

14/04/2020

## Representing an Underline Diacritic in Telugu

Vinodh Rajan [vinodh@virtualvinodh.com](mailto: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/ & /æ:/ that do not have a standard representation in the Telugu orthography. The former is analogous to *Devanagari Candra E* एँ and *Sinhala AE* ඇ and the latter to *Sinhala AAE* ඇ. 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 క k̄ or క్య kyā.

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

4.3. *Distinctive sounds without symbols*: Telugu orthography has not provided for two important distinctive sounds in Modern Standard Telugu, viz., *EE* and *f*.

EE is pronounced like the English vowel in *cat*, but is approximately twice as long in duration; e.g.

ఘో [mEEka] 'goat', తాగేను [taagEEnu] 'I drank'.

When it occurs as the past tense suffix, EE is generally represented in writing by → [aa] in its secondary form; e.g.

Written	Pronounced	
కొట్టాడు	koTTEEDu	'he beat'
వచ్చాడు	waccEEDu	'he came'.

But sometimes ఁ in its secondary form is used, e.g. కొట్టెడు, వచ్చెడు. In most cases EE occurring in the first syllable of a word is represented by the symbol ఁ [ee] or its secondary form; e.g.

Written	Pronounced	
ఘోష	mEEDa	'mansion'

When there is a need to represent [EE] as a separate sound, we shall transcribe it as EE.

Krishnamurti et al. (1985)

14/04/2020

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

ఎ ఏ లకు ఈ యుగంలో తాలవ్య హల్లుల తర్వాత ఎ ఏ (ఐ ఐ) అని సవర్ణాలుండినట్లు కొన్ని స్పష్టమైన సూచనలు కనిపిస్తాయి.

Vowels AE and AAE

ఇదికాక వివృతమైన ఏ ఉచ్చారణ క్రియల్లో ఇ + ఆ అచ్చుల సంయోగం వల్ల కూడా ఏర్పడుతుంది. ఉదా : ఆగి + ఆడు > ఆగేడు, చెప్పి + ఆడు > చెప్పేడు, వెళ్ళి + ఆడు > వెళ్ళేడు.

పదాంతంలోనూ, మకారాంత పదాల్లో చివరి 'మ' కరానికి పూర్వం వచ్చినప్పుడు 'ఎ' వర్ణానికి వివృతీచ్ఛారణ ఉంది. ఉదా : గిన్నె, ప్తెం.

Vowel Signs AAE (Red) and AE (Green)

(24) ఏ → ఏ / # —హ ఆ

పరాక్షరంలో అత్యంత ఉన్నప్పుడు దాతుగత 'ఏ' వర్ణం ఉచ్చారణలో వివృతమై ఏ ('Cat' లో ఉన్న అచ్చులాగా) అవుతుంది. ఇది 15, 16 శతాబ్దాల శాసన రచనల్లో కనిపిస్తుంది. ప్రత్యేక వర్ణత్వం లేకపోవటంవల్ల 'యా' గాను, కొన్ని చోట్ల 'ఆ' గాను ఇది రాతలో సూచించబడ్డది ; ఉదా. పండితుల చాత (క్రి.శ. 9548), (< సేవ ; క్రి. శ. 1574), మేరకు (క్రి శ. 1688) (గ్రి 5.9. 6.8(c)). పశ్యాలు (క్రి. శ. 1518) పదాన్ని ఈనాడు కోస్తా జిల్లాల్లో ఉచ్చరించి నట్లు (పక్ తేలు అని) ఉచ్చరించి ఉండవచ్చు (గ్రి 5.7).

Vowel AAE (Red) and Vowel Sign AAE (Green)

14/04/2020

## 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.

ఖ > క  
 దన్ ఖ క దస్కతు 'సంతకం'

*/kh/ with underline to denote the phoneme /x/*

ప్ర > ప  
 ప్ర స క పురునతు 'శీరిక'

*/ph/ with underline denoting /f/*

## 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.

తో = TA + Nukta + Vowel Sign O (*Modifies the consonant*)

తో = TA + Vowel Sign O + Nukta (*Modifies the vowel*)

Similarly, we could do:

ఘ = PHA + Telugu Sign Underscore + Vowels Sign U (*Modifies the consonant*)

ఘో = 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)

14/04/2020

(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: ण ॡ ॢ ॣ, 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.

Role	Example	Current Encoding Order	Ideal Encoding Order
Vowel Modifier	ॐ –	KA + Vowel Sign EE + Anusvara + Anudatta	KA + Vowel Sign EE + Anudatta + Anusvara
Consonant Modifier	ॡ –	PHA + Vowel Sign U + Anudatta	PHA + Anudatta + Vowel Sign U

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.

*R10 Other modifying marks, in particular bindus and svaras, apply to the orthographic syllable as a whole and should follow (in the memory representation) all other characters that constitute the syllable. The bindus should follow any vowel signs, and the svaras should come last. The relative placement of these marks is horizontal rather than vertical; the horizontal rendering order may vary according to typographic concerns.*

KA<sub>n</sub> + AA<sub>VS</sub> + CANDRABINDU<sub>n</sub>

क + ा + ँ → काँ

**Unicode Chapter 12.1**

It could just be recommended to use to 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**

1. Krishnamurti, B (Ed.) (1979). Telugu Bhāṣācaritra. Andhara Pradesh Sahitya Academy, Hyderabad.
2. Krishnamurti, B., & Gwynn, J. P. L. (1985). A grammar of modern Telugu. Oxford University Press, USA.