

SCRIPT GRAMMAR FOR TELUGU LANGUAGE

Prepared by

Technology Development for Indian Languages (TDIL) Programme

Department of Information Technology, Government of India

in association with

Centre for Development of Advanced Computing (C-DAC)

Table of Contents

0.	INTRODUCTION	3
1.	OBJECTIVES OF SCRIPT GRAMMAR	4
2.	END USERS FOR SCRIPT GRAMMAR	5
3.	SCOPE	6
4.	TERMINOLOGY	7
5.	PHILOSOPHY AND UNDERLYING PRINCIPLES	11
6.	SCRIPT GRAMMARSTRUCTURE	12
6.1.	PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR	13
6.2.	CONFORMITY TO THE SYLLABLE STRUCTURE	14
6.3	SCRIPT GRAMMAR PROPER.....	18
6.3.1.	The Character Set of Telugu.....	18
6.3.2.	Consonant Mātrā Combinations.	25
6.3.3.	The Ligature Set of Telugu.....	37
6.3.4	The Collation Order of Telugu.	43
7.	REFERENCES	45
8.	ANNEXURES	46
	Annexure 1: Names of experts who have contributed to the script grammar.....	46
	Annexure 2: Unicode Table of Telugu	47

0. INTRODUCTION

The term **script grammar** refers to the behaviour pattern of the writing system of a given language. Languages which have written representations do not use a haphazard manner of storing the information within the system, but use a coherent pattern which is similar to the linguistic grammar of a given language. With the help of specialists (not necessarily linguist) who work in the area of the written representation of the language, the manner in which the shapes of the characters of the language and the representation of the conjunct forms is provided. In other words the Script Grammar deals with the surface structure of the language and tries to provide the best possible “fit” for shapes and their representation. Since this is a highly subjective issue, the shapes provided here are recommendations at the best and conform to the perception of the mandating body/evaluators who consensually arrive at the “best possible fit” which is acceptable to a majority of users. An example from the Devanāgarī script will make the above clear. Although Marathi and Nepali share the same script Devanāgarī, not only do they not share the same character inventory but in addition the representation of certain characters is different. Thus the Nepali /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned Nepali ल Marathi ल. This ensures that the Script Grammar conforms to the language in question and provides the character shapes acceptable to a given user community. It should be noted that this does not mean monotony. The Marathi and Nepali /la/ can have a variety of forms once the intrinsic structure of the character is determined.

Script Grammar is the term used to define:

- the writing system used to inscribe a given language
- the history of the script and language (wherever available)
- the syllabic structure of the writing system of the language
- the rule ordering of the characters within the syllable (akshar)
- description of the syllabic clusters
- collation order of the characters: lexical / dictionary sorting order

1. OBJECTIVES OF SCRIPT GRAMMAR

The Objectives of the script grammar for each language can be divided into two major parts:

Societal:

- Provide a visual representation of shapes that are deemed to be in conformity with the perception of a given community
- Ensure thereby that this perception is safe-guarded
- Through wide-spread dissemination and creation of appropriate tools ensure that within the given linguistic community, all media tries to adopt the given shape.

Technical:

- Classify the language in terms of its ISO and also whether it belongs to the Abjad, Akshar (Alphasyllabary) class.
- Provide an inventory of the characters pertinent to the language and classify the same in terms of their taxonomy.
- As a corollary determine whether the inventory is in conformity to the Syllable formalism as stipulated in ISCII'91 and subsequently adopted by Unicode.
- Since Brahmi is written from left to right, and since certain characters do not follow the linear L to R order, provide an inventory of displaced catenators i.e. characters such as Mātrās that concatenate to the Consonant
- Propose the best shape representation of the individual characters as well as of the ligatures used within a given script. As a corollary request the expert(s) to identify the largest possible strings of such ligatures.
- Finally provide the collation order pertinent to that script/language, which would be of great utility to high-end NLP as well as to CLDR's in the pertinent language. The collation order for Nepali is different from Hindi although both languages share the same script. Thus in Nepali ख, ज are placed at the end of the consonant inventory i.e. after ह in the sort order. In Hindi ख is sorted along with क and ज with ज

2. END USERS FOR SCRIPT GRAMMAR

The script-grammar specific to a given language can be used by a large number of users.

- Most importantly it can be used by font developers desirous of developing a font which is compliant with the perception of the characters and ligatures of a language by its user community.
- Certain features of the script grammar such as the shapes can also be used for testing OCR and OHWR. Similarly information regarding Ligatures as well as collation order can help in high-end NLP work such as detecting invalid combinations, correct implementation of syllable structure, prediction routines to name a few. Information regarding collation and character sets can be also used for CLDR.
- They allow the font designer to design a font which is in compliance with the norms and standards of that particular script. A major problem which will be dealt with in the template is one of ligatures. The final list of ligatures defined by the script grammar allows the font designer to write specific rules for such glyphs.
- It permits the software developer to design and implement the keyboard and the input mechanism which will meet the requirement of the particular linguistic community.
- The collation or sort order as described in a Script Grammar permits the software developer to write software functions/ routines for sorting data in all applications.
- Script Grammars are equally important for keyboard design, especially when supplemented by frequency data from a corpus.

As can be seen the script grammar has a wide range of use and can be of utility to font developers, Indian language developers and linguists in the area of computation.

3. SCOPE

This script grammar document contains following information about the language and the script used for writing the language.

1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1 & 639.3 standard.
2. Script used to inscribe the given language
3. The structure of the script used for writing the language
 - Rule ordering of the characters within the syllable formation is a language
 - Description of the syllabic clusters of the script
 - Collation order of the characters: lexical / dictionary sorting order
 - Compliance of the script with Unicode.

These will be treated within the relevant sections of the script grammar

4. TERMINOLOGY¹

Abjad: A writing system in which each symbol always or usually stands for a consonant. The long vowels are indicated. However the short vowels are rarely marked and the reader needs to supply these. Example: Urdu written in Perso-Arabic Script is an example of this writing system.

Abugida: also called an alphasyllabary, is a segmental writing system in which consonant–vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel notation is obligatory but secondary²

Akshar: see **Abugida**

Allographs: Variants of the representation of a character. Thus ae and æ [U+00E6] in Latin alphabet are allographs.

Allo-Script: The term relates to languages which share a common script. Thus Devanāgarī is used to write 9 official languages. However these languages do not use the same set of characters. Thus Marathi uses the retroflex lla ळ [U+ 0933] which Hindi does not use. Flaps used in Hindi ः [U+095C] ः[U+095D] are not used in Konkani. These subsets of scripts based on a single “matricial” script are termed as allo-scripts.

Alphabet: A set of letters used in writing a language. Example: The English Alphabet.

Aspirated consonant: A consonant which is pronounced with an extra puff of air coming out at the time of release of the oral obstruction. This has a sound of an extra "h".

Basic alphabet: The minimal set of letters which can be used for uniquely encoding every word of a language. The basic alphabet for English consists of only the upper-case letters A-Z

Catenators: Also termed as Concatenators are characters which are concatenated to another character. In the Brahmi script these are the Mātrās or Vowel modifiers which are adjoined to the consonant and add a vocalic value to the consonant.

Conjunct: The Indic scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms. This abbreviation takes place only in the context of a consonant cluster....Under normal circumstances, a consonant cluster is depicted with a conjunct glyph if such a glyph is available in the current font. In the absence of a conjunct glyph, the one or more dead

¹ As in the case of the BIS Document, in order to make the terminology accessible for all readers, examples have been chosen from English/Latin scripts, wherever possible. Some definitions have been excerpted from the BIS ISCHII91 document and suitably modified where necessary.

² Wikipedia definition

consonants that form part of the cluster are depicted using half-form glyphs. In the absence of half-form glyphs, the dead consonants are depicted using the nominal consonant forms combined with visible virama signs.³

Consonant: A letter representing a speech sound in which the breath is at least partly obstructed,

Diacritic: A mark added to a letter which distinguishes it from the same letter without a mark, usually having a different phonetic value or stress.

