

TeluguTeX*

Lakshmi V. S. Mukkavilli

There are seven sections in this article. The first section explains the Telugu script. For someone who can speak Telugu this section should help learn Telugu script. On most of the computers there is no facility for inputting Telugu text. We propose a romanization scheme for inputting Telugu. This facilitates text entry in Telugu. This is the subject of the second section. We provide examples of Telugu text typeset using our system in the third section. In the fourth section we explain the problems involved in typesetting telugu text. The fifth section deals with implementation and the sixth section contains samples of telugu text typeset using various sizes/styles. Last section summarizes our work.

1 Telugu Script

Like English, Telugu is written from left to right. Telugu text consists of sentences. A sentence is a sequence of words. A word is a sequence of syllables called Aksharas. An Akshara can be of two types.

1. a stand-alone vowel
2. Consonant + Consonant + ... + Consonant + Vowel

Let C_1 be the first consonant, C_2 be the second consonant and C_n be the n^{th} consonant. Let the vowel be denoted by V .

An Akshara is composed as follows: C_1 is considered the base consonant. C_2 to C_n are called consonant conjuncts. Consonant conjuncts are optional. The Vowel modifies the base consonant. The vowel modifier may be absent. Such a situation is indicated by a symbol called *pollu* or *halant* (𑌖). When this symbol is attached to a consonant we get what is called the half consonant or the pure consonant (so called because of the absence of any vowel sound). The form of Base Consonant + Vowel (henceforth referred to as C+V) varies depending on both the consonant and the vowel. The vowel modifiers can appear to the right, on the top or at the bottom of the base consonant. In many cases a completely different symbol is needed. After Consonant + Vowel is formed, consonant conjuncts are attached in the order they appear in the syllable. Consonant Conjuncts are placed to the right or below C+V. Consonant Conjuncts have a different form than either base consonant or C+V's.

An Akshara can optionally have accent(s). Accents could appear at the top or on the side or at the bottom of a syllable.

Table 1 lists stand-alone vowels and corresponding modifiers. The blob of ink stands for consonant base. The table indicates the placement of vowel modifiers in relation to base consonant. But one should understand that there are very many exceptions.

*©1991 Lakshmi V. S. Mukkavilli

Vowel	Vowel (in roman)	Vowel Modifier
అ	a	■
ఆ	A	■
ఇ	i	■
ఈ	I	■
ఉ	u	■
ఊ	U	■
ఋ	ro	■
ౠ	Roo	■
ౡ	lo	■
ౢ	Loo	■
ఎ	e	■
ఏ	E	■
ఐ	y	■
ఔ	o	■
ౠ	O	■
ౡ	ow	■

Table 1: Vowels in Telugu

Table 2 presents various forms of consonants. In the first column consonants appear in pure form. In the second column consonants appear with an implicit అ. The last column indicates the shape of consonant conjuncts and the positioning of consonant conjuncts. One may recall that a syllable can have several consonant conjuncts. There is no consonant conjunct form for ఌ.

Table 3 presents the accents used in Telugu text. The first column indicates the shape of accents. The second column indicates where the accents are placed relative to the syllable. The accents that go at the top or at the bottom appear over/below the base consonant. A syllable can have multiple accents. Many of the accents presented in the table do not appear in ordinary Telugu text. They occur mostly in Sanskrit text transliterated in Telugu.

These days most Telugu texts use Hindu-Arabic digits. But there are digits that are native to Telugu. These are presented in Table 4.

As we have said when a base consonant combines with a vowel the base consonant undergoes a transformation and takes a new look. There are no rules that are universally applicable. In table 5 we provide the complete picture. In this table you can find the entry for every C+V combination.

Consonant (without “a” modifier)	Consonant (with “a” modifier)	Consonant (in roman)	Consonant Conjunct
క	కా	k	కా
ఖ	ఖా	kh	ఖా
గ	గా	g	గా
ఘ	ఘా	gh	ఘా
ఙ	ఙా	NG	ఙా
చ	చా	c	చా
ఛ	ఛా	ch	ఛా
జ	జా	j	జా
ఝ	ఝా	jh	ఝా
ఞ	ఞా	nx	ఞా
ట	టా	T	టా
ఠ	ఠా	Th	ఠా
డ	డా	D	డా
ఢ	ఢా	Dh	ఢా
న	నా	N	నా
త	తా	t	తా
థ	థా	th	థా
ద	దా	d	దా
ధ	ధా	dh	ధా
ప	పా	p	పా
ఫ	ఫా	ph	ఫా
బ	బా	b	బా
భ	భా	bh	భా
మ	మా	m	మా
య	యా	Y	యా
ర	రా	r	రా
ఱ	ఱా	R	ఱా
ల	లా	l	లా
ళ	ళా	L	ళా
వ	వా	v	వా
శ	శా	S	శా
ష	షా	Sh	షా
స	సా	s	సా
హ	హా	H	హా
ఖ	ఖా	Xh	ఖా

