

Proposal for a Telugu Script Root Zone Label Generation Ruleset (LGR)

LGR Version: 3.0

Date: 2018-08-08

Document version: 2.6

Authors: Neo-Brahmi Generation Panel [NBGP]

1. General Information/ Overview/ Abstract

This document lays down the Label Generation Rule Set for the Telugu script. Three main components of the Telugu Script LGR, viz. Code point repertoire, Variants and Whole Label Evaluation Rules have been described in detail here. All these components have been incorporated in a machine-readable format in the accompanying XML file: "Proposal-LGR-Telu-20180808.xml".

In addition, a list of test labels has been provided in the following file, which covers the repertoire, variant code points and the whole label evaluation rules, providing examples for valid and invalid labels: "telugu-test-labels-20180808.txt".

2. Script for which the LGR is proposed

ISO 15924 Code: Telu

ISO 15924 Key N°: 340

ISO 15924 English Name: Telugu

Latin transliteration of native script name: teluguu

Native name of the script: తెలుగు

Maximal Starting Repertoire [MSR] version: 3

The Unicode Standard, Version: 6.3

Telugu Unicode Range: 0C00–0C7F

3. Background of the Script and Principal Languages Using It

The Telugu language uses the Telugu script which is written in the form of sequences of orthographic syllables. Each orthographic syllable is formed of one or more Telugu characters placed from left to right and top to bottom. Telugu is one of the 22 scheduled languages of India. The Telugu script is immediately related to Kannada and closely related to the Sinhala script.

3.1 The Evolution of the Script

The origins of the Telugu script can be traced to the Brahmi alphabet of ancient India, often known as Asokan Brahmi. Historically the script is derived from the Southern Brahmi or Bhattiprolu Brahmi alternatively known as the Telugu Brahmi alphabet of 3rd century BCE. Later, by 5th century during the *Chalukyan* period, it developed into a common alphabet used for Telugu and Kannada. The Telugu-Kannada common alphabet split into two separate alphabets during the 12th and 13th centuries AD to be called the Telugu and Kannada scripts. In addition to the common origin, a longer period of shared political and cultural confederation of the Telugu and Kannada speaking regions has ultimately resulted in the considerable proportion of the shared identical character signs between the two scripts (34 out of 63 characters, see Table 10) .

The earliest known inscriptions containing Telugu words appear on the bilingual coins of *Satavahanas* that date back to 2nd century AD [104]. The first inscription entirely in Telugu was made in 575 AD and was probably made by *Renati Cholas*, who started writing royal proclamations in Telugu instead of Sanskrit. Telugu developed as a poetical and literary language during the 11th century AD. Until the 20th century Telugu was written in *Granthic* style very different from the colloquial language. During the second half of the 20th century, a modern written style emerged based on the modern colloquial language. In 2008 Telugu was designated as a classical language by the Indian government.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | అ | ఇ | ఈ | ఊ | ఋ | ౠ | ఎ | ఐ | ఒ | ఓ | ఔ | క | ఖ | గ | ఘ | ఙ | చ | ఛ | ఞ | ట | ఠ | డ | ణ | త | థ | ద | ధ | న | ప | ఫ | బ | భ | మ | య | ర | ల | వ | శ | ష | స | | | | |
| 2 | సాతవాహన ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 3 | భట్టిప్రోలు రామనందమూర్తిని ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | | | | | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 4 | సతవాహన కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 5 | చిత్తూరు కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 6 | గుమ్మల కాలపు సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 7 | కాంచీ కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 8 | చిత్తూరు కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 9 | కాంచీ కాలపు సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 10 | భట్టిప్రోలు కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 11 | కాంచీ కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 12 | కాంచీ కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 13 | కాంచీ కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 14 | కాంచీ కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |
| 15 | కాంచీ కాలపు ప్రాకృతం సం. ౩వ శతాబ్దం | 𑀅 | 𑀆 | 𑀇 | 𑀈 | 𑀉 | 𑀊 | 𑀋 | 𑀌 | 𑀍 | 𑀎 | 𑀏 | 𑀐 | 𑀑 | 𑀒 | 𑀓 | 𑀔 | 𑀕 | 𑀖 | 𑀗 | 𑀘 | 𑀙 | 𑀚 | 𑀛 | 𑀜 | 𑀝 | 𑀞 | 𑀟 | 𑀠 | 𑀡 | 𑀢 | 𑀣 | 𑀤 | 𑀥 | 𑀦 | 𑀧 | 𑀨 | 𑀩 | 𑀪 | 𑀫 | 𑀬 | 𑀭 | 𑀮 | 𑀯 | 𑀰 |

Figure 1: Evolution of Telugu script

3.2 Notable Features

The Telugu orthography superficially appears as a series of circles and semi-circles. Most consonants carry a tick mark called *Talakattu*. The writing system is classified as abugida type that employs alpha-syllabaries. The alphabet consists of vowels, consonants and modifiers. Each of these vowels and consonants has one or more secondary allographs. The secondary allographs always appear as dependent symbols on the first character of a syllable. Each syllable is formed of a single standalone vowel or one or more consonants. Each of these consonants may occur with an inherent vowel or modified by a secondary vowel. A Consonant cluster may be formed with a single standalone character followed

by one or more secondary forms of consonants. The order of composition of syllabaries does not match with the reading order. There are rules to learn to read orthographic sequences into phonetic sequences whether simple or complex syllables.

3.3 The Telugu (తెలుగు) Language

The Telugu language is a Dravidian language spoken by about 75 million (ca. 2001) people mainly in the southern Indian states of Andhra Pradesh and Telangana where it is the official language. It is also spoken in such neighboring states as Karnataka, Tamil Nadu, Orissa, Maharashtra and Chattisgarh, and is one of the 22 scheduled languages of India. There are also quite a few Telugu speakers in Canada, the USA, South Africa, Malaysia, Mauritius, Myanmar, Sri Lanka and Réunion

3.4 Languages that Use the Telugu Script

The script is also used for ten other languages, viz. Gondi, Koya, Konda, Kuvi, Kolavar or Kolami, Yerukala, Banjara or Lambadi, Savara or Sora, Adivasi Odiya and also Sanskrit. In the Telugu speaking region, the tradition of writing Sanskrit in the Telugu script has remained a common practice. During the last few decades, a considerable number of publications in the form of text books, dictionaries and other reading material has been produced in the Telugu script in Gondi, Koya, Konda, Kuvi, Kolami, Yerukala, Banjara, Savara and Adivasi Odiya.

| no. | Name of the language (ISO639 Code) | Language family | Status | EGIDS Scale |
|-----|------------------------------------|-----------------|-------------------------|-------------|
| 1 | Telugu (tel) | Dravidian | Scheduled and Classical | 2 |
| 2 | Gondi (gon) | Dravidian | Modern Tribal | 5 |
| 3 | Koya (kff) | Dravidian | Modern Tribal | 5 |
| 4 | Konda (knd) | Dravidian | Modern Tribal | 6b |
| 5 | Kuvi (kxv) | Dravidian | Modern Tribal | 5 |
| 6 | Kolavar or Kolami (kfb) | Dravidian | Modern Tribal | 5 |
| 7 | Yerukala (yeu) | Dravidian | Modern Tribal | 6 |
| 8 | Banjara or Lambadi (lmn) | Indo-Aryan | Modern Tribal | 5 |
| 9 | Savara or Sora (srb) | Austro-Asiatic | Modern Tribal | 5 |
| 10 | Adivasi Odiya (ort) | Indo-Aryan | Modern Tribal | 5 |

| no. | Name of the language (ISO639 Code) | Language family | Status | EGIDS Scale |
|-----|------------------------------------|-----------------|-------------------------|-------------|
| 11 | Sanskrit (san) | Indo-Aryan | Scheduled and Classical | 4 |

Table 1: Main languages considered under Telugu LGR

3.5 The Structure of Written Telugu

The Telugu script as it is used for the Telugu language consists of a total of 72 characters [102] comprising 40 consonants, 16 characters representing vowels that can stand alone and 16 dependent signs, each corresponding one of the sixteen vowels excepting /a/ అ; no explicit dependent symbol exists for that sound, instead it is inherent with the consonants in the absence of a dependent sign. Besides these, there are six additional dependent symbols, of which five always occur with the vowels, as extensions. The sixth, the halant sign ి U+0C4D, occurs with consonants. The following subsections give further details.

