet.
D6	വ്	ര, ്, വ	None	നിവാണം
D7	രവ	ര, വ	None	പാർവതി /bha:rya/
D8	രവ	ര, ്, വ	രവ	വരോ /varvo/ (colloquial)
D9	ർവ്വ	ര, ്, വ, ്, വ	None	പാർവതി /bha:rya/
D10	രവ്വ	ര, വ, ്, വ	രവ്വ	സർവ്വം /sarvvam/
D11	ഷ്	ര, ്, ഷ	രഷ	വർഷം /varsham/

Table 6	Visuals related to Samvruthokaram	Storage Representation	Fallback	Sample Words	Notes
U1	ന്	ന, ZWNJ, ു, ്	U2	അവനും /avanɪ/	
U2	ന	ന, ZWJ, ു, ്	U1	അവനും /avanɪ/	Same reading as U1
U3	ന്	ന, ്	None	അവന് /avanɪ/	
U4	നു OR ന	ന, ു	U5	മനു /manu/	Rendering engine decides the visual
U5	നു	ന, ZWNJ, ു	U6	മനു /manu/	
U6	ന	ന, ZWJ, ു	U5	മന /manu/	Same reading as U3
ന = NA ് = Virama ു = vowel sign U ു = samvruthokaram					

Table 7	Visuals related conjuncting ഖ	Storage Representation	Fallback	Sample Words	Notes
V1	തവ	ത, ്, വ	തവ	തമ്പി /tanvi/	
V2	തവ	ത, ്, വ	V3	താവാരം /ta:r'va:ram/	Same reading as V3. Fallback cannot be തവ.
V3	തവ	ത, ്, ZWNJ, വ	None	താവാരം /ta:r'va:ram/	
V3	തവ	ത, ZWJ, ്, വ	V3	താവാരം /ta:rvo:/	
V4	വവ	വ, ്, വ	വവ	ഉവ /uwwa/	
V5	വവ	വ, ZWJ, ്, വ	വവ	ഉവോ /uw'wo:/	
ത = NA ് = Virama ു = vowel sign U ൃ = samvruthokaram					

Conjuncts, as in V2, is formed only in following regular expression context:
/[തവ]്(വര)/

Fallbacks for these forms will not be with വര subjoining form. Instead, it will be with explicit virama as in V3.

Table 8	Common Conjuncts	Storage Representation	Sample Words
C1	ക	ങ, ്, ക	പങ്ക /panka/
C2	ഞ്ച	ഞ, ്, ച	ഇഞ്ച /incha/
C3	ണ്ട	ണ, ്, ട	മണ്ട /manda/
C4	ന്ത	ന, ്, ത	മൊന്ത /monta/
C5	മ്പ	മ, ്, പ	പമ്പ /pampa/