Background Document for PRI #250: Proposal to Specify Optional Conjuncts in Malayalam

This is the background document on the proposal to specify the further use of ZWJ and ZWNJ in sequences in the Malayalam script to indicate preferences for optional display of conjuncts. Such sequences are intended to indicate the preferences, both for rendering systems that support the reformed Malayalam orthography and for systems that support the traditional Malayalam orthography. Sequences involving ZWJ, ZWNJ and/or Virama are proposed. The UTC would be interested in learning of any current implementations which would be adversely affected by any of the proposed modifications. UTC would also likes to know if there are any related cases that are not considered here. Alternate sequence proposals to achieve this objective are also welcome.

Introduction

In Malayalam there are two prevailing orthographies - traditional and reformed - both are written using same Malayalam character set. The difference between them is typically manifested only by the font. Traditional orthography accomodate lot more full conjuncts, while reformed orthography would use visibile virama (Chandrakkala) separated sequences for many of those full conjuncts. For the vowel signs of U, UU, and Vocalic R, RR and also for the RA-sign, reformed orthography font would use visually separate forms.

However, the difference between the orthographies, comes with spelling differences at times. That is, if something is intended in traditional (both in style and spelling), when viewed in a reformed script font, it would be awkward. For example, consider the word:

കായ്യവും, (U+0D15 U+0D3E U+0D4E U+0D2F U+0D4D U+0D2F U+0D35 U+0D41 U+0D02 = /kaaryyavum/)

If the font is made for reformed orthography, then one cannot avoid the ransom-note effect of orthography mixing as in:

So, 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 of the above scenario is also needed - a traditional orthography font might want to display reformed orthography grapheme clusters optionally.

Current status

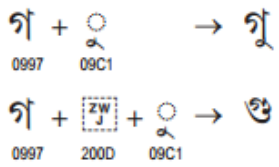
Indic conjunct formation logic currently favors the full conjunct for a given set of characters. Half forms are produced by ZWJ, which acts an invisible consonant that would always try to form a ligature with the consonant on the otherside of the Virama.

So <Consonant + VIRAMA + ZWJ> provides the half form of the initial consonant. Similarly the C2-conjoining forms is specified as <ZWJ + VIRAMA + Consonant>. ZWNJ works on the contratry as consonant that *never* ligates with the consonant on the otherside of the Virama. See the Figure 9-7 in the standard version 6.2 for examples:

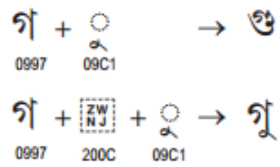
Rendering fallback sequence employed is:

1. Full conjunct
2. Conjunct with half-forms
3. Consonants in nominal form with explicit (visible) Virama.

There is precedence in Bengali for the usage of ZWJ and ZWNJ as in Latin to influence <Consonant + Vowel sign> ligatures. The ZWJ is used as an glue to request a fuller ligature; while ZWNJ blocks the fuller ligature. See Figures 9-12 and 9-13 from the standard:



(fig. 9-12)



(fig. 9-13)

Proposal

This proposal uses ZWJ and ZWNJ insertions to achieve the need described above. Following cases list the additional semantics to be associated to the sequences involving ZWJ and ZWNJ to achieve that. Potentially Chillu forming sequences (<Consonant + VIRAMA + ZWJ>) are not used for any of the cases listed below.

Case 1

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

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

Most of the full conjuncts in Malayalam involve C2-conjoining form. However, there are cases where full conjunct is not by C2-conjoining; also for these conjuncts there are no more fuller form like that of 'KSA' example for Devanagari. So the presence of ZWJ request full conjunct.

Thus, this proposal deviates from the current Devanagari behavior. There, <Consonant1 + ZWJ + VIRAMA + Consonant2> does not form full conjunct, rather would produce the C2-conjoining form, if available. Following examples are with reformed orthography font:

SA + VIRAMA + KA → സ്ക

SA + **ZWJ** + VIRAMA + KA → സ്ക

TA + VIRAMA + SA → ത്സ

TA + **ZWJ** + VIRAMA + SA → ത്സ

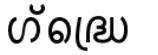
SA + VIRAMA + RA → സ്ര

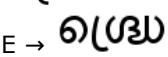
SA + **ZWJ** + VIRAMA + RA → സ്ര

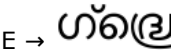
LLLA + VIRAMA + VA → ഴ്വ

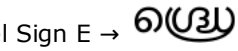
LLLA + **ZWJ** + VIRAMA + VA → ഴ്വ

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 for reformed orthography font:

GA + VIRAMA + DA + VIRAMA + DHA + VIRAMA + RA + Vowel Sign E → 

GA + **ZWJ** + VIRAMA + DA + VIRAMA + DHA + VIRAMA + RA + Vowel Sign E → 

GA + VIRAMA + DA + VIRAMA + DHA + **ZWJ** + VIRAMA + RA + Vowel Sign E → 

GA + **ZWJ** + VIRAMA + DA + VIRAMA + DHA + **ZWJ** + VIRAMA + RA + Vowel Sign E → 

Case 2

<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

This is the reverse of the previous case. See the example below for traditional orthography font:

SA + VIRAMA + RA → 

SA + **ZWNJ** + VIRAMA + RA → 

LLLA + VIRAMA + VA → 

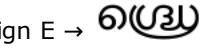
LLLA + **ZWNJ** + VIRAMA + VA →

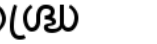
YA + VIRAMA + YA → 

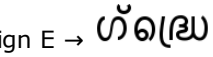
YA + **ZWNJ** + VIRAMA + YA → 

In the general sequence,
... Consonant + VIRAMA + Consonant1 + ZWNJ + VIRAMA + Consonant2 + VIRAMA + Consonant + … breaking of the full conjunct is requested only for Consonant1 + ZWNJ + VIRAMA + Consoant2. Breaking remaining consonants and viramas on the respective sides is under the descretion of the font.

Examples for 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 3

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

1. Stacking conjunct
2. Chillu followed by the consonant

See following example with reformed or traditional orthography font:

LL Chillu + MA → 

LL Chillu + **ZWJ** + MA → 

Case 4

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

1. Chillu followed by the consonant
2. Stacking conjunct

See following example with traditional orthography font with archaic support:

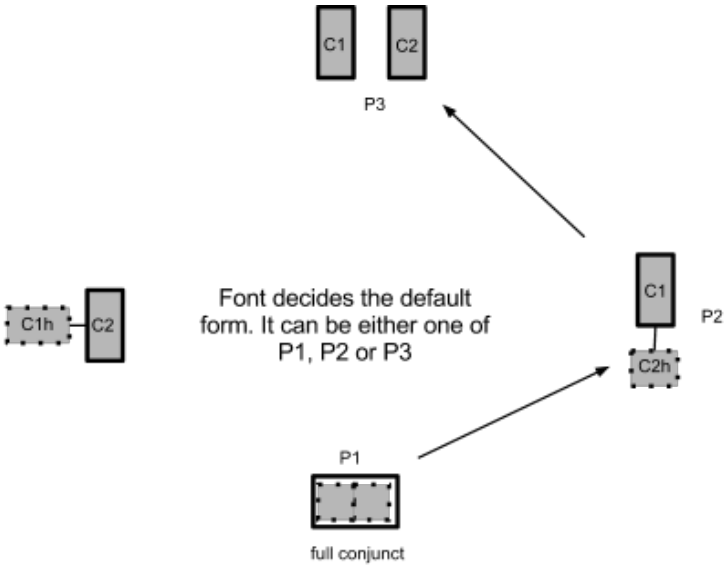
LL Chillu + MA → 

LL Chillu + **ZWNJ** + MA → 

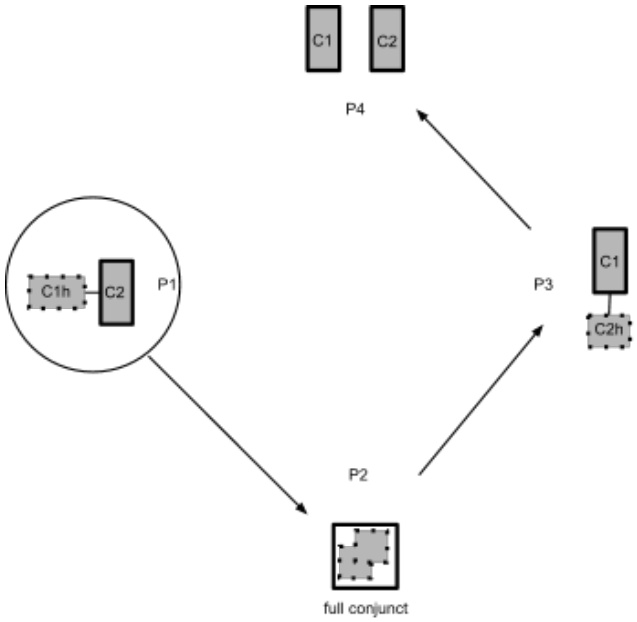
Grahical representation

Above proposal is represented graphically below to show the fallback priority order (P1 → P2 → P3 → P4) for various Joiner-Virama combinations.

