0C48	ై	TELUGU VOWEL SIGN AI = 0C46 ె 0C56 ౖ
0C49		<reserved>
0C4A	ొ	TELUGU VOWEL SIGN O
0C4B	ో	TELUGU VOWEL SIGN OO
0C4C	ౌ	TELUGU VOWEL SIGN AU

Various signs

0C4D	్	TELUGU SIGN HALANT To take out hidden vowel sound of consonant.
0C4E		<reserved>
0C4F		<reserved>
0C50		<reserved>
0C51	ౖ	TELUGU STRESS SIGN UDATTA
0C52	ౖ	TELUGU STRESS SIGN ANUDATTA
0C53	ౖ	TELUGU GRAVE ACCENT
0C54	ౖ	TELUGU ACUTE ACCENT
0C55	ౖ	TELUGU LENGTH MARK
0C56	ౖ	TELUGU AI LENGTH MARK

Generic additions

0C60	ౠ	TELUGU LETTER VOCALIC RR
0C61	ౡ	TELUGU LETTER VOCALIC LL
0C64	l	TELUGU SIGN PURN VIRAMA
0C65	ll	TELUGU SIGN DEERGH VIRAMA

Digits

0C66	౦	TELUGU DIGIT ZERO
0C67	౧	TELUGU DIGIT ONE
0C68	౨	TELUGU DIGIT TWO
0C69	౩	TELUGU DIGIT THREE
0C6A	౪	TELUGU DIGIT FOUR
0C6B	౫	TELUGU DIGIT FIVE
0C6C	౬	TELUGU DIGIT SIX
0C6D	౭	TELUGU DIGIT SEVEN
0C6E	౮	TELUGU DIGIT EIGHT
0C6F	౯	TELUGU DIGIT NINE

4.5.2 Telugu Script Details

Introduction

The Telugu language: Historically the Telugu language is also known by the names, āṃdhraṃ, tenu(ṃ)gu, and Gentoo.

Demographic information: Telugu is one of the major Scheduled languages of India. It has the second largest number of speakers mainly concentrated in South India. It is the official language of Andhra Pradesh and second widely spoken language in Tamilnadu, Karnataka. Considerable number of Telugu speaking minorities live in Maharashtra, Orissa, Madhya Pradesh and West Bengal. Considerable number of Telugu language speakers have migrated to Mauritius, South Africa, and recently to USA, UK, and Australia.

Genetic affiliation and History: Telugu belongs to the South-Central branch of the Dravidian family of languages. It is most widely spoken Dravidian language. It is the only literary language outside the South-Dravidian Branch. Its literature goes back to 11th century A.D. Its ancient forms were attested through inscriptions dating back to 200 A.D.

In the early days of 20th century there was a controversy over the use of a particular variety of Telugu as a medium of instruction. There were two varieties, one the literary or grānthika style, and the other the spoken or colloquial style popularly known as vyāvahārika style. Finally the controversy ended in favour of the colloquial standard and the government of Andhra Pradesh has issued a G.O. to use only the colloquial standard as medium of education at all levels.

The domains of use: The Telugu language is formally taught in Schools, Colleges and Universities. It is the popular medium of instruction at primary, Secondary and Higher-Secondary and early University education. The Telugu Academy on behalf of the Government of Andhra Pradesh regularly publishes text books in Telugu medium in various subjects for use at Intermediate and graduate levels. It is not used as medium of education in

Technical and Professional courses. In lower levels of administration it is popularly used but in higher levels of administration it is sparingly used. It is the common language in the transactions of the assembly of Andhra Pradesh. As a medium of journalese, it flourishes in Telugu language newspapers, radio-broadcasts, and TV-telecasts. It is also one of the popular medium of feature films produced in the South India.

The Telugu Script

Origin & Development: Telugu is written in Telugu script which is derived from Ashokan Brahmi used in the South India cerca 2nd A.D. The Southern Brahmi also known as drāviḍi-brāhmi of 2nd c. A.D. gave rise to vēṅgi-cālukyan script also known as Telugu-Kannada script. By the end of 13th Century A.D., the Telugu and Kannada scripts got separated. In the early combined Telugu-Kannada script, no orthographic distinction was made between the short mid [e, o] /ఎ, ఒ/ and long mid [ē, ō] /ఏ, ఓ/. However, distinct signs were employed to denote the special consonants viz. the trill [ṟa] /ఱ/ the retroflex lateral [ḷa] /ళ/ and the retroflex palatal [ẓa] /ఴ/ found only in South Indian languages, by 5th c.

Telugu Alphabet: The Primary units of Telugu alphabet are syllables, therefore, it should be rightly called a syllabary and most appropriately a mixed alphabetic-syllabic script. Unlike in the Roman alphabet used for English, in the Telugu alphabet the correspondence between the symbols (graphemes) and sounds (phonemes) is more or less exact. However, there exist some differences between the alphabet and the phonemic inventory of Telugu. Since writing habits change slowly and speech changes faster, the script has preserved some symbols which have become otherwise obsolete now in speech. Telugu script is written from left to right and consists of sequences of simple and/or complex characters.

Common Core and Overall pattern: The Telugu alphabet can be viewed as consisting of more commonly used inventory, a common core, and an overall pattern comprising all those symbols that are used in all domains. The overall pattern consists of 60 symbols, of which 16 are vowels, 3 vowel modifiers, and 41 consonants.