Table 2: Consonants in Telugu

Accent	Usage	Comments
◁	◁	arasunna
○	◉	sunna
◌ː	◌ː	visarga
◌ˑ	◌ˑ	ardhavisarga
◌॒	◌॒	avagraha
◌̣	◌̣	ఙ̣, ఙ̣̣
◌̤	◌̤	nakarapollu
◌̥	◌̥	valapalagilaka
◌́	◌́	udatta(one bar)
◌̂	◌̂	udatta(two bars)
◌̄	◌̄	anudatta

Table 3: Accents (used to modify syllables)

౦	zero
౧	one
౨	two
౩	three
౪	four
౫	five
౬	six
౭	seven
౮	eight
౯	nine

Table 4: Telugu Digits

There are two symbols (| & #) that are used as sentence/line delimiters in poetry and in Sanskrit text.

Table 6 displays all the symbols in the font. The font is available in several styles and sizes. All the symbols that also exist in ASCII character set are in the same position as in ASCII. The codes for non-ASCII symbols are determined by us because there are no established standards. But that does not mean that they can not be changed. All the codes are referred to by using variable(symbolic) names. All these variable assignments are placed in one file. If a different code assignment is desired, then it is a simple matter of changing the numbers in this file. Then METAFONT programs should be rerun. As explained later, when we generate a font, a file containing character codes (among other things) is generated. Text composition software and TeX read character codes from this file. In summary, if we want to change the codes for symbols in the font, we need to do that in one place and our implementation propagates the changes across the system.

2 Transliteration of Telugu Text

Since most computers do not have facilities for entering Telugu text and are designed to work with English input we have developed a scheme for inputting Telugu text in English. Typically text entry for Indian languages (most of them have similar phonetic structure) is done in one of two ways. One is called the graphical approach. And the other is called phonetic approach. We do not provide a detailed discussion of these approaches because these approaches are widely debated and documented. In the graphical approach, each syllable is viewed as a collage of various symbols. Keystrokes are used to place constituent symbols on the screen. This method is clumsy, inefficient and language specific. This is particularly true of Telugu. In the phonetic approach, text is entered the way we speak. We do not worry about composition at all. That is left to the computer. In other words we key in constituent consonants/vowel of the syllables that we speak. This approach to inputting Indian(Asian) languages is very elegant, simple and language independent. We have chosen the phonetic approach to inputting Telugu text. This approach makes life easy for the user but for the software developer there is considerable work.

We modified slightly the transliteration scheme for inputting Telugu proposed by Prof. Donald Becker at the University of Wisconsin, Madison. The modifications result mainly from our decision not to use any non-alphabetic characters in the scheme and inclusion of many new characters. Table 7 presents the transliteration system used by us for inputting Telugu text. The English letters are chosen so that they are close to their Telugu equivalents in pronunciation. But one should remember that English is not a phonetic language and hence the same letter can be pronounced in several ways. No transliteration is provided for some rarely used symbols. They can be obtained by using control sequences given in the table. In the next section we provide several examples of Telugu text.

In the transliteration scheme, some letters require inputting two or more characters. But this does not mean that multiple keystrokes are really needed. Most editors/word processors provide facilities for keyboard macros. We can define macros with one letter names. In Telugu, we have 36 consonants and 16 vowels. That means a total of 52 letters. We can remap the alphabet part of a keyboard to facilitate Telugu input. One can buy a keyboard overlay (or skin) and mark the keytops with Telugu letters. Thus we can have a Telugu keyboard (almost!). One can map control characters to accents. By using a pair

