

Unicode for KANNADA

Subject: Use of Unicode Control Characters ZWJ and ZWNJ in Indic scripts

Ref:

- Letter No.13(3)-2003/TDIL, dated 20.7.2004, written by Sri Manoj Jain, Scientist, Department of Information Technology, Government of India, New Delhi
- Public Review Issue Proposal #37 `Proposal on Clarification and Consolidation of the Function of ZERO WIDTH JOINER in Indian Scripts` written by Peter Constable, Microsoft

The important concern in Kannada is about displaying a consonant cluster in two different forms whenever RA (ರ, 0CB0) appears as the first consonant. Having gone through the above proposal (which deals with this issue) in detail, I feel that that the proposal is well thought out, nicely formulated.

However, I would like to add the following comments and suggestions
The proposal in the document is as follows:

| <u>Display form</u> | <u>Unicode sequence</u> |
|---------------------|-------------------------|
| ರಫ | <RA, virama, C> |
| ರೃ | <RA, zwj, virama, C> |

It is to be noted that ರೃ is obtained by following the general rule for the formation of consonant conjuncts with respect to any combination of more than two consonants. The first option where the reph sign (ಫ) is used is the special case whenever RA appears as the first consonant.

Hence my suggestion is that it would be more appropriate to assign the Unicode sequence as follows:

| <u>Display form</u> | <u>Unicode sequence</u> |
|---------------------|-------------------------|
| ರಫ | <RA, zwj, virama, C> |
| ರೃ | <RA, virama, C> |

The effect of zwj would then be to display the consonant conjunct not in the normal form but in the alternate form.

Exception:

The rendering engine (from Unicode to display) should take care of the following exception.
The special case of using reph sign (ಫ) is not applicable for the conjunct formation of RA with RA itself.
i.e. the rendering of (RA, virama, RA) as ರಫ is not acceptable.

The general rule of consonant conjunct formation has to be applied for this special case and (RA, virama, RA) has to be rendered as ರೃ

More Examples:

More examples are given here for the 'Proposal on Clarification and Consolidation of the Function of ZERO WIDTH JOINER in Indian Scripts' written by Peter Constable, Microsoft, whenever a consonant cluster is formed with RA as the first consonant followed by more than one consonant and a vowel.

Case 1: Leave the consonant RA, follow the general rule for the formation of a consonant cluster if there are more than one consonant other than RA and then append the reph sign (ꣳ) in front of the consonant cluster.

| <u>Stored Unicode sequence</u> | <u>Example</u> | <u>Resultant conjunct</u> |
|---|---|---------------------------|
| <RA, virama, C> | <0CB0, 0CCD, 0C95 (ರ ಿ ಕ)> | ಕ಼ |
| < RA, virama, C1, vowel> | <0CB0, 0CCD, 0CCB, 0C93 (ರ ಿ ಕ ೀ)> | ಕೋ಼ |
| <RA, virama, C1, virama, C2, virama, , Cn, vowel> | <0CB0, 0CCD,0CA1, 0CCD, 0CAF, 0CCB (ರ ಿ ಡ ಯ ೀ)> | ಡ್ಯೋ಼ |

Case 2: Including the first consonant RA, follow the general rule for the formation of consonant cluster.

| | | |
|---|---|---------|
| <RA, zwj, virama, C> | <0CB0, zwj, 0CCD, 0C95> | ರೃ |
| < RA, zwj, virama, C1, vowel> | <0CB0, zwj, 0CCD, 0C95, 0CCB> | ರೃೋ಼ |
| <RA, zwj, virama, C1, virama, C2, virama, .. , Cn, vowel> | <0CB0, zwj, 0CCD, 0CA1, 0CCD, 0CAF, 0CCB> | ರೃಡ್ಯೋ಼ |

Note: zwj to be used after RA in **Case 1** and **zwj** to be removed in **Case 2** if the suggestion given by me is accepted and adopted.

Written by:
C V Srinatha Sastry
Bangalore