Vowels

అ	ఆ	ఇ	ఈ	ఉ	ఊ	బు	ఎ	ఏ	ఐ	ఒ	ఓ	ఔ
a	ā	i	ī	u	ū	ṛ	e	ē	ai	o	ō	au

Vowel Modifiers

అఁ	అం	అః
am̐	aṃ	aḥ

Consonants

క	ఖ	గ	ఘ	ఙ
ka	kha	ga	gha	ṅa
చ	ఛ	జ	ఝ	ఞ
ca	cha	ja	jha	ña
ట	ఠ	డ	ఢ	ణ
ṭa	ṭha	da	ḍha	ṇa
త	థ	ద	ధ	న
ta	tha	da	dha	na
ప	ఫ	బ	భ	మ
pa	pha	ba	bha	ma
య	ర	ల	ళ	వ
ya	ra	la	ḷa	va
శ	ష	స	హ	
śa	ṣa	sa	ha	

Special/obsolete graphemes

obsolete:

బుూ	ఌ	ౡ	ఱ	ఴ
ṝ	ḷ	ḹ	ṟa	ẓa

deprecated :

క్ష	ౘ	జ̂	ऩ	ऱ
kṣa	ča	ĵa	ṉa	ṟa

Chart-1: The Telugu alphabet- overall pattern

Technical Characteristics-Vowels

Primaries: There are thirteen vowel signs which occur as stand alone characters viz. [a ā i ī u ū ṛ e ē ai o ō au] /అ ఆ ఇ ఈ ఉ ఊ బు ఎ ఏ ఐ ఒ ఓ ఔ/ in the common core. Each of these is assigned a Hexa decimal code point in ISCII Standard (BIS 1993), from A4 164 to B1 177 (except for AD 173 and B0 176) or 0C05-0C14 and 0C0A-0C0F in UNICODE Standard 3.0. An additional vocalic/ఌ/ occurs in the UNICODE Standard with a code point 0C0C. These vowels are also referred to as primaries or independent vowel signs.

Secondary vowel signs (guNiMtaM gurtulu): When a vowel occurs immediately after a consonant it is always represented by a dependent or secondary sign i.e. as a corresponding diacritic on the consonant. In the case of the vowels /ṛ, ṝ, lo ,lo ‾ / the corresponding secondary vowel signs [○ꜗ, ○ꝭ, ꝭ] are placed on the right side of the consonant. Since, the Telugu alphabet is a syllabary the primary consonant always has an inherent vowel [a] /ˇ/. The secondary vowel sign is attached to the consonant after removing the vowel /a/. In Telugu, normally (some exceptions exist) the secondary vowel signs are attached to consonants in the place where the vowel sign for /a/ would occur as /ˇ/ (talakattu) in unmodified primary consonants. This is unlike in Devanagari where the consonant does not drop or lose explicitly the vowel sign /a/ (the top bar or lakir ?) but simply get attached with a secondary vowel sign. Therefore the phenomenon in Telugu and Kannada allows us to interpret secondary vowel signs as not just contextually determined vowel allographs but functionally distinct and complex ones from those of the primary or stand-alone vowel characters. This distinction between the primary vowel signs and secondary vowels signs can be expressed in the following illustrations:

ki=k+i — 1 (where [k] is pure consonant /క/ and [i] /ఇ/ is primary vowel) or

ki=ka+_i — 2 (where [ka] /క/ is consonant plus the inherent vowel [a] /ˇ/, and

[_i] /ి/ is secondary vowel)

The right hand part of the equation (2) can be rewritten as in (3)

ki=ka+_+i — 3

By combining 3 and 2 we get (4)

ka+_+i=ka+_i — 4

The difference between the two vowel signs can made explicit by effecting a transformation of the deletion of the common element /ka/ from the equation (4):

_+i=_i — 5

These transformational relationships between the two vowel signs in the context of consonants indicate that there is an inherent vowel omission sign in the secondary vowel. Similar interpretation may be made with respect to secondary consonantal signs since they occur always a pure consonant, in other words, a secondary consonantal sign removes an inherent vowel from the preceding consonant sign.

క + ్ → క్
ka+ka → k*k*a

Vowel Modifiers: Though, phonetically they share the properties of both Vowels and consonants are known as ambivalents (ubhayākṣaras). Since they follow always a vowel in their usage, traditionally they are treated as part of Vowel inventory. They are three in number. Unlike in Devanagari, where these Vowel modifiers always occur as diacritics or dependent character signs, in Telugu they occur linearly as independent signs facilitating their use independent of vowels or vowel signs. pūrṇānusvāra [ṃ] / ం/, the only archiphoneme in the language has one of the highest frequencies in usage. The ardhānusvāra [m̐] / ఁ/ as a character does not represent a phoneme but has a phonetic function in the language. It is often used to transcribe certain nasalized expressions, as in a delayed yes, exclamation, o.k. etc. Though, tradition proscribes these from writing in word initial positions they can be used independent of vowel signs as in the case of word initial pre-consonantal nasal, ex. Mkhrumo / ఒఖ్రుమో/ 'Mr. Nkhrumo, a leader of an African country', / Mpala/ 'Impala, a brand name of a car, but a Swahili word, originally pronounced as /ṃpālā/'. The character sign visarga / ః/ is used only to transcribe Sanskrit words, and it is the least frequently used character representing an array of allophanes of various phonemes. Unlike vowels and consonants they are invariants i.e. they do not have contextually dependent character realization.

Consonants: There are 41 consonants in the common core inventory (excluding _(kṣa)/క్ష/, (a) /ఴ/, (a) /ఴ̐ /, (n̠) /ఴ/, (ṟ) /ఴ/ and including [ẓa] /ఴ. Currently they occupy in ISCII Hexa Decimal coding the points beginning with B3 179 and ending D8 216, where as in Unicode Standard 3.0 they begin with 0C15 to 0C39 and 0C1A to 0C2F. The

character set for consonants in Telugu and Kannada is more complex and peculiar in their function. Their character signs have often three or more than three distinct shapes depending on whether they are used as stand alone characters called Primaries or Base consonants, or when used with a vowel other than the inherent vowel /a/ functioning as a hanger or a pure consonant, or when used as a constituent of a conjunct called as secondary or dependent consonant characters called as *ottulu* (ఒత్తులు).

The basic character set for consonants are called as primaries or stand alone characters as they occur in the alphabet. Each of which has an inherent vowel /a/ which often is explicitly indicated by sign /˘/. This graphic sign indicating the vowel /a/ is normally deleted and replaced with another explicit mark for a different vowel except in the case of /u/ and /ū/ and /r̥/ /r̥̄/ and /l̥/ which however are attached at the right side lower corner of the consonant.