3.5.1 The vowels and vowel modifiers

There are fourteen vowel characters viz. అ [a], ఆ [ā], ఇ [i], ఈ [ī], ఉ [u], ఊ [ū], ఋ [r̥], ఎ [̄], ఎ [e], ఏ [ē], ఐ [ai], ఒ [o], ఓ [ō], ఔ [au], in the common inventory [103] for all the languages using Telugu script [111] specified above and two (ఋ [r̄], ౠ [ī]) to write Sanskrit loan words. For these vowels, there are corresponding fifteen marks, except for అ [a] (which is inherent). These are listed in Table 2 below. There are six modifiers for vowels: ఁ [~], ం [m̄], ః [h̄], ఌ [~] (a special symbol not common in standard Telugu writings), ఽ [∴] (the avagraha sign, commonly used to indicate doubling the vowel length and follows only long vowels), and ి [H] (the halant sign, when appended to a consonant, deducts the inherent vowel /a/ from it). The halant sign has similar characteristic as that of a secondary vowel sign in that both of them delete the inherent vowel [a] when added to consonants.

R1. Inherent vowel deletion rule: An inherent vowel of a consonant gets deleted either before a *matra* sign or before the *halant* sign.

C[ca] + M [ఁ, ం ...] | H [ి] -> C [cఁ, ం] | H [ి]
C[ca] + M [0C3E-3F, 0C40-44, 0C62-63, 0C46-48, 0C4A-4C][0C4D] ->
C[c]M [0C3E-3F, 0C40-44, 0C62-63, 0C46-48, 0C4A-4C][0C4D]

C = Consonant, ca= a consonant with an inherent 'a', M =Secondary vowel;

| No. | Independent vowels primary allographs with code points | Dependent vowels secondary allographs with code points |
|-----|--|--|
| 1. | అ U+0C05 | No explicit sign recognized or encoded |
| 2. | ఆ U+0C06 | ౌ U+0C3E |
| 3. | ఇ U+0C07 | ి U+0C3F |
| 4. | ఈ U+0C08 | ీ U+0C40 |
| 5. | ఉ U+0C09 | ు U+0C41 |
| 6. | ఊ U+0C0A | ూ U+0C42 |
| 7. | ఋ U+0C0B | ౠ U+0C43 |
| 8. | ౠ U+0C60 | ౡ U+0C44 |
| 9. | అ U+0C0F | ౢ U+0C62 |
| 10. | అ U+0C61 | ౣ U+0C63 |
| 11. | ఎ U+0C0E | ె U+0C46 |
| 12. | ఏ U+0C0F | ే U+0C47 |
| 13. | ఐ U+0C10 | ై U+0C48 |
| 14. | ఒ U+0C12 | ొ U+0C4A |
| 15. | ఓ U+0C13 | ో U+0C4B |
| 16. | ఔ U+0C14 | ౌ U+0C4C |

Table 2: Vowels and the corresponding dependent signs

| No. | Modifier signs | Code Points | Common name |
|-----|----------------|-------------|--------------------------|
| 1. | ँ | U+0C00 | Candrabindu |
| 2. | ॠ | U+0C01 | Ardhānusvāra or Arasunna |
| 3. | ॡ | U+0C02 | Pūrṇanusvāra or Sunna |
| 4. | ॢ | U+0C03 | Visarga |
| 5. | ॣ | U+0C3D | Avagraha |
| 6. | ̣ | U+0C4D | Halant |

Table 3: Vowel modifiers and the consonantal modifiers

3.5.2 The Anusvāra or sunna (◌ᳵ - U+0C02)

The Anusvāra or *sunna* represents a homorganic nasal before the corresponding consonant and as a substitute to transcribe word final /mu/. Essentially it substitutes a cluster of a Nasal Consonant + Halant before a consonant. Writing alternatively with a nasal consonant + Halant + Consonant is rare and often occur while transcribing Sanskrit words. Otherwise the writing practice with nasal consonant + Halant + Consonant of the later type is virtually absent in Telugu.

| No. | Homorganic nasal = Archiphoneme /M/ | Homorganic nasal + Halant |
|-----|-------------------------------------|---------------------------|
| 1. | లంక /laMka/ | లఙ్క /laŋka/ 'island' |
| 2. | కంచె /kaMce/ | కఞ్చె [kaŋce] 'fence' |
| 3. | పంట /paMTa/ | పఱ్ణ /pa ṅ Ta/ 'harvest' |
| 4. | కంత /kaMta/ | కఱ్ణ /kanta/ 'hole' |
| 5. | కంప /kaMpa/ | కఱ్ప /kampa/ 'thornybush' |
| 6. | కంస /kaMsa/ | కఱ్స /kansa/ 'king Kansa' |
| 7. | సింహ /siMha/ | సిఱ్హ /simha/ 'lion' |

Table 4: Homorganic nasal and Homorganic nasal + Halant

3.5.3 Nasalization: Candrabindu (◌ᳶ U+0C00) or arasunna (◌᳷ U+0C01)

Candrabindu, which denotes nasalization of the preceding vowel, is used in the Prakrit texts transcribed in the Telugu script and the *arasunna* as in old Telugu తెలుగు /telūgu/ 'telugu'. Present-day Telugu users do not use the candrabindu frequently unless to bring special emphasis as in hãã, hũũ, etc.

3.5.4 The Consonants

The Telugu consonants have an implicit vowel /a/ included in them. As per the traditional classification they are categorized according to their phonetic properties. There are 5 *varga* groups (classes) and one non-*varga* group. Each *varga* corresponds to a particular set of stops characterized by particular place of articulation. Each *varga* contains four oral stops and one nasal stop ordered by the complexity of their manner from left to right as [-vd,-asp, -nas], [-vd, +asp, -nas], [+vd, -asp, -nas], [+vd, +asp, -nas], [+vd, -asp, +nas] (where, vd = voiced, asp = aspirated, nas = nasal). Each feature set defines the character by the *varga*. Each *varga* from top to bottom are defined by an additional place feature of articulation. The non-*varga* set is again divided into two subsets, each is characterized by absence or presence of sonority, i.e. [+/- son]. The obstruents characterized by [-son] are

fricatives, viz. శ[ś], ష[ṣ], స[s], హ[h], while the remaining carry the feature of sonority i.e. [+son].

| No. | Place of Articulation | -asp -vd -nas | I S O | +asp -vd -nas | I S O | -asp +vd -nas | I S O | +asp +vd -nas | I S O | -asp +vd +nas | I S O |
|-----|-----------------------|---------------------|-------------|---------------------|-------------|---------------------|-------------|---------------------|-------------|---------------------|-------------|
| 1. | Velar | క | k | ఖ | kh | గ | g | ఘ | gh | జ | ñ |
| 2. | Palatal | చ | c | ఛ | ch | జ | j | ఝ | jh | ఞ | ñ |
| 3. | Retroflex | ట | ṭ | ఠ | ṭh | డ | ḍ | ఢ | ḍh | ణ | ṇ |
| 4. | Dental | త | t | థ | th | ద | d | ధ | dh | న | n |
| 5. | Bilabial | ప | p | ఫ | ph | బ | b | భ | bh | మ | m |

Table 5: Classification of stop consonants

| | | | | | | | | | | | | |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|
| Sonorants | య | y | ర | r | ఱ | ṛ | ల | l | ళ | ḷ | వ | v |
| Fricatives | శ | ś | ష | ṣ | స | s | హ | h | | | | |

Table 6: Non-stop consonants

4. The Development Process and Methodology

The Neo-Brahmi Generation Panel involves a number of different scripts with distinct Unicode blocks. Each of these scripts usually will have a separate LGR. However, a common thread runs through the neo-Brahmi scripts in the process of LGR development.

A number of guiding principles that are laid out will be used in the development of the scheme. As specified elsewhere, the NBBP adopts the following principles in the selection of code-points from the code-point repertoire for the Telugu language script. A principle, like the Inclusion principle, deals with whether the character is regularly used in the language, besides its unambiguous nature.

The second important principle, the exclusion principle, deals with the use of the code point repertoire for root zone and does not allow every character that is tabulated in the Unicode chart. A baseline layer of restriction is set for the Domain Name System by the protocol known as IDNA (Internationalized Domain Names in Applications). IDNA excludes some characters from the Unicode repertoire for the concerned script. An additional layer is added for the root zone, called the Maximal Starting Repertoire (MSR). Telugu does not have many such characters that are restricted. One such character for

example is, the Avagraha " ౨ " (U+0C3D), which is restricted by MSR even if allowed by the IDNA protocol.

