NBGP IP Feedback

Date:	5 July 2018

1. Devanagari

<u>Latest Feedback:</u>

- IP-Response-Devanagari-LGR -20180520-3.docx (31 May 2018)
- Variant-Recommendations-Devanagari-9.docx (2 July 2018)

<u>Summary:</u>

IP responded to the Devanagari LGR proposal dated 2018-05-20, including XML and test file. IP found technical issue regards variant handling. However, the remainder of the Devanagari LGR proposal is nearing finalization.

<u>Open Issues:</u>

# Issue	Options	NBGP Decision
1. Kashmiri vowel 0973 (अ̐) **Variant Set 1** 0902 ◌̇ ←→ 093A ◌̇ 0902 ◌̇ ←→ 0A02 ◌̇ 093A ◌̇ ←→ 0A02 ◌̇ **Variant Set 2** 0905 0902 ←→ 0973 अं अ̐	***Option 1:*** Do nothing The variant mappings as defined happen to give the correct result in index variant generation. ***Option 2:*** Extend the *code point* variant set for the sequence (Set 2) Add all mappings that result from the other partition of 0905 0902: <table><tr><td>Variant Set 2</td><td>0905 0902 ←→ 0973 *0905 093A* ←→ **0973** *0905 0A02* ←→ **0973** **0905 0902** ←→ *0905 093A* **0905 0902** ←→ *0905 0A02* *0905 093A* ←→ *0905 0A02*</td></tr></table> This option seems a lot of effort to define things that can "never happen"; If there is a reliable way to verify that adding these additional mappings cannot change the result of index variant computation, it may be preferable to avoid these mappings. (This needs to be reviewed further).	

# Issue	Options	NBGP Decision
2. Vowel plus Nukta Variants **Variant Set 1** | 0906 ← → 0906 093C 0906 → 0906 093C 0906 093C → 0906 093C 093C … As a result, it is not possible to create variant label sets that are transitive. ***Case of cross-script variant with*** Gurumukhi Nukta (0A3C) is also provided by IP analysis and recommends 'Do nothing'	***Option 1:*** Do nothing The variant label 0906 093C 093C is never enumerated and as it is not a valid label. In principle, doing nothing is a valid option. The downside of it is that it requires this kind of analysis to prove that it works. Other scripts may have formally equivalent mappings that give different results. ***Option 2:*** Add a *variant* context rule 0906 ← → 0906 093C | not-when(followed-by-093C) Because of the added context rule, the partition {0906} {093C} can no longer generate a variant 0906 **093C** 093C. This makes the variant *label* set transitive.	
3. Kashmiri vowel 0974 (आ) **Variant Set 1** | 0902 ◌ ← → 093A ◌ 0902 ◌ ← → 0A02 ◌ 093A ◌ ← → 0A02 **Variant Set 2** | 0906 0902 ← → 0974 आं आं **Variant Set 3** | 0906 ← → 0906 093C : not-when(followed by 093C) **Variant Set 4** | 093C ← → 0A3C ◌	**Option 1:** do nothing This relies on the fact that the numbers just work out. It is however, not satisfying to rely on LGR **Option 5-2:** Extend the *code point* variant set for the sequence (Set 2) **Variant Set 2** | 0906 0902 ← → 0974 *0906 093A ← → 0974* *0906 0A02 ← → 0974* **0906 093C 0902 ← → 0974** *0906 093C 093A ← → 0974* *0906 093C 0A02 ← → 0974* **0906 0902 ← → *0906 093A*** **0906 0902 ← → *0906 0A02*** *0906 093A ← → 0906 0A02* **Option 5-3:** also extend to account for Variant Set 4	

2. Gurmukhi

Latest Feedback:

- Response of IP to-Gurmukhi-LGR-20180525 (31 May 2018)

Summary:

The IP would consider this ready to go to the next stage – once certain fixes have been made. The IP recommends not submitting this to public comment until the Devanagari LGR is ready as well.