Anchors and pure consonants: Anchors are pure consonantal characters carrying explicit secondary vowel signs and differ in their shape in three important ways often without the diacritic for the implict vowel /a/ as in i., with out any modification in the basic character shape as in the case of group ii. and the iii. consists of a set of consonant with special bases.

i. క గ ఘ చ ఛ ఝ ం డ ఢ థ ధ ద ధ న ప ఫ భ మ య ర
    ఴ వ శ ష స హ
    k g gh c ch jh ṭh ḍ ḍh t th d dh n p ph bh m y r
    ḷ v ś ṣ s h

ii. ఖ ఙ జ ఞ ట ణ బ ఴ ల
     kh ṅ j ñ ṭ ṇ b r̥ l

iii. ప ఫ ష స
      p ph ṣ s

Pure consonants are consonant characters obtained by replacing the explict vowel sign /a/ by a ₕₐₗₐₙₜₐ sign as in group (i) or in case where an explicit vowel /a/ sign is absent as in group (ii) then the ₕₐₗₐₙₜₐ sign is placed on the right hand tip of the character.

(i) క్ ఖ్ గ్ ఘ్ చ్ ఛ్ ఝ్ ఠ్ డ్ ఢ్ థ్ ధ్ ద్ ధ్ న్ ప్ ఫ్ భ్ మ్
     k kh g gh c ch jh ṭh ḍ ḍh t th d dh n p ph bh m y r
     ḷ v ś ṣ s h

(ii) ఙ్ జ్ ఞ్ ట్ ణ్ ఴ్ ల్
      ṅ j ñ ṭ ṇ r̥ l

Secondary or dependent consonant characters: They are rightly called as consonant modifiers or conjunct formatives. Unlike in Devanagari Telugu does not use half consonants but uses consonant character allographs which function as constituents of conjuncts. In a majority of the consonants the allographs which are used as secondary consonants are derived historically from the corresponding consonant characters after dropping the explicit sign for the vowel /a/ (talakattu) and placed below or at the right lower corner of the preceding consonant character.

i. ్క ్గ ్ఘ ్చ ్ఛ ్ఝ ్ఠ ్డ ్ఢ ్థ ్ధ ్ద ్ధ ్న ్ప ్ఫ ్భ ్మ ్య
    ్ర ్ల ్వ ్శ ్ష ్స ్హ
    k g gh c ch jh ṭh ḍ ḍh t th d dh n p ph bh m y
    r ḷ v ś ṣ s h

ii. ్ఖ ్ఙ ్జ ్ఞ ్ట ్ణ ్బ ్ర ్ల
     kh ṅ j ñ ṭ ṇ b r̥ l

Consonant Conjuncts : Among the scripts of Brahmi origin we can see that there are two distinct ways of forming the conjuncts: i) Consonants are conjoined linearly from left to right; and ii) arranged as a cascade of consonants in top down manner. Though both types occur in all scripts but only one of the two types are represented predominantly by any script. The former type was predominantly represented by Devanagari where as the latter is represented by Telugu and Kannada. Oriya and Malayalam belongs to the Telugu-Kannada group, while Assamese-Bengali, Gujarati form part of the Devanagari group. In a cluster of consonants either the rightmost one in the case of type (i) or logically the bottom most one as in type (ii) should carry the vowel sign. However, contrary to this in Telugu and Kannada, the top or the left most consonant is idiosyncratically marked for the vowel.

$$C_1 C_2 C_3 ... C_n V = C_1' C_2' C_3' ... C_n V \quad \text{(type i.)}$$
ex. Hi. स्त्री

ex. $C_1 C_2 C_3 ... C_n V = C_1 V C_2' C_3' ... C_n' \quad$ (type ii.)
ex. Te. స్త్రీ

Special/ancient Characters: Out of the 16 symbols for vowels, the vowel signs for long vocalic /r̄/ [ఋూ] and short and long vocalic /l, l̄/ [lV,LV] are almost obsolete or occur rarely in Sanskrit writings in Telugu script. Of the 41 consonants the post-dental africates [ča] / ఛ /, [ǰa] /ఙ and the trill /ṛa/ [ఱ] do not occur in Modern Standard Telugu and the palatal retroflex lateral [ẓ a] /ఴ/ have become obsolete long ago. In other words, these are characters no longer in current use, but which have been used historically. Their use is dependent on domain. For the purpose of inter-transliterability across Indian languages, some of these may be assigned code points. The signs /kṣa/, /ca/ and /ja/ are distinct from other characters, in the nomenclature of UNICODE they may be termed as deprecated i.e. coded characters whose use is strongly discouraged. Such characters are retained in the standard, but should not be used. The sign for /kṣa/ was inherited as part of the inventory as a default allographic representation which is phonologically equivalent to the composite symbol /ka+ṣa/. The signs for [ca] and [ja] are allographs of /ca/ and /ja/ when followed by back vowels hence need not have character encoding as in the case of other consonants. The most essential signs that occur in common core alphabet are 51.

Number signs: Telugu has inherited number signs as part of the Brahmi script. As in the case of characters, Telugu and Kannada share maximum number of similarities and have derived from the same common branch. Unlike in Devanagari using languages where compound numbers from 11 to 19 and the components involving them in higher numbers are written from left to right but their number names are counted or read from right to left, of course as in other Indo European languages. However, in Telugu as in other Dravidian languages, number signs are arranged in the same order as their corresponding number names, ex. paxihedu 'seventeen' lit. 'ten and seven'. Numbers [0-9] were given code points beginning from F1 241 to FA 250 in ISCII coding, and 0C66 to 0C6F in Unicode standards 3.0. Number signs are not very commonly used in Modern Telugu except for regularly in Telugu Calendar, astronomical guides and in non-secular literature. Recently the Road Transport Corporation of Andhra Pradesh has started using Telugu numbers signs on the number plates of their vehicles.

౦	౧	౨	౩	౪	౫	౬	౭	౮	౯
0	1	2	3	4	5	6	7	8	9

Punctuation marks: Modern Telugu uses punctuation marks which are borrowed from English. However, in the domain of religious texts Telugu may use single and double vertical bars to indicate a comma and a fullstop. Any Telugu font today must make available symbols for various punctuation marks as part of the Telugu font only and not leave them to be obtained from roman since these punctuation symbols must conform to the Telugu font style.

Hyphenation: As mentioned in section 2.2.0, Telugu uses syllabary whose constituents are primitive units like vowels, consonants with inherent vowels, and vowel or consonant modifiers, used in the formation of larger units of text. In English, hyphenation is sensitive to graphemic syllables constituted by independent alphabetic-characters. Since, in Telugu, graphemic syllables do not always constitute independent characters there is the difference between English and Telugu graphic syllable constituency. In Telugu hyphens when need to be inserted they are inserted only at the end of a graphemic syllable. A graphemic syllable is the one which ends in a vowel optionally followed by a vowel modifier and may be preceded optionally by one or more consonants as in C0-3V(M). In a sequence of graphemic syllables hyphens are inserted as in #C*V(M)-C*V(M)-C*V(M)# where a C* indicates null to any number of consonants.