	'0	'1	'2	'3	'4	'5	'6	'7
'00_		”	“	—	—	,	=	-
'01_	<	o	:	:	2	o	o	9
'02_	3	4	5	6	7	8	9	0
'03_	అ	ఇ	ఈ	ఉ	ఊ	ఋ	ౠ	ౡ
'04_	ౣ	!	”	#	\$	%	&	'
'05_	()	*	+	,	-	.	/
'06_	0	1	2	3	4	5	6	7
'07_	8	9	:	;	<	=	>	?
'10_	@	1	2	3	4	5	6	7
'11_	అ	ఐ	ఋ	ౠ	ఌ	ౡ	ఔ	ౣ
'12_	ఇ	ఋ	ౠ	ఌ	ౡ	ఔ	ౣ	౤
'13_	఑	ఒ	ఓ	[\]	^	_
'14_	ఘ	ఙ	ఞ	ట	ఠ	డ	య	ఠ
'15_	ఋ	ౠ	ఌ	ౡ	ఔ	ౣ	౤	౥
'16_	ఐ	ఋ	ౠ	ఌ	ౡ	ఔ	ౣ	౤
'17_				{		}	~	
'20_	ఁ	ం	ః	ఄ	అ	ఆ	ఇ	ఈ
'21_	ఉ	ఊ	ఋ	ౠ	ఌ	ౡ	ఔ	ౣ
'22_	఑	ఒ	ఓ	ఐ	ఋ	ౠ	ఌ	ౡ
'23_	ఔ	఑	ఒ	ఓ	ఐ	ఋ	ౠ	ఌ
'24_	ౡ	ౢ	ౣ	౤	౥	౦	౧	౨
'25_	౩	౪	౫	౬	౭	౮	౹	౺
'26_	౻	౼	౽	౿	౾	౿	౾	౿
'27_	క	ఖ	గ	ఘ	ఙ	ఞ	ట	ఠ
'30_	డ	ఢ	ణ	ఠ	డ	ఢ	ణ	ఠ
'31_	డ	ఢ	ణ	ఠ	డ	ఢ	ణ	ఠ
'32_	డ	ఢ	ణ	ఠ	డ	ఢ	ణ	ఠ
'33_	డ	ఢ	ణ	ఠ	డ	ఢ	ణ	ఠ
'34_	డ	ఢ	ణ	ఠ	డ	ఢ	ణ	ఠ
'35_	డ	ఢ	ణ	ఠ	డ	ఢ	ణ	ఠ
'36_								
'37_								

Table 6: Font Table

అ	a	ఆ	A	ఇ	i	ఈ	I		
ఉ	u	ఊ	U	ఋ	rx	ఋ	Rx		
ఋ	lx	ౠ	Lx	ఎ	e	ఏ	E		
ఌ	y	ఒ	o	ఓ	O	ఔ	ow		
ఱ	Mh	ఱ	M	ఱః	ah	ఱః	\zzBB	ఱః	\zzFF
క	k	ఖ	kh	గ	g	ఘ	gh	ఙ	NG
చ	c	ఛ	ch	జ	j	ఝ	jh	ఞ	nx
ట	T	ఠ	Th	డ	D	ఢ	Dh	ణ	N
థ	t	థ	th	ణ	d	ణ	dh	న్	n
ప	p	ఫ	ph	బ	b	భ	bh	మ్	m
య	Y	ర	r	ల్	R	ల్	l	ల్	L
వ	v	శ	S	ష	Sh	ష	s	హ	H
ఋ	Xh	ఱ	\zzCC	ఱః	\zzCF	ఱః	\zzCJ		
ఱ	\zzBF	ఱ	\zzBJ	ఱ	\zzFJ				
ఱ	\zzFB	ఱ	\zzFC	ఱ	X				
ఱ	\zzJB	ఱ	\zzJC	ఱ	\zzJF	ఱ	\zzQB		

Table 7: Transliteration Scheme for inputting Telugu text

of macros we can switch between Hindu-Arabic and Telugu digits. So regardless of which digits one wants one would use the usual numeric keys.

TeX ignores spaces after control words. So if a word ends in a control word then we should type a control space (the escape character followed by a blank space) following the control word. This rule is useful for us particularly when we are entering Sanskrit text in Telugu. Many words in Sanskrit end in control sequences denoting various accents.

Use the macro \QQQ to enable transliteration and the macro \Q to disable transliteration. Use the macro \zzCB to switch to roman digits(this is the default) and use the macro \zzBC to switch to telugu digits.

What if somebody does not like the scheme that we have used? This is really no problem, since the software that interprets the user's input is separate from text composition software. So it is easy to adapt to another unambiguous transliteration scheme.

3 Examples

In this section we will give several examples of Telugu input/output. We use (,) to delimit syllables or Aksharas. We will also provide transliteration of each example. Table 8

హరిః ఓమ్॥ శతం జీవ శరదో వర్ధమానశతం హేమనాశ్చతము వసనాః। శతమీన్ధాగ్ని నవితా బృహనృతిశ్చ
తామషా వ్రావిషేమం పునర్దుః॥ ఋక్సంహితా॥ ౮-౮-౨౦ వగః॥ అస్య మన్తస్య నిరుక్తమ్. —
శతం జీవ శరదో వర్ధమాన ఇత్యపి నిగమో భవతి। శతమితి శతం దీఘమాముమరుత ఏనా వధమన్తి
శతమొనమొన శతాత్కానం భవతి శతమనన్తం భవతి శతమైశ్వర్యం భవతి శతమితి శతం దేఘమాముః।
హరిః ఓమ్॥