Similarly, certain punctuation marks that were used in the traditional texts are not assigned any code points and hence not necessary to be included here. Other cases such as symbols and abbreviations are not permitted. In addition to the above, rare and obsolete characters though recognized in the Unicode chart of Telugu will not be permitted in the root zone LGR.

4.1 Zero Width Joiner and Zero Width Non-Joiner in Telugu Domain Names

MSR excludes invisible characters like Zero Width Non-Joiner (U+200C) and Zero Width Joiner (U+200D), as they require ad hoc representation in different ways. These are required in certain cases where a typical visual shape of an akshar is desired.

There are contrastive usages of written forms derived from the use of Zero Width Joiner (ZWJ) and Zero Width Non-Joiner (ZWNJ). They have special roles in the writing system of Telugu.

ZWNJ is used in sequences like Consonant (C) + Halant (U+0C4D) + Consonant, where the second C is prevented from taking the usual dependent allograph (vattu) form after (below) the first consonant, as in the following example:

1. క (U+0C15) + ీ (U+0C4D) + స (U+0C38) + ీ (U+0C4D) + వ (U+0C35) + ా (0C3E) =
కాస్వ- without using ZWNJ

Example: వాకాస్వతంత్ర్యం

2. క (U+0C15) + ీ (U+0C4D) + ZWNJ (U+200C) + స (U+0C38) + ీ (U+0C4D) + వ (U+0C35) + ా (0C3E) = క్సాస్వ- using the ZWNJ

Example: వాక్సాస్వతంత్ర్యం

Both forms of the words though written with different graphic signs may mean the same and they are also same even in their pronunciation. Though the second form was not previously common, its usage is gaining ground due to the influence of English and Hindi. It is frequently used in transcribing many English words into Telugu, such as 'software' (సాఫ్ట్వేర్, using ZWNJ). The word 'software' will become సాఫ్ట్వేర్ if ZWNJ is not used.

4.2 How to Avoid Duplicate Domain Names Involving ZWJ and ZWNJ?

ZWJ and ZWNJ are used mainly to write two distinct displays of the same consonant cluster or sequence which do not have any semantic and phonetic significance. When ZWJ and ZWNJs are allowed in domain names for Telugu, they create two distinct forms of the same domain name. To make the browsers and DNSs to treat them as equal, we

have to ignore ZWJ and ZWNJs for comparing two words. The same procedure is usually followed by the spell-checkers of the language.

Accepting ZWJ and ZWNJ in domain names creates confusion to a majority of the linguistic community and joiner characters are prohibited for the Root Zone, hence this is explicitly prohibited by the NBGP.

5. The Repertoire

In this section, we present the discussion on the code points that would form the repertoire of code points licensed by the [MSR-3] to be validated and used in the root zone label generation rules. Section 5.1 provides the section of the [MSR-3] applicable to the Telugu script on which the Telugu code point repertoire is based. Section 5.2 details the code point repertoire that the Neo-Brahmi Generation Panel proposes to be included in the Telugu LGR.

5.1 Telugu section of Maximal Starting Repertoire [MSR] Version 3

| | 0C0 | 0C1 | 0C2 | 0C3 | 0C4 | 0C5 | 0C6 | 0C7 |
|---|---------------|-----------|-----------|-----------|-----------|----------------|-----------|-----------|
| 0 | ◌̣ 0C00 | బ 0C10 | ల 0C20 | ర 0C30 | ఠ 0C40 | | ఋ 0C60 | |
| 1 | ◌̣̣ 0C01 | | ళ 0C21 | ఱ 0C31 | ౠ 0C41 | | ౡ 0C61 | |
| 2 | ◌̣̣̣ 0C02 | జ 0C12 | ఱ 0C22 | ల 0C32 | ౡ 0C42 | | ౢ 0C62 | |
| 3 | ◌̣̣̣̣ 0C03 | ఙ 0C13 | ఱ 0C23 | ళ 0C33 | ౢ 0C43 | | ౣ 0C63 | |
| 4 | | ఞ 0C14 | ఱ 0C24 | ఱ 0C34 | ౣ 0C44 | | | |
| 5 | ఠ 0C05 | ఠ 0C15 | ఠ 0C25 | ఠ 0C35 | | ◌̣̣̣̣̣ 0C55 | | |
| 6 | ఠ 0C06 | ఠ 0C16 | ఠ 0C26 | ఠ 0C36 | ఠ 0C46 | ఠ 0C56 | ఠ 0C66 | |
| 7 | ఠ 0C07 | ఠ 0C17 | ఠ 0C27 | ఠ 0C37 | ఠ 0C47 | | ఠ 0C67 | |
| 8 | ఠ 0C08 | ఠ 0C18 | ఠ 0C28 | ఠ 0C38 | ఠ 0C48 | ఠ 0C58 | ఠ 0C68 | ఠ 0C78 |
| 9 | ఠ 0C09 | ఠ 0C19 | | ఠ 0C39 | | ఠ 0C59 | ఠ 0C69 | ఠ 0C79 |
| A | ఠ 0C0A | ఠ 0C1A | ఠ 0C2A | | ఠ 0C4A | | ఠ 0C6A | ఠ 0C7A |
| B | ఠ 0C0B | ఠ 0C1B | ఠ 0C2B | | ఠ 0C4B | | ఠ 0C6B | ఠ 0C7B |
| C | ఠ 0C0C | ఠ 0C1C | ఠ 0C2C | | ఠ 0C4C | | ఠ 0C6C | ఠ 0C7C |
| D | | ఠ 0C1D | ఠ 0C2D | ఠ 0C3D | ఠ 0C4D | | ఠ 0C6D | ఠ 0C7D |
| E | ఠ 0C0E | ఠ 0C1E | ఠ 0C2E | ఠ 0C3E | | | ఠ 0C6E | ఠ 0C7E |
| F | ఠ 0C0F | ఠ 0C1F | ఠ 0C2F | ఠ 0C3F | | | ఠ 0C6F | ఠ 0C7F |

Color convention¹:

All characters that are included in the [MSR] are highlighted in Yellow background

PVALID in IDNA2008 but excluded from the [MSR] are highlighted in Pinkish background

Not PVALID in IDNA2008 are in White background

Figure 2: Telugu Code Page from [MSR-3]

¹This document needs to be printed in color for this to be read correctly.

5.2 Code Points Repertoire

In the following, the Telugu Script Unicode Code points have been presented and discussed with reference to the Principles that constrain the label generation rules. It is important to note that the purpose of this document is to state unambiguously the Telugu code points that can be used in the root zone repertoire.

The following table lists 63 code points for the Telugu LGR, out of a total number of 67 code points listed in MSR-3, excluding four code points which are obsolete.

| No. | Unicode Code Point | Glyph | Character Name | EGIDS status | Indic Syllabic Category | Reference |
|-----|--------------------|-------|-------------------------|---|-------------------------|-----------|
| 1. | 0C02 | ◌ం | TELUGU SIGN ANUSVARA | 2 Tel 4 San 5 Others ² | ANUSVĀRA | 102, 103 |
| 2. | 0C03 | ◌ః | TELUGU SIGN VISARGA | 2 Tel 4 San 5 Others | VISARGA | 102, 103 |
| 3. | 0C05 | అ | TELUGU LETTER A | 2 Tel 5 Others | Vowel | 102, 103 |
| 4. | 0C06 | ఆ | TELUGU LETTER AA | 2 Tel 5 Others | Vowel | 102, 103 |
| 5. | 0C07 | ఇ | TELUGU LETTER I | 2 Tel 5 Others | Vowel | 102, 103 |
| 6. | 0C08 | ఈ | TELUGU LETTER II | 2 Tel 5 Others | Vowel | 102, 103 |
| 7. | 0C09 | ఉ | TELUGU LETTER U | 2 Tel 5 Others | Vowel | 102, 103 |
| 8. | 0C0A | ఊ | TELUGU LETTER UU | 2 Tel 5 Others | Vowel | 102, 103 |
| 9. | 0C0B | ఋ | TELUGU LETTER VOCALIC R | 2 Tel 5 Others | Vowel | 102, 103 |
| 10 | 0C0E | ఎ | TELUGU LETTER E | 2 Tel 5 Others | Vowel | 102, 103 |
| 11. | 0C0F | ఏ | TELUGU LETTER EE | 2 Tel 5 Others | Vowel | 102, 103 |

