

## Neo-Brahmi Generation Panel Meeting Notes from meeting on 30 November 2017

Meeting Attendees (in alphabetical order)

GP members:

1. Ajay Data
2. Akshat S. Joshi
3. Bal Krishna Bal
4. Gangadhar Panday
5. Gurpreet Singh Lehal
6. Hempal Shrestha
7. Jay Paudyal
8. Neha Gupta
9. Prasad PK
10. Rajiv Kumar
11. Shanmugam R
12. Shubham Saran
13. U.B. Pavanaja
14. Udaya Narayana Singh

Staff:

15. Akanksha
16. Pitinan Kooarmornpatana
17. Sarmad Hussain

The following communicated their inability to attend

1. Harish Chowdhary
2. Raiomond Doctor

### Meeting Notes

The GP discussed the following agenda items:

- 1. Status of Devanagari proposal.** Akshat informed the GP that most of the IP feedback was incorporated in the proposal and he can complete and submit to the mailing list on 8 December. **The GP was requested to review the proposal carefully** because it is most likely to be the final version of the Devanagari proposal.

The GP set the submission date to the IP at 16 December, so it provides IP possibility to complete a round of revision before the year-end holidays.

The GP discussed IP Major Feedback:

- (a) Whether the Halanta is needed?

*Background:* The previous proposal defined a label and a label with Halanta ending as variants. The GP then revised and reached the conclusion that such labels are visually distinguishable, therefore the GP agreed not to define them as variant labels.

*IP Feedback:* IP has no objection to the update. However, while permitting Halanta may cause possible confusion, the IP raised that it should be restricted. If there is a good reason to not restrict this instance in spite of possible confusion to the “majority of Devanagari users” the GP should make it explicit.

Conclusion: The GP will elaborate the point that Halanta-ending and non-Halanta labels are explicitly needed by some community to indicate different sounds. And they are visually distinguishable; therefore, such case should be handled at the string similarity step rather than as variant labels. **The GP will document this conclusion, and this can be used as a guideline for other scripts.**

- (b) The Nukta character needs additional references.

Background: Previously IP had feedbacked that the restrictions on Nukta in Hindi may not be valid for Devanagari as a whole and raised the Konkani cases. The GP discussed with Konkani language expert from Konkani Academy and has updated the proposal.

IP Feedback: The IP noted the addition of two code points permitted as left-context of Nukta, However, the Rule “1” in section 7 or the HTML/XML; the purported Konkani usage appears to not be supported – which may be fine, but hard to tell. Section 3.3.6 could point to source material describing the use of the Nukta or it could describe how the GP came to identify the needed code points allowed with Nukta.

Conclusion: **The GP will add more references for the discussion on the use of Nukta in Konkani.**

- 2. Progress on Gurmukhi proposal.** Dr. Lehal informed the GP that he has uploaded the document and has received the XML and HTML files. There are around 20 cross script variants with Devanagari. The GP discussed and clarified the following points.

- (a) The cross-script variant between Devanagari and Gurmukhi will be included in the Devanagari proposal.
- (b) Defining a code point variant as “blocked” does not disallow that code point. The blocking happens on the label level. The variant label is a label that comprise of one or more variant code points. If a user registers a label, only then its cross-script variant label will be blocked.
- (c) The XML file and the document file shared may be different due to multiple reasons. First, there are variants defined between a code point and a sequence of code points. To be able to define such variants, we need to add the code point sequence in the repertoire. Another reason is that if the document defines A as a variant of B, it means that B is also a variant of A, and this symmetric mapping must be captured, and it doubles number of mappings. The transitivity must also be captured. If A and B are variants and B and C are variants, then A and C are also variant code points.
- (d) **In case of any discrepancies in the WLE rules or other parts should be shared and ICANN staff will update accordingly.**

- 3. Progress on Kannada proposal.** Dr. Pavanaja updated that the code point analysis column was filled. The following points were discussed.

- (a) Historically used characters.

Background: There are two code points which are used in the classical text:

0CDE ☞ KANNADA LETTER FA (out of MSR-2)

0CB1 ☞ KANNADA LETTER RRA (in MSR-2)

They should be handled the same way. Currently the 0CDE is out of MSR-2 while 0CB1 is in. Dr. Pavanaja discussed this among the Kannada community and the community felt they both should be in the MSR. From the corpus of size 224,000 words, total words containing 0CB1 ಃ is 1448 and total words containing 0CDE ಃ is 680.

