

Response to PRI 250 on mechanisms to specify optional conjuncts in Malayalam

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

2013-Mar-09

I don't agree with this proposal which would cause deviation of the Malayalam rendering model from the existing and successful pan-Indic rendering model. I discuss this per-topic:

§1. ZWJ + VIRAMA for consonant ligatures

I don't agree with the re-specification of the ZWJ + Virama sequence for consonant ligatures. I explain my reasons below:

§1.1. ZWJ + VIRAMA is already used for C2-conjoining forms

The sequence C1 + ZWJ + VIRAMA + C2 is accepted in existing Indic practice to produce C2-conjoining forms. See for instance TUS 6.1 p 299 for Bengali R·YA or p 315 for Telugu R·MA. See also the original detailed (and accepted) proposal L2/04-249 p 14. (This was PRI 37.)

IIUC even in contemporary Malayalam orthography, there are at least four consonants that regularly take C2-conjoining forms – YA, RA, LA and VA. In a traditional orthography font which might ligate these with the C1 of the cluster, the sequence C1 + ZWJ + VIRAMA + C2 will *break* that ligature and produce C2's post-conjoining form. It is inappropriate to reassign that sequence to *produce* a ligature in a modern orthography font as that would mean inconsistent semantics. That is, users would no longer be able to rely on a particular sequence to always request a particular written form, which defeats the whole point of specifying these mechanisms.

For example, a traditional style font would present GA + VIRAMA + RA as a ligature 𑌒 , and GA + ZWJ + VIRAMA + RA will *break* that ligature and cause RA to take its C2-conjoining form, so: 𑌒 . The present proposal requests that GA + ZWJ + VIRAMA + RA should be *promoted* to a ligature in a modern style font which directly contradicts the earlier use.

If ZWJ + VIRAMA is thus given a new meaning, a user would no longer be able to relate this sequence to a given kind of written form throughout all kinds of fonts, which was the intention of PRI 37 which proposed the current Indic virama/joiner model.

§1.2. No need for a mechanism to request consonant ligatures

There is no real need for a mechanism to request consonant ligatures. The current situation where the font is expected to provide all desirable ligatures by default is quite satisfactory. Therefore, Indic has never had any sequence to request consonant ligatures.

While it is true that Bengali uses ZWJ to request vowel sign ligatures, the virama is not involved here. Combinations of virama with ZWJ/ZWNJ have specific existing semantics all of which are related to *prevention* of consonant ligatures to various degrees.

So far this model has evidently worked out very successfully since there have been no complaints from the Indic user community at large related to absence of mechanisms to *request* ligatures across Indic scripts.

Even in Malayalam, there has been no demand from the native user community *at large* for such a mechanism. There are evidently several native Malayalis active on encoding/rendering related matters (based on the traffic in the unicode/unicore/harfbuzz lists) and one has not observed any dissatisfaction voiced on how Malayalam written forms are supported. (I say this with all due respect to the native Malayali status of the author of the PRI background doc.)

The argument “it would be useful to have a sequence to request consonant ligatures” is entirely hypothetical and to my knowledge unlike the need/demand that was there from at least one section of the user community in the case of the Malayalam chillus or the Bengali Khanda TA there is no such need/demand in the case of consonant ligatures. Therefore, I do not consider it wise to modify the existing and successful virama/joiner model based on a hypothetical need.

§1.3. Involvement of Sanskrit and adverseness to it of current proposal

My own locus standi in the present matter is that of a Sanskrit scholar, and it is well known that Sanskrit is the only Indic language which regularly employs a high degree of consonant clusters even up to the extent of very heavy clusters involving 5/6 consonants. In fact the cluster g-dh-re provided in the PRI background doc for an example would only occur in Sanskrit (as in jagdhre, “to the eater”) and I am not aware of any other Indic language which would have such (heavy) clusters.

As a Sanskrit scholar, I am concerned with how the Indic encodings are able to represent Sanskrit, since Sanskrit is (again) the one language which is written in the most

number of Indic scripts, and I don't want Sanskrit text converted from one Unicode Indic script to another to suddenly produce dotted circles or weird (to say the least) rendering (aka awkwardly broken up clusters, misjoined vowel signs etc) because the joiner rules for one Indic script in Unicode are different from those of another.

§1.4. Let us not further complicate the joiner situation

The existing Indic joiner situation is complicated enough. Rendering software and font programmers already have to work very carefully to correctly support the existing joiner rules and there are always corner case bugs being fixed in popular software like Uniscribe and Harfbuzz. In my judgment, the hypothetical benefit to adding such a mechanism to request consonant ligatures is too little (both in the potential number of takers and in the potential number of situations where such a mechanism would be actually required) to justify the costs of rendering engines and font makers being required (well, in some way) to support the additional requested mechanisms at the risk of introducing new bugs in existing software.

For these reasons, I seriously disagree with specifying either this or any other contrived sequence (perhaps involving multiple joiners) to request consonant ligatures in any Indic script.

