Response of IP on Tamil LGR Proposal of 2018-05-30

DATE: 2018-06-07

# Overview

In its response [2018-05-30] to the feedback of IP on Tamil LGR proposal dated 2018-03-02, NBGP appear to have accepted and responded to all IP requests, and in addition to have completed the unfinished WLE for U+0BB3 Visarga (aka Aytham or X), viz CP 0B83.

# Conclusion

In the previous round of comment, it was noted that the biggest unresolved issue in reviewing the Tamil LGR proposal draft was the fact that the Tamil script is closely related to Malayalam. Hence, agreement would be needed on a set of cross-script homoglyphs between the scripts, which might result in blocked variants.

At that date, there had been no corresponding proposal draft for the Malayalam script. This has now changed. The Malayalam proposal (.docx) of 2018-05-01 (in section 6.2.1), proposed six cross-script variants with Tamil; and this same set of six pairs is now recognized in this new proposal for Tamil, in section 6.4.

Since the changes proposed by IP to the Malayalam proposal did not affect the substance of these proposed cross-script variants, this matter may now be considered closed.

# Comments on main document (.docx)

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Issue** | **IP Comment** | **PK** |
| 1 | 0BC6 is part of 2 variant sets, but part of a sequence for one of them0BCD 🡨🡪0D46and0BCD 0BB3 🡨🡪0BCCThe IP is currently reviewing technical issues related to such cases. (Until that review is concluded, it is unknown whether this particular example presents any issues).More importantly, the general issue would seem to be thatU+0BB3 ள TAMIL LETTER LLA is effectively indistinguishable fromU+0BD7 ௗ TAMIL AU LENGTH MARK.If this is the case, defining a blocked variant0BB3 🡨 🡪 0BD7 would seem to uncouple the two variant definitions. | No immediate action, see attached analysis by the IP.This analysis identifies two options, of which one is to make no change.The IP is still discussing whether to recommend any changes.If any further action is required, IP will communicate this separately.In the meantime the GP is encouraged to review the analysis. | 0BD7 is rarely used, it is excluded from the LGR  |
| 2 | Reference 1003 can no longer be accessed. | Please replace by suitable alternative. | Done. |
| 3 | Please note here that reference 1004 is entirely in Tamil and give its Tamil title in addition to any translated title. | Please amend. | Done. |
| 4 | Minor:A marked up version of the LGR docx file is attached, with some comments and occasional small typo. Please review. | Please review attached DOCX file for minor issues (change-tracked). | Done. |

# Comments on LGR specification (.xml)

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Issue** | **IP Comment** | **PK** |
| 1 | The <description> should be a bit more clear about the two different cases for variants. | See sample XML attached to this feedback, where suggested edits have been carried out. (As well as some small typos fixed).Please review, adopt or further change all suggested changes. | Adopted.  |
| 2 | Comments on named classes Named classes should be commented as to their definition. | Please add comment attributes in the XML (see attached) | Adopted. |
| 3 | Named classes V, M are unused. As they are not used, they can be deleted. (The definition of V and M in the <description> is fine and can remain). | Please remove V and M (see attached) | Adopted. |
| 4 | <references>:Reference 1003 needs to be fixed (see comment for DocX.) also, data for [1003] and [1004] seem to have been conflated.As neither appears used in the XML file, both could simply be deleted.here.As Unicode 11.0.0 will be the most current version when the LGR is published, perhaps reference 1002 could be updated to the equivalent for Unicode 11.0.0 | please fix (NOT yet fixed in XML – this requires GP action) | Updated. |
| 5 | The following shows WLE rule 3 and 4 as currently implemented in the XML (shown is HTML equivalent) |  |  |
|  |  |
|  | There are several issues that are worth correcting:1. naming of special consonants:

Across NeoB LGRs, the convention is to use a letterfor a class, and adjacent letters mean two instances, one for each class, that are adjacent: Therfore XC would mean: “X followed by C”.The preferred naming would be C11. Naming of rules“succeeded by” seems to be confused with “succeeds”. Rule 4 should be “followed-by-C1”.
2. “follows-X” should be renamed: “not-following-X”, as the rule does not describe what comes after X, but what must not come before. (Because of the simultaneous prescription of what must follow a Visarga, the context rule must be of type “when”)
3. The implementation of the compound rule currently implements “OR” not “AND”. It is also not complex enough to warrant splitting into two. It is easier to write:

([:^X:])<- ->([:C1:])in a single line and get the correct “AND” logic. | Please change XC to C1 in the XML and make the changes to rule names and implementation.(see suggested fix). | 1) Adopted.2) Adopted.3) and 4) Updated the rule. |

# Comments on Test Labels

|  |  |  |
| --- | --- | --- |
| **Item** | **Issue** | **IP Comment** |
|  | No new comments |  |
|  |  |  |