² Others are the EGIDS 5 languages, listed in Table 1: Main languages considered under Telugu LGR

| No. | Unicode Code Point | Glyph | Character Name | EGIDS status | Indic Syllabic Category | Reference |
|-----|--------------------|-------|-------------------|-------------------|-----------------------------------|-----------|
| 12. | 0C10 | ఐ | TELUGU LETTER AI | 2 Tel 5 Others | Vowel | 102, 103 |
| 13. | 0C12 | ఒ | TELUGU LETTER O | 2 Tel 5 Others | Vowel | 102, 103 |
| 14. | 0C13 | ఓ | TELUGU LETTER OO | 2 Tel 5 Others | Vowel | 102, 103 |
| 15. | 0C14 | ఔ | TELUGU LETTER AU | 2 Tel 5 Others | Vowel | 102, 103 |
| 16. | 0C15 | క | TELUGU LETTER KA | 2 Tel 5 Others | Consonant | 102, 103 |
| 17. | 0C16 | ఖ | TELUGU LETTER KHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 18. | 0C17 | గ | TELUGU LETTER GA | 2 Tel 5 Others | Consonant | 102, 103 |
| 19. | 0C18 | ఘ | TELUGU LETTER GHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 20. | 0C19 | ఙ | TELUGU LETTER NGA | 2 Tel 5 Others | Consonant, Nasal- Consonant | 102, 103 |
| 21. | 0C1A | చ | TELUGU LETTER CA | 2 Tel 5 Others | Consonant | 102, 103 |
| 22. | 0C1B | ఛ | TELUGU LETTER CHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 23. | 0C1C | జ | TELUGU LETTER JA | 2 Tel 5 Others | Consonant | 102, 103 |
| 24. | 0C1D | ఝ | TELUGU LETTER JHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 25. | 0C1E | ఞ | TELUGU LETTER NYA | 2 Tel 5 Others | Consonant, Nasal- Consonant | 102, 103 |
| 26. | 0C1F | ట | TELUGU LETTER TTA | 2 Tel 5 Others | Consonant | 102, 103 |

| No. | Unicode Code Point | Glyph | Character Name | EGIDS status | Indic Syllabic Category | Reference |
|-----|--------------------|-------|--------------------|-------------------|-----------------------------------|-----------|
| 27. | 0C20 | ఠ | TELUGU LETTER TTHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 28. | 0C21 | డ | TELUGU LETTER DDA | 2 Tel 5 Others | Consonant | 102, 103 |
| 29. | 0C22 | ఢ | TELUGU LETTER DDHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 30. | 0C23 | ణ | TELUGU LETTER NNA | 2 Tel 5 Others | Consonant, Nasal- Consonant | 102, 103 |
| 31. | 0C24 | త | TELUGU LETTER TA | 2 Tel 5 Others | Consonant | 102, 103 |
| 32. | 0C25 | థ | TELUGU LETTER THA | 2 Tel 5 Others | Consonant | 102, 103 |
| 33. | 0C26 | ద | TELUGU LETTER DA | 2 Tel 5 Others | Consonant | 102, 103 |
| 34. | 0C27 | ధ | TELUGU LETTER DHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 35. | 0C28 | న | TELUGU LETTER NA | 2 Tel 5 Others | Consonant, Nasal- Consonant | 102, 103 |
| 36. | 0C2A | ప | TELUGU LETTER PA | 2 Tel 5 Others | Consonant | 102, 103 |
| 37. | 0C2B | ఫ | TELUGU LETTER PHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 38. | 0C2C | బ | TELUGU LETTER BA | 2 Tel 5 Others | Consonant | 102, 103 |
| 39. | 0C2D | భ | TELUGU LETTER BHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 40. | 0C2E | మ | TELUGU LETTER MA | 2 Tel 5 Others | Consonant, Nasal- Consonant | 102, 103 |
| 41. | 0C2F | య | TELUGU LETTER YA | 2 Tel 5 Others | Consonant | 102, 103 |

| No. | Unicode Code Point | Glyph | Character Name | EGIDS status | Indic Syllabic Category | Reference |
|-----|--------------------|-------|------------------------------|-------------------|-------------------------|-----------|
| 42. | 0C30 | ర | TELUGU LETTER RA | 2 Tel 5 Others | Consonant | 102, 103 |
| 43. | 0C32 | ల | TELUGU LETTER LA | 2 Tel 5 Others | Consonant | 102, 103 |
| 44. | 0C33 | ళ | TELUGU LETTER LLA | 2 Tel 5 Others | Consonant | 102, 103 |
| 45. | 0C35 | వ | TELUGU LETTER VA | 2 Tel 5 Others | Consonant | 102, 103 |
| 46. | 0C36 | శ | TELUGU LETTER SHA | 2 Tel 5 Others | Consonant | 102, 103 |
| 47. | 0C37 | ష | TELUGU LETTER SSA | 2 Tel 5 Others | Consonant | 102, 103 |
| 48. | 0C38 | స | TELUGU LETTER SA | 2 Tel 5 Others | Consonant | 102, 103 |
| 49. | 0C39 | హ | TELUGU LETTER HA | 2 Tel 5 Others | Consonant | 102, 103 |
| 50. | 0C3E | ా | TELUGU VOWEL SIGN AA | 2 Tel 5 Others | Matra | 102, 103 |
| 51. | 0C3F | ి | TELUGU VOWEL SIGN I | 2 Tel 5 Others | Matra | 102, 103 |
| 52. | 0C40 | ీ | TELUGU VOWEL SIGN II | 2 Tel 5 Others | Matra | 102, 103 |
| 53. | 0C41 | ు | TELUGU VOWEL SIGN U | 2 Tel 5 Others | Matra | 102, 103 |
| 54. | 0C42 | ూ | TELUGU VOWEL SIGN UU | 2 Tel 5 Others | Matra | 102, 103 |
| 55. | 0C43 | ృ | TELUGU VOWEL SIGN VOCALIC R | 2 Tel 5 Others | Matra | 102, 103 |
| 56. | 0C44 | ౄ | TELUGU VOWEL SIGN VOCALIC RR | 2 Tel 5 Others | Matra | 102, 103 |
| 57. | 0C46 | ె | TELUGU VOWEL SIGN E | 2 Tel 5 Others | Matra | 102, 103 |

| No. | Unicode Code Point | Glyph | Character Name | EGIDS status | Indic Syllabic Category | Reference |
|-----|--------------------|-------|----------------------|-------------------|-------------------------|-----------|
| 58. | 0C47 | ః | TELUGU VOWEL SIGN EE | 2 Tel 5 Others | Matra | 102, 103 |
| 59. | 0C48 | ఌ | TELUGU VOWEL SIGN AI | 2 Tel 5 Others | Matra | 102, 103 |
| 60. | 0C4A | ఋ | TELUGU VOWEL SIGN O | 2 Tel 5 Others | Matra | 102, 103 |
| 61. | 0C4B | ౠ | TELUGU VOWEL SIGN OO | 2 Tel 5 Others | Matra | 102, 103 |
| 62. | 0C4C | ౡ | TELUGU VOWEL SIGN AU | 2 Tel 5 Others | Matra | 102, 103 |
| 63. | 0C4D | ౢ | TELUGU SIGN VIRAMA | 2 Tel 5 Others | Matra | 102, 103 |

Table 7: Included code points

5.3 Code Points Not Included

Referring to the principle in section 4, the code points to be excluded from the repertoire are the following, for the reasons listed.

The following code points are not in widespread use.

- 0C00 ీ TELUGU LETTER CANDRABINDU
- 0C01 ఁ TELUGU LETTER ARASUNNA
- 0C0C ఱ TELUGU LETTER VOCALIC L
- 0C31 ఱ TELUGU LETTER RRA

Various signs: Allographs of vowel diacritics /a:/ and part of a diacritic specific to particular consonant /h/.

- 0C55 ా TELUGU LENGTH MARK
- 0C56 ా̄ TELUGU AI LENGTH MARK

Historic phonetic variants: Phonological variants shall not be permitted. They are not in MSR-3.

- 0C58 ఱ TELUGU LETTER TSA

- 0C59 ఙ TELUGU LETTER DZA

The two additional vowels listed below to transcribe Sanskrit are not permitted. They are not in MSR-3.

- 0C60 ఋ TELUGU LETTER VOCALIC RR
- 0C61 ౠ TELUGU LETTER VOCALIC LL

The following two dependent vowels used to transcribe Sanskrit sounds are not permitted. They are not in MSR-3.