Displaced Catenator: (see Catenator) Within the Brahmi script, the writing system is linear and moves from left to right. However in the case of some catenators this rule is not observed and the catenator (wholly or partially) is placed to the right of the consonant to which it relates. The short vowel I /i:/ in Devanāgarī is an example of a displaced catenator.

Display composing: The process of organizing the basic shapes available in a font in order to display (or print) a word.

Display rendition: The process by which a string of characters is displayed (or printed). In this process several consecutive characters may combine with each other on the screen. The sequence of display of the characters may become different.

Eyebrow repha: (See Eyelash ra)

Eyelash ra: The eyelash ra is used in Konkani, Nepali and Marathi. It is treated as different from the Ṛ(repha) by certain linguists. While the former is treated as a flap, the latter is a continuant trill (*cf.*, Kalyan Kale and Anjali Soman. 1986).

Font: A set of symbols used for display or printing of a script in a particular style.

International numerals: The conventional 0 to 9 digits used in English for denoting numbers. These are also known as Indo-Arabic numerals (to differentiate them from the Roman numerals like IX for 9).

Latin alphabet: The alphabet used for writing the language of ancient Rome. Also known as the Roman alphabet. The alphabet is used today for writing English and European languages.

Letter: A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

Ligature: (see Conjunct)

³Unicode ver. 6.0 Chapter 9.0 pp 6-7

Nasal consonant: A consonant pronounced with the breath passing through the nose.
Example *m* in English.

Nasalized vowel: A vowel pronounced with the breath passing both through the nose and the mouth. In Indian scripts this is denoted by a Chandrabindu and gives the vowel/vowel modifier over which it placed a nasal value. Example:

Phonetic alphabet: An alphabet which has direct correspondence between letters and sounds
Example: The International Phonetic Alphabet..

Pure consonant: A consonant which does not have any vowel implicitly associated with it.

Rafar: A special case of a ligature constituted by the adjunction of ra followed by a halanta to consonant. The resultant combination places the ra on top of the consonant to which it is adjoined. In case the consonant itself is adjoined to another consonant, the rafar is placed above the consonant e.g.

Rakar: A special case of a ligature constituted by the adjunction of a consonant followed by a halanta to ra. In a large number of Brahmi scripts the ra is adjoined to the stem of consonant to which it relates. In the case of consonants which have no stem such as the dental retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates.

Repha: (see Rafar)

Roman script: The script based on the ancient Roman alphabet, with the letters A-Z and additional diacritic marks. Used for writing a language which is not usually written in the Roman alphabet.

Script: A distinctive and complete set of characters used for the written form of one or more languages.

Script numerals: The 0 to 9 digits in a script, which have shapes distinct from their international counterparts.

Syllable: A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel or diphthong sound optionally surrounded by one or more consonants

Transliteration: Representation of words with the closest corresponding letters in an alphabet of a different language.

Vowel: A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction

Vowel sign: A graphic character associated with a letter, to indicate a vowel to be associated with that character (Mātrā in Hindi).

5.PHILOSOPHY AND UNDERLYING PRINCIPLES

The script grammar is based on the following principles:

1. The Grammar aims to depict the surface grammar of the written language: the manner in which characters as well as conjuncts are depicted
2. Where a given script admits many languages, it is pre-suppose that such languages will prescribe different representations for a given shape or conjunct according to the perception of the native users of that language
3. Corollary to the above the result is a script and allo-scripts i.e. a given script shared by many languages is not uniformly deployed across all the languages but is subject to variations and modulations.
4. The term Grammar is used here in a non-normative sense: what is prescribed is in the form of recommendations provided by experts who visualize the shape of the given script in their mother tongue in a specific manner. Subjective variations may occur⁴
5. The Grammar is limited to its synchronic use i.e. the manner in which a given language as of today admits a character set within the script used to write it. It is not diachronic or historical in nature and does not study the evolution of the given script across centuries.

⁴ It is recommended that such variations be culled by placing the Grammars of different scripts in public review.

6. SCRIPT GRAMMARSTRUCTURE

The script grammar provided below has the following parts.

Part 6.1. deals with peripheral elements such as the ISO of the language, the writing system used: (Alphasyllabic) Abugida or Abjad.

Part 6.2. treats of the syllabic structure. It verifies whether the character set of the language complies with the ISCII syllabic structure and if not which cases are not compliant.

Part 6.3 is the script grammar proper and describes the character set as well as the conjunct shapes of the given script along with the collation order

6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR

These constitute the elements that are peripheral to the Script Grammar. The main parameters considered are the mnemonic and name of the language (needed for CLDR and also for language tags), the writing system used to inscribe the language and wherever possible a short history of the language.

6.1.1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1. & 639.3

Name of the Language: TELUGU

ISO Mnemonics: tel

This refers to a one line description of the language and its mnemonic representation as per the ISO.

6.1.2. Identification of the writing system(s) used to inscribe the given language Telugu is written using the Telugu script. It is an alphasyllabary with the akshar as its core.

This is a one line description of the script used to write the language. However in case the language uses more than one script, all the scripts in question are specified, provided these constitute the official language of the given state.

All scripts derived from Brahmi are Abugidas i.e. syllabary driven systems. The main features of Abugidas are as under:

- The consonant has an implicit vowel built-in which is normally the schwa.
- The inherent vowel can be modified by the addition of other vowels or muted by a diacritic termed as a Virama or Halanta
- Vowels can be handled as full vowels with a vocalic value
 - When two or more consonants join together they form ligatures which can be recognized partly or wholly by their shape
 $\ddot{\text{ప}} + \ddot{\text{ర}} + \ddot{\text{మ}} = \ddot{\text{ప్రమ}}$ or alternatively form an entirely new shape
 $\ddot{\text{క}} + \ddot{\text{ర}} + \ddot{\text{మ}} = \ddot{\text{క్రమ}}$.

Abugidas/Alphasyllabaries because of their syllabic structure require a special description which is the subject of the discussion in 6.2. below.

6.1.3. Amendments needed in Unicode for Telugu language

The Telugu Akademi which has mandated the script grammar has proposed that the following character whose image and description is given below be added:

Telugu Letter LLLA 

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

Telugu language complies with the syllable (akshar) structure described above. It can admit up to 4 consonant clusters.

Alphasyllabaries are determined by the notion of the syllable or the Akshar. The compositional grammar of the syllable determines its well-formedness. This is through a series of formal constraints based on a Backus-Naur Formalism which is given below. The syllable (akshar), first defined in the ISCII document (1991), identifies the following character ‘sub-sets’ for the purposes of identifying the syllable (akshar). In what follows the syllable analysis will be restricted to Telugu.

(C) Consonants

క	ఖ	గ	ఘ	జ
చ	ఖ	జ	ధు	శ
ట	ఠ	డ	ధ	ణ
త	ధ	ద	ధ	న
ప	ఫ	బ	భ	మ
య	ర	ఱ	ల	ళ
వ	శ	ష	స	హ

(V) Vowels

అ	ఆ	ఇ	ఈ	ఉ	ఊ	ం	ఒ	ఔ	ఎ	ఏ
ఐ	ఒ	ఔ	ఔ							

(M) Mātrās or Vowel Modifiers

ం	ి	ం	ం	య	ం	ం	ం		ం	ం	ం	ం	ం
---	---	---	---	---	---	---	---	--	---	---	---	---	---

(D) Diacritics

ଓ: Anuswar	Anuswara, a nasal, is denoted by a dot placed to the right of the letter to which it is adjoined .This falls under Nasal
------------	--

	category.
ଓঁ:Chandrabindu /Arasunna	Chandrabindu, a nasal, is denoted by a dot placed to the right of the letter to which it is adjoined .This falls under Nasal
ঁঁ:Visarga	Visarga, denoted by two dots placed above the other.
ৰ: Avagraha	For extra length with long vowels seen in the Sanskrit text.

(H):Halanta - Halanta used in most writing systems to signify the lack of an inherent vowel.

(N)⁵ Nukta  -is Not used in Telugu

Each of these sub-types has its restrictions in terms of what can precede or follow it within a syllable (akshar), as shown in the table below:

PRECEDED BY	SUBTYPE	FOLLOWED BY
-, H	C	M,D,H
-	V	D
C	M	D
C,V,M	D	-
C	H	C

C can be preceded by H or no subtype and followed by any one of the following: M,D,H
V can be preceded by no subtype and followed by D but not by another sub-type.

