# Response to PRI 250 of Unicode

**Swathanthra Malayalam Computing**

Prepared by: Anivar Aravind, Ashik S, Rajeesh Nambiar, Santhosh Thottingal and Suresh P

Email: contact@smc.org.in

Date: 2013 – April – 27

For ease, our response has been divided into 5 sections that provide a different perspective on the PRI 250 and why it solves a fictional, non existing problem and introduces real, unnecessary confusion into existing systems, ranging from encoding to input.

## The Orthographical argument

Addressing the example in the background document for citing the need of new proposal "കാര്യവും", the 'ransom-note' effect is fully prevalent in the current printing scenarios. Many newspapers and journals use a mix of traditional and reformed orthography since the language reformation in 1970's. So far, not a single party cited this as a blocking reason or inconvenience that jeopardize orthography system, though the language reformation (which many considers as a grave mistake) itself resulted in this cited issue. The cited issue is a hypothetical one at best, but the proposal to 'fix' it would introduce wide ranging undesirable effects.

The background document also says, this requirement is for displaying an old orthography ligature using a new orthography font. This will happen when proposal is accepted and fonts and rendering engines has to interpret the zwj/zwnj semantics. In that case, to interpret the zwj usage to create old orthography ligature, that new orthography font must have the ligatures rules and glyphs required for old orthography. But then that font is no longer a new orthography font, invalidating the base argument.

In addition to this, the proposal actually provides more ways to create RaNsOMe EfFeCt i.e. within a word, it becomes possible to use multiple orthographies. This effect and the /kaaryavum/ ("കാര്യവും") example are very hypothetical ones and never demands a change in way of writing Malayalam, possibly affecting it in a large scale.

# The Chillu argument

Case #3 in PRI 250 proposes usage of zwj in between chillu and consonant to create a stacked ligature.

ൾ + ZWJ + മ => ൾമ (stacked).

The ൾമ (stacked ligature) is never formed from a chillu. Chillus will never form conjuncts. They are vowel less pure consonants. ൾമ(stacked) is rare in usage, but if at all it is appearing in text, it is formed by ള് + മ. ള is base of chillu ൾ.

This example is similar to its ന counterpart- the stacked form ന്റ is never formed from chillu ൻ, but from ന.

ന്റ = ന+ ◌്+ റ

This is how ന്റ is defined and used widely in Malayalam, and this is how its defined in popular fonts.

But now that the chillus are encoded there a problem arose – ൻ+റ was discussed as possible combination for ന്റ, but again the linguistical argument that chillus will not form conjunct a came up. To avoid this issue, a virama was proposed and as of now the official, but never used sequence for ന്റ is ൻ + ◌് + റ

The existing nta(ന്റ) encoding in Unicode is wrong as indicated by its zero usage. But PRI 250 brings another way of writing chillu with stacking for the next consonant. This opens another possibility of writing ന്റ as ൻ+ zwj+ റ.

Case#4 of PRI 250 is the reverse of the above case and the arguments that are equally applicable.

To summarize, considering that we already have two different mechanisms that produce the same visual output with regards to chillus, this proposal paves the way for another unnecessary, third mechanism in the name of stacked ligature.

# The Opentype markup argument

If part of a text has to be displayed in a specific format, it warrants the use of markup to indicate such formatting. This requirement does not arise from any content consumer, but from a content producer – only in very specific and rare cases – such as while preparing a document explains differences in traditional and reformed orthography. All such scenarios can currently be handled using traditional and reformed fonts appropriately. This must not be marked at the encoding level, but at a higher level, and facility for this is provided by the Opentype spec. This requirement is equivalent to the demand that "there is a definite requirement to display the Latin ligature fi using fonts that does not have this ligature, thus ligature fi should be encoded as f + ZWJ + i".

# The Input method argument

As per the diagrams in PRI 250, this is what we understand about the order of joiners and virama.

1. C1+VIRAMA+ZWJ =  Chillu
2. C1+ ZWJ + VIRAMA+C2 =  C2 to conjoin to C1
3. C1+ VIRAMA+ZWNJ+C2 =  non joining C2
4. C1+ZWNJ+VIRAMA+C2 = C1+ Visually separate conjoining form of C2

To be fair to the content producers, we shouldn't expect anybody to remember these complications to write some rarely used conjunct. We would rather use different fonts with different orthography – as we can do now.

# The Rendering engine argument

Another perspective on this is the effort required to get all possible combinations of the above get bug free in rendering engine and fonts. The proposed sprinkling of ZWJ and ZWNJ poses great danger to the current state of improving rendering systems, such as Harfbuzz, where a lot of effort has been put over the past many years to correct the rendering. Adding additional nuances which even contradicts the current usages of ZWJ and ZWNJ in Indic languages, points to the future of incorrect rendering which cannot be corrected.

# Conclusion

Just stick to the well known adage, "if it ain't broke, don't fix it".