Figure 1: Text from *rigveda-* in Telugu, transliteration in Figure 2

```
{\tlc\QQQ
Hariah OmX\zzFC\ Sa\zzFJ taM jI\zzBFva Sa\zzFJ radO\zzFJ\ vardha\zzBF mAnaSSa\zzFJ taM
HE\zzBFma\zzFJ ntAnxcha\zzFJtamu\zzBF\ vasa\zzFJ ntA\zzCF\zzFB\
Sa\zzFJ tami\zzBF ndhrA\zzFJ gnI na\zzBF vi\zzFJ tA brx\zzBF Ha\zzFJ npatiSSa\zzFJ\
tAYu\zzBF ShA Ha\zzFJ viShE\zzFJ maM puna\zzBF rduah\zzFC\
rxksaMHita\zzFC\
{\zzBC 8-8-20} vaga\zzCJ\zzFC\
asYa mantrasYa niruktamX. ---

SataM jIva SadadO vardhamAna itYapi nigamO bhavati\zzFB\
Satamiti SataM dIgha\zzCJ mAYuma\zzCJruta EnA vadha\zzCJ Yanti
Satamonamona SatATmAnaM bhavati SatamanantaM bhavati SatamySvarYaM
bhavati Satamiti SataM dEgha\zzCJ mAYuah\zzFB\ Hariah OmX\zzFC
}
```

Figure 2: Text from *rigveda-* transliteration, Telugu version in Figure 1

contains several examples of transliteration of Telugu words using the proposed transliteration scheme.

వేమన శతకము
(Selected poems from "Vemana Satakamu")
చిత్తశుద్ధి

ఆత్మశుద్ధిలేని యాచార మది మేల
భాండ శుద్ధిలేని పాకమేల
చిత్తశుద్ధిలేని శివపూజలేలర
విశ్వాదాభిరామ విసురవేమ.

తా. మనస్సు నిర్మలముగా వుంచుకొనక పాటించిన ఆచారములు, పాత్రలు శుభ్రపరుచక చేసిన వంటకములు ప్రయోజనములేనివి. అట్లే భక్తి లేని పూజలు మెన్ని చేసిననూ ప్రయోజనము ఉండదు.

అల్పుడు - శ్లోకుడు

ఆఅల్పుడెప్పుడు పల్కు నాడంబరముగాను
సజ్జనుండు పల్కు చల్లగాను
కంచు మ్రోగినట్లు కనకంబు మ్రోగునా
విశ్వాదాభిరామ విసురవేమ.

తా. నేచుడు డాంబికములు పల్కును. ఉత్తముడు శాంతముగ మాటలాడును. ఏట్లన కంచు బిగ్గరగా మ్రోగును కానీ బంగారము అట్లు మ్రోగదుకదా?

గుణము

ఆపూజకన్న నెంచ బుద్ధి నిదానంబు
మాటకన్న నెంచ మనస్సు దృఢము
కులముకన్నా మిగుల గుణము ప్రధానంబు
విశ్వాదాభిరామ విసురవేమ.

తా. పూజకంటే బుద్ధి, మాటకన్నా మనస్సు ప్రధానము అయినట్లు, కులముకన్న గుణమే ముఖ్యమయినది.

దుర్జనుడు

ఆతప్పు లెన్నువారు తండోప తండంబు
లర్వి జనులకెల్ల నుండు తప్పు
తప్పు లెన్నువారు తమ తప్పులెరుగరు
విశ్వాదాభిరామ విసురవేమ.

తా. ఈ లోకములో తప్పులెన్నెడి వారధిక సంఖ్యలో వుండురు. ప్రపంచమున గలవారందరికీ తప్పులుండును. కానీ తప్పులు ఏన్నెడివారు తమతప్పు తాము తెలుసుకొనజాలరు.

Figure 3: Sample Text - Telugu version, transliteration in Figure 4

```
\setlength{\parindent}{0pt}
\begin{center}
```

```
                {\tlcb\QQQ vEmana Satakamu}\}
                (Selected poems from ‘‘Vemana Satakamu’’)\}
{\tlcb\QQQ cittaSuddhi}\}
\end{center}
{\tlcs\QQQ
```

```
A\zzFC      AtmaSuddhilEni YAcAra madi YEla\}
  \ bhAMDa SuddhilEni pAkamEla\}
  \ cittaSuddhilEni SivapUjalElara\}
  \ viSvadAbhirAma vinuravEma.\}
```

tA. manassu nirmalamugA vuMcukonaka pATiMcina AcAramulu, pAtralu
Subhraparucaka cEsina vaMTakamulu praYojanamulEnivi. aTLE bhakti lEni
pUjalu Yenni cEsinanU praYojanamu uMDadu.

```
{\Q\begin{center}
{\QQQ alpuDu - SrEShThuDu}\}
\end{center}}
```

```
A\zzFC      alpuDeppuDu palku nADaMbaramugAnu\}
  \ sajjanuMDu palku callagAnu\}
  \ kaMcumrOginaTlu kanakaMbu mrOgunA\}
  \ viSvadAbhirAma vinuravEma.\}
```

