# Issues with Sinhala symbols Rakaaraansaya and Yansaya.

The ICT Agency of Sri Lanka wishes to bring a major problem with the current rendering of two Sinhala symbols to the attention of the UTC.

i. *Introduction:*

In Sinhala writing, there are two common modifier symbols named Rakaransaya and Yansaya, which represent a "ra" or "ya" respectively following a pure (i.e. without a vowel) consontant.

These are currently represented by:

SINHALA MODIFIER YANSAYA; 0DCA 200D 0DBA  (virama + zwj + ya)
SINHALA MODIFIER RAKAARAANSAYA; 0DCA 200D 0DBB (virama + zwj + ra)
and must follow a consonant.
Use of these symbols is mandatory in Sinhala orthography, and a "ya" or a "ra" following a pure consonant **must** be represented by the appropriate symbol (except in compound words).

ii. *Issues with these symbols*

Although a large amount of Unicode encoded Sinhala text is now available, one major issue is the breaking up of these two symbols by websites and applications either

1. deleting the zero-width joiner or

2. rendering the zero-width joiner in another font.

In either case, the result is that the explicit virama is shown instead of the conjunct form, making the text incorrect – as shown below – and resulting in major criticisms of Unicode.

| Correct representation | Representation with zwj dropped | with zwj seen |
|---|---|---|
| ශ්‍රී ලංකා | ශ්රී ලංකා | ශ්‍රී ලංකා |
| විද්‍යා | විද්යා | විද්‍යා |

Figure 1: renderings of rakaaraansaya and yansaya

One major site which drops the zwj is Google, though other sites also do this. Some applications (e.g. Adobe InDesign) too incorrectly display text containing these symbols.
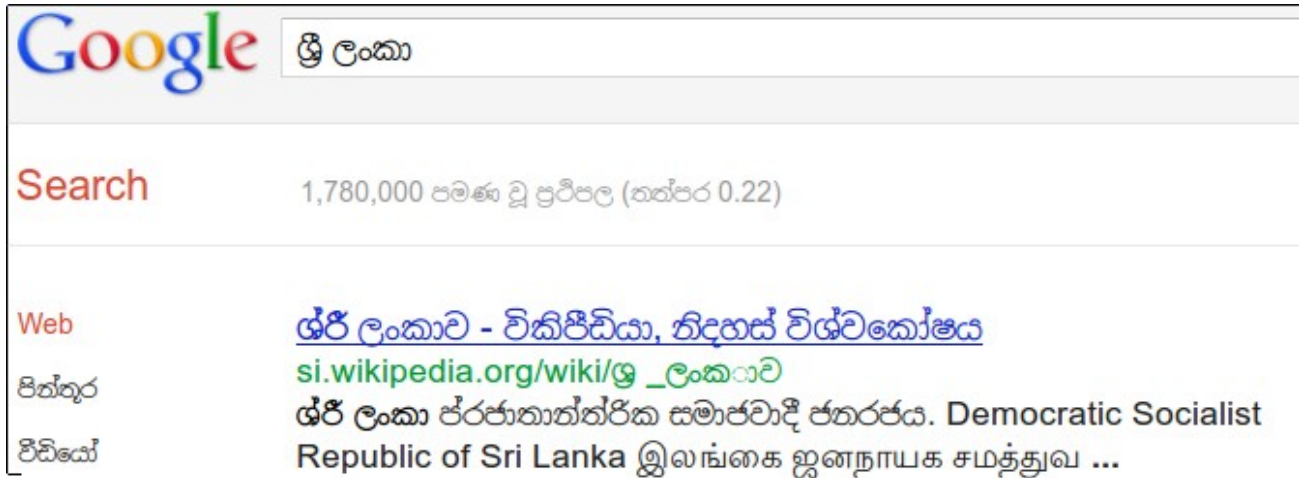


Figure 2: "Sri" breaking up in Google search engine

Note that in the above screenshot, the search query is represented correctly. Only the output is mangled.
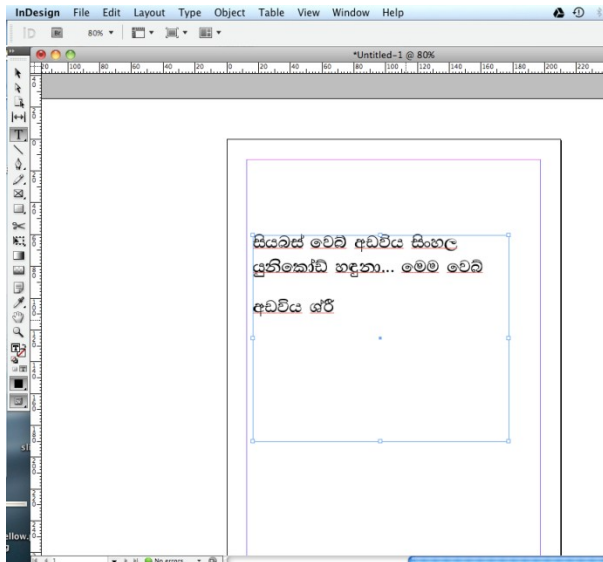


Figure 3: "Sri" breaking up on InDesign.

    iii. *Possible solution for breaking up of rakaaraansaya and yansaya*

When implementing the unicode encoding of Sinhala in 2002, we were informed that rendering engines will render these symbols correctly "after the next OS upgrade". Ten years later, it has

now become evident that the problem is still present. In today's context, breaking up of Medial Signs has become unacceptable. The signs work well in certain applications, such as word processors, but when saving a file, the operating system may render the name of the file in a broken form.

Since breaking up of these forms has become offensive and unacceptable, Sri Lanka plans to submit a proposal on *encoding of two the two modifiers Rakaransaya and Yansaya - as has been done in other scripts such as Malayalam and Burmese.*

iv. *Recommendation:*

Use of the ZWJ had become major issue in the formation of two conjunct forms in Sinhala since it may not be possible to use these with IDNs etc. Sri Lanka will submit a proposal for encoding of the rakaaraansaya and yansaya as separate code points, and a special joiner in the Sinhala area where the joiner will act either with Consonant+0DCA+Virama Joiner+ Consonant or Virama Joiner+ 0DCA+Consonant for touching letters in place for ZWJ.

The Local Language Working Group (LLWG) of the Information and Communication Technology Agency of Sri Lanka (ICTA) discussed the issue at its meeting held on 2012-10-31, and recommended that Sri Lanka table a comprehensive document highlighting the issues discussed briefly above, and the requirement for encoding the rakaaransaya and the yansaya in the Sinhala page of the Basic Multilingual Plane.

For the all the above recommendations, Sri Lanka requires backward compatibility in the font files.