

Response of NBGP to IP Response: Devanagari LGR Proposal on 2018-01-02

DATE: 2018-01-30

Overview

This response is based on the IP's feedback (dated 2018-01-16). First three columns in the tables below have been kept as it is and fourth column is added that provides the NBGP response.

General comments

Item	Issue	IP Response	NBGP Response
Valid and Invalid Labels	The GP supplied labels in the requested plain text format and the IP performed verification of label status.	The IP has verified that all valid labels are accepted by the LGR and all invalid labels are rejected. The IP appreciates getting the test files in plain text format, which simplifies automated test protocols.	Noted
Test Coverage	The IP collected data on test coverage. The new files reduce the code point coverage somewhat, but add new contexts related to candrabindu. No test file covers the variant definitions.	A file for verifying variant resolution should be added.	Added.
Variant Labels File	There is no file that tests the defined variants.	The missing file needs to be supplied.	Added as said above.
Cross script variants	The IP understands that an update is pending	Waiting for update	Added
Language coverage	Earlier proposals were accompanied by labels for 11	This reduction does not affect the use of the label files for regression	They were inadvertently excluded. Have

	languages (about 1100 labels) this has been reduced to 5 or six languages and 630 labels	testing, but the omission seems curious, as the data clearly exist and could be seen as useful in demonstrating that the LGR caters to typical akshara patterns in ALL of these languages. As these labels existed in the earlier spreadsheet, there seems no good reason to exclude these from the transcription to plain text files.	been added back.
File date for XML	The internal and external data do not match (the filename contains the latest date while the <date> element contains an earlier date	The internal date should always match the date in the filename to avoid confusion.	The same has been corrected.

Comments on main document (.docx)

Item	Issue	IP Response	NBGP Response
Title (p1)	in the title of the Devanagari LGR proposal there's an "[LGR]", which should be changed to "(LGR)" as it would otherwise be confused with a citation	Done	Noted
§3.3.6 (p. 8, l. 2)	"Use of nukta in other languages should be described here."	The IP notes that the discussion of Nukta has been revised; generally it appears much improved, however, see additional comments below. Nothing more is required.	Noted
§5.2 (p. 17)	Former §6.2 about labels ending in Halant. it doesn't answer the question whether it would be a possible alternative to disallow trailing Halant altogether.	This is now moved to §3.3.2, which answers the question raised.	Noted
Spelling of	The IP notes some inconsistency in	The spelling "Halant" is	Noted

Halant	spelling “Halant”. IP suggests eliminating “halant” (lowercase) and “Halanta	now universal in the document.	
§7 Whole Label Evaluation Rules (WLE)	We note in the latest XML file, that there is a new rule called “follows-only-C” (in contrast to “follows-only-C-or-CN”), which is applied to the sequence U+093E U+093C. This new rule is not described in this section.”	This rule had been present but unused in favor of “follows-only-C-or-CN”. In the IP-supplied XML it has been removed. "follows-only-C" has now been eliminated in favour of "follows-only-C-or-CN"	Noted
§3.3.3 (p 9)	Is there a single or multiple reference documenting the specific set of code points to be used with Nukta? If so, it would be nice to annotate the description.	No reference appears to be given in the .docx file.	The specific sets of permissible Nukta preceding characters has been finalized from various works of C-DAC GIST Research Labs and Omniglot. As the procedure requires Online verifiable sources, the same was not mentioned till now. However, now a footnote to state the same has now been added.

Comments on main document (XML)

Item	Issue	IP Response	NBGP Response
Under <description> Element in XML	The document provides a number of suggestions for supplementing this part of the XML file.	All of these changes (with some trivial modifications, e.g. of pronouns) are incorporated in the most recent XML, with two exceptions:	Noted
Exception 1	[Under Candrabindu] It can follow a vowel, matra, consonant or Nukta.	This is enforced by a tag in the repertoire re U+0901. Therefore the explicit comment in the text is unnecessary.	Noted as no explicit comment is needed.
Exception 2	[Under Nukta] Vowels that are followed by nukta	Variant rules for Nukta (U+093C) which require	Noted as no explicit

	may not be reliably distinct from vowels without Nukta by a large part of the user community. They should therefore be mutually exclusive in the same position in the label (See Variants below).	this are in place: namely, sets 3, 6, 9, 11. Therefore the explicit comment in the text is unnecessary – unless it is intended as the rationale for the restriction implemented via the defined variants.	comment is needed. However we note that this comment is probably not properly understood by NBGP. If there is explicit action is needed, NBGP requests that IP explicitly states that.
§7 . (pp . 26-7)	Previously, the IP had written: “It appears that the restrictions on <i>nukta</i> in Hindi may not be valid for Devanagari as a whole: e.g., in Konkani, as the Unicode standard 9.0, vol. 2, p. 462, states that <i>nukta</i> (U+093C) may be used after U+091A <i>ca</i> ; this is confirmed by https://www.omniglot.com/writing/konkani.htm , which also adds use of <i>nukta</i> after U+091D <i>jha</i>). However, other sources (e.g. http://tdil-dc.in/tdil-dcMain/articles/285368Konkani%20Script%20Grammar.pdf) suggest that <i>nukta</i> is not used in Konkani at all.”	The IP notes the addition of two code points permitted as left-context of nukta, the full set now includes 091A but not 091D. See Rule “1” in section 7 or the HTML/XML; the purported Konkani usage appears not to be supported – which may be fine, but hard to tell based on the evidence presented, which includes only the result, not the data leading up to it. Perhaps section 3.3.6 could point to source material describing the use of Nukta, or (if there are no written sources), could describe how the GP came to identify the needed code points allowed with Nukta).	091D is not a valid Nukta preceding character as it is not required so by any language. Earlier Omniglot had probably flagged it so but the same has been corrected by Omniglot. This was corrected by inputs received by Dr. Shantaram Walawlikar who is a member of NBGP and Ex Head of Konkani Academy. A footnote about references for valid nukta characters has been added as well.
		The XML appears now to be explicit about incidence of Nukta. But in this review, no formal tests have tested this.	

Additional comments referencing earlier communication

Item	Issue	IP Response	NBGP Response
Spelling	The word Karoshti occurs once in .docx	It has still not been re-spelt (correctly) as Kharoshthi.	Corrected.
Spelling	In the documents as they stand, the character	Does this difference serve any	Chandrabind

	U+0901 is named CANDRABINDU, but the tag referring to it is consistently named chandrabinu.	purpose?	u how it is pronounced by some Indian users. However, to avoid confusion, it has been changed.
Label validation	As for: Validation: येये (092F 0946 092F 093C 0946): INVALID - (093C) invalid context (follows-only-specific-C-or-V-or-M)	This evidently fails because, in the LGR as it stands, (Section 7, WLE 1: Rule for Nukta), 092F is not a character that Nukta can follow (even though 092F is tagged as a consonant). Dealt with by removal of the label in question from the new "valid labels" file submitted.	Removed as it was an invalid case.
Label validation	As for: From the "valid" column in the WLE-Examples spreadsheet, this label fails: Label ख्वाब (0959 094D 0935 093E 092C != 0916 093C 094D 0935 093E 092C) does not round trip IDN mapping Validation: ख्वाब (0959 094D 0935 093E 092C): INVALID - (0959) not in repertoire	This evidently fails because, in the LGR as it stands, (Section 7, WLE 1: Rule for Nukta), 0959 is not a character that Nukta can follow (even though 0959 is tagged as a consonant). Dealt with by removal of the label in question from the new "valid labels" file submitted.	Removed as it was an invalid case.