- 0C62 ౡ TELUGU VOWEL SIGN VOCALIC L
- 0C63 ౢ TELUGU VOWEL SIGN VOCALIC LL

Starting from the MSR-3, There are four code points to be excluded.

| No. | Unicode Code Point | Glyph | Character Name | EGIDS status | Indic Syllabic Category | Reference | Note |
|-----|--------------------|-------|-------------------------|-----------------------------|-------------------------|---------------|--|
| 1. | 0C0C | అ | TELUGU LETTER VOCALIC L | 2 Telu 5 Gon 6b other | Vowel | 103, 108, 109 | It is not used in modern Telugu |
| 2. | 0C31 | ఱ | TELUGU LETTER RRA | 2 Telu 5 Gon 6b other | Consonant | 103, 108, 109 | It is not used in modern Telugu |
| 3. | 0C55 | ం | TELUGU LENGTH MARK | 2 Telu 5 Gon 6b other | Matra | 103, 108, 109 | It is not available on general keyboard. |
| 4. | 0C56 | ి | TELUGU AI LENGTH MARK | 2 Telu 5 Gon 6b other | Matra | 103, 108, 109 | It is not used in modern Telugu |

Table 8: Excluded code points

6. Variants

Telugu code points representing the basic simple stand-alone characters and some dependent characters may enter into different combinations to form syllables. There are no characters in the Telugu Unicode chart that - either in simple form or in combined form

are deemed similar by NBGP. However, Telugu has a small number of variants that have identical values but derive from different character combinations. The NBGP categorizes these confusingly similar variants in two groups.

6.1 Type 1: Similarity within the Script

Certain vowels [o, ō] display different shapes in combination with certain consonants, though they have shared sound and code point values. For example:

- i. Ca + e + u(:) -> mo(:)
- ii. Ca + o(:) -> ko(:)

The variants, which are often confusing and of variable acceptance are due to the display of their rendering differently due to the identical code points.

These cases are interesting in that they present no similarity in their forms but have similar phonetic output. It is not unusual to find such regional variations and they are regularly used by Telugu users. These may not cause confusion but become annoying to learners.

However, ీ + ు (U+0C46 + U+0C41) is matra + matra sequence, which is not allowed in the WLE rules in section 7. Therefore, these are not defined as variant sequences by NBGP.

| Class | Character seq. [Ca+e+u] | -> Co <- | | Ca+o |
|-------|--|---------------|---------------|---------------------|
| 1 | [క + ీ + ు] -> 0C15+0C46+0C41 (This class includes other consonants like, kha, ga, nga, ca, cha, ja, nya, ta, tha, da, dha, na, ta, tha, da, dha, na, pa, pha, ba, bha, ra, la, va, Sa, sha, sa, and ha) | కౌ Blocked | కొ | క+ౌ 0C15+0C4A |
| 2 | [మ + ీ + ు] -> 0C2E+0C46+0C41 | మౌ | మొ Blocked | మ+ ౌ 0C2E +0C4A |
| | [య + ీ + ు] -> 0C2F+0C46+0C41 | యౌ | యొ Blocked | య+ ౌ 0C2F +0C4A |
| | [రు + ీ + ు] -> 0C1D+0C46+0C41 | రౌ | రొ Blocked | రు+ ౌ 0C1D +0C4A |

| | | | | |
|--|-------------------------------------|----|------------|--------------------|
| | [ము + ీ + ు] -> 0C18+0C46+0C41 | ము | ము Blocked | ము+ ీ 0C18+0C4A |
|--|-------------------------------------|----|------------|--------------------|

Table 9a: Similarity within the script

6.2 Type 1: Variants within Script due to Alternative Spelling

Similar to the above, there are a set of representations in Telugu syllable formations where a homorganic nasal (anusvāra) in a syllable has alternate spelling which is represented visually different, as shown below.

| No. | Homorganic nasal (anusvāra)+ consonant | Homorganic nasal consonant + halant + consonant |
|-----|--|---|
| 1. | లంక /laMka/ | లఙ్క /lan̄ka/ 'island' |
| 2. | కంచె /kaMce/ | కఞ్చె [kaŋce] 'fence' |
| 3. | పంట /paMTa/ | పఱ్ణ /paNTa/ 'harvest' |
| 4. | కంత /kaMta/ | కఱ్ణ /kanta/ 'hole' |
| 5. | కంప /kaMpa/ | కఱ్ప /kampa/ 'thornybush' |
| 6. | కంస /kaMsa/ | కఱ్స /kansa/ 'king Kansa' |
| 7. | సింహ /siMha/ | సిఱ్హ /simha/ 'lion' |

Table 9b: Variants with anusvāra alternating with nasal consonants

Writing alternatively with a nasal consonant + halant + consonant is rare in Telugu and often occur while transcribing Sanskrit words. Since the variants have exactly the same pronunciation, the rarer representation of nasal consonant + halant + consonant is disallowed in order to avoid the source of confusion.

Nasal Consonants are:

1. U+0C19 TELUGU LETTER NGA (ఙ)
2. U+0C1E TELUGU LETTER NYA (ఞ)
3. U+0C23 TELUGU LETTER NNA (ణ)
4. U+0C28 TELUGU LETTER NA (న)
5. U+0C2E TELUGU LETTER MA (మ)

Similarly and very frequently, the word final ము [mu] is often represented alternatively by the variant anusvāra ు [M] as in the following:

కలం kalaM

కలము kalamu

'pen'

| | | |
|------------------|--------------------|--------------|
| పుస్తకం pustakaM | పుస్తకము pustakamu | 'book' |
| ఆముదం a:mudaM | ఆముదము a:mudamu | 'castor oil' |
| దేశం deSaM | దేశము deSamu | 'country' |

In such cases, one of the confusable variants must be disallowed. This can be disallowed by the WLE rule: H cannot follow a nasal consonant.

6.3 Type 2: Shared Similarity with the Other Related Scripts.

There are many Brahmi derived scripts particularly in the Southern part of India, Sri Lanka, and South East Asia. Some of the characters of these scripts display similarity with each other. Such cases, relevant for Telugu script, are given below.

6.3.1 Type2: Cross-Script Variants for Telugu and Kannada

A number of characters of the Kannada script are almost similar to characters of Telugu script, except for the flattened head-stroke in Kannada contrasting with a tick mark on the top of the character in Telugu. Out of the total, there are 34 such cases which are categorized as variant sets, as shown in the following table.

| Variant Set | Telugu Code Point | Kannada Code Point |
|-------------|-------------------|--------------------|
| 1 | ం (0C02) | ం (0C82) |
| 2 | ః (0C03) | ః (0C83) |
| 3 | అ (0C05) | అ (0C85) |
| 4 | ఆ (0C06) | ఆ (0C86) |
| 5 | ఇ (0C07) | ఇ (0C87) |
| 6 | ఈ (0C08) | ఈ (0C88) |
| 7 | ఐ (0C10) | ఐ (0C90) |
| 8 | ఒ (0C12) | ఒ (0C92) |
| 9 | ఓ (0C13) | ఓ (0C93) |
| 10 | ఔ (0C14) | ఔ (0C94) |
| 11 | ఖ (0C16) | ఖ (0C96) |
| 12 | గ (0C17) | గ (0C97) |
| 13 | జ (0C1C) | జ (0C9C) |
| 14 | ఝ (0C1D) | ఝ (0C9D) |

| Variant Set | Telugu Code Point | Kannada Code Point |
|-------------|-------------------|--------------------|
| 15 | ಇ (0C1E) | ಇ (0C9E) |
| 16 | ಉ (0C1F) | ಉ (0C9F) |
| 17 | ಠ (0C20) | ಠ (0CA0) |
| 18 | ಠ (0C21) | ಠ (0CA1) |
| 19 | ಠ (0C22) | ಠ (0CA2) |
| 20 | ಠ (0C23) | ಠ (0CA3) |
| 21 | ಠ (0C25) | ಠ (0CA5) |
| 22 | ಠ (0C26) | ಠ (0CA6) |
| 23 | ಠ (0C27) | ಠ (0CA7) |
| 24 | ಠ (0C28) | ಠ (0CA8) |
| 25 | ಬ (0C2C) | ಬ (0CAC) |
| 26 | ಭ (0C2D) | ಭ (0CAD) |
| 27 | ಮ (0C2E) | ಮ (0CAE) |
| 28 | ಯ (0C2F) | ಯ (0CAF) |
| 29 | ರ (0C30) | ರ (0CB0) |
| 30 | ಲ (0C32) | ಲ (0CB2) |
| 31 | ಳ (0C33) | ಳ (0CB3) |
| 32 | ಠ (0C3F) | ಠ (0CBF) |
| 33 | ಠ (0C41) | ಠ (0CC1) |
| 34 | ಠ (0C43) | ಠ (0CC3) |

Table 10: Cross-script variant code points for Telugu and Kannada

The Telugu and Kannada variant sets in Table 10 are cross-script variant code points. The details of various *akshar* combinations and variant disposition can be found in section 6.4

Code points which have been analyzed and found to be similar, but not considered as variants, are listed in Appendix A.

