

Sinhala Generation Panel:

Analysis of comments for Sinhala script LGR Proposal for the Root Zone

Revision: June 30, 2019

Sinhala Generation Panel (GP) published the Sinhala script LGR Proposal for the Root Zone for [public comment](#) on 2 October 2018. This document is an additional document of the public comment [report](#), collecting all comments and GP analyses as well as the concluded responses. There are 3 (three) comment analyses as follow:

No.	1	From	Thin Zar Phyo, Myanmar GP Chair																																						
Subject	Feedback for Sinhala																																								
Comment	<p>Dear Sinhala Generation Panel members,</p> <p>Myanmar GP would like to congratulate on the complete work of the Sinhala LGR proposal. We are currently developing the Myanmar Script LGR proposal. During the Sinhala and Myanmar cross-script variant analysis, Myanmar GP defines the following code points as confusable code points (not variant code points) and listed in the appendix of the Myanmar LGR proposal.</p> <p>Confusable code Points:</p> <table border="1"><thead><tr><th>No.</th><th>Glyph</th><th>Code Point</th><th>Myanmar Character Name</th><th>Glyph</th><th>Code Point</th><th>Sinhala Character Name</th></tr></thead><tbody><tr><td>1</td><td>꠵</td><td>U+1025</td><td>MYANMAR LETTER U</td><td>උ</td><td>U+0D8B</td><td>SINHALA LETTER UYANNA</td></tr><tr><td>2</td><td>꠵꠵</td><td>U+102C</td><td>MYANMAR VOWEL SIGN AA</td><td>උඹ</td><td>U+0DCF</td><td>SINHALA VOWEL SIGN AELA-PILLA</td></tr><tr><td>3</td><td>꠵</td><td>U+1031</td><td>MYANMAR VOWEL SIGN E</td><td>උඹ</td><td>U+0DD9</td><td>SINHALA VOWEL SIGN KOMBUVA</td></tr><tr><td>4</td><td>꠵꠵꠵</td><td>U+1031, U102C</td><td>MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E</td><td>උඹඹ</td><td>U+0DDC</td><td>SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA</td></tr></tbody></table>						No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Sinhala Character Name	1	꠵	U+1025	MYANMAR LETTER U	උ	U+0D8B	SINHALA LETTER UYANNA	2	꠵꠵	U+102C	MYANMAR VOWEL SIGN AA	උඹ	U+0DCF	SINHALA VOWEL SIGN AELA-PILLA	3	꠵	U+1031	MYANMAR VOWEL SIGN E	උඹ	U+0DD9	SINHALA VOWEL SIGN KOMBUVA	4	꠵꠵꠵	U+1031, U102C	MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E	උඹඹ	U+0DDC	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA
No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Sinhala Character Name																																			
1	꠵	U+1025	MYANMAR LETTER U	උ	U+0D8B	SINHALA LETTER UYANNA																																			
2	꠵꠵	U+102C	MYANMAR VOWEL SIGN AA	උඹ	U+0DCF	SINHALA VOWEL SIGN AELA-PILLA																																			
3	꠵	U+1031	MYANMAR VOWEL SIGN E	උඹ	U+0DD9	SINHALA VOWEL SIGN KOMBUVA																																			
4	꠵꠵꠵	U+1031, U102C	MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E	උඹඹ	U+0DDC	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA																																			
Analysis	SINHALA LETTER UYANNA and MYANMAR LETTER U is distinguishable Set 2,3,4 similar but they are combining marks, therefore, they are not defined as variants.																																								

Response	No action required in the normative part. The set 2,3,4, will be added to the confusable table.
----------	---

No.	2	From	Prof Udaya Narayana Singh, Neo-Brahmi Generation Panel (NBGP) Co-Chair
-----	---	------	--