tA. nIcuDu DAMbikamulu palkunu. uttamuDu SAMtamuga mATalADunu. ETlana
kaMcU biggaragA mrOgunu kAnI baMgAramu aTlu mrOgadukadA?

```
{\Q\begin{center}
{\QQQ guNamu}\}
\end{center}}
```

```
A\zzFC      pUjakanna neMca buddhi nidAnaMbu\}
  \ mATakanna neMca manassu drxDhamu\}
  \ kulamukanna migula guNamu pradhAnaMbu\}
  \ viSvadhAbhirAma vinuravEma.\}
```

tA. pUjakaMTE buddhi, mATakanna manassu pradhAnamu aYinaTlu,
kulamukanna guNamE mukhYamaYinadi.

```
{\Q\begin{center}
{\QQQ durjanuDu}\}
\end{center}}
```

```
A\zzFC      tappu lennuvAru taMDopa taMDaMbu\}
  \ lurvi janulakella nuMDu tappu\}
  \ tappu lennuvAru tama tappulerugaru\}
  \ viSvadAbhirAma vinuravEma.\} 11
```

tA. I lOkamuLO tappulenneDi vAradhika saMkhYalo vuMduru.
prapaMcamuna galavAraMdarikI tappuluMDunu. kAnI tappulu EnneDivAru
tamatappu tAmu telusukonajAlaru.

Figure 4: Sample Text - transliteration, Telugu version in Figure 3

Next we will reproduce two stories from an ancient collection of stories for children.

పంచతంత్ర కథలు
(Stories from the “Panchatantra”)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

ఇంద్రపాలితుడు అనే వర్తకుడు వ్యాపారం నిమిత్తం వొరుగుదేశానికి వోవలసి వచ్చింది. అతడు తన వద్దనున్న వెయ్యి బంగారు నాణాలను, తన మిత్రుడయిన ధనగుప్తుని వద్ద దాచి దేశాంతరం వోయాడు. కొంతకాలం అయిన తరువాత ఇంద్రపాలితుడు స్వగ్రామం తిరిగి వచ్చాడు. అతడు ధనగుప్తుని ఇంటికివోయి తన బంగారు నాణాలను తిరిగి యిమ్మని అడిగాడు. ధనగుప్తుడు విచారంగా “అన్నా! ఏమి చెప్పను, నీ బంగారు నాణాలన్నింటినీ మా ఇంటిలో ఎలుకలు తినివేశాయి” అని చెప్పాడు. ధనగుప్తుని మోసాన్ని అర్థం చేసుకొన్న ఇంద్రపాలితుడు “వోనీలే జరిగినదానికి చింతించి ఏమి పలము” అని సర్ది చెప్పి, ఇదివరలోలాగానే మిత్రునితో స్నేహం కనబరస్తూ ఉన్నాడు. కొన్నాళ్ళువోయ్యాక, ఇంద్రపాలితుడు ధనగుప్తుని ఇంటివద్ద కబుర్లు చెబుతూ ఉన్న సమయంలో ప్రక్కవీధి లో రేగుపళ్ళా అమ్మకానికి వచ్చాయి. పళ్ళు అమ్మేవాని కేకవిని ధనగుప్తుని పిల్లవాడు పళ్ళకొసం వోరుపెట్టాడు. “నేను కొని ఇస్తాను రా నామనా!” అంటూ పిల్లవానిని ఎత్తుకొని ఇంద్రపాలితిడు ప్రక్కవీధికి బతులుదేరాడు. పళ్ళు కొని, పిల్లవానికి ఇచ్చిన ఇంద్రపాలితుడు ధనగుప్తుని ఇంటికి తిరిగి రాకుండా తన ఇంటికి వోయి పిల్లవానిని లోపల దాచి ఉంచాడు.

ఎంతసేపటికీ పిల్లవాడుగానీ, ఇంద్రపాలితుడుగానీ తిరిగి రాకవోమేసరికి ధనగుప్తుడు ఆదుర్దాగా వారిని వెతుకుతూ మిత్రుని ఇంటికి చేరాడు. అతనిని చూడగానే ఇంద్రపాలితుడు దుఃఖించిస్తూ “అన్నా పళ్ళుకొందామని మీ పిల్లవానిని చంకనుంచి దింపానో లేదో, ఒక పాడు డేగ వచ్చి మీ వాణ్ణి ఎత్తుకు వోయింది, ఈ విషాదవార్త నీకు చెప్పలేక నేను ఇంటికి వచ్చి దుఃఖిస్తున్నాను” అన్నాడు. ఆది విని ధనగుప్తుడు “నీచుడా, పిల్లవాని వంటివైననున్న బంగారు వస్తువులకాశపడి ఇంత అన్యాయం చేస్తావా” అంటూ దూషిస్తూ న్యాయాధిపతికి ఫిర్యాదు చేసాడు.