For example the word śāstrīyata /శాస్త్రీయత/ can be hyphenated as shown below:

i.   śā-   not śās-
        #striyata.   triyata.
or
ii.   śāstrī-   not śāst-
        #yata.   riyata.
or
iii.   śāstrīya-   not śāstrīy-
        #ta.   ata.

However, when one or more of the consonant group involves consonant characters with an explicit halanta, then a hyphen can be inserted after such a consonant as in #C*V(M)C'*-C*V(M)C'*-CV(M)C'*#.

ex. ek_sṭrā → ek_-sṭrā [ఎక్ -స్ట్రా]
ex. en_rān_ → en_-rān_ [ఎన్_-రాన్]

Character set considerations

Collation sequence: The collation of units of textual information unambiguously has always been the source of contention. Generally, the sort order or the alphabetic sorting is the order of the position of characters in the alphabet. It is usually specific to a particular language. Though Indian languages agree in having structural similarity in the organization of characters in the alphabet they do differ in certain minor ways forcing different sort orders. Telugu follows the standard sort order as shown below:

అ ఆ ఇ ఈ ఉ ఊ ఋ ౠ ఌ ౡ ఎ ఏ ఐ ఒ ఓ ఔ ం ఁ ః _

a ā i ī u ū ṛ ṝ ḷ ḹ e ē ai o ō au m̐ ṃ ḥ _

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 ఱ ల ళ ఴ వ
శ ష స హ

ka kha ga gha ṅa
ca cha ja jha ña
ṭa ṭha ḍa ḍha ṇa
ta tha da dha na
pa pha ba bha ma
ya ra ṛa la ḷa ẓa va
śa ṣa sa ha

The allographic variations do not call for any attention here.

Alternate Collation sequences: The above order is considered as standard since it is the order that is followed in compiling lexica, dictionaries, thesauri and glossaries by most of the academic institutions and other agencies. The major dictionaries such as sūryarāya āṃdhra nighaṇṭuvu ((SAN) 6 vols., Ist issued 1936, 3rd reprint 1988, Telugu Univ.), telugu vyutpatti

padakōśaṃ; Telugu Etymological Dictionary (TED), 8 vols. 1981-95, Andhra Univ.) and the recently published telugu-telugu nighaṇṭuvu (TTN) (2001, Telugu Academy, Hyderabad) all conform to the above order. The order specified here as the standard order of characters in the Telugu alphabet differs from other non-standard orders. In elementary education when children are introduced to primary reading material in Telugu, the number and order of the characters in the alphabet deviate in a specific manner where the characters / ḷa/ [ళ] and /kṣa/ [క్ష] and sometimes /ṛa/ [ఱ] occur at the end of the alphabet. With in the standard sort order mentioned above there is a minor variation with regard to the ordering of /ṃ/ [ం]. There are at least three sort orders in vogue with respect to the sorting of /ṃ/ as exemplified by the four words in the following three orders as represented by the dictionaries mentioned above:

i.)	(SAN)	కంక,	కంస,	కక,	కస
		kaṅka,	kaṃsa,	kaka,	kasa
ii.)	(TTN)	కక,	కంక,	కంస,	కస
		kaka,	kaṅka,	kaṃsa,	kasa
iii.)	(TED)	కంస,	కక,	కంక,	కస
		kaṃsa,	kaka,	kaṅka,	kasa

The order in i.) as in sūrya rāya āṃdhra nighaṇṭuvu (SAN) is rather the standard one and conforms strictly with the alphabetical order of the character / ṃ/[ం]. The order in ii.) is based on the interpretation that the symbol /ṃ/[ం] is a cover symbol for the phonemes /n/ with its allophones {[ṅ,ñ] : [జ్ఞ], [క్ష]}, /m/ with its allophones {[m], [w̃] and /ŋ/ [with its allophone [ṇ]}. Here the sequence is interpreted as a cluster of the phonemes /ms/ phonetically [w̃ష]. In the case of iii.) the order is a mixture of interpretations of the character /ṃ/[ం] as a vowel modifier, and as an archiphoneme covering preconsonantal homorganic nasal sounds. The latter two sort orders mix up the order of graphemes in the syllabary, and phonemes of the Telugu inventory. A pure alphabetic sort order represented by i.) as in SAN is to be the preferred order.

Statistical Properties of Telugu Characters: The following are frequencies of characters in Telugu texts. A corpus of 3 million words of running texts covering

a wide range of genre viz. modern fiction, short stories, novels, science writing, childrens stories, and journalese in Telugu forms the basis for the analysis of character frequencies. Word frequencies are dropped in order to avoid their skewing effect on the results of character frequencies. There are a little over 5 lakh disticnt wordforms involved. Information regarding Phonemic frequencies are relevant particularly in keyboard layout, in designing roman notation, and even making decisions in assigning code points to certain characters (similar and comparable studies on Phonemes and Character frequencies of Telugu may be found in Kostic, Mitter, and Krishnamurti, 1977; Narasimham, et al 1981).

cumulative percentage		character frequency coverage
16.19	a	16.19
23.57	u	07.38
30.37	i	06.80
36.57	ā	06.27
42.32	n	06.25
46.95	r	04.63
51.49	l	04.54
55.54	k	04.05
59.32	ṃ	03.78
62.38	v	03.06
65.42	t	03.04
68.37	m	02.95
71.31	p	02.94
73.69	d	02.38
75.93	s	02.24
78.09	y	02.16
80.14	c	02.05
82.18	ē	02.04
84.26	ṭ	01.98
86.13	g	01.97
88.07	ḍ	01.94
89.96	ō	01.89
91.13	e	01.17
92.13	ī	01.00
92.95	ū	00.82
93.72	b	00.77
94.37	j	00.65
95.01	ś	00.64
95.62	o	00.61
96.17	ai	00.55
96.70	ṣ	00.53
97.19	dh	00.49
97.64	ṇ	00.45
98.06	h	00.42
98.49	bh	00.43
98.86	ḷ	00.37
99.09	th	00.23
99.27	ṟ	00.18
99.43	Z	00.16
99.55	ph	00.13
99.68	kh	00.13
99.76	au	00.08
99.81	gh	00.05
99.85	ṭh	00.04
99.89	ñ	00.04
99.92	ṛ	00.03
99.95	ch	00.03
99.97	ḥ	00.02
99.99	ḍh	00.02
100.00	m̆	00.01
100.00	jh	00.00
100.00	ṅ	00.00
100.00	_	00.00
100.00	ṝ	00.00

Coverage of Vowel characters: Vowels alone have a coverage of 43.16% i.e. a considerably higher percentage when compared to those in other languages.

