Representing an Underline Diacritic in Telugu

Vinodh Rajan vinodh@virtualvinodh.com

Introduction

Apart from the Nukta proposed in L2/20-085, Telugu also uses an underline diacritic that acts both as a vowel and consonant modifier. This document discusses the appropriate Unicode representation of this diacritic mark in Telugu.

Underline as Vowel Modifier

Telugu phonology displays two additional vowels: /æ/ & /æː/ that do not have a standard representation in the Telugu orthography. The former is analogous to Devanagari Candra E े and Sinhala AE ඇ and the latter to Sinhala AAE ඍ. They have been represented in a number of inadequate ways as a stop-gap measure. For instance, the syllable /kæ/ is mostly represented in Telugu as కా kā or కాా kyā.

Modern Telugu linguists, however, use an underscore-like diacritic with Telugu character E and E to unambiguously denote the two phonemes.

Krishnamurti et al. (1985)
Below are attestations from Krishnamurti (1979), a edited volume of articles written by various Telugu linguists.

Vowels AE and AAE

Vowel Signs AAE (Red) and AE (Green)

Vowel AAE (Red) and Vowel Sign AAE (Green)
Underline as Consonant Modifier

Though, the Nukta is already in use by Telugu scholars. At least, one scholar in Krishnamurti (1979) assigns the underline diacritic a Nukta-like role to extend the aspirated (those with tear-drop like glyphic component near the base) consonants and denote Perso-Arabic phonemes.

Possible Unicode Representations

(i) Encode a Telugu Sign Underscore

This is similar to the Oriya Sign Overline U+0B55 (See L2/19-005) encoded in the Oriya block. If a new Telugu Sign Underscore is encoded and given the combining class 7, it can be used as a consonant modifier and a vowel modifier. Devanagari Sign Nukta already does this double role.

\[
\text{తు = TA + Nukta + Vowel Sign O (Modifies the consonant)}
\]

\[
\text{తు = TA + Vowel Sign O + Nukta (Modifies the vowel)}
\]

Similarly, we could do:

\[
\text{ఫా = PHA + Telugu Sign Underscore + Vowels Sign U (Modifies the consonant)}
\]

\[
\text{ఫా = KA + Vowel Sign EE + Telugu Sign Underscore + Anusvara (Modifies the vowel)}
\]

However, this would amount to re-encoding the Anudatta again (albeit with a different combining class)
(ii) **Unification with Anudatta**

Sometimes, an *underline* is what it is, an *underline* and nothing more. The UCS already has an underlining character, the Stress Sign Anudatta U+0952 that can be readily used to represent the underline diacritic. This is quite similar to the Sindhi Implosives in Devanagari: ढ ढः ढ़, which initially used *Nukta + Anudatta*, but were eventually encoded as separate characters.

The placement of the Anudatta in the character stream might be an issue. As a consonant diacritic, it would have to come immediately after the consonant and as a vowel diacritic immediately after the vowel. But the Anudatta is currently placed after a syllable.

<table>
<thead>
<tr>
<th>Role</th>
<th>Example</th>
<th>Current Encoding Order</th>
<th>Ideal Encoding Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel Modifier</td>
<td>කు viewModel</td>
<td>KA + Vowel Sign EE + Anusvara + Anudatta</td>
<td>KA + Vowel Sign EE + Anudatta + Anusvara</td>
</tr>
<tr>
<td>Consonant Modifier</td>
<td>ఫు.viewModel</td>
<td>PHA + Vowel Sign U + Anudatta</td>
<td>PHA + Anudatta + Vowel Sign U</td>
</tr>
</tbody>
</table>

It is not a major problem anyway. Even though the Anudatta is placed at the end of the syllable, the rendering engines/fonts are supposed to place it appropriately near the base.

It could just be recommended to use Anudatta at the end of the orthographic syllable in both the cases (as vowel and consonant modifier) as long as the placement is done properly. A section should be added to the Telugu chapter of the core specification commenting on the representation of the underline diacritic.

*This would be the recommended option as it allows the usage of existing characters without encoding new characters.*

**References**