Response to IP on second round of GJ LGR feedback received on 25th March 2018 (On proposal of 2018-03-03)

DATE: 2018-04-19

# Overview

This document is the Neo-Brahmi GP response to the second version of the feedback received by NBGP on 25th Mar. 2018.

# Conclusion

The NBGP has modified the document as per the suggestions from IP. following table gives NBGP response to specific suggestions.

In addition, as per the internal discussions, two changes which were not suggested by IP.

1. A major change has been done to Gujarati LGR in WLE section. The rule No. 6 in WLE section has been removed as having explicit halant in Gujarati text is very rare and a dying practice. This enables an average Gujarati reader distinguish the same word with and without halant. Hence it is presumed that there would be no confusion to the reader. Hence the said rule is being removed.

2. The WLE rules in Section 7 (Rule 2 and Rule 5) have been modified to make N independent from C.

All the common issues that were flagged by IP have also been taken care of in this revision

The test label file is added.

## Comments on main document (.docx)

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Issue** | **IP Comment** | **NBGP Response** |
|  |  |  |  |
| §3.4.6, pp. 11-12 | There are two text passages in Gujarati. | The points made would be clearer for non-Gujarati readers if the text is accompanied by a romanized transliteration, and an English translation. | Added. |
| Clarification in fn 25 , p. 22 | The possibility of a Visarga following an Anusvara is ruled out, since it is used only in Vedic and in Bengali script. | The possibility of a Visarga following an Anusvara is ruled out, since **this combination** is used only in Vedic and in Bengali script. | Corrected |
| Contributors | Not completed | Please complete | Added. |
| References | The references section is incomplete (by admission in a note)  Both MSR-1 and MSR-2 are mentioned. At this point, these can be replaced by a single reference to MSR-3. This reference should have a URL  The XML uses [101], the DOCX uses [Omniglot]. Normally we would recommend to align these by changing the DOCx file, but with only a single script-specific reference it is perhaps OK to leave as is (that is 101 in XML and [Omniglot] in the document) | Please complete  Please update | MSR reference has been modified. References will be updated till the document gets finalized. Hence marked as “To be updated further” |

## Comments on LGR specification (XML)

|  |  |  |
| --- | --- | --- |
| **Item** | **Issue** | **IP Comment** |
| <description> | <p>Variant Disposition: As variants are of confusingly similar, albeit of a peculiar nature, it is proposed that they be considered of "blocking" nature. There is no preference among these variants. Whichever label…blocked .</p>  There are two issues. (1) It is unclear what “peculiar nature” implies and their natures has already been described in the paragraph prior.  (2) The term used. “blocking” is not defined, so it would be better to use phrasing that allows the term “blocked” to be used.  Finally, it might be useful to gloss the meaning of “blocked”.  This change was already made in the 2017-12-07 LGR for Devanagari, but not propagated to any LGR that used the Devanagari LGR as template.  THIS ISSUE APPLIES ACROSS A NUMBER OF NEO BRAHMI LGRS THAT USE THIS LANGUAGE | <p>Variant Disposition: All variants are of type &ldquo;blocked&rdquo;, making labels that differ only by these variants mutually exclusive: whichever … blocked. There is no preference among these variants.</p> |
| <rules> | The named classes D and X are not used in any rules. | Please remove the definitions; but it is strongly preferred to retain the tag values on the <char> element even if they are not used to define a named class. |

## Comments on Test Labels

Various tools concur that in identifying 179 out of 120K labels as invalid in the corpus file. Almost all of these instances are "out-of-repertoire", leaving 20 instances across four context rules that were not matched.

Some of the out-of-repertoire code points are from the excluded set of Gujarati code points. [which is desirable for a test file].

Test coverage is generally good, but does not include 2 code points. However, all 8 named classes and tag values are covered as are all 4 context rules (each has at least one match and one fail). WLE rules: "leading-combining-mark" does not have a match case (i.e. no invalid label starts with one).

There is no dedicated file of test labels. A test label file would contain a few hundred labels and attempt to cover most code points and combinations of context. This is used in regression testing by the IP, in contrast to a corpus file, which takes a long time to process and is used in understanding whether the WLE rules match the script.

Conclusion: high confidence that context/WLE rules are appropriate (definitely not too restrictive). A test label file is missing.

A quick check of the XML indicates that all our comments appear to have been addressed.

## Comments on Justification of terms Halanta/Virama

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Issue** | **IP Comment** | **NBGP Response** |
| Halanta/Virama | GP statement:  Whether the community calls it Halant or Virama is dependent on it's shape in the particular script.  The word Halant is actually made up of Hal+Ant where Hal means Plough and Ant meant end. So it describes the shape of character like a plough. However in Tamil as an example languages it does not look like a plough so it is called as Virama which describes its functional aspect (Viram i.e. Pause) | The IP respectfully thinks this is wrong. If halanta were a derivative of Sanskrit hala "plough" (Therefore "hala" + "anta"), the word would be "halānta", with a long second a.  As "halanta" with a short second a, it is clearly "hal" + "anta", where the former part is Panini's grammatical word for "consonant" (one of the techncal terms defined in the Shivasūtras), and the second part the Sanskrit for "end". It means therefore "consonant-end", which is precisely the effect of "halanta". It is appropriate as the name of a character which functions as a vowel-killer in any Indic script, and has nothing to do with the shape of the glyph. (virāma does indeed mean literally cessation or pause.)  However, in practice, IP can go along with the GP's approach to Gujarati; and in fact, to retain the word "halanta" in any Indic script, for any letter with this function, as seems to be modern practice in India. It is clear to the users of the script, and non-users can be told that this is equivalent to "DEVANAGARI VIRAMA". (VIRAMA was originally used in Sanskrit at the end of a sentence, and hence never word-internally, but that is a dead letter nowadays.) | NBGP acknowledges IP comments on the same. A footnote has already been added which draws a parallel between the two terms. |
|  |  |  |  |
|  |  |  |  |