16.19%	a	16.19%
23.57%	u	07.38%
30.37%	i	06.80%
36.64%	ā	06.27%
38.68%	ē	02.04%
40.57%	ō	01.89%
41.74%	e	01.17%
42.35%	o	00.61%
42.90%	ai	00.55%
43.08%	ṛ	00.18%
43.16%	au	00.08%

Vowel modifiers have a coverage of 3.81%. The addition of the coverage of vowel modifiers to the total coverage of vowels raise the total to 46.97%.

Coverage of consonants based on their frequencies amount to little more than fifty per cent i.e. 53.07%. The sonorants /n, r, l/ constitute more than 15% of

the total coverage. The next set of consonants /k,v, t, m, p/ has the coverage of 15%. The remaining 28 consonant characters have only a coverage of 23%. The contribution of character signs of aspirate phonemes account only for 1.51%.

Keyboard

Layout: If one looks at the way the existing keyboard layouts for Telugu are presented, they reveal that they are not ergonomically planned, since the distribution of characters are not balanced according to the well known keyboard layout recommendations. Government of Andhra Pradesh has approved (vide. G.O. 391, GA.OL-2, 1987) a Standard keyboard for Telugu/English for use in Electronic typewriters, wordprocessors, Teleprinters, Phototype setters, and in computers. Though, the designers of the keyboard claim to combine phonetic order and frequency of use of characters, the actual recommended keying-in operation sequence itself is not in phonetic order instead follows the idiosyncratic convention practiced in writing Telugu orthography.

Ministry of Information Technology (then DoE), Government of India has brought out a keyboard standard for keying in of Indian characters known as Inscript keyboard (report of the Committee, SKLISBC, DOE 1986) Later it was revized in 1988 for the consideration of proper sorting of 'nukta' characters and certain special set of front and back mid vowels. It allows the funtioning of an Inscript overlay on any Querty keyboard by pressing caps-lock key. The inscript overlay combines to a certain extent the logical order of Indian alphabets and the frequency of use of these characters.

Besides these a considerable number of keyboard layouts for Telugu are in use in the market, many of them are adopted from "typewriter" keyboard-layouts (Ramington, Facit etc.) and they do not match one another. One such keyboard developed under the joint project of CMC, Secunderabad and RIND, Madras (Narasimham, Ramakrishna Rao, 1981) for Telugu is claimed to be based purely on frequency of occurrence of characters in written texts without regard to their logical order. However it has never become popular.

Recently, a sub-committee, appointed by The Government of Andhra Pradesh has recommended Inscript Keyboard layout for Telugu as it is popularly used in most Indian languages, since it also incorporates both the logical order of the Indian characters plus certain amount of optimization with respect to frequency of use of the characters, and the phonetic order of keying. It also envisages that the use of Telugu keying-in method will evolve and settle in favour of inscript keyboard layout with minimum modifications, since the users are expected to use the same for multilingual purposes.

Transliteration: Wide spread bilingualism among Indian language mother tongue speakers, and English being the most popular among the literates, justifies the demand that Indian languages are often transcribed in roman. A roman transliteration has also evolved one representing the South and the other the North. Examples can be drawn from one of the major industry the film production which transliterates the titles from the respective languages into roman for the benefit of other Indian language speakers. Particularly for the creation of corpora, where a particular Indian language fonts are not available, having a recognized standard roman transliteration is of great help. There are already a couple of roman transliteration schemes employed by Telugu linguists, Technology Resource Centre for Telugu, and Anusaraka Machine Translation group. The latter group has used a roman transliteration scheme popularly known as WX-notation where in /t, d/ are used to transcribe retroflexes [ṭa,ḍa] and /w, x/ to denote dental stops [t, d]. The other group follows a more commonly used notation which uses T and D for voiceles and voiced retroflex stops and t and d for dentals.

WX-notation:

a A i I u U q Q eV e E oV oO z M H
అ ఆ ఇ ఈ ఉ ఊ ఋ ౠ ఎ ఏ ఐ ఒ ఓ ఔ ం ఁ ః

k K g G f c C j J F t T d D N w W x X n p P b B m
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

y r rY l lY lYY v S R s h _
య ర ఱ ల ళ ఴ వ శ ష స హ ఁ

TD-notation:

a aa i ii u uu r' r" e ee eY o oo oW z M H
అ ఆ ఇ ఈ ఉ ఊ ఋ ౠ ఎ ఏ ఐ ఒ ఓ ఔ ం ః

k kh g gh ng c ch j jh nY T Th D Dh N t th d dh
క ఖ గ ఘ జ్ఞ చ ఛ జ ఝ ఞ ట ఠ డ ఢ ణ త థ ద ధ న

n p ph b bh m
న ప ఫ బ భ మ

y r R l L LY v S Sh s h _
య ర ఱ ల ళ ఴ వ శ ష స హ క్

There are added advantages and disadvantages with both the notations but the WX-notation has minimum number of successive strokes for the same character, and same letter is never used again to represent another one even as a component stroke.

Character composition: Telugu text like in any script of Brahmi origin, is composed of a series of syllables composed of stand alone characters listed in the alphabet and the corresponding secondary symbols. Telugu (like Kannada) and unlike any other Indian script of Brahmi origin, composes compound graphic syllables consisting of two or more consonants and followed by a vowel modifier in a very idiosyncratic manner. Compound syllables involving conjunct consonants of two or more have a vowel modifier which is always attached to the first (primary) consonant symbol of the conjunct cluster as shown in the examples here: strī /sītr/ (స్త్రీ), pārku /pAruk/ (పార్కు), pārk /pArk/ (పార్క్), pārts /pArts/ (పార్ట్స్), pārṭsu /pāruṭs/ (పార్ట్సు), kārl mārks /kārl mārks/ (కార్ల్ మార్క్స్), harṣ /harṣ/ (హార్ష్), klēśaṃ /klēśaṃ/ (క్లేశం), spriṅg /spriṅg/ (స్ప్రింగ్) etc. (the words spelled in roman within the slashes display Telugu character sequence).