M can be preceded by C and followed by D.

D can be preceded by C, V, M and followed by no other subtype. It closes the syllable (akshar).

H can be preceded by C alone and followed only by C and no other sub-set.

6.2.1.Syllable (akshar) Types

The formalism defines the syllable (akshar) in terms of both what can constitute a syllable (akshar) and what cannot. A valid syllable (akshar) as per this definition can be of only two types:

⁵ The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving 5 other consonants required for representing Urdu loan words ক, খ, গ, জ, ফ

1. A vowel syllable (akshar): a full vowel.
2. A consonant syllable (akshar): a full consonant (having a weak vowel or a mātrā)

The three other subsets viz. Mātrās, Diacritics, Halanta cannot constitute a syllable (akshar) by themselves or in combination among themselves.

6.2.1.1. The Vowel syllable (akshar) is of the following types:

6.2.1.1.1. A pure vowel all by itself: ఓ, /a/ ఏ /ā/ etc.

6.2.1.1.2. A vowel followed by a modifier which is a nasal marker : ఉఃచ /ū/

Visarga or avagraha are not permitted after a nasal

6.2.1.2. The Consonant syllable (akshar) can be of the following types:

6.2.1.2.1. A full consonant i.e. with the inherent vowel : క : /ka/

6.2.1.2.2. A consonant followed by a mātrā i.e. the inherent vowel being substituted by another vowel: కి /ki/

6.2.1.2.3. A consonant followed by a modifier: కో /kə/, హాః /haH/

6.2.1.2.4. A consonant followed by a mātrā and a modifier: కుం /kū/, దుః /duH/.

6.2.1.2.5. A consonant cluster i.e. a dead or half consonant (Consonant+Halanta) followed by a full consonant followed optionally by a mātrā, a modifier or a combination of both. These result in a ligature or what is often termed as *yuktakshara*.

కు- /tkā/ కు-ఽ /tkā/, కు-ః /tkaH/ తు-ఽ /tkū/, తు- /tdu/.

The above permutations and combinations result in 7 major syllable (akshar) types. Of these the last type introduces the problem of the number of consonant clusters. ISCII (91, p.23) provides for up to three consonant clusters as the “worst case” i.e. the largest possible string. This is functional for modern Prakrits where the largest consonantal cluster rarely exceeds three consonants. In Telugu, with the exception of one possibility, the largest permissible cluster is 3 consonants

This means that theoretically the following forms can be postulated:

1. Vowel Set: With the Vowel as the node.
V VD (only Nasal)
2. Consonant set: With the Consonant as the node (an implicit or modified vowel is pre-implied).

Node	Mātrā	Modifier	Mātrā+Modifier
C	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD

CHCHCHC	CHCHCHCM	CHCHCHCD	CHCHCHCM
---------	----------	----------	----------

A total number of 16 theoretical syllables is therefore possible. It will be seen that the written syllable (akshar) is not very different in structure from the phonetic syllable and that the movement from the written to the spoken levels is made feasible by application of certain rules.

This formal structure of the syllable (akshar) explained above is common to all Brahmi based scripts (with a few variations). It will form the basis of an exhaustive description of the characters as well as their ligatural representations.

6.3 SCRIPT GRAMMAR PROPER

This section lays down in detail the different parameters of the Script Grammar for Telugu. These are:

- 6.3.1. The Character Set of Telugu.
- 6.3.2. The Consonant mātrā combinations of Telugu.
- 6.3.3. The Ligature Set of Telugu.
- 6.3.4. Collocation Order of Telugu

6.3.1. The Character Set of Telugu.

This section provides detailed information about the characters in the language and the list of the same and also more importantly shows the manner in which the character is to be written. Each subsection comprises therefore two parts: the basic character set and the shape each character should have, as mandated by the experts who have designed the script grammar of Telugu.

This comprises the following:

- 6.3.1.1. The Consonant Set
- 6.3.1.2. The Vowel Set
- 6.3.1.3. The Mātrā Set
- 6.3.1.4. Displaced Catenators
- 6.3.1.5. Shape of the combination of ra (rakar,repha)
- 6.3.1.6. The Set of Diacritics
- 6.3.1.7. Numerals
- 6.3.1.8. Punctuation marks
- 6.3.1.9. Other symbols

Each of these will be analysed in detail:

6.3.1.1. The Consonant Set

The Consonant set of Telugu comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	క	ఖ	గ	ఘ	ঢ
Palatal	চ	ছ	ঝ	ঠ	ঢ়
Retroflex	ট	ঠ	ঢ	ণ	জ
Dental	ঢ	ণ	দ	ধ	ন
Bi-labial	প	ষ	ব	ধ	ম

Other consonants

య	ర	గ	ల	ఝ
వ	శ	ష	స	హ

Note: క్ష was earlier listed as a consonant is now treated as a ligature.

The exact shapes as desired by the experts are provided in the table below:

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	క	భ	గ	ఘ	ঢ
Palatal	চ	ছ	জ	ঝ	ঢ
Retroflex	ট	ঠ	ড	ঢ	ণ
Dental	ঢ	ধ	ঢ	ধ	ন
Bi-labial	ঞ	ঞ	ঞ	ঞ	ম

Other consonants

య	ర	గ	ల	ఝ
వ	శ	ష	స	హ

6.3.1.2. The Vowel Set

The Vowel set of Telugu is as under:

అ	TELUGU LETTER A
ఆ	TELUGU LETTER AA
ఇ	TELUGU LETTER I

శః	TELUGU LETTER II
ఁ	TELUGU LETTER U
ుః	TELUGU LETTER UU
ము	TELUGU LETTER VOCALIC R
మూ	TELUGU LETTER VOCALIC RR
౦	TELUGU LETTER VOCALICL
౨	TELUGU LETTER VOCALIC LL
ఎ	TELUGU LETTER E
ఏ	TELUGU LETTER EEE
ఐ	TELUGU LETTER AI
ఒ	TELUGU LETTER O
ఓ	TELUGU LETTER OO
ఔ	TELUGU LETTER AU

As per expert recommendations the character set should be written as under:

అ ఆ ఇ శః ఁ ఊ ము మూ ఊ ఎ
ఏ ఐ ఒ ఓ ఔ

6.3.1.3. The Mātrā Set

The Mātrā (Vowel Modifier Set) of Telugu is as under:

Mātrā Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
-------------	-------------	--------------------	-------------------------

1. Telugu sign AA	ా	అ	ణ + ా = ణ్ణ
2. Telugu sign I	ి	ఇ	ణ + ి = ణి
3. Telugu sign II	ిం	ఇం	ణ + ిం = ణిం
4. Telugu sign U	ఁ	ఁ	ణ + ఁ = ణు
5. Telugu sign UU	ఁఁ	ఁఁ	ణ + ఁఁ = ణుఁ
6. Telugu sign vocalic R	ఋ	బు	ణ + బు = ణు
7. Telugu sign vocalic RR	ఋ	బురు	ణ + బురు = ణురు
8. Telugu sign E	ఎ	ఎ	ణ + ఎ = ణె
9. Telugu sign E	ఎ	ఎ	ణ + ఎ = ణె
10. Telugu sign AI	ఐ	ఐ	ణ + ఐ = ణై
11. Telugu sign O	ఒ	ఒ	ణ + ఒ = ణొ
12. Telugu sign OO	ఒఁ	ఒఁ	ణ + ఒఁ = ణొఁ
13. Telugu sign AU	ఔ	ఔ	ణ + ఔ = ణౌ

As per expert recommendations the character set should be written as under:

ా ి ిం ఁ ఁఁ ఁఁ ఇ ఎ ఎ ఐ ఒ ఒఁ ఔ
 ా ఒ ఒఁ ఔ

6.3.1.4. Displaced Catenators

Under normal circumstances Vowel Modifiers also known as catenators (since they concatenate to the preceding consonant) in Brahmi based scripts are written from left to

right in linear order (with the exception of Consonant stacks). Telugu has no displaced catenator

6.3.1.5. Shape of the combination of ra (rakar,repha)

Telugu has only rakars (ra vattu) which can admit three possible combinations: to the left to the right and adjoined to the bottom of the consonant to which they relate.