6.3.2 Type2: Cross-Script Variants for Telugu and Devanagari

Visarga is the only identical code point that exhibits shape similarity between the Telugu and Devanagari scripts. However, as there are no other variant code points between the two languages, it is not defined as a variant code point.

| Devanagari Code Point | Telugu Code Point |
|-----------------------|-------------------|
| ◌ः (0903) | ◌ః (0C03) |

Table 11: Candidate cross-script variant code point for Telugu and Devanagari

6.3.3 Type2: Cross-Script Variants for Telugu and Gujarati

Visarga is the only identical code point that exhibits shape similarity between the Telugu and Gujarati scripts. However, as there are no other identical code points between the two languages, it is not defined as a variant code point.

| Gujarati Code Point | Telugu Code Point |
|---------------------|-------------------|
| ◌ઃ (0A83) | ◌ః (0C03) |

Table 12: Candidate cross-script variant code point for Telugu and Gujarati

6.3.4 Type2 Cross-Script Variants for Telugu and Oriya

The following code points exhibit similarity between the Telugu and Oriya scripts.

| Telugu Code Point | Oriya Code Point |
|---------------------------|--------------------------|
| ◌◌ (0C02) ANUSVĀRA | ◌ (0B20) LETTER TTHA |
| ◌ః (0C03) SIGN VISARGA | ◌ (0B03) SIGN VISARGA |
| ◌ఠ (0C30) LETTER RA | ◌ (0B20) LETTER TTHA |

Table 13: Candidate cross-script variant code points for Telugu and Oriya

The first two (U+0C02 – U+0B20 and U+0C03 – U+0B03) are dependent signs and U+0C30 is a stand-alone character in Telugu. NBGP discussions concluded that there is no need to recognize the cross-script variant code points between the Oriya and the Telugu scripts. This is because U+0C30 and U+0B20 are distinguishable and there are not enough other variant code points in each script to form labels that look the same. Therefore, these are not defined as variant code points.

6.3.5 Type2: Cross-Script Variants for Telugu and Malayalam

The two code points, viz. the anusvāra and the visarga are the only identical signs between the Telugu and Malayalam scripts. However, as there are not enough other

variant code points to form labels, they are not defined as variant code points between the two languages.

| Telugu Code Point | Malayalam Code Point |
|-------------------|----------------------|
| ◌o (0C02) | ◌o (0D02) |
| ◌ọ (0C03) | ◌ọ (0D03) |

Table 14: Candidate cross-script variant code points for Telugu and Malayalam

6.3.6 Type2: Cross-Script Variants for Telugu and Sinhala

The following three pairs of characters represented by the corresponding code points between the Telugu and Sinhala which may be considered as having only similarity if the similarity between 0C30 and 0DBB is not sustainable. However NBGP, in consultation with Sinhala, concludes that 0C30 and 0DBB could cause confusion from the script user point of view. Therefore, they are proposed as cross script variants between the two scripts and the disposition is blocked.” This analysis follows the NBGP Cross-script Variant inclusion policy available in Appendix C.

| Telugu Code Point | Sinhala Code Point |
|-------------------|--------------------|
| ◌o (0C02) | ◌o (0D82) |
| ◌ọ (0C03) | ◌ọ (0D83) |
| ఠ (0C30) | ච (0DBB) |

Table 15: Cross-script variant code points for Telugu and Sinhala

6.4 Cross Script Variants of Various Akshar Combinations

6.4.1 Conjunct Consonant Combinations

Cross script variants of various Akshar combinations (consonant-consonant-dependent characters) common between the Telugu and Kannada scripts include the following:

| Variant Set | Telugu Code Point | Kannada Code Point |
|-------------|-------------------|--------------------|
| 1 | ◌o (0C02) | ◌o (0C82) |
| 2 | ◌ọ (0C03) | ◌ọ (0C83) |
| 3 | ఋ (0C16) | ఋ (0C96) |
| 4 | ౠ (0C17) | ౠ (0C97) |
| 5 | ఌ (0C1C) | ఌ (0C9C) |
| 6 | ౡ (0C1D) | ౡ (0C9D) |

| Variant Set | Telugu Code Point | Kannada Code Point |
|-------------|-------------------|--------------------|
| 7 | ಇ (0C1E) | ಇ (0C9E) |
| 8 | ಉ (0C1F) | ಉ (0C9F) |
| 9 | ಠ (0C20) | ಠ (0CA0) |
| 10 | ಡ (0C21) | ಡ (0CA1) |
| 11 | ಢ (0C22) | ಢ (0CA2) |
| 12 | ಣ (0C23) | ಣ (0CA3) |
| 13 | ಢ (0C25) | ಢ (0CA5) |
| 14 | ದ (0C26) | ದ (0CA6) |
| 15 | ಢ (0C27) | ಢ (0CA7) |
| 16 | ನ (0C28) | ನ (0CA8) |
| 17 | ಬ (0C2C) | ಬ (0CAC) |
| 18 | ಭ (0C2D) | ಭ (0CAD) |
| 19 | ಮ (0C2E) | ಮ (0CAE) |
| 20 | ಯ (0C2F) | ಯ (0CAF) |
| 21 | ರ (0C30) | ರ (0CB0) |
| 22 | ಲ (0C32) | ಲ (0CB2) |
| 23 | ಳ (0C33) | ಳ (0CB3) |
| 24 | ಿ (0C3F) | ಿ (0CBF) |
| 25 | ಱ (0C41) | ಱ (0CC1) |
| 26 | ೈ (0C43) | ೈ (0CC3) |

Table 16: Cross-script variants between Telugu and Kannada for conjunct consonant combination analysis

Table 16 includes 26 distinct Telugu code points that occur in the formation of conjunct consonant combinations in Telugu and Kannada. Excluding the stand alone vowels from the total common *Akshar* combinations of cross script variants, there are a set of 21 consonants (C), three vowel matras (M) and two vowel modifiers that enter into the formation of the following combinations:

| Sl. No. | Akshar combinations | Number |
|---------|---------------------|-----------------|
| 1. | CM | = 21*3=63 |
| 2. | CB | = 21*1=21 |
| 3. | CX | = 21*1=21 |
| 4. | CHCM | =21*21*3=1323 |
| 5. | CHCB | =21*21*1=441 |
| 6. | CHCX | =21*21*1=441 |
| 7. | CHCMB | =21*21*3*1=1323 |
| 8. | CHCMX | =21*21*3*1=1323 |
| 9. | All combinations: | =4956 |

Table-17 total number of Akshar combinations

There occurs a total of 4956 conjunct consonant combinations modified by matras and vowel modifiers that are identical and can be labeled for variant labels between Telugu and Kannada scripts. These combinations are covered by the variant code points in Section 6, Table 10 and Table 15.

6.4.2 Other Combinations

NBGP creates the possible combinations of Telugu code points and cross check with other Neo-Brahmi scripts for candidate variants. The possible combinations are:

1. CHCMB, CHCMX
2. CHCM, CHCB, CHCX
3. VB, VX, V
4. CHC, CM, CB, CX, C

Where,

- | | | |
|---|---|------------------|
| C | → | Consonant |
| M | → | Matra |
| V | → | Vowel |
| B | → | Anusvāra (Bindu) |
| X | → | Visarga |
| H | → | Halant / Virama |

NBGP concludes that beside those identical code points defined as variants in Section 6, Table 10 and Table 15, there are no other variant code points between Telugu combinations and other scripts code points or code point combinations.

6.5 Variant disposition

As variants mentioned in Section 6, Table 10 and Table 15 can result in whole label variants, they may be considered for "blocked" disposition. There is no preference among these variants. Whichever label containing either of these variants is chosen earlier, the other equivalent variant label should be blocked.