Character values and sequence validation: Syllable composition assumes four distinct approaches with respect to character-key relationship and keyboarding as following: i) Both primary and secondary symbols of vowels and consonants have separate keys on the keyboard; ii) Primary and secondary symbols vowels and only the primary symbols are assigned keys and the secondary symbols of consonants are obtained contextually; iii) Only primary symbols of vowels and consonants are assigned keys and secondary symbols are obtained

through context; and iv) primary and secondary symbols of consonants and only primary symbols of vowels are assigned keys. Besides the above key assignments, there are certain sequencing generalizations that are assumed in actual implementation scheme: a) Every consonant character has an inherent vowel /a/ and it is retained only before a space or a punctuation symbol, in otherwords, it is always deleted before a vowel or a consonant, which is statable as rule1.

Rule1. $a \longrightarrow 0 /C\_ +\{V,C\}$

Alternatively, scheme b. involves the keying in of consonant which introduces only pure consonant with an explicit halanta sign hence does not require the operation of rule 1. The following exemplifies the implementation in terms of number of keys and the resulting display.

Scheme a.

క+ఇ -→ కి	ka+i → ki
క+అ+ఇ → కఇ	ka+a+i → kai
క+క -→ క్క	ka+ka → kka
క+అ+క -→ కక	ka+a+ka → kaka

Scheme b.

క+ఇ -→ కి	k+i → ki
క+అ+ఇ → కఇ	k+a+i → kai
క+క+అ → క్క	k+k+a → kka
క+అ+క+అ →కక	K+a+k+a → kaka

Obviously, the implementation of sheme a. has an advantage over scheme b. in removing the redundancy in keying by 20% utilizing the distributional generalizations among vowels and consonants particularly in Telugu and in all Indian languages in general. Midway to the schemes a. and b. there is another scheme of implementation which assumes keying in of consonant with an inherent vowel /a/ but that is not automatically deleted by a following vowel or a consonant, but it has to be removed by pressing a halant key.

Scheme c.

క+్+ఇ → కి	ka+i → ki
క+ఇ → కఇ	ka+a+i → kai
క+్+క → క్క	ka+ka → kka
క+క → కక	ka+a+ka → kaka

Schemed.

క+ ి → కి   ka+ _i → ki

క+ఐ → కఐ   ka+ I → kai

క+○+క → క్క   ka+_+ka → kka

క+క → కక   ka+ka → kaka

The schemes are generally implemented in combination with the assumed character values and sequence validation. Scheme a. is currently implemented in ILEAP with phonetic keyboard, scheme d. is implemented in Ileap using inscript kb which assumes different keys for primary and secondary vowels. In the last scheme the number of key strokes are same as in scheme a. Scheme c. which is implemented in a Telugu editor called Telugu-Lekha (a product of BCT Ltd.) is a mirror image of scheme a. Scheme b. is implemented in an editor called Telugu-Lipi (Srinivas, S and Anuradha, K.). Of all the four implementations, the scheme a. (Ileap with phonetic keying) optimally utilizes the best possible features for keyboarding.

Keying in Sequence: There are two orders of composing or keying in the components of the graphic syllables in Telugu as described below:

Conventional Method: As handed down to us in traditional practice of writing saMyuktAksharas (graphic syllables consisting of conjuncts - type the first consonant of the conjunct cluster as a primary sign with the relevant vowel modifier and the rest of the consonants in their secondary form, placed one after another in sequence to the bottom right of the primary consonant. In this method one need to conceptualize the correct formation of the syllable and then write it on the paper or compose it or key in on the machine. Typing frequently occurring syllables may not require any time but syllables of less frequently occurring ones may take more than a moment of thinking, compare:

ex.   స్త్రీ = స - సీ - సీత్ - సీత్ర
      strī = sa - sī - sīt - sītr

keying sequence	→	display sequence	::	phonetic equivalence
స	→	స		[స]
sa	→	sa		[sa]
స + ఈ	→	సీ		[సీ]

sa +ī	→	sī		[sī]
సీ + త	→	సీత్'		[స్తీ]
sī + ta	→	sīt		[stī]
సీత్' + ర	→	సీత్'ర'		[స్త్రీ]
sīt' + ra	→	sīt'r'		[strī]

స్తూక్ర్ = స - సూ - సూక్ - స్తూక్ర్

sūkr = sa - sū - sūk' - sūkr

[skrū] = [sa] - [sū] - [skū] - [skrū]

2.6.5.2 Phonetic Method: The composition of saMyuktAksharas follow from linear order of utterance or pronunciation. Here one need not learn or put a special effort in composing of conjunct clusters. There is naturalness involved here and one need only pay attention to one's pronunciation.

@ex. స్త్రీ = స - స్త - స్త్ర - స్త్రీ

      strī = sa - sta - stra - strī

keying sequence	→	display sequence	=	phonetic equivalence
స	→	స		[స]
sa	→	sa		[sa]
స + త	→	స్త		[స్త]
sa + ta	→	sta		[sta]
స్త+ ర	→	స్త్ర		[స్త్ర]
sta+ ra	→	stra		[stra]
స్త్ర + ఈ	→	స్త్రీ		[స్త్రీ]
stra + ī	→	strī		[strī]

స్తూక్ర్ = స - స్క - స్క్ర - స్తూక్ర్

skrū = sa - ska - skra - skrū

Following method 1) results in the reduction of speed, since composers/writers have to recall from their memory for a moment. It also results in wrong sorting of words, and create problems for spell checking and searching operations. Therefore method 2) which is based on phonetic order is preferred.