There is no repha (rafar) in Telugu

RAKARS/RA VATTU

1. LEFT లక్, లచ
2. RIGHT ష్టో
3. BOTTOM ఱ

The eyelash ra is appended only to two consonants: య and హ. In Telugu unlike Marathi (where such cases are rare), the eyelash “ra” can be also at the beginning of a word.

Examples of words where Rakars / Ra vattus are used in Telugu language are given below:

Left	చలకం
Right	రాష్టోం
Bottom	వలకం

6.3.1.6. Diacritics

These are as under in the case of Telugu:

ଓ- Anuswar e.g. పున్నకం

ଓ- Candrabindu/Arasunna e.g. గావ్

ঃ- Halanta e.g. ভাৰত

ଓ- Visarga e.g. দুঃখ

ৰ - For extra length with long vowels e.g.

6.3.1.7. Numerals

Following are the numbers used in Telugu language.

As per the experts, Latino-Arabic set: (0,1,2,3,4,5,6,7,8,9) is used in official documents for Teugu. The practice of using Telugu numerals which was earlier prevalent has been discontinued. The numerals are as follows. However the experts are of the opinion that these be preserved.

○ ఉ న్ ర్ ౩ ఉ ద్ ఎ స్ గ్ ఏ

Numeral Shapes	Explanation
○	Telugu Digit Zero
ఉ	Telugu Digit One
న్	Telugu Digit Two
ర్	Telugu Digit Three
౩	Telugu Digit Four
ద్	Telugu Digit Five
౯	Telugu Digit Six
ఎ	Telugu Digit Seven
స్	Telugu Digit Eight
గ్	Telugu Digit Nine
ఏ	

Telugu admits in addition to the above the following fractions which are rarely used but form part of the Unicode set. These are:

Telugu fractions and weights

0C78	೯	TELUGU FRACTION DIGIT ZERO FOR ODD POWERS OF FOUR
0C79	೧	TELUGU FRACTION DIGIT ONE FOR ODD POWERS OF FOUR
0C7A	೨	TELUGU FRACTION DIGIT TWO FOR ODD POWERS OF FOUR
0C7B	೩	TELUGU FRACTION DIGIT THREE FOR ODD POWERS OF FOUR
0C7C	೪	TELUGU FRACTION DIGIT ONE FOR EVEN POWERS OF FOUR
0C7D	೫	TELUGU FRACTION DIGIT TWO FOR EVEN POWERS OF FOUR
0C7E	೬	TELUGU FRACTION DIGIT THREE FOR EVEN POWERS OF FOUR
0C7F	೦	TELUGU SIGN TUUMU

As per expert opinion, these are not in everyday use.

6.3.1.8. Punctuation Markers

Telugu uses punctuation markers from the Latin set. such as . , ; : “ ‘ () [] etc.

However the abbreviation marker is often used in Telugu.

English fullstop [.] is used. The use of Purna and Deergha Virama (full-stop/danda)

Devanagari code block: 0964, 0965 |, || is limited to poetry.

A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
1.	Full Stop or Period	.
2.	Question Mark	?
3.	Comma	,
4.	Exclamation Mark	!
5.	Apostrophe	'
6.	Semi Colon	;
7.	Colon	:
8.	Hyphen	-
9.	Dash	--
10.	Ellipsis mark	...
11.	Oblique	/
12.	Double quotation mark	""
13.	Single quotation mark	‘ ’
14.	Cross	XXX
15.	As Above	--" --
16.	Round Brackets	()
17.	Square Brackets	[]
18.	Curly Brackets	{ }
19.	Devanagari Danda	

	Devanagari Double Danda	
--	-------------------------	--

6.3.1.9 Other Symbols

These are religious, currency markers etc. included in Unicode:

₹:Rupee Sign as mandated by Government of India.

In addition the Unicode chart for Telugu has certain characters deemed obsolete. These are as under

Historic phonetic variants

0C58 త్స TELUGU LETTER TSA

0C59 త్ఱ TELUGU LETTER DZA

6.3.2. Consonant Mātrā Combinations.

These refer to the shapes generated when a Mātrā is adjoined to the Consonant. The layout of these is in the shape of a matrix where the first horizontal row refers to the active consonant and the first vertical column refers to the vowel-modifier.

Due to constraints of space and also for reasons of clarity, for each class a series of 3 tables are provided.

Table 1: క ఖ గ ఘ మ ఙ చ ఞ జ ర్మ ఞ

Table 2: ట ఠ డ ఢ ణ త థ ద ధ న

Table 3: ప ఫ బ భ మ య ర ల ళ
వ శ ష స హ

e.g. The table accommodates all possible theoretical combinations, although some might not be used in the language. This is because such shapes need to be handled at the font level in the anticipation that a user could type this combination. Thus although ఔ+ ర్మ

cannot exist in the language the font would show the following: ఔర్మ

The classes are as under:

6.3.2.1. refers to a simple concatenation of Consonant and Mātrā combinations.

6.3.2.2. refers to a concatenation of Consonant and Mātrā + Nasal marker combinations. Other diacritics such as avagraha and visarga have been avoided, since these are linear in

nature, are adjoined to the combination and do not in any way modify the structure of the shapes.

6.3.2.1 Consonant and Mātrā combinations.

This set refers to a simple concatenation of Consonant and Mātrā.

Consonant and Mātrā combinations Set 1

	క	ఖ	గ	ఘ	జ	చ	ఛ	జ	రు	శ
ర్మ	ర్మ	ఖ్రూ	గ్రూ	ఘ్రూ	జ్రూ	చ్రూ	ఛ్రూ	జ్రూ	ర్మూ	శ్రూ
హీ	హీ	వీ	గీ	ఘీ	జీ	చీ	ఛీ	జీ	రీ	శీ
త్లీ	త్లీ	వ్లీ	గ్లీ	ఘ్లీ	జ్లీ	చ్లీ	ఛ్లీ	జ్లీ	ర్లీ	శ్లీ
య	కు	ఖు	గు	ఘు	జు	చు	ఛు	జు	రు	శు
యూ	కూ	ఖూ	గూ	ఘూ	జూ	చూ	ఛూ	జూ	రూ	శూ
ఎ	కృ	ఖృ	గృ	ఘృ	జృ	చృ	ఛృ	జృ	రృ	శృ
ఎ్లో	క్లో	ఖ్లో	గ్లో	ఘ్లో	జ్లో	చ్లో	ఛ్లో	జ్లో	ర్లో	శ్లో
ఉ	క్ల	ఖ్ల	గ్ల	ఘ్ల	జ్ల	చ్ల	ఛ్ల	జ్ల	ర్ల	శ్ల
ఉ్లో	క్లో	ఖ్లో	గ్లో	ఘ్లో	జ్లో	చ్లో	ఛ్లో	జ్లో	ర్లో	శ్లో
ఏ	కె	ఖె	గె	ఘె	జె	చె	ఛె	జె	రె	శె
ఎం	కెం	ఖెం	గెం	ఘెం	జెం	చెం	ఛెం	జెం	రెం	శెం
ఎప్పు	కెప్పు	ఖెప్పు	గెప్పు	ఘెప్పు	జెప్పు	చెప్పు	ఛెప్పు	జెప్పు	రెప్పు	శెప్పు
ఏం	కెం	ఖెం	గెం	ఘెం	జెం	చెం	ఛెం	జెం	రెం	శెం
ఏప్పు	కెప్పు	ఖెప్పు	గెప్పు	ఘెప్పు	జెప్పు	చెప్పు	ఛెప్పు	జెప్పు	రెప్పు	శెప్పు
ఏం	కెం	ఖెం	గెం	ఘెం	జెం	చెం	ఛెం	జెం	రెం	శెం

Consonant and Mātrā combinations Set 2

This set is in continuation of set 1 which shows Consonant and Matra combinations.

