IP Response to Revised Malayalam LGR Update

DATE: 2020-05-02

The GP reviewed and agreed with the REVISED recommendation. Please see response after the REVISED recommendation section.

# Overview

The IP has reviewed the updated Malayalam LGR proposal dated 2020-03-26.

The IP tried to answer the following questions:

1. Does the proposed update apply the previous IP suggestions?
2. Do the proposed fixes introduce other problems?

The following recommendations address issues found. The IP is looking forward to a final revised update reflecting our REVISED recommendations. The IP has created an XML matching the REVISED recommendation and used that file to perform testing. We invite the GP to verify the suggestions before accepting the REVISED recommendation. We feel that with the final tweaks suggested here, the proposed update would be ready for public comment or submission.

# Review of GP Response to IP Recommendations

The following items lists the prior Integration Panel recommendations together with the GP response. In one case there is a REVISED IP RECOMMENDATION:

A1: The IP reviewed the text of the revised LGR and found no changes unrelated to the issue being updated. The text of the revised LGR needs some fixes (in some cases, the final text would depend on other recommendations made here, for example item A.2 and A.3). An annotated copy of the LGR document with notes and proposed language in redline is attached.

*The IP requests that the GP review the notes and suggestions in the attached document and make any changes appropriate (and, once resolved, delete any notes from the text).*

***Response: Noted.***

A2: Conflicts due to overlapped variants. The IP notes that the new sequence introduces an effective Null-variant (see section 6.6. in LGR-3 for a description of that). This does not seem to result in any issues. However, all the variant sequences in that set also end in 0D31 റ which introduces an **overlapped variant** with the existing variant sequences starting with 0D31 റ. *The IP has investigated these and determined that additional constraints are needed to ensure that the corresponding variant label sets are well-behaved.* (For details see Section 3 “Variant Issues” below).

***IP Recommendation:***

*(1) remove trailing 0D31 from the sequences defined for variant mappings for “nta”*

*(2) add when(followed-by-0D31) as a context rule on the variants*

*(3) add when(followed-by-0D31) as a context rule on the sequence 0D7B 0D4D (see also A 2 above)*

***Response: Done***

1. ***Edits in the document:*** 
   1. ***Used “(0D31)” in Table 7a.***
   2. ***Accepted the suggested paragraphs under Table 9.***
   3. ***Inserted V1 contextual rule for variant, the former V1, V2 become V2, V3.***
2. ***Edits in the XML*** 
   1. ***Removed trailing 0D31 from three sequences***
   2. ***Add rule “followed-by-0D31” and updated relevant descriptions***
   3. ***Add when(followed-by-0D31) as a context rule on the variants***

***\*Did not apply recommendation (3). The context rule on sequence of 0D7B 0D4D was edited as per******recommended in A4.***

*IP Verification: no further issues*

A3: New cross-script variants in LGR-4. Once integrated with the other scripts slated for LGR-4, the Malayalam LGR will also have an overlapped cross-script variant issue with Myanmar and Georgian (based on cross-script variants for 0D31 expected from the Myanmar proposal). This issue affects the sequence 0D31 0D31 and its variants. For details see Section 3.1, “ Cross-script overlapped variants” below.

***IP Recommendation on new cross-script variants***

*(1) Define cross script variants for 0D31 to match other expected LGR-4 LGRs (1002, 1077 and 10D8)*

*(2) Define equivalent mappings for the pair 0D31 0D31 (and its variant 0D31 0D4D 0D31), e.g. 1002 1002, and so on.*

***Response: Done***

1. ***Edits in the document:*** 
   1. ***Added 1077 and 0D31 mapping in section 6.2.3.***
   2. ***Added section 6.2.4 for cross-script variant with Georgian.***
   3. ***Added section 6.2.5 for overlapped cross-script variant of 0D31. These sequences also added to table 7a.***
2. ***Edits in the XML*** 
   1. ***Added 1002, 1077, 10D8 code point and added variant mapping for each code point.***
   2. ***Add mapping of 0D31 0D31 and its variant.***
   3. ***Add mapping of 0D31 0D31 0D4D 0D31 and its variants.***

*IP Verification:*

*In reviewing the GP’s modifications, the IP noted two issues.*

*(1) A fully consistent implementation of these cross script variants would require additional edits to context rules (we are omitting the details, since item (2) would render them moot).*

*(2) The only labels that would ever have actual variant labels in Myanmar and Georgian would be the two labels റ and ററ. All other labels would by necessity contain some Malayalam code point that does not have a variant with Myanmar or Georgian, making the whole label unique.*

*Because of (2), the IP feels the complexity of trying to account for these cross script variants is not in keeping with the Principles, and another, simpler solution must be sought. The simplest solution would be to disallow the two labels റ and ററ. This represents a minuscule restriction of the available name space, in exchange for removing dozens of variant mappings, not to mention avoiding further complex context rules that would be necessary (as stated in (1), we are omitting the details of what would have been required here, but the complexity cost would have been significant; with the new recommendation we also avoid introducing that complexity to the Georgian and Myanmar LGR via transitivity).  
  
