

# DEVELOPING A TRANSLITERATION SYSTEM FOR TELUGU SCRIPT

Vinodh Rajan, [vinodh@virtualvinodh.com](mailto:vinodh@virtualvinodh.com)

## 1.0 Introduction

Transliteration is a method of representing the text written in one writing system in another writing system. It does not reproduce the pronunciation of the letters in the source, but only attempts to replicate the way a text is written in the source. During transliteration, the source is preserved in the target script, by replacing it with the characters of the target script - ideally, in a one-to-one manner. But, the level of preservation of the source generally varies. In a fully reversible transliteration, the source spelling is completely preserved in the target script, which enables to recreate the original without any loss of information.

## 1.1 Indic Scripts

All modern Indic scripts are ultimately derived from Brāhmi, the script of the Aśōkan inscriptions. Being Brāhmic descendants, the modern Indic scripts still maintain – for the most part - a one-to-one correspondence with each other. But apart from the common repository of characters, each script had also developed several other additional characters for representing the phonemes of the regional language. These include the perso-arabic extended characters present in the North Indic scripts, and the characters such as short e, short o, Trill ṛa etc present in the South Indic scripts.

## 1.2 Modern Telugu Script

The modern Telugu script is a descendant of the old Kannada-Telugu script that was employed for writing Telugu & Kannada languages[1]. The old script is in turn a derivation of the older Kadamba script which evolved from the southern Brāhmic variant. Telugu Script has a common Brāhmic character stock along with several characters required to write the native language.

## 1.3 Transliteration Standards

Transliteration standards such as IAST, ISO 15919 exist to standardize the way in which Indic scripts are represented using Latin characters. IAST primarily deals with Sanskrit, and ISO 15919 with all the Indic Languages. ISCII prescribes a standard that enables Dēvanāgarī script to represent all the characters of the Indic scripts.

In all of these cases, special diacritic marks are introduced to extend the native character set and enable it to represent a foreign character set.

## 1.4 Transliteration System

A Transliteration System is composed of two units. First, is the static constituent - which is an established *Transliteration Standard* for representing the non-native characters in the native script. The next is the actual working element – a converter module which will actually implement the standard, by transliterating the input text to the native script's characters in accordance with the standard.

Both of these components integrate together to form a working Transliteration System.



ऐ	ओ	र	य	न	ळ
ఎ	ఒ	ఱ	ఝ	ణ	ఱ
e	o	ṛ	ỵ	ṇ	ḷ

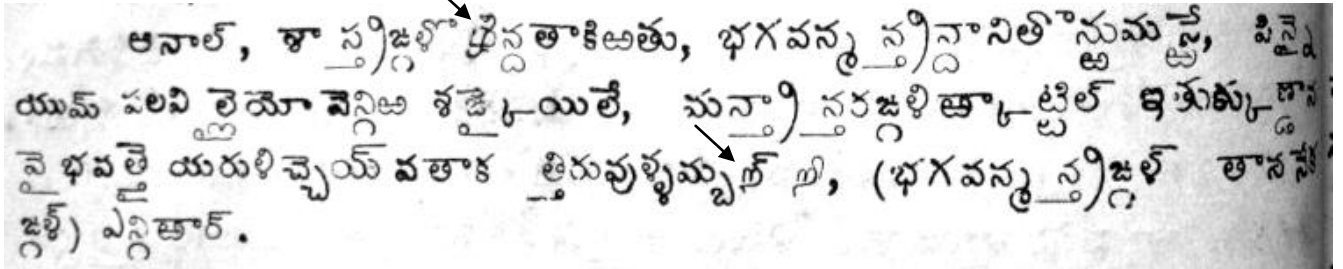
As seen above several new characters were created to represent the missing characters. In most of the cases a diacritic mark *Nukta* has been used to invent the new characters, in some other chases existing shapes were modified to create new ones (for instance, note the wavy stroke in ओ as compared to औ )

### 2.3 Transliteration to Telugu

At times, mostly in scholarly works and also in dictionaries, it is necessary to express other Indic Languages in Telugu script, with the originality of the source preserved. In English, IAST/ISO enables such lossless transliteration into Latin script. However, such a situation is not possible for Telugu.

Also, Telugu still has its presence and influence among the Vaiṣṇava group surrounding Tamil Nadu, with a need for transliteration. With Urdu as the second official language of Andhra Pradesh, it may be necessary to accurately represent the perso-arabic consonants in Urdu just like the north Indic scripts.