§2. ZWNJ + VIRAMA to break consonant ligatures

To complement the proposed ZWJ + VIRAMA model for *requesting* consonant ligatures in a new style font, the PRI doc proposes ZWNJ + VIRAMA as a mechanism to *prevent* consonant ligatures in an old style font. That is, if GA + VIRAMA + RA would be presented as $\text{GA} \text{RA}$, the sequence GA + ZWNJ + VIRAMA + RA is proposed to be presented as $\text{GA} \text{RA}$. In the current accepted encoding model, the sequence to be used in any style font for preventing the G-RA ligature and causing RA to take a C2-conjoining form is ZWJ + VIRAMA as in GA + ZWJ + VIRAMA + RA. There is no requirement for an additional sequence to request C2-conjoining forms in old-style forms, given that it is only advisable for ZWJ + VIRAMA should retain its semantics of doing the same as discussed in the preceding section.

Furthermore, the behaviour of ZWNJ has always been (and not just in Indic scripts) to merely act as an invisible obstacle to the rules which would cause conjoining or ligating behaviour and not to be itself included in glyph manipulation rules of fonts (TUS 6.1 p 551):

*For modern font technologies, font vendors should add ZWJ to their ligature mapping tables as appropriate. ... In contrast, ZWNJ will normally have the desired effect naturally for most fonts without any change, **as it simply obstructs the normal ligature/cursive connection behaviour.***

To request that ZWNJ + VIRAMA should produce C2-conjoining forms would hence be a deviation from existing practice of fonts and rendering engines of handling ZWNJ at all. If there were a justifiable need to include ZWNJ in glyph substitution rules, it could of course be done, but there is no such justification to deviate from existing practice.

§3. Stacks of chillu-s with consonants

The PRI doc requests the use of ZWJ to produce stacks of chillu-s with consonants and ZWNJ to prevent these stacks. This is inappropriate since as per TUS 6.1 p 321, chillu-s do not produce stacks with following consonants and the mechanism recommended for this is to insert a virama 0D4D between them. The example provided in the standard is: NA-CHILLU + RRA to be presented as ṚṚ and NA-CHILLU + VIRAMA + RRA as ṚṚ . It is recommended that the same mechanism be used throughout in the case of all stacked/non-stacked combinations of chillu-s with consonants such as the LLA-CHILLU + MA combination given in the PRI doc.

§4. Usage of ZWJ for vowel-sign ligatures

At the very outset, the PRI doc discusses sequences such as KA + VS-AA + DOT REPH + YA + VIRAMA + YA + VA + VS-U + ANUSVARA which involve a “spelling difference” between two orthographies i.e. a different encoded sequence to denote the same linguistic content (in the present case the word kāryyavum). It is to be noted that such a “spelling difference” will only occur (in the current encoding) for consonant clusters with /r/ in initial position, as these would be represented in the old orthography by the separately encoded DOT REPH character and in the modern orthography by RA-CHILLU.

The author notes that a modern style font would present such sequences as ṚṚ which is perceived as awkward, as it mixes up the old-style dot reph and the modern-style unligated VS-U. This is cited as justification for the proposal of a mechanism to request “conjuncts”. However, the author goes on to discuss only consonant clusters (which are not the issue in the given example) and no mechanism to help with the given example is considered at all.

What is actually required for this particular case is a mechanism to request vowel sign ligatures (and not consonant ligatures), so that the /vu/ is presented as a ligature. And naturally the sequence would be CONSONANT + ZWJ + VOWEL SIGN as used in Bengali (TUS 6.1 p 297). In fact, the author has quoted the Bengali situation, but omitted to request that the same model may be applied to Malayalam. However, Grantha also used such vowel-sign ligatures in old orthography and the same mechanism would apply there as well. (See my Grantha proposal L2/09-372 §5.2 on p 20 and §6.1 on p 34.)

Hence to additionally cater to the Malayalam/Grantha situation, I request that the description of the mechanism for Bengali vowel-sign ligatures be generalized in wording and moved to the general Indic rendering guidelines under TUS chap 9.1 on Devanagari.

§5. Conclusion

- 1) There is no requirement for a sequence to *request* consonant ligatures in any Indic script and the existing practice of requiring the font to provide the desirable ligatures by default is most satisfactory.
- 2) There is insufficient justification for Malayalam to deviate from such a successful pan-Indic model, especially considering that the consonant clusters under question are important chiefly to Sanskrit which should ideally have a uniform encoded representation across all Indic scripts including Malayalam.
- 3) The hypothetical need for a “would-be-useful” sequence to request consonant ligatures does not justify the costs in terms of upgrading and debugging rendering software and fonts to support the same.
- 4) Therefore it is recommended to avoid specifying new semantics for ZWJ + VIRAMA for Malayalam. This also makes the complementary counter-mechanism ZWNJ + VIRAMA unnecessary.
- 5) 0D4D MALAYALAM VIRAMA and not ZWJ should be used to cause stacks of chillu-s with consonants as in CHILLU + VIRAMA + CONSONANT. Without a virama, no stack would be produced, and ZWNJ is not necessary for the same.
- 6) The usage for ZWJ/ZWNJ in the case of vowel-sign ligatures currently documented for Bengali should be moved to the general Indic rendering discussion as being common to Bengali, Malayalam and other scripts.

-o-o-o-