7. Whole Label Evaluation Rules (WLE)

In this section we provide the WLEs that are required by the language. A number of rules have been formulated so that they can be adopted for LGR specification. Below are the symbols used in the WLE rules, for each of the "Indic Syllabic Category" as mentioned in the Table 7: Code point repertoire and the details of syllable formation, see Appendix B.

| | | |
|---------|---|------------------|
| C | → | Consonant |
| M | → | Matra |
| V | → | Vowel |
| B | → | Anusvāra (Bindu) |
| X | → | Visarga |
| H | → | Halant / Virama |
| Nasal-C | → | Nasal Consonant |

- Rule 1. H must be preceded by C (Ref. Appendix B: Syllable formation Rule 4)
- Rule 2. M must be preceded by C (Ref. Appendix B: Syllable formation Rule 6)
- Rule 3. X must be preceded by V or M or C (Ref. Appendix B: syllable formation rule 3c, 5c and 7c)
- Rule 4. B must be preceded by V or M or C (Ref. Appendix B: syllable formation rule 3b, 5b and 7b)
- Rule 5. H cannot follow Nasal-C (Ref. Section 6.2 Type 1)
- Rule 6. V cannot be preceded by H

For Rule 6, there could be cases involving multi-word domains where V may need to be allowed to follow an H. This is the case where two different words are joined together but first of which ends with a Halant and the second word begins with a Vowel. Some sections of the linguistic usage require the explicit presence of H for full representation of the sound intended. However, by and large, the form of the first word without the H is considered enough for full representation of the sound intended as in the following examples:

Example:

'house of knowledge': दार् अल्लुलाम् da:rH alHulu:mH / दार् अलुलाम् da:rH alulu:mH

'The Qor'an': खुरहान् KhurHa:nH / खुरान् Khura:nH

'in Telangana Rashtra Samiti': तीअरहसलो तीअरहसलो ti:a:rHesHlo / तीअरहसलो ti:a:rHesHlo

'Y.S. R. C. party': व्वायेशसिपि व्वायेशसिपि vaiesHa:rHsi:pi / व्वायेशसिपि vaiesHa:rHsi:pi

'British India': ब्रिटीशमिडिया ब्रिटीशमिडिया bHritiShHiMdiya / ब्रिटीशमिडिया britiShiMdiya

The representations where there are cases with V preceded by H against where V is not preceded by H, the latter is awkward and the former is in demand in modern usage.

This is a unique situation necessitated by the lack of hyphen, space or the Zero Width Non-joiner character in the permissible set of characters in the Root zone repertoire. Otherwise, V is never required to be allowed to follow an H. However, permitting this

may create a perceptually dissimilar but phonetically and semantically similarity between the two labels (with and without H) for majority of the linguistic community, hence this is explicitly prohibited by the NBGP.

8. Contributors

Gangadhar Panday
Uma Maheshwara Rao, G.
NBGP members

9. References

- [MSR-3] Integration Panel, "Maximal Starting Repertoire — MSR-3 Overview and Rationale", 28 March 2018
<https://www.icann.org/sites/default/files/packages/lgr/msr/msr-3-wle-rules-28mar18-en.html>
- [101] Disanayaka, J.B. 2017. Encyclopedia of Sinhala Language and Culture. Colombo: Sumitha Publishers. First edition 2012.
- [102] Krishnamurti, Bhadriraju, Ed., 2000. *Telugu bhaashaa charitra*. Hyderabad: P.S. Telugu University. First edition 1974.
- [103] Krishnamurti, Bhadriraju and J P L Gwynn. 1985. *A Grammar of Modern Telugu*. New Delhi: Oxford University Press. *ISBN 978-0-19-561664-4*. Delhi.
- [104] Sarma, I. K. 1980. *Coinage of Satavahana Empire*. Delhi : Agam Kala Prakashan.
- [105] Sridhar, S.N. 1980. *Kannada*. New York: Routledge.
- [106] Suresh, Kolichala. 2012. Proposal to encode Telugu LLLA, Telugu ళ: <http://eemaata.com/unicode-proposal/telugu-llla-proposal.pdf>. Accessed on 9 July 2018.
- [107] Suresh, Kolichala. 2012. Divergent developments of alveolar stop *ʈ in Telugu [http://kolichala.com/dravidian/Divergent developments of alveolar stop in Telugu.pdf](http://kolichala.com/dravidian/Divergent%20developments%20of%20alveolar%20stop%20in%20Telugu.pdf). Accessed on 9 July 2018.
- [108] Telugu Unicode Chart, Telugu Range: 0C00–0C7F. The Unicode Standard, Version 10.0. <http://www.unicode.org/Public/10.0.0/charts>. Accessed on 9 July 2018.
- [109] Uma Maheshwara Rao, G. 2012. *Telugu bhaasha-saMgaNanaM*. Hyderabad: P.S. Telugu University. ISBN: 81-86073-372-9.
- [110] Uma Maheshwara Rao, G. 2003. *Standard Telugu Written Language*. VIDYULLIPI-4. pp. 1-14. Hyderabad: SCIL.
- [111] Usha Devi, A. and Chandra Sekhara Reddy. D. 2015. *Peoples Linguistic Survey of India. Andhra Pradesh and Telangana rAshtraala bhaashalu*, vol.3, part 1. ISBN: 978-93-85231-05-6. Hyderabad: emesco.

Appendix A: Confusable Code Points Analysis

A-1. Telugu and Kannada

The following table defines Telugu and Kannada code points which are confusable.

| No. | Telugu | | Kannada | |
|-----|--------|-------|---------|-------|
| | CP | Glyph | CP | Glyph |
| 1 | 0C35 | వ | 0CB5 | ವ |
| 2 | 0C36 | శ | 0CB6 | ಶ |
| 3 | 0C38 | స | 0CB8 | ಸ |

Table A-1: Confusable code points of Telugu and Kannada script

The following table lists other code points which have been analyzed and concluded that they are distinguishable.

| No. | Telugu | | Kannada | | NBGP resolution |
|-----|--------|-------|---------|-------|-----------------|
| | CP | Glyph | CP | Glyph | |
| 1 | 0C0E | ఎ | 0C8E | ಎ | distinguishable |
| 2 | 0C18 | ఘ | 0C98 | ಘ | distinguishable |
| 3 | 0C19 | ఙ | 0C99 | ಙ | distinguishable |
| 4 | 0C1A | చ | 0C9A | ಚ | distinguishable |
| 5 | 0C1B | ఛ | 0C9B | ಛ | distinguishable |
| 6 | 0C2A | ప | 0CAA | ಪ | distinguishable |
| 7 | 0C2B | ఫ | 0CAB | ಫ | distinguishable |
| 8 | 0C37 | ష | 0CB7 | ಷ | distinguishable |
| 9 | 0C4C | ౌ | 0CCC | ೌ | distinguishable |

Table A-2: Other NBGP resolutions on Telugu and Kannada script

A-2. Telugu and Malayalam

Beside those identical code points defined as variants in Section 6, there are no other similar code points between Telugu and Malayalam.

A-3. Telugu and Sinhala

Beside those identical code points defined as variants in Section 6, there are no other similar code points between Telugu and Sinhala.

Appendix B: Syllable formation in the Telugu Script

The Telugu script grammar allows us to state the nature and structure of the graphic syllables in the formation of words. The extended notion of syllable is often used to characterize orthographies of South-Asian scripts especially Brahmi derived scripts where words are composed of sequences of one or more orthographic *aksharas* or syllables. These aksharas are again composed of sequences of certain characters from the alphabet. The Telugu alphabet has the following types of characters (encoded into the Unicode) that either on their own or by entering larger combinations form *aksharas* as shown here. There are 12 different types of syllables possible in Telugu:

The following Variables are involved in the formation of syllable [\$]:

- C = Consonants, that are standalone characters or graphemes with an inherent vowel `a' can function as syllables;
Stops: క ఖ గ ఘ జ చ ఛ జ ఝ ఞ ట ఠ డ ఢ ణ త థ ద ధ న ప ఫ బ భ మ;
Fricatives: శ ష స హ
Sonorants: య ర ల ళ వ
- V = Vowels, that stand alone and represented by the graphic signs of the following may function as syllables;
అ ఆ ఇ ఈ ఉ ఊ ఎ ఏ ఐ ఒ ఓ ఔ ఋ
- M = Matras or the dependent vowel signs when occur with a consonant may function as syllables (characteristically delete the inherent vowel of the consonant);
Example. కా కి కీ కు కూ కె కే కై కొ కో కౌ; etc.
- H = Halant or virama = ీ; It may occur with one of the consonants represented by C to form CH syllables;
Example. క్ క్ ఖ్ గ్ ఘ్ జ్ చ్ ఛ్ జ్ ఝ్ ఞ్ ట్ ఠ్ డ్ ఢ్ ణ్ త్ థ్ ద్ ధ్ న్ ప్ ఫ్ బ్ భ్ మ్
- B= Pūrṇānusvāra, the homorganic nasal and an Archiphoneme = ం, may occur with one of the C, V, and the combined CM to form CB, CMB, VB, and C([HC]*)B
- X= visarga or the glottal check= ె, may occur with one of the C, V, and the combined CM to form CX, CMX, VX

The operators used: The following four operators are employed to define the delimitation of the graphic syllables in Telugu.

| No. | Symbol | Function; |
|-----|--------|-----------------------------|
| 1. | | Alternative; |
| 2. | [] | encloses optional elements; |
| 3. | * | Variable occurrence; |
| 4. | () | The sequence cluster; |

Table B-1 symbols and functions

An Akshara in Telugu can be defined as any C or V and a combination of M (dependent vowels), and the vowel modifiers as in the following:

The following syllable formation rules derive all possible graphic syllables in Telugu.

