

Dot-Reph and Subscript YA/VA Sequences in Malayalam Unicode

Naga Ganesan (naa.ganesan@gmail.com) May 2, 2008

Malayalam script in Unicode is normally treated as an Indic script with Reph (Cf. Table 9-23, TUS 5.0, pg. 336, Malayalam Conjuncts). For example, "rpa" conjunct <ra, virama, pa> is defined in Table 9-23 which produces the dot-reph.

Table 3 in L2/07-402 gives a list of possible Visuals with Dot-Reph over YA or VA for Malayalam. This document proposes a simple set of dot-reph sequences in Unicode to represent all the Visuals in Table 3, L2/07-402 for UTC consideration. Approved Dot-Reph sequences should be documented in the next version of the Unicode Standard explicitly. Right now, the developers and users of Malayalam Unicode are not given clear instructions on how to handle Reph sequences

While making the following list of Malayalam reph sequences in Tables 1 through 3, particular attention has been made to treat the Reph in Malayalam in the usual manner as listed in TUS 5.0 or earlier. So, minimum changes, if any, are needed for dot-reph sequences.

Table 1. Dot Reph RA + YA

	Visual	Suggested Encoding in TUS	Notes
1	ര്യ	CHILLU RR, YA	
2	ര്യ	RA, ZWJ, VIRAMA, YA	As in Kannada. Pg. 334, TUS 5.0
3	ര്യ	RA, VIRAMA, ZWNJ, YA	
4	ര്യ	RA, VIRAMA, YA	Regular reph
5	ര്യ	RA, VIRAMA, YA, VIRAMA, YA	

Table 2. Dot Reph RA + VA

	Visual	Suggested Encoding in 5.1	Notes
1	ರೞ	CHILLU RR, VA	
2	ರೞ	RA, ZWJ, VIRAMA, VA	As in Kannada. Pg. 334, TUS 5.0
3	ರೞ	RA, VIRAMA, ZWNJ, VA	
4	ರೞ	RA, VIRAMA, VA	Regular reph
5	ರೞ	RA, VIRAMA, VA, VIRAMA, VA	

Table 3. Dot Reph RA + SSA

	Visual	Suggested Encoding in 5.1
1	ರೞ	RA, VIRAMA, SSA

In Row 2, Tables 1 and 2, <ra, zwj, virama, ya/va> is used for {RA_SUBSCRIPT_YA/VA} sequence. The use of ZWJ for ರೞ in **Malayalam** is parallel with **Kannada Unicode** in a similar situation. Page 334, TUS 5.0

Consonant Clusters involving RA. Whenever a consonant cluster is formed with the U+0CB0 ರ KANNADA LETTER RA as the first component of the consonant cluster, the letter ra is depicted with two different presentation forms: one as the initial element and the other as the final display element of the consonant cluster.

U+0BC0 റ ra + U+0CCD ി̇ halant + U+0C95 റ ka -> റ്റ rka

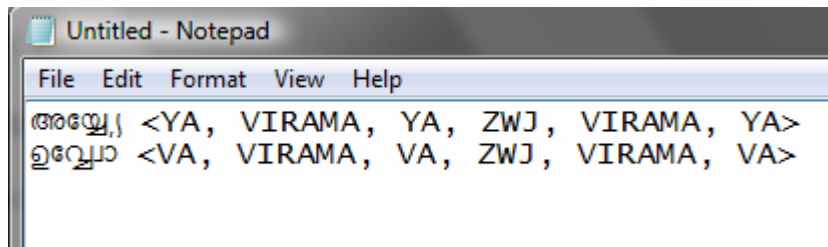
U+0BC0 റ ra + ZWJ + U+0CCD ി̇ halant + U+0C95 റ ka -> റ്റ̣ rka”

It is noteworthy that in the Malayalam script user community, there is wide consensus for Row 2, Tables 1 and 2 to employ <ra, zwj, virama, ya/va> sequence. Both Cibu Johny and Rachana team members like Rajeev Sebastian indicated that this is the proper sequence for in Indic and South Asia Unicode lists.

In traditional style fonts, {REPH_OVER_YA} can be <RA, VIRAMA, YA>. And, {RA_SUBSCRIPT_YA_SIGN} = <RA, ZWJ, VIRAMA, YA>.

On 3/21/08, Cibu C J <cibucj@gmail.com> in South Asia list wrote: >> <RA, ZWJ, VIRAMA, YA> representing {REPH_OVER_YA} is not >>that logical. It should be {RA_SUBSCRIPT_YA_SIGN}.

For rare subscripted YA/VA following YYA and VVA also, <YYA/VVA, ZWJ, VIRAMA, YA/VA> sequence is working in Microsoft Vista SP1 (Santhosh Pillai, email in South Asia list, 3/21/2008).



MS Vista SP1 representation

Summary: Malayalam dot-reph sequences are proposed in Tables 1, 2, and 3. Use of ZWJ for Malayalam {RA_SUBSCRIPT_YA/VA} recommended sequences:

രൂപ്യ = <RA, ZWJ, VIRAMA, YA> and

രൂപ്യ = <RA, ZWJ, VIRAMA, VA>.

N. Ganesan