Subject	Comments from the NBGP group on Sinhala LGR Proposal
---------	--

Comment	<p>The Devanagari LGR, in the appendix, has listed substantial number of Cross-script variants which are consonants, vowels and vowel signs. Among all of those, Sinhala Visarga character also is included. As it is the dependent vowel sign which in isolation cannot form any valid label, it is mentioned only for the sake of completeness as a confusable.</p> <p>Devanagari-Sinhala</p> <table border="1"> <thead> <tr> <th>Devanagari Code Point</th> <th>Sinhala Code Point</th> <th>In Devangari LGR Proposal</th> <th>In Sinhala LGR proposal</th> </tr> </thead> <tbody> <tr> <td>◌ः (U+0903)</td> <td>◌: (U+0D83)</td> <td>Confusable</td> <td>Confusable</td> </tr> </tbody> </table> <p>Devanagari LGR team would like to place on record it's agreement with the Sinhala GP.</p>	Devanagari Code Point	Sinhala Code Point	In Devangari LGR Proposal	In Sinhala LGR proposal	◌ः (U+0903)	◌: (U+0D83)	Confusable	Confusable
Devanagari Code Point	Sinhala Code Point	In Devangari LGR Proposal	In Sinhala LGR proposal						
◌ः (U+0903)	◌: (U+0D83)	Confusable	Confusable						

Analysis	The GP notes that both LGR proposals agree.
----------	---

Response	No action required.
----------	---------------------

Comment	<p>The Gujarati LGR has not listed any cross-script variants as the team did not find any Consonants/Vowels which can be confused with the Consonants/Vowels of any other scripts. As the dependent signs cannot form any valid label, such signs were not listed as the GJ LGR confusables. The discrepancy between cross-script confusable analysis between Gujarati (0A83) and Sinhala (0D83) Visarga character is seen because of that. As the pair is not found the normative Sinhala LGR, it is assumed that it is listed only for the sake of reference and as a confusable.</p> <p>Gujarati-Sinhala</p> <table border="1"> <thead> <tr> <th>Gujarati Code Point</th> <th>Sinhala Code Point</th> <th>In Gujarati LGR Proposal</th> <th>In Sinhala LGR proposal</th> </tr> </thead> <tbody> <tr> <td>◌ઃ (U+0A83)</td> <td>◌: (U+0D83)</td> <td>This is not mentioned</td> <td>Confusable</td> </tr> </tbody> </table> <p>NBGP would like to place on record that this being part of the confusables only, the discrepancy is a non-issue and hopes that the Sinhala GP is in consonance with the same.</p>	Gujarati Code Point	Sinhala Code Point	In Gujarati LGR Proposal	In Sinhala LGR proposal	◌ઃ (U+0A83)	◌: (U+0D83)	This is not mentioned	Confusable
Gujarati Code Point	Sinhala Code Point	In Gujarati LGR Proposal	In Sinhala LGR proposal						
◌ઃ (U+0A83)	◌: (U+0D83)	This is not mentioned	Confusable						