Glyphs: There have been a number of attempts at reforming Telugu script (Andhra Pradesh Government, G.O.194:1961), particularly to suit to the needs of the letter press printing. Telugu script composers and typists are expected to memorize hundreds of glyphs and scores of their combinations to form meaningful graphic syllables. The attempts

to reform the Telugu script were basically aimed at reducing the burden on the part of the composers and typists by reducing the variation in the number of glyphs and eliminating the combinations which are considered to be illogical.
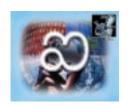
Variation in glyph combinations: In Telugu, certain compound graphic syllables beginning with [మ, య, ష, రు] and modified by a vowel mātrā [ొ or ో] are alternatively written as [మొ, యొ, షి, రొ or as మా, యా , షా , రుా] and [మో, యో , షా, రో or మో , యో , షో రుో]. This variation shall be resolved by preferring to write in the second alternate way since it involves the same glyph [ొ or ో] = secondary symbols for short /o/ and long /ō / used with all other consonants. It has one immediate advantage besides regularity. The use of first alternate combination are interpretable glyptographically as combinations of [ఎ+ు] /e+u/ [ఎ+ూ] /e+ū/ and may create problems in sorting and searching operations.

Glyphs in Character Composition: The Telugu script is distinct from the other scripts of Brahmi origin not only by the rounded shape of the letters but also by the sheer number of glyphs used in Character formation. To illustrate, the mātrā for long [A] /ఆ/ i.s. |ా| is represented at least in five distinct ways due to the difference in size, shape of the consonant and its placement as shown below:

Consonant class	Glyph composition	Glyph placement
1. [గ ర డ ఢ త థ ద ధ న ర వ శ]	కా	Represents a regular substitution of [ˇ] of the anchor consonant by the mAtrA [ా].
2. [ఘ రు మ య]	ఘా	The placement of [ా] mAtrA does not replace the [ˇ] of the anchor consonant.
3. [ఖ ఞ జ ఙ ట ణ బ ల ఞ]	ఖా	The [ా] mAtrA is placed on the top right half of the anchor consonant.
4. [చ ఛ భ]	భా	The mAtrA [ా]is placed on the top right edge of the anchor substituting [ˇ].
5. [ప ఫ స ష]	పా	A change in the shape of the anchor consonant and the [ా] mAtrA runs across the consonant.
6. [హ]	హా	A change in the shape of the mAtrA [˚] and its placement does not substitute [ˇ].

More or less such variation in mātrā size may be found with all other vowels (except in case of /a/) and the associated idiosyncracies dependent on the anchor consonant are a common phenomenon in Telugu (and also in Kannada) but not commonly attested in other Indian scripts. The outcome of this phenomenon is the glyph chart for Telugu will be more complex than that of any other Brahmi derived script. Traditionally the composers or typists of Telugu script are expected to memorize these mātrā variants (which roughly range on an average five per each vowel mātrā amounting to sixty glyphs) besides scores of the associated consonant variants and their combinations. The average number of glyphs for each consonant character is not less than 3. The total number of glyphs for all characters would be well beyond 150, which is much higher than for any other Indian language. Because of this reason it is not easy to achieve glyph standard for Telugu. The mechanical constraints of the type-writers which lack glyph matching and composition management facility have forced us to live with the poor quality of mechanical type faces in Telugu.

However, with the advent of Computers and their appropriate adaption for Telugu Script, the composing of Telugu writing and printing have become much easier. The entire burden of dealing with the complexity of glyph selection, matching and composition has now shifted from man to machine.

The entire scenario i.e. the number of glyphs and their idiosyncratic combinations that the Telugu composers/writers had to remember, now has transformed into keying in the relevant characters representing the sounds in the phonetic sequence (linear) in a word. The computer (driven by the relevant

print/display rendering routines) puts the relevant glyphs together and renders them into an appropriate graphic display on the screen. This is one of the best examples of the application of computer technology in eliminating the complex problems associated with nonlinear scripts as in the case of Telugu script.

Fonts: Characters are represented visually as glyphs accopanied by a set of parameters viz. size, posture, thickness etc. One of the major differences between roman and Indic alphabets is that the latter has glyphs whose size is highly variable and characters are compositional. In roman fonts we see often one to one association between a character from the alphabet and the corresponding glyph. Not only it is hard to find such an association between a character and a glyph but in actual practice, as in Telugu, glyphs are often reduced to fewer number of primitive geometric shapes from which characters can be composed. So the size of the glyph inventory of Telugu from one font to another is so vast that it is often fruitless to accomplish.

Storage considerations: Among the available systems that are in the market for rendering Telugu characters on Computers, only one i.e. ISCII (@BIS 1991, its earlier version was ISSCII-1983) has been somewhat widely used both in Unix and MS-DOS world. All others are based on commonly available methods employed commercially for working with Telugu language fonts and DOS based graphics programs. No other Character coding scheme than ISCII is known to exist. Most of the commercially available Telugu software program store their own table of glyph/character pictures or by installing characters in the computer's video memory. The lack of common standard in glyph/character coding of Telugu is a potential source of frustration in Telugu computing due to mismatch between the glyphs/ characters and print handlers when accessed on screen

and printed using other than the one used in the creating the document.

The theoretical basis for the analysis of Telugu script: Words in Telugu script are composed of one or more graphic syllables of either (a) simple or (b) compound type. The graphic syllables are unlike phonetic/phonological syllable may not have a vowel as a compulsory component but will have a secondary vowel/halant symbol as a necessary component. This assumption is necessitated by the occurrence of the graphic representations of halantākṣaras as /ś/ [ఴ̄] in [āṃdhrapradēś] /Andhra Pradesh/, /d/ [ద̄] in [prasād] 'Prasad', /rl/ [ర̄ఴ] and /rks/ [ర̄ఴ̄] in /kārl mārks/[కార్̄ మార్̄ఴ̄] 'Karl Marx'. a.) A simple graphic syllable is always comprised of one of the signs for primary vowels /a ā i ī u ū ṛ ṝ e ē ai o ō au/ [అ ఆ ఇ ఈ ఉ ఊ ఋ ౠ ఎ ఏ ఐ ఒ ఓ ఔ] or one of the three primary symbols: the pūrṇānusvāra [ం], the ardhānusvāra [ఁ] and the visarga [ః] or one of the symbols for primary consonants with implicit vowel /a/ [̆]. Also recall that the traditional varṇamāla groups these three signs along with vowels. As stated above a graphic syllable here may be composed of non-vocalic signs but which have independent and linear representation in Telugu writing. All other graphic syllable representations are interpreted as compounds. b) A compound graphic syllable may be interpreted as one or more consonant signs (there is no upper limit) plus an inherent or covert (implicit) vowel [Ã] or an overt (explicit) secondary vowel symbol /ā i ī u ū e ē ai o ō au and _ / [ా ి ీ ు ూ ృ ె ే ై ొ ో ౌ and ్].