REVISED IP RECOMMENDATION*

*(A) Remove all mention of 1002, 1077 and 10D8 from the XML (data and description)  
(B) Add a new WLE rule “only-RRA” to the XML (rule 10)  
(C) Add an action that makes invalid any label matching “only-RRA”  
(D) Add the new rule to the <description> w/ brief note for its rationale*

*(E) In the document, explain how 1002, 1077 and 10D8 do represent a set of cross- script homoglyphs for 0D31, but state that because in the case of these two scripts there’s only a single Malayalam code point with a cross-script variant, the simple solution is to disallow any labels that consist only of that code point (RRA). Further explain that because of other restrictions in the LGR, only the two labels റ and ററ are affected.*

*(F) Add the new rule to Section 7 (See attached proposal DOCx file with proposed edits)*

GP Response: Agreed with the revised recommendation and GP thanks IP for providing the updated files. GP has finalized the LGR package to be version 2.4, preparing for being published for public comment on 7 May 2020.

A4: WLE rule update: After detailed review of the issue, we now believe that the chosen resolution for the inconsistencies in the formulation of the WLE may be overly broad. Supporting a generic pattern   
<0D7B ൻ CHILLU N, 0D4D ◌് VIRAMA, **[any consonant here]**>   
when the intent appears to be to support a single exception for   
<0D7B ൻ CHILLU N, 0D4D ◌് VIRAMA, 0D31 റ RRA>   
seems unnecessary when a focused exception for <0D7B, 0D4D, 0D31> is desired.

By defining a sequence <0D7B, 0D4D, 0D31> the revised LGR already provides an implicit override for the original WLE rule 1. Because a WLE for a code point is never applied to that code point when it appears inside a sequence, the definition of sequence <0D7B, 0D4D, 0D31> would effectively override WLE 1 as it applies to 0D4D in the context of this sequence; therefore, *no modification for the rule* would be needed. The inconsistency in the published formulation of the rules could be resolved by making the XML match the document’s WLE as published, not the other way around.

However, as part of the recommendations for variants, IP suggest a change of <0D7B, 0D4D, 0D31> to <0D7B, 0D4D>, but with a context rule for the code point sequence: when(followed-by-0D31). Even though 0D31 would no longer be part of the sequence, the effect of adding it would still be to allow 0D4D to follow 0D7B if and only if 0D4D is followed by 0D31.

***IP Recommendation:***

*(1) The IP suggests that the GP remove the “or 0D7B” from both WLE1 and the comment attribute for the XML version of the rule. Also to rename the XML version of the rule to drop reference to 0D7B and finally to fix the XML rule itself (by removing 0D7B).*

***Response: see below, after (3)***

This recommendation assumes that a sequence <0D7B, 0D4D> is defined with context rule when(follows-0D31). However, this sequence already has a context rule, not-when(follows-B-X-or-H). As a result, a new merged context rule needs to be created: when(follows-other-than-B-X-or-H-and-followed-by-0D31). Note that the “follows-other-than-B-X-or-H” part would be equivalent to “follows-C-L-M-or-V” because those four classes complete the repertoire, and the rule would probably best be implemented in the XML using the union of those four sets anyway. Alternatively, all the context rules for chillus could be changed from “not-when” contexts to “when” rules with B-X-or-H changed to C-L-M-or-V. That’s a bit of an editorial choice. Whichever choice the GP makes, all documentation in Section 7 and the <description> need to agree.

***IP Recommendation:***

*(2) Create a context rule “follows-other-than-B-X-or-H-and-followed-by-0D31” and apply it to sequence <0D7B, 0D4D>. (See note above for alternative naming suggestions).  
(3) Document this merged rule in the <description>*

***Response:***

1. ***Edits in the document:*** 
   1. ***Removed ‘or 0D7B’ from the WLE Rule 1, section 7.1.2.***
   2. ***Added Rule 9 and its description***
2. ***Edits in the XML*** 
   1. ***Removed ‘or 0D7B’ from the rule name and updated the code for Rule1***
   2. ***Added Rule 9, and assigned to 0D7B 0D4D sequence: when (follows-C-L-M-or-V-and-followed-by-0D31***
   3. ***Added Rule 9 in the description***

*IP Verification: no further issues*

A5: Changes needed in the XML: The XML file needs a couple of fixes as well.

* The <date> element needs to be updated to the current date **– Updated to 26 March 2020**
* The <version> needs to be set to 4 (because the update will likely be part of LGR 4).**- Done**
* The <description> in its “Overview” section needs a sentence acknowledging that this is an update. - **Done**
* The section on “Variants” needs to mention any new sequence. (Any counts of sequences needs to be checked). **– Done – Added two paragraphs in Variant section.**
* Additional context rules on variants need to be documented in the <description>. **- Done**
* The change in WLE 1 (naming and content) will need to be documented **– Done**

And, of course, the changes to the actual specification of the sequences, the context rules and WLE1 and or variants need to be reflected in the <data> and <rules> sections.

*The IP requests the GP to make these updates to the XML.*

*IP Verification: except as affected by the REVISED IP RECOMMENDATION above, there are no issues.*

A6: Stability of “nta” conjunct encoding: In light of the fact that implementers seem to disagree on how to support this conjunct and the further fact that it is in current discussion in the Unicode Technical Committee, a case could be made asking for removal of all support for the stacked "nta" conjunct in the LGR for the reason that the encoding of that conjunct is "unstable".

***IP Recommendation:***

*The IP is monitoring this issue and has not come to a final conclusion. No action is requested at this time; however, we provide a summary of our best understanding of the issue below (For details see Section 4 “*Background on “nta”*”).*

**Response: Noted.**

*IP Verification: no further issues*