Analysis	These defined as confusable not in the normative part.																																
Response	No action required.																																
Comment	<p>In addition, we may be able to submit the comments of the Kannada and the Telugu teams to strongly argue for inclusion of the normative variant pairs mentioned in their respective LGRs to be included in the Sinhala normative variant pairs. As the Kannada and Telugu mentions those code points as normative variants and the Sinhala LGR does not, this is a breaking discrepancy and needs to be addressed.</p> <p>Kannada-Sinhala</p> <table border="1"> <thead> <tr> <th>Kannada Code Point</th> <th>Sinhala Code Point</th> <th>In Kannada LGR Proposal</th> <th>In Sinhala LGR proposal</th> </tr> </thead> <tbody> <tr> <td>◌◌ (U+0C82)</td> <td>◌◌ (U+0D82)</td> <td>Variant</td> <td>Confusable</td> </tr> <tr> <td>◌ಃ (U+0C83)</td> <td>◌ಃ (U+0D83)</td> <td>Variant</td> <td>Confusable</td> </tr> <tr> <td>◌ಠ (U+0CB0)</td> <td>◌ಠ (U+0DBB)</td> <td>Variant</td> <td>Not mentioned</td> </tr> </tbody> </table> <p>Telugu-Sinhala</p> <table border="1"> <thead> <tr> <th>Telugu Code Point</th> <th>Sinhala Code Point</th> <th>In Telugu LGR Proposal</th> <th>In Sinhala LGR proposal</th> </tr> </thead> <tbody> <tr> <td>◌◌ (0C02)</td> <td>◌◌ (U+0D82)</td> <td>Variant</td> <td>Confusable</td> </tr> <tr> <td>◌ಃ (0C03)</td> <td>◌ಃ (U+0D83)</td> <td>Variant</td> <td>Confusable</td> </tr> <tr> <td>◌ಠ (0C30)</td> <td>◌ಠ (U+0DBB)</td> <td>Variant</td> <td>Not mentioned</td> </tr> </tbody> </table>	Kannada Code Point	Sinhala Code Point	In Kannada LGR Proposal	In Sinhala LGR proposal	◌◌ (U+0C82)	◌◌ (U+0D82)	Variant	Confusable	◌ಃ (U+0C83)	◌ಃ (U+0D83)	Variant	Confusable	◌ಠ (U+0CB0)	◌ಠ (U+0DBB)	Variant	Not mentioned	Telugu Code Point	Sinhala Code Point	In Telugu LGR Proposal	In Sinhala LGR proposal	◌◌ (0C02)	◌◌ (U+0D82)	Variant	Confusable	◌ಃ (0C03)	◌ಃ (U+0D83)	Variant	Confusable	◌ಠ (0C30)	◌ಠ (U+0DBB)	Variant	Not mentioned
Kannada Code Point	Sinhala Code Point	In Kannada LGR Proposal	In Sinhala LGR proposal																														
◌◌ (U+0C82)	◌◌ (U+0D82)	Variant	Confusable																														
◌ಃ (U+0C83)	◌ಃ (U+0D83)	Variant	Confusable																														
◌ಠ (U+0CB0)	◌ಠ (U+0DBB)	Variant	Not mentioned																														
Telugu Code Point	Sinhala Code Point	In Telugu LGR Proposal	In Sinhala LGR proposal																														
◌◌ (0C02)	◌◌ (U+0D82)	Variant	Confusable																														
◌ಃ (0C03)	◌ಃ (U+0D83)	Variant	Confusable																														
◌ಠ (0C30)	◌ಠ (U+0DBB)	Variant	Not mentioned																														
Analysis	NBGP and Sinhala GP met and agreed that there are no variant relationships between Kannada - Telugu - Sinhala scripts.																																
Response	No action required for Sinhala GP.																																

No.	3	From	Liang Hai
Subject	A quick review of the Sinhala proposal		
Comment	§3.2: Unclear what exact scope of writing systems have been decided. For example, if Sanskrit is included in the scope then the full set of four vocalic liquids (vocalic r/rr/l/ll) probably need to be included in §5.2. However the text only vaguely says “considered”.		

Analysis	We have considered that Sinhala used to be used to write Sanskrit but not common, therefore, it is not in the scope of Sinhala LGR. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.2, “In addition, Myanmar script is also related.”: This statement seems to come from the comparison with Myanmar in §6.3.4, Sinhala and Myanmar. However those pairs are similar only by accident, and can’t suggest the two script are related in the same sense of how Sinhala is related to the scripts covered by the NBGP.
Analysis	Sinhala Myanmar and Neo-Brahmi script derived from the same root and because of the rounded shape of Sinhala and Myanmar and Buddhism. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.1, “In addition, conjunct characters and touching letters are features of Sinhala text, but do not require representation in the root-zone for labels.”: Either provide the reasoning in place or simply don’t talk about the usage in labels at this stage.
Analysis	There is a technical issue which GP would like to point out in this context. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.1, “... ඥ (jna) the symbol is considered as representing ජ+ඥ (j+na), identical to the consonant in contemporary Sinhala ඥ which has a code point U+0DA5.”: The existence of U+0DA5 SINHALA LETTER TAALUJA SANYOOGA NAAKSIKYAYA is simply because the j.nya conjunct was analyzed as a structure eligible to be encoded atomically. This character is meant to represent the conjunct j.nya and the conjunct also must be represented/encoded with this character. The phrasing “... identical to ...” here is misleading.
Analysis	The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.1, “When modifiers are added to any of the above categories, including ... they will be formed as follows ...”: Unclear what is suggested here. Are the following examples meant to illustrate how conjunct