There have been several attempts to transliterate other languages into Telugu script. Consider the sample below printed at 1928.



*Tamil written in Telugu Script [3].*

Note how Tamil letters ఱ ḷ & ఱ ṛ have been imported as Telugu characters by using Telugu mātras (The typesetter for some reasons didn't use the native archaic Telugu equivalents when they do exist).

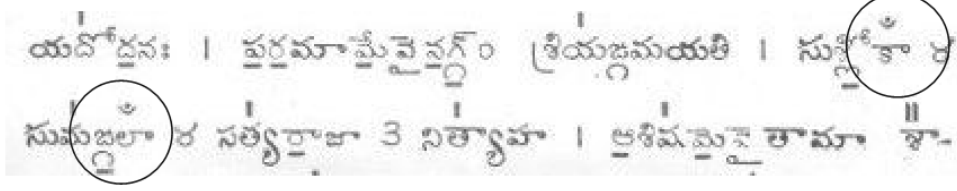
### 3.0 Parivrddha Telugu (పరివృద్ధ తెలుగు)

As in Dēvanāgarī, there is a need and necessity to develop *Parivrddha Telugu* for reasons akin. Based on the overview of characters seen earlier, the following characters need to be added and/or re-introduced to the standard Telugu script to establish *Parivrddha Telugu*.

The important part is to get these characters encoded in the Telugu Unicode Block. Detailed proposals must be made to the UTC to make these characters included in the Telugu Unicode standard based on the feedback and discussions from the user community.

### 3.1 Telugu Sign Combining Chandrabindu

The Telugu Unicode block has the sign named Telugu Sign Chandrabindu encoded at U+0C01 already. But it is not a true equivalent of the north Indic Chandrabindu but a Telugu prosody character named “*ara sunna*” ు. A true Chandrabindu is needed for transliteration. It must be noted that several Vedic Sanskrit books in Telugu script already employ this character.



*Vedic Sanskrit in Telugu Script [4]*

A Unicode proposal for encoding this Chandrabindu has already been submitted and accepted by the UTC to be encoded at U+0C00.

### 3.2 Telugu Sign LLLA

ᱚ is currently encoded in Unicode blocks of the three south Indic languages - *Tamil* ழ, *Kannada* ಳ & *Malayalam* ഴ. Telugu Unicode doesn't have this letter currently encoded. But Old Telugu Script had this archaic letter included before it was dropped out of usage from Telugu. The letter also found its usage in the śāsanas written in the Telugu language.



*Telugu script notes, TDIL [5]*

Telugu Sign LLLA looks very similar to the equivalent Kannada character ಳ. The Telugu Sign LLLA must be included at U+0C34

### 3.3 Telugu Sign Nukta

*Nukta* (Arabic for *Dot*) is a dot shaped character used to extend the basic character set to represent new sounds. It is present in all the north Indic scripts and also Kannada. (In Tamil ∴ U+0B83 plays the de-facto role of the Nukta). The north Indic scripts use the Nukta to extend the basic characters to represent the perso-arabic consonants. Kannada uses the Nukta to represent /f/ and /z/.


क ख ग ज ङ ढ ढ ण ञ

Basically, the presence of the Nukta can enables a script to extend to denote any non-native sound.

There have been efforts of using Nukta in the Telugu script such as below [6].

మాయ ప్పొరుపడై వాణనై\* ఆయిరందోఁకుం పొఁకికురుది -  
పాయ\* ప్పళ్ళియి ఆఁకివల్లసుక్కు\* ప్పల్లాండు కూఱుదుమే

However Nukta at the center of the consonant overlaps with the mahāprāṇa mark in some consonants. A better solution is provided by the TDIL (Technology Development for Indian Languages) which has proposed a Telugu Nukta placed at the left of the consonant [7].

0C3C		TELUGU SIGN NUKTA • Placed at bottom left corner of the letter
------	---	---

However a formal proposal was never made to the Unicode consortium.

The Nukta sign can be used to create the equivalent of all the Indic characters in Telugu. It completely eases the transliteration of North Indic scripts into Telugu

क ख ग ज ड ढ फ ण य  
క ఖ గ జ డ ఢ ఫ న య

It can have a variety of other uses also – for instance, if there arises a need to represent the /w/ sound it can be done by placing the Nukta on వ - va <sup>◌̣</sup>. Any required sound can be represented by combining the Nukta and the required consonant.