1. The syllable formation rule-1, a \$= V;

Every standalone vowel character can function as a syllable, Ex.

అ, ఆ, ఇ, ఈ, ఊ, ఊ, ఎ, ఏ, ఐ, ఒ, ఓ, ఔ, ఔ, ఋ;

After the exclusion of obsolete vowels 13 syllables are possible.

2. The syllable formation rule-2, a \$= C;

Every standalone consonant character can function as a syllable, Ex.

క ఖ గ ఘ ఙ,
 చ ఛ జ ఝ ఞ,
 ట ఠ డ ఢ ణ,
 త థ ద ధ న,
 ప ఫ బ భ మ,
 య ర ల ళ వ,
 శ ష స హ;

There are 35 such syllables are possible.

3. Syllable formation rule-3, \$=VB|X;

Example:

3a=V+B=\$; అం ఆం ఇం ఈం ఊం ఎం ఏం ఐం ఒం ఓం ఔం;

3b=V+X=\$; అః ఆః ఇః ఈః ఊః ఎః ఏః ఐః ఒః ఓః ఔః;

In combination with V and one of the two B or X, a total 36 syllables are possible.

Syllable combinations with vocalic R are not used.

4. Syllable formation rule-4, a \$= CH;

A standalone consonant may be appended by the halant marker H to form the corresponding graphic syllables as shown here.

Example:

క ఖ గ ఘ జ్
చ ఛ డ డ్ ణ్
ట ఠ డ్ డ్ ణ్
త థ ద ధ న్
ప ఫ బ భ మ్
య ర ల్ ల్ ల్ వ్
శ ష స హ్

There are 35 such graphic syllables are possible.

5. Syllable formation rule-5, \$=CB|X; Ex.

Standalone consonants can take one of the three vowel modifiers and form the corresponding syllables as shown below:

Example:

5a. \$=CB: కం ఖం గం ఘం జం చం ఛం డం ణం టం ఠం etc.

5b. \$=CX: కః ఖః గః ఘః జః ఛః డః ణః టః ఠః etc.

There are $2*35=70$ graphic consonant modifier syllables are possible.

6. Syllable formation rule-6, \$=CM;

A consonant may get attached with a vowel modifier or the dependent vowel diacritic to form the corresponding syllables;

Example:

కా కి కీ కు కూ కృ క కె కే కై కొ కో కౌ; etc.

A total of $35*13$ consonant + vowel diacritic combinations may derive 455 graphic syllables in Telugu.

7. Syllable formation rule-7, \$=CMB|X;

A consonant with a dependent vowel when followed by one of the three modifiers may derive the following graphic syllables;

Example:

7a. కాం కిం కీం కుం కూం కౌం కేం కైం కొం కోం కౌం

7b. కాః కిః కీః కుః కూః కౌః కేః కైః కొః కోః కౌః

A total of $35*12*2$ consonant plus a dependent vowel and one of the three modifiers derive 840 possible graphic syllables in Telugu.

8. Syllable formation rule-8, \$=CH[(C)*C];

Any consonant followed by the halant marker may combine with another consonant or consonants to form complex graphic syllables;

Example:

2 consonant clusters: ఖ్ గ, ఘ్ జ, చ్ ఛ, డ్ ణ, ట్ ఠ, డ్ డ్, ణ్ ణ్, etc.

3 consonant clusters: ధ్ ధ్, ణ్ ణ్, ణ్ ణ్, ణ్ ణ్, ణ్ ణ్, ణ్ ణ్, etc.

4 consonant clusters: త్స్స; ;

A total of $35*1*35 = 1225$ CHC syllables involving two consonant clusters are possible; Further, a total of $35*1*35*1*35 = 42,875$ CHCHC syllables involving three consonant clusters are possible; Though four consonant clusters are extremely rare but theoretically possible as shown above.

9. Syllable formation rule-9, \$=CH(CH[CH])CM;

Any consonant followed by the halant marker and a consonant or consonants may be appended by one of the dependent vowels to form complex graphic syllables involving two to three consonant clusters;

Example:

కు, ఖ్వి గో, ప్పి, జ్జో, చ్చా, చ్చూ, జ్జే, య్మి, య్మో, ట్టీ, ర్మా, డ్డా, డ్డూ, ణ్ణా, etc.

ర్దు, ప్పి, స్తీ, న్ని, జ్జి, ప్పా, త్త్యా, త్తు, etc.

తాన్మి

A total of $35*1*35*1*12 = 14,700$ complex syllables involving two consonant clusters followed by dependent vowels are possible.

A total of $35*1*35*1*35*12 = 5,14,500$ complex syllables involving three consonant clusters followed by dependent vowels are possible.

The following is a summary of possible syllable types with the glyphs in Telugu:

$$\$ = V([B|X])|CM([B|X])|CH(CH[C])M([B|X])$$

As per our definition the following 21 subtypes of graphic syllables are possible which however can be grouped under 8 rules as discussed above.

$$\begin{aligned} \$ = & V|VB|VX| \\ & C|CB|CX|CM|CH| \\ & CHC|CHCB|CHCX|CHCM\text{re } CHCH|CHCHC|CHCHCB|CHCHCX|CHCHCM \end{aligned}$$

Therefore, typologically 8 distinct types of graphic syllables can be derived in the language.

Appendix C: NBGP Cross-script Variant Inclusion Policy

If, in any two given scripts, all the potential cross-script variants consist of dependent (e.g. Vowel Signs, Anusvara, Visarga, Chandrabindu etc.) characters **ONLY**, then that entire set can be ignored and no cross-script variants be proposed between those two scripts.

If, in any two given scripts, there is **AT LEAST ONE** non-dependent (e.g. Consonant, Vowel etc.) cross-script variant character/sequence present, all the potential cross-script variants be considered and proposed between the two scripts.

This cross-script analysis has been restricted to the scripts that have descended from the Brahmi as most of them share similar usage patterns. By and large, all of these scripts have a common set of characters that existed in Brahmi script and bear the same identities. However, as the scripts branched out from the Brahmi, depending on various factors, the shapes of the characters changed. This change in the shape was not uniform across all the characters and the scripts. Some characters shapes did change significantly whereas some of them still retained similarity. The cross-script similarity analysis also aims to identify such cases where the same character retained almost the same shape despite being part of the different scripts. These set of characters are variants of each other in true sense than merely of co-incidental visual similarity.

Since, having such labels is a realistic possibility and the corresponding labels look almost exactly alike, NBGP has proposed them as blocked variants.

NBGP acknowledges the concern that this shape is quite generic and may have parallels in other scripts not under its ambit. However, as NBGP does not have any exposure about actual usage of those characters in those particular scripts, NBGP desisted from including them in the analysis. As NBGP has already considered all the related scripts under the cross-script variant analysis, the similarity of the characters belonging to NBGP scripts with other scripts not under the NBGP ambit, may be of a mere co-incidental visual nature.

Additionally, this concern is not limited to these two characters but for all the characters in all the scripts under the scope of the Root LGR procedure. Carrying out this analysis can practically be done only with the Generation Panels that exist while the NBGP is active. This still leaves out those scripts out of the scope which may not have a Generation Panel established yet. Hence, carrying out this exercise in entirety is quite impracticable. This conundrum can be resolved if all the such cases are handled by the "String Similarity Assessment Panel" of ICANN.