Open Issue: none

Current Status: Preparing for public comment, waiting for Devanagari

3. Gujarati

Latest Feedback:

- IP-response-to-Gujarati-LGR-20180605-IP-2.docx (2 July 2018)

Summary:

This version appears very close to readiness for public comment. Two significant changes noted by the Gujarati GP have been commented specifically below.

- Suggest more explanation on Halant
- Suggest change CN to C1

Open Issue: none

Current Status: Incorporating IP feedback

4. Telugu

Latest Feedback:

- Response of IP on Telugu-2018-06-25_IP.docx (2 July 2018)

Summary:

The documents approach readiness for public review, but there are two significant content issues that still need review and updates to the documents as appropriate.

- simplification that appears possible for the context rules
- suggested change in cross-script variants. The latter affects Kannada and possibly Sinhala LGR as well.

Open Issue:

# Issue	Options	NBGP Decision
Should the singleton cross-script code point be defined as variant?	Option 1 – Remove cross-script code point Option 2 – Maintain the cross-script Telugu-Kannada-Sinhala | Telugu	Kannada	Sinhala | |---|---|---| | ಂo (0C02) | ಂo (0C82) | o (0D82) | | ಃ (0C03) | ಃ (0C83) | ඃ (0D83) | | ర (0C30) | ರ (0CB0) | ර (0DBB) | Oriya-Malayalam | Oriya	Malayalam | |---|---| | ଠ (0B20) | ഠ (0D20) |	

Current Status: Incorporating IP feedback. And resolving the open issue.

5. Kannada

Latest Feedback:

- Response of IP on Kannada-v1.8-20180615-2.docx (2 July 2018)

Summary:

The documents approach readiness for public review. However, there remains a question about the cross-script variants with Sinhala.

Open Issue: Please see Telugu

Current Status: Resolving the open issue.

6. Oriya

Latest Feedback:

- IP-reponse-to-Oriya-LGR-2018-05-29-IP.docx (27 June 2018)

Summary: The proposal appears substantially complete. Consider removing the lone variant with Malayalam.

Open Issue: Please see Telugu

Current Status: Resolving the open issue.

7.	Malayalam

Latest Feedback:

- IP-Response-Malayalam-LGR-20180511-IP.docx (25 May 2018)
- Variant-Recommendations-Malayalam.docx (18 June 2018)

Summary: There were still number of feedback as it was the feedback on first submission.

Open Issue:

# Issue	Options	NBGP Decision
1. Variant of 0D33 0D33 <-> 0D33 **0D4D** 0D33 ഉള <--> ഉള Desirable label is ഉളള	**Option 1** – Restricting the variant so it cannot occur following an 0D33 ള or Halant. In particular ഉള്ള (with a visible Halant 0D33 **0D4D** 0D33 **0D4D** 0D33) would not become a variant of ഉളള, etc. **Option 2*** – Restricting valid labels to exclude ളള Restricting labels from containing two 0D33 ള that are not joined by a Halant would robustly prevent any spoofing. However, it would also disallow a small number of potentially meaningful labels. (About 0.0015% of the words in the test file are affected - or 1 in 60K).	
2. The singleton cross-script variant with Oriya	Please see Telugu	

Current Status: Incorporating IP feedback. Resolving the open issues.

8.	Tamil

Latest Feedback:

- Response of IP on Tamil 2nd version 2018-06-07-IP.docx (27 June 2018)
- Variant-Analysis-Tamil-0BCC-4.docx (27 June 2018)
- Email: Tamil - addendum about allocatable variants (2 July 2018)

Summary: The proposal appears substantially complete. Consider removing the lone variant with Malayalam.

- Update the reference and edit the class name
- Add a rule to limit the number of Srii and Shrii in a label

Open Issue: None

Current Status: Incorporating IP feedback.

| **9. Bengali** | 6 |

Latest Feedback:

- I IP-response-Bengali-LGR-20180511-IP.docx (2 July 2018)

Summary: The feedback based on the previous version dated 11 May 2018. Most of the feedback has been incorporate. It is planned to be finalized during Bangla F2F meeting at 10 July 2018.