Differentiating Script from Language: Script is not coterminus with language. It is only a means for visual representation of a spoken language. Since a script is associated with a given language for over a millennium the grammar of a script is often measured with the same yardstick as that is used for the language. It is true that the historical development of Telugu language (the writing conventions and the tools used for writing) has reflections on the Telugu script. However developing languages change faster than the script that is being used for that language. Therefore phonological rules that are proper to a language shall not be imposed upon the script.

The computer technology has rendered all earlier arguments for reforming Telugu script into vaccuous and needless. The standard argument now remaining is the maintenance of strict correspondence between the phonetic order and the keying-in order, thus breaking the requirement of the correspondence between the language(script) dependent displays and the order of the keying-in. This one particular feature has a great significance allowing us to view all Indian scripts as mere 'font' variants. Far from reducing the level of script to a font, it shows the grand unification among all Indian scripts. The computer technology has also one interesting aspect. Unlike the reformists who would like to drop certain infrequent and obsolete characters in order to save manual labour, the current technology not only makes allowances for such characters but also makes available new representations in the Indian scripts particularly to represent those special sounds borrowed frequently from Perso-Arabic and European languages hence the extended alphabet.

Localization of Data

Calendar : There are standard ways of expressing units of temporal space. The date is usually expressed in the order of month date, year i.e. month precedes date, and date precedes years, as in August 15, 1947 the Indian Independence day. However, the same when stated in the form of numbers the order is dd-mm-yy i.e. 15-08-1947.

Months : The names of months borrowed from English are commonly employed in Telugu in all purposes. These names are often abbreviated and prefixed to the date:

janavari	ja.	January
phibravari	phi.	February
mārci	mā.	March
ēpril	ē.	April
mē	mē.	May
jūn	jū.	June
julai	ju.	July
āgaṣṭu	ā.	August
sepṭembaru	se.	September
akṭōbaru	a.	October
navambaru	na.	November
ḍisembaru	ḍi.	December

ex. ā. 15 12:15:15 bhā.kā. 1947 'August 15th 12:15:15 IST.'

Week : Usually week days are abbreviated in date expressions and precede the month as in ādi. āgaṣṭu 15, 1947.

ādi.	for	ādivāraṃ	'Sunday'
sō.	for	sōmavāraṃ	'Monday'
maṃ.	for	mangaḷavāraṃ	'Tuesday'
bu.	for	budhavāraṃ	'Wednesday'
gu.	for	guruvāraṃ	'Thursday'
śu.	for	śukravāraṃ	'Friday'
śa.	for	śanivāraṃ	'Saturday'

Time : English expressions AM/PM do not have equivalent expressions in Telugu. In the expressions of time, generally, the hours are usually prefixed with abbreviations standing for different time periods in a day. A day of 24 hours is divided into five divisions (pUtalu) as shown below with a roughly corresponding english equivalent.

udayaṃ:	u.	Morning	06:00:01-11:59:59
madhyāhnaṃ:	ma.	Noon	12:00:01- 03:59:59
sāyantraṃ:	sā.	Evening	04:00:01- 06:59:59
rātri:	rā.	Night	07:00:01-12:59:59
tellavārujāmu:	te.	midnight	01:00:01- 05:59:59

The date Sat Nov 17 14:44:05 IST 2001 may be expressed in Telugu as śa. na. 17 14:44:05 bhā.kā. (bhārata kālamānaṃ) 2001.

Currency : There is no specific currency symbol current in Telugu. However, the word రూపాయి is prefixed to the amount in an abbreviated form 'rū. as below:

rū. 5,431.12

Rs. 5,431.12

Appendices

References:

Kostic, D., A. Mitter, Bh. Krishnamurti. 1979. A Short Outline of Telugu Phonetics (particularly pp. 202-204 on Phoneme frequencies): Calcutta: ISI.

Krishnamurti, Bh. and J.P.L. Gwynn. 1985. A grammar of Modern Telugu (particularly Chapter 5: The structure of Telugu Orthography: Problems of Reform). Delhi:OUP.

krishnamurti, Bh. 1997. telugu lipi nirmāṇaṃlō unna niyamālu. Telugu, 4.1-5. Telugu Akademi.

Narasimham, P.V.H.M.L., G.L. Narasimham, and G. Ramakrishna Rao.1981. Design Information on Text Composition in Telugu (a report of a joint project of Computer Maintenance Corporation Limited (CMC), Secunderabad, and Research Institute for Newspaper Development (RIND), Madras.

Uma Maheshwar Rao, G., Rajeev Sangal, P.V.H.M.L. Narasimham, S. C. Babu, J. Satyanarayana. 2001. Subcommittee report on Standards for the Implementation of Telugu in Information Technology. Govt of Andhra Pradesh.

(Courtesy : Prof. K. N. Murthy
University of Hyderabad,
Hyderabad-500046
Tel. 040-3010500 Extn. 4056
E-mail : knmcs@uohyd.ernet.in
&
Dr. G. Umamaheshwar Rao
Reader,
Centre for Applied Linguistics
and Translation Studies
University of Hyderabad.
E-mail: guraosh@uohyd.ernet.in)

4.5.3 Typical Colloquial Sentences in Telugu

GREETING

◗ Hello
హలో
हलो
halō

◗ Good Morning
నమస్కారం
नमस्कारं
namaskāraṃ

◗ Good Afternoon
నమస్కారం
नमस्कार
namaskāraṃ

◗ Good Night
నమస్తే
नमस्ते
namastē

◗ Good Bye
నమస్తే
नमस्ते
namastē

◗ Thanks
కృతజ్ఞతలు
कृतज्ञतलु
kṛtajñatalu

◗ How are you
ఎలాఉన్నారు?
ऐलाउन्नारु?
elāunnāru?

◗ I am fine thank you
బాగానే ఉన్నాను, కృతజ్ఞుణ్ణి.
बागाने उन्नानु, कृतज्ञुणि.
bāgānē unnānu, kṛtajñuṇṇi.

◗ Sorry
క్షమించండి
क्षमिंचंडि
kṣamiñcaṇḍi