	ట	ఠ	డ	ఢ	ణ	త	ధ	ద	ధ	న
ా	టా	ఠా	డా	ఢా	ణా	తా	ధా	దా	ధా	నా
ి	టి	ఠి	డి	ఢి	ణి	తి	ధి	ది	ధి	ని
ీ	టీ	ఠీ	డీ	ఢీ	ణీ	తీ	ధీ	దీ	ధీ	నీ
ం	టు	ఠు	డు	ఢు	ణు	తు	ధు	దు	ధు	ను
ూ	టూ	ఠూ	డూ	ఢూ	ణూ	తూ	ధూ	దూ	ధూ	నూ
ృ	టృ	ఠృ	డృ	ఢృ	ణృ	తృ	ధృ	దృ	ధృ	నృ
్య	ట్య	ఠ్య	డ్య	ఢ్య	ణ్య	త్య	ధ్య	ద్య	ధ్య	న్య
ఊ	ట్లు	ఠ్లు	డ్లు	ఢ్లు	ణ్లు	త్లు	ధ్లు	ద్లు	ధ్లు	న్లు
్లు	ట్లూ	ఠ్లూ	డ్లూ	ఢ్లూ	ణ్లూ	త్లూ	ధ్లూ	ద్లూ	ధ్లూ	న్లూ
్ల్లు	ట్ల్లు	ఠ్ల్లు	డ్ల్లు	ఢ్ల్లు	ణ్ల్లు	త్ల్లు	ధ్ల్లు	ద్ల్లు	ధ్ల్లు	న్ల్లు
్రె	ట్రె	ఠ్రె	డ్రె	ఢ్రె	ణ్రె	త్రె	ధ్రె	ద్రె	ధ్రె	న్రె
్రే	ట్రే	ఠ్రే	డ్రే	ఢ్రే	ణ్రే	త్రే	ధ్రే	ద్రే	ధ్రే	న్రే
్ర్లె	ట్ర్లె	ఠ్ర్లె	డ్ర్లె	ఢ్ర్లె	ణ్ర్లె	త్ర్లె	ధ్ర్లె	ద్ర్లె	ధ్ర్లె	న్ర్లె
్లో	ట్లో	ఠ్లో	డ్లో	ఢ్లో	ణ్లో	త్లో	ధ్లో	ద్లో	ధ్లో	న్లో
్రో	ట్రో	ఠ్రో	డ్రో	ఢ్రో	ణ్రో	త్రో	ధ్రో	ద్రో	ధ్రో	న్రో

Consonant and Mātrā combinations Set 3

This set is in continuation of set 2 which shows consonant and Matra combinations.

	ప	ఫ	బ	భ	మ	య	ర	ణ	ల	ళ	వ	శ	ష	స	హ
రా	పా	ఫా	బా	భా	మా	యా	రా	ణా	లొ	ళొ	వా	శా	షో	సో	హా
ఱ	పి	ఫి	బి	భి	మి	యి	రి	ణి	లి	ళి	వి	శి	షి	సి	హి
తీ	పీ	ఫీ	బీ	భీ	మీ	యీ	రీ	ణీ	లీ	ళీ	వీ	శీ	షీ	సీ	హీ
లు	పు	ఫు	బు	భు	ము	యు	రు	ణు	లు	ళు	వు	శు	షు	సు	హు
లూ	పూ	ఫూ	బూ	భూ	మూ	యూ	రూ	ణూ	లూ	ళూ	వూ	శూ	షూ	సూ	హూ
ంగ్	పృ	ఫృ	బృ	భృ	మృ	యృ	రృ	ణృ	లృ	ళృ	వృ	శృ	షృ	సృ	హృ
ంగ్రౌ	ప్రగ్	ఫ్రగ్	బ్రగ్	భ్రగ్	మ్రగ్	య్రగ్	ర్గ	ణ్గ	ల్గ	ళ్గ	వ్గ	శ్గ	ష్గ	స్గ	హ్గ
ను	ప్రస్తు	ఫ్రస్తు	బ్రస్తు	భ్రస్తు	మ్రస్తు	య్రస్తు	ర్సు	ణ్సు	ల్సు	ళ్సు	వ్సు	శ్సు	ష్సు	స్సు	హ్సు
నూ	ప్రస్తూ	ఫ్రస్తూ	బ్రస్తూ	భ్రస్తూ	మ్రస్తూ	య్రస్తూ	ర్సూ	ణ్సూ	ల్సూ	ళ్సూ	వ్సూ	శ్సూ	ష్సూ	స్సూ	హ్సూ
నో	ప్రస్తో	ఫ్రస్తో	బ్రస్తో	భ్రస్తో	మ్రస్తో	య్రస్తో	ర్సో	ణ్సో	ల్సో	ళ్సో	వ్సో	శ్సో	ష్సో	స్సో	హ్సో
రె	పే	ఫే	బే	భే	మే	యే	రె	ణే	లె	ళె	వె	శె	షె	సె	హె
ంరీ	పీ	ఫీ	బీ	భీ	మీ	యీ	రీ	ణీ	లీ	ళీ	వీ	శీ	షీ	సీ	హీ
ంర్మై	పై	ఫై	బై	భై	మై	యై	ర్మై	ణై	లై	ళై	వై	శై	షై	సై	హై
ంంరో	పో	ఫో	బో	భో	మో	యో	రో	ణో	లో	ళో	వో	శో	షో	సో	హో
ంంరో	పో	ఫో	బో	భో	మో	యో	రో	ణో	లో	ళో	వో	శో	షో	సో	హో
ంరో	పో	ఫో	బో	భో	మో	యో	రో	ణో	లో	ళో	వో	శో	షో	సో	హో

6.3.2.2 Consonant and Mātrā +Nasal combinations.

This set refers to a Consonant and Mātrā + Nasal marker combinations.

Consonant and Mātrā +Nasal combinations - Set 1

	క	ఖ	గ	ఘ	జ	చ	ఛ	జ	రు	ఇ
ం०	క०	ఖ०	గ०	ఘ०	జ०	చ०	ఛ०	జ०	రు०	ఇ०
రాం०	కా०	ఖా०	గా०	ఘా०	జా०	చా०	ఛా०	జా०	రుా०	ఇా०
ిం०	కి०	ఖి०	గి०	ఘి०	జి०	చి०	ఛి०	జి०	రుి०	ఇి०
తిం०	క్తి०	ఖ్తి०	గ్తి०	ఘ్తి०	జ్తి०	చ్తి०	ఛ్తి०	జ్తి०	రుతి०	ఇతి०
యిం०	క్యి०	ఖ్యి०	గ్యి०	ఘ్యి०	జ్యి०	చ్యి०	ఛ్యి०	జ్యి०	రుయి०	ఇయి०
ంయిం०	క్యా०	ఖ్యా०	గ్యా०	ఘ్యా०	జ్యా०	చ్యా०	ఛ్యా०	జ్యా०	రుయా०	ఇయా०
ంయిం०	క్యి०	ఖ్యి०	గ్యి०	ఘ్యి०	జ్యి०	చ్యి०	ఛ్యి०	జ్యి०	రుయి०	ఇయి०
ంయిం०	క్యా०	ఖ్యా०	గ్యా०	ఘ్యా०	జ్యా०	చ్యా०	ఛ్యా०	జ్యా०	రుయా०	ఇయా०
ఉ०	X	X	X	X	X	X	X	X	X	X
ఊ०	X	X	X	X	X	X	X	X	X	X
ఎ०	కె०	ఖె०	గె०	ఘె०	జె०	చె०	ఛె०	జె०	రె०	ఇె०
ఏ०	కే०	ఖే०	గే०	ఘే०	జే०	చే०	ఛే०	జే०	రే०	ఇే०
ఔ०	కై०	ఖై०	గై०	ఘై०	జై०	చై०	ఛై०	జై०	రై०	ఇై०
ఓ०	కొ०	ఖొ०	గొ०	ఘొ०	జొ०	చొ०	ఛొ०	జొ०	రొ०	ఇొ०
ఔ०	కో०	ఖో०	గో०	ఘో०	జో०	చో०	ఛో०	జో०	రో०	ఇో०
ఏ०	కౌ०	ఖౌ०	గౌ०	ఘౌ०	జౌ०	చౌ०	ఛౌ०	జౌ०	రౌ०	ఇౌ०