న్యాయాధిపతి ఇంద్రపాలితుని తో “ఏమయ్యా పుష్టిగా ఉండే పిల్లాడిని డేగ తన్నుకు వోయిందన్నావుంట. ఏమిటీ విడ్డూరం?” అని ప్రశ్నించాడు. దానిపై ఇంద్రపాలితుడు జరిగినదంతా వివరించి “అయ్యా ఈ ధనగుప్తుని ఇంట నేను దాచిన వెయ్యి బంగారు నాణాలు ఎలుకలు తినివేసినప్పుడు, ఓ చిన్న బిడ్డను డేగ తన్నుకు వోవడంలో ఆశ్చర్యమేముంది?” అన్నాడు. జరిగినదంతా గ్రహించిన న్యాయాధికారి, ఇంద్రపాలితుని ఇంటనున్న పిల్లవానిని రప్పించి ధనగుప్తునికి అప్పజెప్పి, మిత్రునినొమ్ము అవహరించి నమ్మక ద్రోహమొనర్చిన ధనగుప్తుని ఆస్తి స్వాధీనం చేసుకొని, అందులోనుంచి ఇంద్రపాలితునికి రావలసిన వేయి నాణాలు ఇప్పించాడు.

Following is the transliteration of the preceding Telugu text.

```
\begin{center}
{\tlcb\QQQ paMcataMtra kadhalu}\
(Stories from the ‘Panchatantra’)\
\end{center}
{\tlcnx\QQQ
```

‘parulanu mosagiMci HANicEYa talapeTTinavAru tAmE ceDipOtAru’.\

iMdrapAlituDu anE vartakuDu vYApAraM nimittaM porugudESAniki pOvalasi
vacciMdi. ataDu tanavaddanunna veYYi baMgAru nANALanu, tana mitruDaYina

dhanaguptuni vadda dAci dESAMtaraM poYADu. koMtakAlaM aYina taruvAta iMdrapAlituDu svagrAmaM tirigi vaccADu. ataDu dhanaguptuni iMTikipOYi tana baMgAru nANAlanu tirigi Yimmani aDigADu. dhanaguptuDu vicAraMga ‘‘anna! Emi ceppanu, nI baMgAru nANAlanniMTinI mA iMTilo elukalu tinivESAYi’’ ani ceppADu. dhanaguptuni mOsAnni ardhMcEsukonna iMdrapAlituDu ‘‘pOnIIE jariginadAniki ciMtiMci Emi phalamu’’ ani sardi ceppi, idivaralOlAgAnE mitrunitO snEHaM kanabarastU unnADu. konnALLupOYAKa, iMdrapAlituDu dhanaguptuni iMTivadda kaburlu cebutU unna samaYaMlO prakkavIdhi lO rEgupaLLU ammakaniki vaccAYi. paLLu ammEvAni kEkavini dhanaguptuni pillavADu paLLakOsaM pOrupeTTADu. ‘‘nEnu koni istAnu rA nAYana!’’ aMTU pillavAnini ettukoni iMdrapAlituDu prakkavIdhiki baYaludErADu. paLLu koni, pillavAniki iccina iMdrapAlituDu dhanaguptuni YiMTiki tirigi rAkumDA tana YiMTiki poYi pillavAnini lOpala dAci uMcADu.

eMtasEpaTikI pillavADugAnI, iMdrapAlituDugAnI tirigi rAkapOYEsariki dhanaguptuDu AdurdAgA vArini vetukucU mitruni iMTiki cErADu. atanini cUDagAnE iMdrapAlituDu duahkhaMnaTistU ‘‘anna paLLukoMdAmani mI pillavAnini caMkanuMci diMpAnO lEdO, oka pADu DEga vacci mI vANNi ettuku poYiMdi, I viShAdavArta nIku ceppalEka nEnu iMTiki vacci duahkhistunnu’’ annADu. Adi vini dhanaguptuDu ‘‘nIcuDA, pillavAni vaMTipynunna baMgAru vastuvulakASapaDi iMta anYAYaM cEstAvA’’ aMTU dUSHistU nYAYAdhipatiki phirYadu cEsADu.

nYAYAdhipati iMdrapAlituni to ‘‘EmaYYA puShTigA uMDE pillADini DEga tannuku poYiMdannAvuTa. EmiTI viDDUraM?’’ ani praSniMcADu. dAnipy iMdrapAlituDu jariginadaMtA vivariMci ‘‘aYYA I dhanaguptuni YiMTa nEnu dAcina veYYi baMgAru nANAlu elukalu tinivEsinappuDu, O cinna biDDanu DEga tannuku pOvaDaMlO AScaYamEmuMdi?’’ annADu.

jariginadaMtA graHiMcina nYAYAdhikAri, iMdrapAlituni YiMTanunna pillavAnini rappiMci dhanaguptuniki appajepi, mitrunisommu apaHariMci nammaka drOHamonarcina dhanaguptuni Asti svAdhInaM cEsukoni, aMdulOnuMci iMdrapAlituniki rAvalasina vEYi nANAlu ippiMcADu.