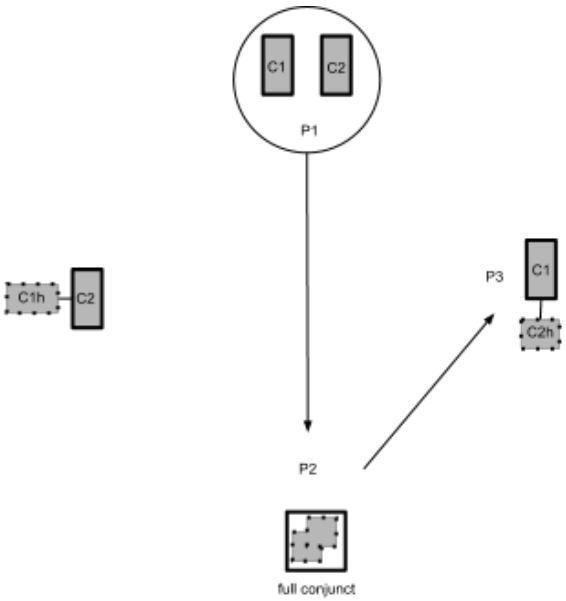
The circle represents the default form. Arrows represent the fallback path to be taken for laying out.



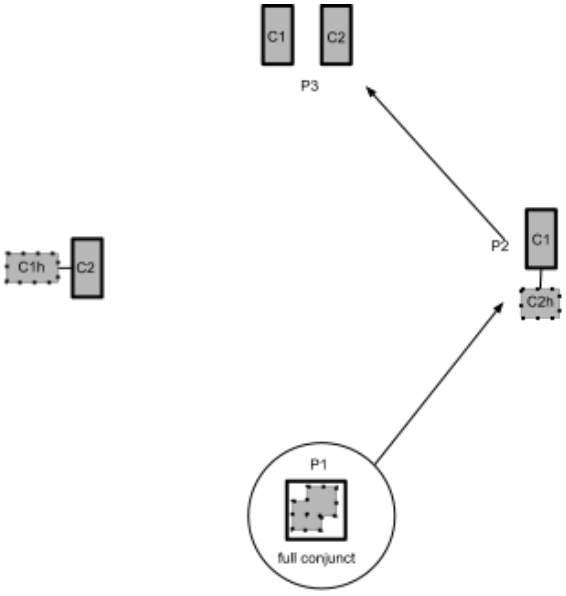
C1 + VIRAMA + C2



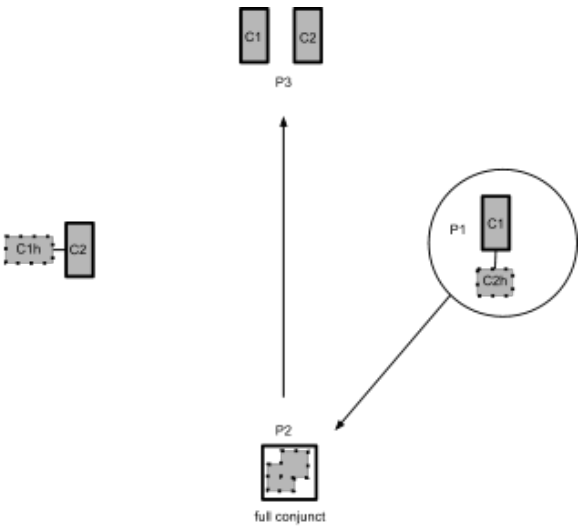
C1 + VIRAMA + ZWJ + C2



C1 + VIRAMA + ZWNJ + C2



C1 + ZWJ + VIRAMA + C2



C1 + ZNWJ + VIRAMA + C2