Consonant and Mātrā +Nasal combinations - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Nasal marker

	ట	ఠ	డ	ఢ	ణ	త	థ	ద	ధ	న
ం०	ట०	ఠ०	డ०	ఢ०	ణ०	త०	థ०	ద०	ధ०	న०
రాం०	టా०	ఠా०	డా०	ఢా०	ణా०	తా०	థా०	దా०	ధా०	నా०
ిం०	టి०	ఠి०	డి०	ఢి०	ణి०	తి०	థి०	ది०	ధి०	ని०
తిం०	ట్తి०	ఠ్తి०	డ్తి०	ఢ్తి०	ణ్తి०	త్తి०	థ్తి०	ద్తి०	ధ్తి०	న్తి०

యం	టుం	రుం	డుం	ధుం	ఱుం	తుం	థుం	దుం	ధుం	నుం
యాం	టూం	రూం	డూం	ధూం	ఱూం	తూం	థూం	దూం	ధూం	నూం
ఎం	ట్యుం	ర్యూం	డ్యూం	ధ్యూం	ఱ్యూం	త్యూం	థ్యూం	ద్యూం	ధ్యూం	న్యూం
ఎం	ట్యూం	ర్యూం	డ్యూం	ధ్యూం	ఱ్యూం	త్యూం	థ్యూం	ద్యూం	ధ్యూం	న్యూం
ఉం	X	X	X	X	X	X	X	X	X	X
ఊం	X	X	X	X	X	X	X	X	X	X
ఇం	టెం	రెం	డెం	ధెం	ఱెం	తెం	థెం	దెం	ధెం	నెం
ఇం	టేం	రేం	డేం	ధేం	ఱేం	తేం	థేం	దేం	ధేం	నేం
ఇం	టైం	రైం	డైం	ధైం	ఱైం	తైం	థైం	దైం	ధైం	నైం
ఓం	టొం	రొం	డొం	ధొం	ఱొం	తొం	థొం	దొం	ధొం	నొం
ఓం	టోం	రోం	డోం	ధోం	ఱోం	తోం	థోం	దోం	ధోం	నోం
ఎం	టోం	రోం	డోం	ధోం	ఱోం	తోం	థోం	దోం	ధోం	నోం

Consonant and Mātrā +Nasal combinations - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Nasal marker

	ప	ఫ	బ	భ	మ	య	ర	ఱ	ల	ళ	వ	శ	ష	న	హ
ం	పం	ఫం	బం	భం	మం	యం	రం	ఱం	లం	ళం	వం	శం	షం	నం	హం
ా	పాం	ఫాం	బాం	భాం	మాం	యాం	రాం	ఱాం	లాం	ళాం	వాం	శాం	షాం	నాం	హాం
ి	పిం	ఫిం	బిం	భిం	మిం	యిం	రిం	ఱిం	లిం	ళిం	విం	శిం	షిం	నిం	హిం
ం	పిం	ఫిం	బిం	భిం	మిం	యిం	రిం	ఱిం	లిం	ళిం	విం	శిం	షిం	నిం	హిం
య	పుం	ఫుం	బుం	భుం	ముం	యుం	రుం	ఱుం	లుం	ళుం	వుం	శుం	షుం	నుం	హుం
ా	పూం	ఫూం	బూం	భూం	మూం	యూం	రూం	ఱూం	లూం	ళూం	వూం	శూం	షూం	నూం	హూం
ి	ప్యుం	ఫ్యుం	బ్యుం	భ్యుం	మ్యుం	య్యుం	ర్యుం	ఱ్యుం	ల్యుం	ళ్యుం	వ్యుం	శ్యుం	ష్యుం	న్యుం	హ్యుం
ం	ప్యుం	ఫ్యుం	బ్యుం	భ్యుం	మ్యుం	య్యుం	ర్యుం	ఱ్యుం	ల్యుం	ళ్యుం	వ్యుం	శ్యుం	ష్యుం	న్యుం	హ్యుం
ా	ప్యూం	ఫ్యూం	బ్యూం	భ్యూం	మ్యూం	య్యూం	ర్యూం	ఱ్యూం	ల్యూం	ళ్యూం	వ్యూం	శ్యూం	ష్యూం	న్యూం	హ్యూం
ి	ప్యొం	ఫ్యొం	బ్యొం	భ్యొం	మ్యొం	య్యొం	ర్యొం	ఱ్యొం	ల్యొం	ళ్యొం	వ్యొం	శ్యొం	ష్యొం	న్యొం	హ్యొం
ం	ప్యొం	ఫ్యొం	బ్యొం	భ్యొం	మ్యొం	య్యొం	ర్యొం	ఱ్యొం	ల్యొం	ళ్యొం	వ్యొం	శ్యొం	ష్యొం	న్యొం	హ్యొం
ా	ప్యోం	ఫ్యోం	బ్యోం	భ్యోం	మ్యోం	య్యోం	ర్యోం	ఱ్యోం	ల్యోం	ళ్యోం	వ్యోం	శ్యోం	ష్యోం	న్యోం	హ్యోం
ి	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

ಖ್ಯಾಂ	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ರೆ ಂ	ಪೆಂ	ಫೆಂ	ಬೆಂ	ಭೆಂ	ಮೆಂ	ಯೆಂ	ರೆಂ	ಟೆಂ	ಲೆಂ	ಫೆಂ	ವೆಂ	ಷೆಂ	ಷೆಂ	ಸೆಂ	ಪೆಂ	ಪೆಂ	
ಕೆ ಂ	ಪೀಂ	ಫೀಂ	ಬೀಂ	ಭೀಂ	ಮೀಂ	ಯೀಂ	ರೀಂ	ಟೀಂ	ಲೀಂ	ಫೀಂ	ವೀಂ	ಷೀಂ	ಷೀಂ	ಸೀಂ	ಪೀಂ	ಪೀಂ	
ತೈ ಂ	ಪೈಂ	ಫೈಂ	ಬೈಂ	ಭೈಂ	ಮೈಂ	ಯೈಂ	ರೈಂ	ಟೈಂ	ಲೈಂ	ಫೈಂ	ವೈಂ	ಷೈಂ	ಷೈಂ	ಸೈಂ	ಪೈಂ	ಪೈಂ	
ಶ್ರೀ ಂ	ಪೊಂ	ಫೊಂ	ಬೊಂ	ಭೊಂ	ಮೊಂ	ಯೊಂ	ರೊಂ	ಟೊಂ	ಲೊಂ	ಫೊಂ	ವೊಂ	ಷೊಂ	ಷೊಂ	ಸೊಂ	ಪೊಂ	ಪೊಂ	
ಕೋ ಂ	ಪೋಂ	ಫೋಂ	ಬೋಂ	ಭೋಂ	ಮೋಂ	ಯೋಂ	ರೋಂ	ಟೋಂ	ಲೋಂ	ಫೋಂ	ವೋಂ	ಷೋಂ	ಷೋಂ	ಸೋಂ	ಪೋಂ	ಪೋಂ	
ರೋ ಂ	ಪೌಂ	ಫೌಂ	ಬೌಂ	ಭೌಂ	ಮೌಂ	ಯೌಂ	ರೌಂ	ಟೌಂ	ಲೌಂ	ಫೌಂ	ವೌಂ	ಷೌಂ	ಷೌಂ	ಸೌಂ	ಪೌಂ	ಪೌಂ	