}]

“అతి లోలత్వం పనికిరాదు”

పూర్వము ఒక చెరువులో రకరకాల వేపలు, ఒక పీత కలసి జీవించేవి. ఒకనాడు ముసలి కొంగ ఒకటి ఆ చెరువు లోకి వచ్చి తాను అహింసా వ్రతమును అవలంబించి ధార్మిక జీవనాన్ని గడపదలుచుకొన్నానని, తన జవధ్యానాదులకు ఆ చెరువు జాగుందని ఆ జలచరములతో చెప్పింది. కొన్ని రోజులపాటు వారికి విశ్వాసం కలిగేటట్లుగా ప్రవర్తించింది. ఆ జలచరములన్నీ కొంగ పట్ట నమ్మకము పెంచుకున్నాయి. తనలో వాటికి నమ్మకం కుదిరిన తరువాత ఒక రోజున ఆ కొంగ కన్నీరు కారుస్తూ “ఓ వారం రోజులకు ఈ చెరువు నీరు ఇంకిపోతుందిట అప్పుడే వేపలను సులభంగా పట్టుకుపోవచ్చునని బెస్తవాళ్ళు అనుకున్న మాటలు నా వెవిన పడ్డాయి. నాకు చాలా విచారంగా ఉంది” అని వేపలతో చెప్పింది. వేపలు హడలిపోయి “స్వామి! మీరే మాకు దిక్కు. ఎట్లాగమనారే మీరు ఈ ఆపదనుండి మమ్ములను కాపాడాలి” అని కొంగను వేడుకొన్నాయి. కొంగ కొంచెంసేపు ఆలోచించినట్లు నటించి “ఇక్కడకు కొంచెం దూరంగా ఎల్లప్పుడూ నీరుండే కొలను

ఉంది. అక్కడకు చేరుస్తాను. నా నోట తడవకు నాలుగు చేపల కన్నా ఎక్కువ పట్టవుగనుక మిమ్మల్ని నాలుగు మంది చొప్పున అక్కడకు చేరుస్తుంటాను” అంది. చేపలు ఆ కొంగకు ధన్యవాదాలు చెప్పుకొన్నాయి. తన దురలోచన ప్రకారం, ఆ కొంగ రోజూ చేపలను నోట కరచుకొని, ఒక కొండ మీద వాలి, వానిని ఆరగించి వచ్చేది. కొన్ని నాళ్ళకు చేపలన్నీ కొంగ కడుపున బడ్డాయి. ఇంక వీత ఒక్కటే మిగిలింది. అ వీతను కూడా తినాలి అనే లోలత్వంతో కొంగ ఆ కర్కటకాన్ని కూడా నోట కరచుకొని కొండవైపు ఎగిరింది.

కొంగ వోతున్న మార్గము, కొండ మీద దూరంగా కన్పిస్తున్న చేపవోలునులా చూచి, వీత అసలు విషయం అర్థం చేసుకుంది. అది కొంగ తో “నాకు కళ్ళు తిరుగుతున్నాయి. నీవు వదలివేస్తావేమోనని భయంగా ఉంది. నా డెక్కలతో నీ మెడను పట్టుకొనిస్తావా పడివోకుండా ఉంటాను” అంది. కొంగ దానికి అంగీకరించింది.

పెంటనే వీత తన వాడిమైన డెక్కలతో కొంగ మెడను కొరికి కొంగను చంపివేసింది. చచ్చిన కొంగ-తోపాటు కొండ మీద పడిన వీత, పెల్లగా లేచి, తన మితులైన చేపల మరణానికి విచారించి వేరొకచోట చేరింది.

