

Comments on L2/04-235 [Sinhala]

Paul Nelson
Microsoft Corporation
10 June, 2004

In document L2/04-235, section 3, there are some statements based on what appear to be invalid assumptions. It is my belief that the Sri Lankan government has correctly defined the use of the ZWJ in their text formation for Sinhala script.

From my understanding, the default form for Sinhala, which is the overt halant form, should happen without other control characters. Sinhala does not follow the rules of the other Indic scripts and should be allowed to use ZWJ/ZWNJ in a way that provides the correct behavior, as documented by the government of Sri Lanka.

1. In J2, the document incorrectly assumes that RA HALANT should form a Reph. The RA with overt halant is the default form. The user should not have to type a ZWNJ every time they type a cluster that begins with a RA HALANT, which is the normal use. The rare case of the rapaya is correct to form using RA HALANT ZWJ.
2. Conjunct forms do not occur by default in most situations. The use of the ZWJ to force the conjunct is appropriate for Sinhala as it is within the general scope of existing uses for ZWJ and satisfies the functional requirements for this script. To force a model suited for the functional requirements of other scripts that do not have the same behavior would be a significant disservice to users.
3. Suggestion of J4 is not good. Using a higher level markup, e.g. writing system in OpenType font, would be a better way to handle this. I find that the recommendation is fine for Indic scripts other than Sinhala. However, from what I have seen in implementing Sinhala support is that the document's recommendation would introduce ambiguous situations that would confuse users.
- 4.

Having implemented Sinhala with considerable interaction with users and government of Sri Lanka, it is clear to me that the correct approach to implementation is to be in accordance with the Sri Lanka government standard. Therefore, I would recommend that the UTC accept the standard that the Sri Lanka government has written as the way Unicode should be used with the Sinhala script. The default character/conjunct shaping is not the same as most of the other Indic scripts. Trying to force Sinhala to operate in the same way as other Indic scripts is not helpful.