Consonant and Mātrā + Nasal combinations: With Chandrabindu/Arasunna - Set 1

	క	ఖ	గ	ఘ	జ	చ	ఖ	జ	రు	ఇ
ం०	క०	ఖ०	గ०	ఘ०	జ०	చ०	ఖ०	జ०	రు०	ఇ०
ఠాం०	కా०	ఖా०	గా०	ఘా०	జా०	చా०	ఖా०	జా०	రుయా०	ఇయా०
ఢిం०	కి०	ఖి०	గి०	ఘి०	జి०	చి०	ఖి०	జి०	ర్మి०	ఇయి०
ఢిం०	కీ०	ఖీ०	గీ०	ఘీ०	జీ०	చీ०	ఖీ०	జీ०	ర్మి०	ఇయీ०
ఱం०	కు०	ఖు०	గు०	ఘు०	జు०	చు०	ఖు०	జు०	రుయు०	ఇయు०
ఱాం०	కూ०	ఖూ०	గూ०	ఘూ०	జూ०	చూ०	ఖూ०	జూ०	రుయూ०	ఇయా०
ఁ०	కృ०	ఖృ०	గృ०	ఘృ०	జృ०	చృ०	ఖృ०	జృ०	రుయృ०	ఇయృ०
ఁ०	క్యా०	ఖ్యా०	గ్యా०	ఘ్యా०	జ్యా०	చ్యా०	ఖ్యా०	జ్యా०	రుయ్యా०	ఇయ్యా०
అ०	X	X	X	X	X	X	X	X	X	X
అా०	X	X	X	X	X	X	X	X	X	X
ంै०	ంै	ఖै	గै	ఘै	జै	చै	ఖै	జै	రై०	ఇै०
ంాै०	ంాै	ఖాै	గాै	ఘాै	జాै	చాै	ఖాै	జాै	ర్మాै०	ఇయాै०
ఁै०	ఁै	ఖై	గై	ఘై	జై	చై	ఖై	జై	రై०	ఇయై०
ఁాै०	ఁాै	ఖ్యాै	గ్యాै	ఘ్యాै	జ్యాै	చ్యాै	ఖ్యాै	జ్యాै	రై०	ఇయ్యాै०
ఁौ०	ఁौ	ఖౌ	గౌ	ఘౌ	జౌ	చౌ	ఖౌ	జౌ	రై०	ఇయౌ०
ఁాौ०	ఁాौ	ఖ్యాौ	గ్యాौ	ఘ్యాौ	జ్యాौ	చ్యాौ	ఖ్యాौ	జ్యాौ	రై०	ఇయ్యాौ०

Consonant and Mātrā +Nasal combinations With Chandrabindu/Arasunna - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Chandrabindu

	ట	ర	డ	ధ	ణ	త	థ	ద	ధ	న
ం	టం	రం	డం	ధం	ణం	తం	థం	దం	ధం	నం
ంా	టాం	రాం	డాం	ధాం	ణాం	తాం	థాం	దాం	ధాం	నాం
ిం	టిం	రిం	డిం	ధిం	ణిం	తిం	థిం	దిం	ధిం	నిం
ిా	టీం	రీం	డీం	ధీం	ణీం	తీం	థీం	దీం	ధీం	నీం
ంయ	టుం	రుం	డుం	ధుం	ణుం	తుం	థుం	దుం	ధుం	నుం
యా	టూం	రూం	డూం	ధూం	ణూం	తూం	థూం	దూం	ధూం	నూం
ం	ట్లం	ర్లం	డ్లం	ధ్లం	ణ్లం	త్లం	థ్లం	ద్లం	ధ్లం	న్లం
ంా	ట్లాం	ర్లాం	డ్లాం	ధ్లాం	ణ్లాం	త్లాం	థ్లాం	ద్లాం	ధ్లాం	న్లాం
ఊ	X	X	X	X	X	X	X	X	X	X
ఊా	X	X	X	X	X	X	X	X	X	X
ఁం	టేం	రెం	డెం	ధెం	ణెం	తెం	థెం	దెం	ధెం	నెం
ఁంయ	టేం	రెం	డెం	ధెం	ణెం	తెం	థెం	దెం	ధెం	నెం
ఁంా	టైం	రేం	డైం	ధైం	ణైం	తైం	థైం	దైం	ధైం	నైం
ఁంా	టైం	రేం	డైం	ధైం	ణైం	తైం	థైం	దైం	ధైం	నైం
ఁంం	టౌం	రోం	డౌం	ధౌం	ణౌం	తౌం	థౌం	దౌం	ధౌం	నౌం
ఁంాం	టౌం	రోం	డౌం	ధౌం	ణౌం	తౌం	థౌం	దౌం	ధౌం	నౌం
ఁంంం	టౌం	రోం	డౌం	ధౌం	ణౌం	తౌం	థౌం	దౌం	ధౌం	నౌం

Consonant and Mātrā +Nasal combinations With Chandrabindu/Arasunna - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Chandrabindu

	ప	ఫ	బ	భ	మ	య	ర	ణ	ల	ళ	వ	శ	ష	స	హ
ం	పఁ	ఫఁ	బఁ	భఁ	మఁ	యఁ	రఁ	ణఁ	లఁ	ళఁ	వఁ	శఁ	షఁ	సఁ	హఁ
ం	పా	ఫా	బా	భా	మా	యా	రా	ణా	లా	ళా	వా	శా	షా	సా	హా
ం	పి	ఫి	బి	భి	మి	యి	రి	ణి	లి	ళి	వి	శి	షి	సి	హి
ం	పీ	ఫీ	బీ	భీ	మీ	యీ	రీ	ణీ	లీ	ళీ	వీ	శీ	షీ	సీ	హీ
ం	ప్ప	ఫ్ప	బ్ప	భ్ప	మ్ప	య్ప	ర్ప	ణ్ప	ల్ప	ళ్ప	వ్ప	శ్ప	ష్ప	స్ప	హ్ప
ం	పూ	ఫూ	బూ	భూ	మూ	యూ	రూ	ణూ	లూ	ళూ	వూ	శూ	షూ	సూ	హూ
ం	ప్పూ	ఫ్పూ	బ్పూ	భ్పూ	మ్పూ	య్పూ	ర్పూ	ణ్పూ	ల్పూ	ళ్పూ	వ్పూ	శ్పూ	ష్పూ	స్పూ	హ్పూ
ం	ప్ప్ప	ఫ్ప్ప	బ్ప్ప	భ్ప్ప	మ్ప్ప	య్ప్ప	ర్ప్ప	ణ్ప్ప	ల్ప్ప	ళ్ప్ప	వ్ప్ప	శ్ప్ప	ష్ప్ప	స్ప్ప	హ్ప్ప
ం	ప్ప్పూ	ఫ్ప్పూ	బ్ప్పూ	భ్ప్పూ	మ్ప్పూ	య్ప్పూ	ర్ప్పూ	ణ్ప్పూ	ల్ప్పూ	ళ్ప్పూ	వ్ప్పూ	శ్ప్పూ	ష్ప్పూ	స్ప్పూ	హ్ప్పూ
ం	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ం	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ం	పె	ఫె	బె	భె	మె	యె	రె	ణె	లె	ళె	వె	శె	షె	సె	హె
ం	పే	ఫే	బే	భే	మే	యే	రే	ణే	లే	ళే	వే	శే	షే	సే	హే
ం	పీ	ఫీ	బీ	భీ	మీ	యీ	రీ	ణీ	లీ	ళీ	వీ	శీ	షీ	సీ	హీ
ం	పై	ఫై	బై	భై	మై	యై	రై	ణై	లై	ళై	వై	శై	షై	సై	హై

ం															
రొ	పొ	ఫొ	బొ	భొ	మెయి	యొ	రొ	తొ	లొ	ళొ	వో	శొ	షొ	సొ	హొ
ం															
రో	పో	ఫో	బో	భో	మో	యో	రో	తో	లో	ళో	వో	శో	షో	సో	హో
ం															
రె	పె	ఫె	బె	భె	మె	యె	రె	తె	లె	ళె	వె	శె	షె	సె	హె
ం															
రు	పు	ఫు	బు	భు	ము	యు	రు	తు	లు	ళు	వు	శు	షు	సు	హు
ం															

6.3.3. The Ligature Set of Telugu.

Telugu has a large set of ligatural forms. These are combinations of Consonant+Halanta+Consonant (CHC) or CHCHC or even rarer CHCHCHC. The CHC combinations which are the most frequent are arranged in the shape of a matrix: the abscissa or horizontal axis refers to the Consonant which constitutes the ligature and the ordinate or vertical axis shows the consonant which forms the ligature and which is followed by a halanta.

As in 6.3.2. the ligature sets are divided into the following

6.3.3.1 CHC (in a matrix)

6.3.3.2 CHCHC

6.3.3.3.CHCHCHC

6.3.3.1. CHC (combination of two consonants)

These ligatures are presented as in the earlier case of Consonant+Mātrā combinations in three sets.

