On Cross-Script Variants

DATE 2018-08-22

In response to the question:

**What is the best way to progress if GP finds a variant character from the script which is beyond the scope of the GP?**

and in consideration of the revised NeoBrahmi GP’s policy on cross script variants the IP offers these observations:

**Inter-temporal processing of cross-script variants.**

Normally, the IP attempts to process all related scripts together, but in some cases cross-script variants may exist without a deeper relationship between the scripts. In such cases, there may be proposed variants between scripts that are not processed concurrently, so that, for example, the two GP's for the affected scripts are not in session at the same time, or do not produce concurrent drafts. A similar case may arise from deferred scripts, for which a GP may no longer be constituted at the time they are finally integrated.

 Proposed variants to scripts not yet in the current LGR version (and not currently processed) cause an issue in integration because the integrated LGR must have transitive closure, but cannot contain code points that are outside the collective repertoire. If an LGR contains such variants to a “future” script, they would have to be removed from the integrated LGR until such time that the future script is finally added.

 To facilitate this process, GPs are encouraged to list such variants, if possible, separately in their LGR document as tentative variants, but not to list them in the XML specification. For the example relevant to the the NeoB GP, Sinhala is planned to be integrated at the same time, but Myanmar is not., so any variants proposed for the latter would be considered tentative.

The IP plans to take any such listings of tentative variants to scripts not in the current LGR version and note them in the "overview" document for the next published LGR. The IP will then work with the GP for the future script based on that list to make them part of the future proposal. If necessary, the IP will raise the issue during public comments on the future script. Where a future script is in the early stages and may already have a GP seated, the GPs are also encouraged to engage in dialog; any seated GP for a future script is encourage to comment on any tentative cross-script variants in an LGR under public comment.

**In case of unrelated scripts** (e.g. out of region, without or with less direct historical derivation) GPs have been reluctant to identify certain critical cross script variants. In order to assure a secure Root Zone, the IP has identified some critical cases. The plan is to add those as part of the "overview" document and enforce them in integration, after first raising notice in public comment for any affected LGR, and if necessary by rejecting proposals that omit variants that are deemed critical for a secure DNS,

The IP reiterates that security-relevant true homoglyphs are in scope for cross-script variants, independent of whether the scripts are related.

In case where **finalized LGR proposals differ in cross-script variants**, the IP will follow the provisions of the Procedure and create the union and transitive set as part of integration - provided that this does not lead to in-script variants in any of the affected scripts (other than in the special case of overlapping repertoires). In all cases, the IP will try to get GPs to resolve any differences, but where that is not possible, the IP will resolve these as prescribed in the Procedure.

**Requirement to have base character**. Generally, the IP agrees with the planned policy not to consider variant code points that are purely dependent signs and for which there is not a standalone variant code point that could provide a suitable base character. However, enforcing transitivity is something that is defined on the code point level and in order to have transitive closure, some mappings may need to be defined on the code point level even if no label could be built. (This can happen if one script has both independent and dependent code points as variants with two scripts, and with the dependent variant the same code point. If those two scripts do not have any independent code point as variant between them, transitivity still requires that the "orphaned" dependent code point is mapped between the two scripts - counter to the policy requiring a base character).

The IP requests the GP's input on whether it is preferable to keep the policy focused as it is on pairs of scripts, and whether to leave it to the IP to enforce transitivity in such edge cases, or whether the GP feels the need to resolve such issue before finalizing proposals.

**Omitted variants.** If for any reason the IP is concerned that security relevant variants were omitted, the IP would first work to get the future LGR to add the critical variant, and then to apply it to the other scripts in the normal way during integration by making the variant set transitive.

If for some reason the GP for the future script does not include such critical variants in their LGR, the the IP will raise that as a comment during the public comment period. (As much as possible, the IP would raise comments in cases of other discrepancies in cross-script variants).

By raising such comments at the time of the future LGR, even if the original GP my no longer be active, all involved communities have a chance to weigh in.

At the conclusion of the public comment, and based in part on any reaction to the identification as critical variants, the IP would request that the GP add these to the LGR before it can be accepted for integration.

In keeping with the procedure, the IP aims to limit actual implementation of variants by the IP to clear cases of "completing" the set to be transitive (transitive closure), or, in other words, "pure mechanical integration". The goal therefore is to make sure that at least one LGR contains the required variant definition.