	consonants and touching consonants behave graphically just like plain/individual consonant letters?
Analysis	The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.1, “Special symbols ෝඃ (rakaranshaya) for ඃ (ra) and ෝය (yanshaya) for ය (ya) ...”: Inline images are needed for representing rakaranshaya and yanshaya if they can’t form properly with text.
Analysis	The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.1, “... used in Sinhala writing when they occur after a consonant (from which the inherent vowel has been removed).”: Since the authors appear to analyze this process as the symbols (rakaranshaya and yanshaya) have their own inherent vowels, the inherent vowel of the base consonant is removed by the very action of attaching a symbol, therefore one can’t exactly say the inherent vowel “has been removed” when a symbol is used.
Analysis	This comment can be incorporated.
Response	Update the proposal as suggested.
Comment	§3.3.2, Table 2: Pronunciations of the vocalic l and vocalic ll are suspicious. I understand the vocalic l and vocalic ll and not really used in Sinhala, but it seems the two letters’ names instead of their actual pronunciations are listed in the table. Note the vocalic r and vocalic rr also appear to have letter names different from their pronunciations, but their pronunciations are correctly listed in this table.
Analysis	The term elu since it refers to pure Sinhala. What I put in the table are /ilu/ /ilu:/. These two characters (among few others) are borrowings from Sanskrit (Devanagari) which are not used in Sinhala language but preserved in the alphabet.
Response	No action required.
Comment	§3.3.3, “This is thus used to join consonants and form conjunct characters.”: U+0DCA doesn’t form conjuncts by itself. The requirement of ZWJ when forming a conjunct can’t be ignored in such statements.

Analysis	0DCA conjuncts while ZWJ is for the rendering. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.3: Actually, as the requirement of ZWJ is not mentioned in the preceding sections (where it's good to discuss the script's behavior independently from its encoding), it should be emphasized here that, ZWJ is required for forming not only typical conjuncts, but also the "special symbols" (rakaranshaya, yanshaya, and rephaya) and touching consonants. And it's not emphasized enough anywhere in the proposal (even in §5.5) that excluding ZWJ is a major problem for Sinhala labels because all the aforementioned consonantal structures rely on it.
Analysis	ZWJ is for the rendering operation. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§3.3.4, "... represents all the nasals": Probably, "represents a general nasal sound" or "represents a context-dependent nasal sound"?
Analysis	This comment can be incorporated.
Response	Update the proposal as suggested.
Comment	§3.3.6, "One constraint for Sannjakas is that they cannot be followed by halanta.": Is this a phonetic (so pre-nasalized stops cannot directly precede another consonant even ya, ra, or va) or a graphic statement (so pre-nasalized stops are not written with an attached vowel killer)?
Analysis	This is a phonetic function. The GP does not see the need to update the proposal.
Response	No action required.
Comment	<p>§5.2, Code Point Repertoire:</p> <p>* For writing the [f] sound, this lately invented structure represented by U+0DC6 ෆ SINHALA LETTER FAYANNA is often considered less used compared to the more popular form "ෆ combined with f". The usage of these f-sound graphemes should be discussed.</p> <p>* About U+0DF2 ෆ SINHALA VOWEL SIGN DIGA GAETTA-PILLA, see the comment below for 5.4.</p>

