

Sinhala Named Sequences

The Sinhala script contains three symbols which represent an R before or after a consonant, or a Y after a consonant.

These symbols have well known names, and considered to be modifiers similar to vowel modifiers. Their non-inclusion as Unicode characters has led to frequent criticism that Unicode does not fully represent Sinhala.

These modifiers are defined in Sections 5.7 and 5.8 of Sri Lanka Standard 1134:2004 Sinhala Character Code for Information Interchange (Second Revision), as follows:

5.7 Rakaaraansaya and Yansaya

The rakaaraansaya and yansaya are forms of **conjunct letters**.

The rakaaraansaya ේ represents a ජ which follows a pure consonant. It can, in turn, be followed by a vowel sign. It is joined to the preceding letter by a zero-width joiner (zwj).

A rakaaraansaya is represented by the character sequence *cons* 0DCA 200D 0DBB (*cons* + ^p + zwj + ජ) where *cons* represents some consonant.

e.g.: කූ = 0D9A 0DCA 200D 0DBB (ක + ^p + zwj + ජ),

කූූ = 0D9A 0DCA 200D 0DBB 0DD9 (ක + ^p + zwj + ජ + ේ)

Similarly the yansaya ය represents a ය which follows a pure consonant.

e.g. කය = 0D9A 0DCA 200D 0DBA (ක + ^p + zwj + ය)

කය් = 0D9A 0DCA 200D 0DBA 0DDD (ක + ^p + zwj + ය + ේ)

NOTES:

1. As the ේ and ය are present on the keyboard, users will not need to key in the above sequences (see 6).
2. The yansaya and rakaaraansaya are required in normal Sinhala text. However, if for some reason, it is desired not to use the rakaaraansaya or yansaya, the zwj should be omitted.

5.8 Repaya

The repaya ේ represents the letter ජ preceding a consonant. It is represented by the sequence 0DBB 0DCA 200D *cons* (ජ + ^p + zwj + *cons*).

e.g. කූූ = 0D9A 0DBB 0DCA 200D 0DB8 (ක + ජ + ^p + zwj + ේ)

A ya followed by a yansaya with a repaya (in words such as කායීයාලය) is represented by the code sequence 0DBB 0DCA 200D 0DBA 0DCA 200D 0DBA (ජ + ^p + zwj + ය + ^p + zwj + ය).

Accordingly, we propose that these modifiers be identified by the following Unicode named character sequences, to confirm that they are supported by Unicode.

SINHALA MODIFIER YANSAYA; 0DCA 200D 0DBA
 SINHALA MODIFIER RAKAARAANSAYA; 0DCA 200D 0DBB
 SINHALA MODIFIER REPAYA; 0DBB 0DCA 200D