To avoid complications in font rendering, As in [6], Consonants with Nukta can be prevented to form saṁyuktākṣaras (Conjuncts).

### 3.4 Other Characters

There are several other characters like *Cillākṣaras* in Malayalam, and *Khanda Ta* in Bengali, we may require unique representation. In such cases, U+02BC ' (Modifier letter apostrophe) can be used alongside with the usual halant-forms to denote these special pure-consonant forms.

### 3.5 Native Conventions

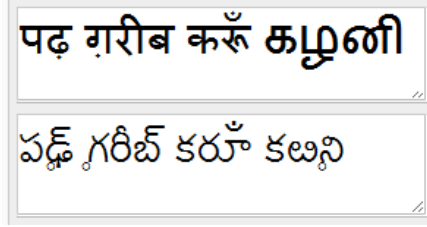
Telugu has some native conventions to write specific phonemes. A transliteration system must also be able to apply these conventions and transform the text (if required).

- The Nasal letters must be converted into Anusvāra (e.g.) पञ्च - పంచ (rather than పచ్చ), वेङ्कट - వేంకట (rather than వేఙ్కట)
- The pure consonant 'm' when occurring as the final syllable, it must be changed into Anusvāra (e.g.) सत्यम् - సత్యం

### 4.0 Aksharamukha – Script Converter

Aksharamukha (<http://www.virtualvinodh.com/aksharamukha>), an open source PHP based web transliteration application that works based on the proposed standard has been developed. This application converts all the Indic orthographies including East-Asian Scripts and also several Roman Transliteration standards such as ISO 15919, IAST, Harvard-Kyoto, Velthuis into Telugu Script & Vice Versa

Parivṛddha Telugu as discussed above has been implemented by the converter. A working Unicode font called “*Parivṛddha Telugu*” [8] has been forked from the open source Telugu font *Lohit Telugu* with all the new glyphs introduced (Telugu Sign Combining Chandrabindu, Telugu Letter LLLA and Telugu Sign Nukta) at the proposed locations. The converter is presented as a model implementation of the proposed system.



### 5.0 Extending Parivṛddha Telugu – Future Developments

This paper as an initial attempt deals only with the wholesome representation of the Indic script characters. In the future, it is undoubtedly necessary to enable Telugu script to represent other phonemes with additional diacritics. Telugu script can be vitalized more by standardizing some diacritics that can be used along with existing Telugu characters to present vital non-native sounds such as /a/ in *apple*.

There exists dedicated Unicode blocks for diacritics – *Spacing Modifier Letters & Combining Diacritical Marks*. A required subset of these diacritics could be standardized to be used as consonantal and vowel modifiers to represent the non-existing sounds in Telugu. One possible usecase is an English-Telugu dictionary, where English phonemes need to be represented in Telugu.

(e.g.) cot [kɒt] - ీృ <U+0C15,U+0C3E,U+0306,U+0C1F,U+0C4D>

### 6.0 Conclusion

In the initial stages, *Parivṛddha Telugu* may be restricted to scholarly usage and possibly in dictionaries, but as time passes it may get adopted to mainstream usage also. Scholars will be able to represent other Indic languages in Telugu script with more ease. It will also enable people to read, learn and represent other Indic languages completely in Telugu script itself. Existence of a complete transliteration system will certainly enrich Telugu.

### 7.0 References

1. <http://www.ancientscripts.com/telugu.html>
2. Indian Script Code for Information Interchange (ISCII) – IS13194 : 1991
3. tirumarṅtrārthamu, āṁdhra drāṁḁa vyākhyāna sahitamu, śrīmān karṁdāḁai śēṣācāryulu, 1928
4. Request to encode South Indian Chandrabindu-s, JTC1/SC2/WG2 N396  
<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3964.pdf>
5. Vishwabharat @ TDIL, Issue 5, April 2002  
<http://tdil.mit.gov.in/ori-guru-telu.pdf>
6. <http://prapatti.com/slokas/telugu/naalaayiram/periyaazvaar/tiruppallaandu.pdf>
7. Proposed changes in Unicode Standards for Indic Scripts – Telugu  
[http://tdil.mit.gov.in/prop\\_uni/Telugu.pdf](http://tdil.mit.gov.in/prop_uni/Telugu.pdf)
8. <http://www.virtualvinodh.com/download/fonts/Parivṛddha%20Telugu.ttf>