Analysis	Now the U+0DC6 ෂ is commonly used. “ෂ combined with f” used to be used in the past, but not now. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§5.4, Code point not included: * Unclear why U+0DF2 ෂා SINHALA VOWEL SIGN DIGA GAETTA-PILLA is included when its independent form U+0D8E ජා SINHALA LETTER IRUUYANNA is excluded is excluded.
Analysis	This used to be used for writing Sanskrit which is not in this LGR. The GP does not see the need to update the proposal.
Response	No action required.
Comment	* It’s inappropriate to simply say “Usage unknown” for U+0D8E, U+0D8F, U+0D90, U+0DDF, and U+0DF3, as they’re apparently used in the standard Sanskrit alphabet as least. So they have known usage in Sanskrit and are probably not used for the Sinhala language.
Analysis	This used to be used for writing Sanskrit which is not in this LGR. The GP does not see the need to update the proposal.
Response	No action required.
Comment	* The exclusion of U+0D9E ක SINHALA LETTER KANTAJA NAASIKYAYA and U+0DA6 ජ SINHALA LETTER SANYAKA JAYANNA is concerning. It seems a stronger case is needed for excluding letters that are considered a part of the standard Sinhala alphabet and already have attestations (the word that uses U+0DA6). However I understand the standard alphabet itself is not as fixed as other Indic languages’.
Analysis	This is not used in modern writing. The GP does not see the need to update the proposal.
Response	No action required.
Comment	§5.5, “One of the most important deficiencies of not being able to have Top Level Domain with Rakar form is that one cannot have “ශ්‍රී” (Shri) in a top level domain name ...”: The systematical necessity of ZWJ in the Sinhala encoding is not emphasized enough. Calling out ශ්‍රී here almost feels like a “fun fact”, while in fact the exclusion of ZWJ affects a great number of common words and those words just cannot be encoded correctly without ZWJ. The exact effect of excluding ZWJ (although not a decision made by

	the Sinhala panel) should be thoroughly analyzed See also the comment above for §3.3.3.
Analysis	The GP does not see the need to update the proposal.
Response	No action required.
Comment	<p>§5.6, Akshar Formation Rules for Sinhala: See the comments below for §7.</p> <p>§6.1, In-Script Variants:</p> <p>* This list is nice (it can be ordered better though, according to either the shapes or code points). Proposals by the NBGP probably should undergo a similar set of criteria for identifying in-script variants. I do feel the criteria are strict (as these pairs are probably not that confusable) though.</p> <p>* “j. ෙ (U+0D95) and ෙ (U+0DB9 U+0DCA)” is already disallowed by the akshar formation rule of that prenasalized stops cannot be followed by a vowel killer.</p> <p>§7, “This section provides the WLE rules that are required by all the languages mentioned in section 3.2 when written in Sinhala Script.”: The authors need to clearly define a scope of languages. “All the languages mentioned” is vague.</p>
Analysis	The GP does not see the need to update the proposal.
Response	No action required.
Comment	§7, “... for each of the "Indic Syllabic Category" as mentioned ...”: The term “Indic Syllabic Category” can cause confusion with the Unicode character property of the same name. Should note this is not the Unicode property mentioned here.
Analysis	This comment can be incorporated.
Response	Update the "Indic Syllabic Category" to “Category”
Comment	§7, Whole Label Evaluation (WLE) Rules: §5.6 basically suggests such a pattern: `V[B X] C[M][B X] CH J[M][B]`. It’s questionable whether it’s necessary to split J from C when the argument for disabling H and X after J is weak. It’s unclear whether it’s necessary to introduce such a restriction based on attestation instead of actual problems. Also the attestation of visarga following a prenasalized stop already exists, according to §5.6.5, then why is it disallowed? Atypical spellings (such as the ones of colloquial words and loan words) should not be considered the second-

	class use cases when underlying technical rules (instead of language policies) are being drafted.
Analysis	J and C are separated because earlier J used to be a half-nasal combining marks. The GP does not see the need to update the proposal.
Response	No action required.