

Specifying Optional Malayalam Conjuncts

Cibu Johny <cibu@google.com>

Roobeh Poornader <roobeh@google.com>

2013-Jan-28

Current status

Indic conjunct formation scheme currently favors the full conjunct for a given set of characters. Example:

ക് + ഘ → ക്ഷ is preferred as opposed to ക്ഷ.

($KA_d + SSA_1 \rightarrow K.SSA_n$)

ക്ഷ can be obtained by ക് + **ZWJ** + ഘ which is $KA_d + \mathbf{ZWJ} + SSA_1 \rightarrow KA_h + SSA_n$

The Need

In Malayalam there are two prevailing orthographies - traditional and reformed - both written with same Malayalam character set. The difference between them is typically manifested only by the font.

Traditional orthography fonts accommodate lot more full conjuncts, while reformed orthography fonts would use visible virama (Chandrakkala) separated sequences for many of those full conjuncts. For the vowel signs of U, UU, and Vocalic vowels and also for the RA-sign, reformed orthography font would use visually separate conjoining form.

However, there is a definite need for the ability in a reformed orthography font to display the traditional full conjuncts on demand. As of now there is no mechanism specified in the standard to suggest a full conjunct of a cluster.

The reverse case is also needed - a traditional orthography font might want to display reformed orthography grapheme clusters optionally.

Following proposal uses ZWJ and ZWNJ insertions to achieve this need. However, potentially Chillu forming sequence <Consonant + Virama + ZWJ> is not used for any of the cases listed below.

Proposal

Case 1

The sequence <Consonant + ZWJ + Conjoining Vowel Sign> has following fallback order for display:

1. Full Conjunct
2. Consonant + non-conjoining vowel sign

Example with reformed orthography font

(in a reformed orthography Malayalam font that can allow optional traditional orthography)

SA + Vowel Sign U → സു

SA + **ZWJ** + Vowel Sign U → സു

Case 2

<Consonant1 + ZWJ + Virama + Consonant2> has following display fallback order:

1. Full conjunct
2. Consonant1 + C2-conjoining
3. Consonant1 + Virama + Consonant2

In the general setting,

... Consonant + Virama + Consonant1 + ZWJ+ Virama + Consonant2 + Virama + Consonant + ... full conjunct is requested only for Consonant1 + Virama + Consoant2. Including remaining consonants and viramas into this is under fonts descretion. See example 4.

Example 1 with reformed orthography font:

SA + Virama + KA → സ്ക

SA + **ZWJ** + Virama + KA → സ്ക

Example 2 with reformed orthography font:

TA + Virama + SA → ത്സ

TA + **ZWJ** + Virama + SA → ത്സ

Example 3 with reformed orthography font:

SA + Virama + RA → (സ

SA + **ZWJ** + Virama + RA → (സ

Example 4 with reformed orthography font:

GA + Virama + DA + Virama + DHA + Virama + RA + Vowel Sign E → ഗ്റെലു
 GA + **ZWJ** + Virama + DA + Virama + DHA + Virama + RA + Vowel Sign E →
 റെലു
 GA + Virama + DA + Virama + DHA + **ZWJ** + Virama + RA + Vowel Sign E →
 ഗ്റെലു
 GA + **ZWJ** + Virama + DA + Virama + DHA + **ZWJ** + Virama + RA + Vowel Sign E →
 റെലു

Note:

This proposal is different from current Devanagari behavior. There, <Consonant1 + ZWJ + Virama + Consonant2> does not form full conjunct, rather the C2-conjoining form, if available.

Case 3

<Chillu + ZWJ + Consonant> has following fallback order for display for its archaic conjuncts:

1. Stacking conjunct
2. Chillu followed by the consonant

Example with reformed or traditional orthography font:

LL Chillu + MA → ശമ
 LL Chillu + **ZWJ** + MA → ശമ

Case 4

<Consonant + ZWNJ + Conjoining Vowel Sign> has following fallback order for display:

1. Consonant + non-conjoining vowel sign
2. Full Conjunct

Example with traditional orthography font:

(in a traditional orthography Malayalam font that allows optional reformed orthography)

SA + Vowel Sign U → സു
 SA + **ZWNJ** + Vowel Sign U → സു

Case 5

<Consonant1 + Virama + ZWNJ + Consonant2> has following fallback order:

1. Consonant1 + Visible Virama (Chandrakkala) + Consonant2
2. Full conjunct

In the general setting,

... Consonant + Virama + Consonant1 + Virama + ZWNJ + Consonant2 + Virama + Consonant + ...
Breaking of the conjunct is requested only for Consonant1 + Virama + Consoant2. Breaking remaining consonants and viramas is under fonts descretion. This infact is the current behavior. This section is added only for completeness.

Example 1 with traditional orthography font:

SA + Virama + KA → ക്ക

SA + Virama + ZWNJ + KA → ക്ക

Example 2 with traditional orthography font:

TA + Virama + SA → ട്ട

TA + Virama + ZWNJ + SA → ട്ട

Example 3 with traditional orthography font:

GA + Virama + DA + Virama + DHA + Virama + RA + Vowel Sign E → റ്റേ

GA + Virama + ZWNJ + DA + Virama + DHA + Virama + RA + Vowel Sign E →

റ്റേ

Case 6



<Consonant1 + ZWNJ + Virama + Consonant2> has following fallback order:

1. Consonant1 + Visually separate conjoining form of Consonant2
2. Full conjunct
3. Consonant1 + Visible Virama (Chandrakkala) + Consonant2



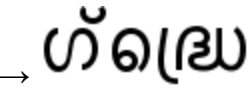
In the general setting,

... Consonant + Virama + Consonant1 + ZWNJ + Virama + Consonant2 + Virama + Consonant + ...
 Breaking of the conjunct is requested only for Consonant1 + ZWNJ + Virama + Consoant2. Breaking remaining consonants and viramas on the respective sides is under the descretion of the fonts.

Example 1 with traditional orthography font:

SA + Virama + RA → 
 SA + **ZWNJ** + Virama + RA → 

Example 2 with traditional orthography font:


GA + Virama + DA + Virama + DHA + Virama + RA + Vowel Sign E → 
 GA + Virama + DA + Virama + DHA + **ZWNJ** + Virama + RA + Vowel Sign E →

 GA + Virama + **ZWNJ** + DA + Virama + DHA + **ZWNJ** + Virama + RA + Vowel Sign E
 → 

Case 7

<Chillu + ZWNJ + Consonant> has following fallback order for display for its archaic conjuncts:

1. Chillu followed by the consonant
2. Stacking conjunct

Example with traditional orthography font with archaic support:

LL Chillu + MA → 
 LL Chillu + **ZWNJ** + MA → 