

Sinhala Text Representation

Decisions of the SLSI Working group on 2004-06-09

1. Use space + pilla to show pilla in isolation. e.g., aelapilla (ා) = 0020 0DCF (spc + ා), kombuva = 0020 0DD9. The pilla should not have any additional space to the left (or right).

[MS uses a zwj between the space and the pilla. We decided not to use a zwj.]

Use no-break space to depict alternate forms of pili – e.g. 00A0 0DD4 (nbsp + papilla) to show the “koku” (hook) form of the papilla. This is an overloading of the NBSP character.

2. Stand-alone yansaya = 0020 0DCA 200D 0DBA (spc + al + zwj + ya),
stand-alone rakaransaya = 0020 0DCA 200D 0DBB (spc + al + zwj + ra),
stand-alone repaya 0DBB 0DCA 200D 0020 (ජ + ් + zwj + spc).
3. For the sequence ෝ (special case): ra + al + zwj + ya + al + zwj + ya (a repaya and yansaya applied to a ya)
4. “Touching” characters

Consider touching letter, e.g., in තසස, to be a distinct form, different from conjunct letters, e.g., ස, and use a different construct to represent them. (Touching letters, while uncommon in contemporary Sinhala, are common in classical texts, and in Bhuddist texts.) This will also give three encodings for variant forms such as ද්ඛ, ඛ and ඛ (normal, “touching” and conjunct forms).

e.g., use ta + sa + al-lakuna + sa to represent තසස and ta + sa + al-lakuna + x + sa for තසස

Question: What character or characters should be used for x? Possibilities: zwnj; zwj+zwnj+zwj. We would like your advice for the selection of x.

5. Letters such as ද් and ද, may be written with the yansaya, aelapilla, etc. subscripted. It was decided not to have a separate encoding for the subscripted form, and allow the depiction to depend on the font (and possibly the point size).