

Request to finalize encoding model for Telugu Reph

Shriramana Sharma, jamadagni-at-gmail-dot-com

2012-Jan-17

L2/06-334 by Nagarjuna Venna provides attestation for usage of the reph in the Telugu script in older orthography and notes that in newer orthography the reph is mostly not used. Thus, the word dharma would be written in the old orthography as:

ధమఁ

... and in the new orthography as:

ధర్మ

Thus in a Telugu sequence RA + VIRAMA + C where C is any consonant, in the older orthography it is RA which takes the C1-conjoining form in the cluster, with the following consonant C retaining its nominal form. In the newer orthography, RA retains its nominal form and the following consonant takes the C2-conjoining form.

In a normal Telugu font catering to the contemporary style, given that:

RA + VIRAMA + CONSONANT → [ra sub_base_consonant]

ఠ + ం + మ → ఠ్మ

the following should be used for selecting the reph:

RA + VIRAMA + ZWJ + CONSONANT → [consonant reph]

ఠ + ం + ZWJ + మ → మఁ

And in an old-style Telugu font the following would happen:

RA + VIRAMA + CONSONANT → [consonant reph]

ఠ + ం + మ → మఁ

whence absence of reph would be selected by:

RA + ZWJ + VIRAMA + CONSONANT → [ra sub_base_consonant]

ఠ + ZWJ + ం + మ → ఠ్మ

The above sequences are appropriate since the reph is a C1-conjoining form and it should be selected using VIRAMA + ZWJ, and when the reph is not desired the sub-base form of the following consonant should be selected using ZWJ + VIRAMA.

I find that L2/06-334 which originally suggested the above has been considered in UTC 109, resulting in:

[109-A79] Action Item for Eric Muller: In preparing the UTR on Indic issues, consider document L2/06-334, and add to the Telugu discussion of alternate forms of "R" analogous to Gurmukhi.

I do not find a UTR on Indic issues at <http://www.unicode.org/reports/index.html#reports> but anyhow the above model should go into TUS under the rendering section for Telugu.

-0-