

To: UTC
 Title: Lepcha Encoding model feedback
 From: Lorna A. Priest (SIL International)
 Date: 7/27/11

Introduction

I submitted the following feedback as an online error report:

Subject: Chapter 10: Lepcha ra and ya

There seems to be an error in two places in the Lepcha section of Chapter 10.

The -ra and -ya codepoints are getting mixed up.

Under Medials they are correct (page 325)

Under Retroflex Consonants the LEPCHA SUBJOINED LETTER RA is incorrectly given the codepoint of U+1C24 in two places. It should be U+1C25.

On page 326, Table 10-2 it lists medial -ra as U+1C24 and medial -ya as U+1C25.

The codepoints (and glyphs) are wrong. Medial -ya is U+1C24 and medial -ra is U+1C25. I think the Class order in the chart is correct according to the original Unicode proposal, it is the Example and Encoding that are incorrect.

Then, a question. In the Unicode proposal L2/05-158R it gives the correct order as C(.) (R)(Y)(V)(^)(F) and Table 10-2 would indicate the order to be: C(.) (R)(Y)(V)(F)(^)

I'm wondering if this was deliberately changed or if this is also an error.

I was requested to submit detailed information about the problem. The comments above are duplicated below in bold with further commentary below each comment.

Discussion

LPriest: Under Retroflex Consonants the LEPCHA SUBJOINED LETTER RA is incorrectly given the codepoint of U+1C24 in two places. It should be U+1C25.

Chapter 10, page 325 of the Unicode book says:

In other words, the *retroflex t* would be represented as <U+1C00 LEPCHA LETTER KA, U+1C24 LEPCHA SUBJOINED LETTER RA>. To distinguish such a sequence representing a *retroflex t* from a sequence representing the actual syllable [kra], it is common to use the *nukta* diacritic sign, U+1C37 LEPCHA SIGN NUKTA. In that case, the *retroflex t* would be visually distinct, and would be represented by the sequence <U+1C00 LEPCHA LETTER KA, U+1C37 LEPCHA SIGN NUKTA, U+1C24 LEPCHA SUBJOINED LETTER RA>.

LEPCHA SUBJOINED LETTER RA is clearly U+1C25. There should not be any question that the above text is an error.

LPriest: On page 326, Table 10-2 it lists medial -ra as U+1C24 and medial -ya as U+1C25.

The codepoints (and glyphs) are wrong. Medial -ya is U+1C24 and medial -ra is U+1C25. I think the Class order in the chart is correct according to the original Unicode proposal, it is the Example and Encoding that are incorrect.

At issue here is what is circled in red in the chart below. The -ra and -ya codepoints and glyphs are mixed up. I suppose you could say that it is the “Class” that is mixed up, but I don’t think there should be any question about the)ra (U+1C25) and)ya (U+1C24) order. Where there is a -ra and a -ya in the same conjunct, the -ra should come first. Plaisier lists conjuncts using both -ra and -ya as:



Jason Glavy’s custom Lepcha font also included these conjuncts:



The transliteration for all of these includes “-rya” at the end. There are no conjuncts of “-yra”

I believe that some of the confusion is because -ya is encoded before -ra. I do not know the rationale for that decision, except that *kya* is sorted before *kra* and it could also be based on frequency. There are approximately 24 -ya conjuncts and perhaps only 8 -ra conjuncts.

Table 10-2. Lepcha Syllabic Structure

Class	Example	Encoding
consonant, letter a	ཅ	[U+1C00..U+1C23, U+1C4D..U+1C4F]
nukta	◌̣	U+1C37
medial -ra	◌ྲ	U+1C24
medial -ya	◌ླ	U+1C25
dependent vowel	◌̣̣	[U+1C26..U+1C2C]
final consonant sign	◌̣̣̣	[U+1C2D..U+1C35]
syllabic modifier	◌̣̣̣̣	U+1C36

LPriest: Then, a question. In the Unicode proposal L2/05-158R it gives the correct order as C(.) (R)(Y)(V)(^)(F) and Table 10-2 would indicate the order to be: C(.) (R)(Y)(V)(F)(^)

I'm wondering if this was deliberately changed or if this is also an error.

This is circled in blue in the table above. The problem addressed in this comment is that the *final consonant sign* (F) is listed as before the *syllabic modifier* or *ran* (^). This is opposite to what was outlined in Everson’s L2/05-158R proposal.

If a final consonant sign is used, the *ran* (or *syllabic modifier*) always positions *above* the final consonant sign. That may be why the editorial committee indicates the *ran* coming last in a sequence. I do not totally understand the rationale for the backing store listed in Everson’s proposal: C(.) (R)(Y)(V)(^)(F), and in fact, Plaisier is somewhat unclear on the function of the *ran*. She says “The *ran* ‘circumflex’ sign is a diacritic flourish written over a consonant sign or over a vowel sign... Although the original function of the *ran* sign is still unclear, it is often present in closed syllables, in which case the circumflex sign should be written above the final consonant sign.” (page 29). If that is true, then it clearly must come *after* the final consonant sign if it indicates a closed syllable. It would not make sense to come before the final consonant sign if it indicates a closed syllable. This would follow the backing store as listed in Table 10-2.

On the other hand, the minutes of the meeting where the Lepcha proposal (L2/05-158) was accepted do not indicate any problems with the encoding model outlined in that proposal:

Scripts - Lepcha (C.1)

[[104-C24](#)] **Consensus:** Accept 73 Lepcha characters at 1C00 - 1C4F, with the new block name "Lepcha" for encoding in a future version of the standard (except for the character U+1C35 LEPCHA CONSONANT SIGN KANG.) [[L2/05-158](#)]

[[104-A77](#)] **Action Item** for Ken Whistler: Update the pipeline to include the acceptance of 73 Lepcha characters, at 1C00 - 1C4F, minus the character U+1C35 LEPCHA CONSONANT SIGN KANG. [[L2/05-158](#)]

[[104-A78](#)] **Action Item** for Michael Everson: Clarify U+1C35 LEPCHA CONSONANT SIGN KANG and submit a document discussing it.

As far as I'm aware the only other Unicode Lepcha implementation was done by XenoType for Mac OS X 10.2. That implementation was done for the original Unicode Lepcha proposal and thus it is based on the Lepcha proposal order. If UTC chooses to keep the order listed in the chart then the XenoType implementation should be encouraged to be changed. If the UTC chooses to change the order listed in the chart, I do not believe it will have negative impact on any existing Lepcha implementation.

This issue must be resolved in order to implement a truly Unicode-compliant Lepcha font.

References:

Plaisier, Heleen. 2006. A Grammar of Lepcha. Universiteit Leiden: doctoral dissertation. <https://openaccess.leidenuniv.nl/handle/1887/4379> (accessed 27 July 2011).

The Unicode Consortium. The Unicode Standard, Version 6.0.0, (Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6). Chapter 10. South Asian Scripts-II. <http://www.unicode.org/versions/Unicode6.0.0/ch10.pdf> (accessed 27 July 2011).

Everson, Michael. 2005. Proposal for encoding the Lepcha Script (WG2 N2947). L2/05-158 <http://www.unicode.org/L2/L2005/05158-n2947-lepcha.pdf> (accessed 27 July 2011).

Everson, Michael. 2005. Analysis of the Indian proposal to encode Lepcha script in the UCS. L2/05-061R <http://www.unicode.org/L2/L2005/05061-lepcha.pdf> (accessed 27 July 2011).

UTC. 2005. Approved Minutes of the UTC 104 / L2 201 Joint Meeting. <http://www.unicode.org/L2/L2005/05180.htm> (accessed 27 July 2011).