Conclusion: LGR procedure mandates that the repertoire should not include historically used characters. In addition, the Kannada LGR can move forward without delaying the process and then see community feedback. If the community decides that these characters need to be included, with the sufficient supporting information, it is possible to include in the next version of Kannada LGR. **The GP agreed to take 0CB1 KANNADA LETTER RRA out from the repertoire in this version.**

(b) Variant section.

If the variant analysis has been done and there is no variant, it is useful to document it into the proposal, so that the IP and the community understand it has been analysed and result is no variant code points. **The GP agreed to document the analysis and result in the proposal.**

**4. Progress on Gujarati proposal.** Akshat updated that the components of Gujarati are already in place. He is also working on the presentation for the meeting in Colombo. After face to face meeting in Colombo, when the Devanagari proposal is finalized, he can complete the Gujarati proposal within a week. **The GP discussed and agreed on the tentative date for the Gujarati proposal on 27 December 2017.**

**5. Progress on Tamil proposal.** Shanmugam updated that the document has been updated for all parts. The historical section has been circulated to the Tamil community for feedback. The code point repertoire analysis is completed with the EGIDS. Variant and WLE rules sections are also updated. The document is still changing day by day. He is looking forward to working with the new member of NBGP who is a Tamil expert from Sri Lanka. He is also preparing the materials for Sri Lanka meeting and looking forward to discussing about variant code points in the face 2 face meeting.

The GP discussed that it might be useful to have another meeting for Tamil script with Sri Lanka community before travelling to the meeting. **ICANN staff will reach out to both Sri Lanka community and NBGP to set up a meeting.**

**6. Progress on Telugu proposal.** Gangadhar informed the GP that Telugu script is used in multiple languages which may need experts for each language. However, most of the languages are not in the scale EGIDS 0-4. Gondi is the only language with EGIDS scale 4. Gangadhar listed three code points that are used traditionally. They are 0C23, 0C31, and 0C33. **The GP agreed that this can be taken out of the repertoire. If it's needed, Gangadhar could sent the list in the mailing list for discussion.**

**7. Progress on Bangla proposal.** Dr. Udaya informed the GP that he has updated the google doc from section one to section five. The following points were discussed.

(a) The conjunct characters in Bangla might be handled differently in India and in Bangladesh. Some characters have been dropped in Bangladesh. **The GP agreed that Dr. Udaya circulate the detail and continue the discussion via mailing list.**

- (b) It was also discussed that there will be WLE rules that are specific for Bangla and have not been covered by the Devanagari analysis.
- (c) It was reminded that the proposal must cover the need for Bangla users in Bangladesh as well **The GP requested ICANN staff to reach out to some experts in Bangladesh to join the Neo-Brahmi GP. Dr. Udaya will suggest a few contacts.**

**8. Progress on Oriya proposal.** Jay updated that there is one resource for Oriya. The initial training has been done and the related materials has been shared. They are on the initial phase of developing repertoire.

The GP discussed that more resources are needed for Oriya. Dr. Udaya suggested Prof. Prakash from Delhi University. Dr. Ajay requested **Jay to contact Ramamurti for more resources.**

**9. Progress on Malayalam proposal.** Prasad informed the GP that the Malayalam document was updated and now he is working on variants and WLE rules. He noted that there are a lot of connections between Malayalam and Tamil. **Parsad will complete the code point repertoire section on Google doc.**

#### Action Items

S. No.	Action Items	Owner
1	Revise the Devanagari proposal for the Halanta and the Nukta case as well as the rest of the IP feedback, and circulate it to the mailing list on <b>8 December 2017</b>	AJ
2	Review the final version of Devanagari proposal and submit it to the IP by <b>16 December 2017</b>	All
3	Review the Gurmukhi proposal and feedback to the ICANN staff to update the XML	GSL
4	Update the Kannada proposal to exclude 0CB1 from code point repertoire and document the detail of variant analysis.	UBP
5	Finalize the first draft of Gujarati proposal on <b>27 December 2017</b>	AJ
6	Reach out to the Tamil resource in Sri Lanka and set up a call <b>before Colombo meeting</b>	SH
7	Update the Telugu proposal to exclude the historically used characters or list those code points to circulate via mailing list for further discussion if needed.	GP
8	Circulate the detail of conjunct characters in Bangla via mailing list	UNS
9	Suggest contact of Bangla resources from Bangladesh for the ICANN staff to follow up	UNS
10	Contact Prof. Prakash and Ramamurti for more resources for Oriya,	JP
11	Complete the code point repertoire and update the Malayalam proposal as discussed by the GP	PPK