CHC Set 1:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “క” and “శ” is ligature “క్షు”.

	అ	ఒ	ఎ	ఇ	ఔ	చ	ఏ	య	జ	ము	శు
క	క్షు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
ఖ	ఖ్సు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
గ	గ్ంసు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
పు	ప్సు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు
చ	చ్చు	ఖ్సు	గ్ంసు	ప్సు	జ్ఞు	చ్చు	శ్చు	ఘ్ను	జ్ఞు	ర్ము	శ్శు

CHC Set 2:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “**କ**” and “**ଶ**” is ligature “**କୁଶ**”.

భ	ట్టు	ర్భు	డ్భు	థ్భు	ణ్భు	త్భు	ధ్భు	ద్భు	ద్భు	న్భు
మ	ట్లై	ర్లై	డ్లై	థ్లై	ణ్లై	త్లై	ధ్లై	ద్లై	ద్లై	న్లై
య	ట్వీ	ర్వీ	డ్వీ	థ్వీ	ణ్వీ	త్వీ	ధ్వీ	ద్వీ	ద్వీ	న్వీ
ర	ట్రీ	ర్టీ	డ్టీ	థ్టీ	ణ్టీ	త్టీ	ధ్టీ	ద్టీ	ద్టీ	న్టీ
ఐ	ట్టె	ర్భె	డ్బె	థ్బె	ణ్బె	త్బె	ధ్బె	ద్బె	ద్బె	న్బె
ల	ట్లె	ర్లె	డ్లె	థ్లె	ణ్లె	త్లె	ధ్లె	ద్లె	ద్లె	న్లె
న	ట్టూ	ర్భూ	డ్బూ	థ్బూ	ణ్బూ	త్బూ	ధ్బూ	ద్బూ	ద్బూ	న్బూ
వ	ట్వూ	ర్వూ	డ్వూ	థ్వూ	ణ్వూ	త్వూ	ధ్వూ	ద్వూ	ద్వూ	న్వూ
ష	ట్టూ	ర్షూ	డ్బూ	థ్బూ	ణ్బూ	త్బూ	ధ్బూ	ద్బూ	ద్బూ	న్బూ
ష్ణ	ట్టొ	ర్షొ	డ్బొ	థ్బొ	ణ్బొ	త్బొ	ధ్బొ	ద్బొ	ద్బొ	న్బొ
స	ట్టొ	ర్భొ	డ్బొ	థ్బొ	ణ్బొ	త్బొ	ధ్బొ	ద్బొ	ద్బొ	న్బొ
పొ	ట్టొ	ర్పొ	డ్బొ	థ్బొ	ణ్బొ	త్బొ	ధ్బొ	ద్బొ	ద్బొ	న్బొ

CHC SET 3:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “**ప**” and “**శ**” is the ligature “**పు**”.

బ	పబ	ఫబ	బబ	భబ	మబ	యబ	రబ	ఱబ	లబ	తబ	వబ	శబ	షబ	సబ	హబ
భ	పభ	ఫభ	బభ	భభ	మభ	యభ	రభ	ఱభ	లభ	తభ	వభ	శభ	షభ	సభ	హభ
మ	పమై	ఫమై	బమై	భమై	మమై	యమై	రమై	ఱమై	లమై	తమై	వమై	శమై	షమై	సమై	హమై
య	పయ్య	ఫయ్య	బయ్య	భయ్య	మయ్య	యయ్య	రయ్య	ఱయ్య	లయ్య	తయ్య	వయ్య	శయ్య	షయ్య	సయ్య	హయ్య
ర	ప్ర	ఫ్ర	బ్ర	భ్ర	మ్ర	య్ర	ర్ర	ఱ్ర	ల్ర	త్ర	వ్ర	శ్ర	ష్ర	స్ర	హ్ర
ణ	పణ	ఫణ	బణ	భణ	మణ	యణ	రణ	ఱణ	లణ	తణ	వణ	శణ	షణ	సణ	హణ
ల	పణ	ఫణ	బణ	భణ	మణ	యణ	రణ	ఱణ	లణ	తణ	వణ	శణ	షణ	సణ	హణ
త	ప్తు	ఫ్తు	బ్తు	భ్తు	మ్తు	య్తు	ర్తు	ఱ్తు	ల్తు	త్తు	వ్తు	శ్తు	ష్తు	స్తు	హ్తు
వ	పవ్వ	ఫవ్వ	బవ్వ	భవ్వ	మవ్వ	యవ్వ	రవ్వ	ఱవ్వ	లవ్వ	తవ్వ	వవ్వ	శవ్వ	షవ్వ	సవ్వ	హవ్వ
శ	పన్న	ఫన్న	బన్న	భన్న	మన్న	యన్న	రన్న	ఱన్న	లన్న	తన్న	వన్న	శన్న	షన్న	సన్న	హన్న
ష	పష్ట	ఫష్ట	బష్ట	భష్ట	మష్ట	యష్ట	రష్ట	ఱష్ట	లష్ట	తష్ట	వష్ట	శష్ట	షష్ట	సష్ట	హష్ట
స	పన్న	ఫన్న	బన్న	భన్న	మన్న	యన్న	రన్న	ఱన్న	లన్న	తన్న	వన్న	శన్న	షన్న	సన్న	హన్న
హ	పహ్మ	ఫహ్మ	బహ్మ	భహ్మ	మహ్మ	యహ్మ	రహ్మ	ఱహ్మ	లహ్మ	తహ్మ	వహ్మ	శహ్మ	షహ్మ	సహ్మ	హహ్మ

6.3.3.2 CHCHC (combination of three consonants)

These are not as frequent as the CHC combinations. Only the major are listed below.

With a few exceptions these are mainly linear in nature

రాష్ట్రం	ఆస్తి	స్వి	అగ్న్య	వస్తి	దృష్టి	పోర్కు	సీట్
ఆస్తి శ్వత	సంస్కృతం	శాస్త్రం	ఉత్కుషష్ట	పొర్చువ	సైర	కక్ష	కోర్టు

6.3.3.3.CHCHCHC (Combination of four Consonants)

This is rare and is found only in the following:

శార్యంపు

6.3.4 The Collation Order of Telugu.

Collation is one of the most important features of a script grammar. It determines the order in which a given culture indexes its characters. This is best seen in a dictionary sort where for easy search words are sorted and arranged in a specific order. Within a given script, each allo-script may have a different sort-order. Different scripts admit different sort orders and for all high-end NLP applications, sort is a crucial feature to ensure that the applications index data as per the cultural perception of that community. In quite a few States, sort order is clearly defined by the statutory bodies of that state and hence it is crucial that such sort order be ascertained and introduced in the script grammar.

In the case of Telugu the following is the traditional sort order as determined by the experts. The order as given below is pertinent to sorting by a computer program and is compliant with CLDR as laid down by Unicode and W3C.

In Tabular format:

ଠ	ଠୁ	ଠୁଁ	ଫ	ଫ	ବୁ	ରୁ	ରୁଁ	ଙ୍କ	ଙ୍କୁଁ	ମୁ
ମୁଁ	ତ	ପୁଁ	ଏ	ଏ	ବି	ବି	ଚ	ଚି	ଦ୍ବୀ	

క	ఖ	గ	ఘ	జ	చ	చ	ఛ	జ	ర్మ	శ
ట	ఠ	డ	ధ	ణ	త	థ	ద	ధ	ఫ	న
ప	శ	బ	భ	మ	య	ర	ల	ల	శ	వ
వ	ళ	స	ష	హ						
ం	ఊ	ఉ	ఁ	ఁ	ఁ	ఁ	ఁ			ఁ
ఁ	ఁ	ఁ	ఁ	ఁ						

7. REFERENCES

1. <http://www.unicode.org>
2. ISCII'91

8. ANNEXURES

Annexure 1: Names of experts who have contributed to the script grammar

Annexure 2: Unicode Table of Telugu⁶

⁶ The Unicode chart provided is for version 5.1 since the Script Grammar was prepared at that time. No considerable change in the script grammar can be seen in the updated versions of Unicode, with the possible addition of the Rupee Sign U+02B9