Following is the transliteration of the preceding Telugu text.

```

\begin{center}
                                {\tlcs\QQQ ‘‘ati l0latvaM panikirAdu’’\}
\end{center}
{\tlcny\QQQ
pUrvamu oka ceruvulo rakarakAla cEpalu, oka pIta kalasi jIviMceDivi. okanADu
musali koMga okaTi A ceruvu l0ki vacci tAnu aHiMsA vratamunu avalaMbhiMci
dhArmika jIvanAnni gaDapadalucukonnAnani, tana japadhYAnAdululaku A ceruvu
bAguMdani A jalacaramulato ceppiMdi. konni r0julapATu vAriki viSvAsaM
kaligETaTlugA pravartiMciMdi. A jalacaramulannI koMga paTla nammakamu
peMcukunnAYi. tana l0 vATiki nammakaM kudirina taruvATA oka r0juna A koMga
kannIru kArustU ‘‘0 vArAm r0julaku I ceruvu nIru iMkip0tuMdiTa appuDE cEpalanu
sulabhaMga paTTukupOvaccunani bestavALLu anukunna mATalu nA cevina paDDAYi.
nAku cAlA vicAraMga uMdi’’ ani cEpalat0 ceppiMdi. cEpalu HaDalip0Yi ‘‘svAmi!
mIrE mAku dikku. eTlAgayinAsarE mIru I ApadanuMDi mammulanu kApADAli’’ ani
koMganu vEDukOnnAYi.

koMga koMceMsEpu Al0ciMcinaTlu naTiMci ‘‘ikkaDaku koMceM dUraMga ellappuDU
nIruMDE kolanu uMdi. akkaDaku cErustAnu. nA n0Ta taDavaku nAlugu cEpala kannA
ekkuva paTTavuganuka mimmalni nAlugu maMdi coppuna akkaDaku cErustuMTAnu’’
aMdi. cEpalu A koMgaku dhanYavAdAlu ceppukonnAYi. tana durAl0cana prakArAm,
A koMga r0jU cEpalanu n0Ta karacukoni, oka koMDa mIda vAli, vAnini AragiMci
vaccEdi. konni nALLaku cEpalannI koMga kaDupuna baDDAYi. iMka pIta okkaTE
migiliMdi. a pItanu kUDA tinAli anE l0latvaMto koMga A karkaTakAnni kUDA
noTa karacukoni koMDavypu egiriMdi.

koMga potunna mArgamu, koMDa mIda dUraMga kanpistunna cEpap0lusulu cUci,
pIta asalu viShaYaM ardhaM cEsukuMdi. adi koMga to ‘‘nAku kaLLu tirugutunnAYi.
nIvu vadaliVEstAvEmOnani bhaYaMga uMdi. nA Dekkalato nI meDanu paTTukOnistAvA
paDipokuMDA uMTAnu’’ aMdi. koMga dAniki aMgIkariMciMdi.

veMTanE pIta tana vADiYyna Dekkalato koMga meDanu koriki koMganu
caMpivEsiMdi. caccina koMgatOpATu koMDa mIda paDina pIta, mellaga lEci, tana
mitrulyna cEpala maraNAniki vicAriMci vErokacOTa cEriMdi.

```

Transliteration	In Telugu	Syllable Decomposition
k	క	(క)
ka	కా	(క + అ)
e	ఎ	(ఎ)
le	లే	(ల + ఎ)
l	ల	(ల)
ravi	రవి	(ర + అ)(వ + ఇ)
telugu	తెలుగు	(త + ఎ)(ల + ఉ)(గ + ఉ)
amerikA	అమెరికా	(అ)(మ + ఎ)(ర + ఇ)(క + ఆ)
strI	స్ట్రీ	(స్ + త + ర + ఊ)
SrI	స్రీ	(స్ + ర + ఊ)
laXhmi	లక్ష్మి	(ల + అ)(క్ష + మ + ఇ)
mukkavilli	ముక్కవల్లి	(మ + ఉ)(క + క + అ)(వ + ఇ)(ల్ + ల్ + ఇ)
laXhmaNXkumArX	లక్ష్మణకుమార్	(ల + అ)(క్ష + మ + అ)(ణ)(క + ఉ)(మ + ఆ)(ర్)
prabhu	ప్రభు	(ప్ + ర + అ)(భ + ఉ)
U\zzBF rdhvemowMjAah	ఉర్దవమౌంజాః	(ఉ + ా)(ర్ + ధ + వ + ఎ)(మ + ఎ + ాం)(జ్ + ఆ + ః)
uppu	ఉప్ప	(ఉ)(ప్ + ప + ఉ)
up\zzJB	ఉప్ప	(ఉ)(ప్ + ప + ఉ)
karpUramu	కర్పూరము	(క)(ర్ + ప + ఊ)(ర్ + అ)(మ + ఉ)
kar\zzJC ramu	కర్పూరము	(క)(ర్ + ప + ఊ)(ర్ + అ)(మ + ఉ)
aM	అం	(అ + ాం)
AMdhra pradESX	ఆంధ్ర ప్రదేశ్	(ఆ + ాం)(ధ + ర + అ)(ప్ + ర + అ)(ద్ + ప)(ష్)

Table 8: Examples of Transliteration

}

4 Problems in Typesetting Telugu Text

When we try to adapt technologies developed for English, to a vastly different language we face many difficulties. And more so in case of Telugu. We are somewhat spoiled by the simplicity of English text. In this section we will enumerate the difficulties faced in developing a typesetting system for Telugu.

Tradition English has a long tradition of typography. English typography is widely studied and researched. There are well established styles of typefaces. This is not the case with Telugu. Much of the composition in Telugu is still done by hand. Use of computers to typeset Telugu is really a recent phenomenon.

Character Set In English the character set is well defined. In fact there are effective standards in place. Telugu character set is still evolving. Some letters that were used a few decades ago are not in use any more. Some symbols have been replaced with new symbols. There is no unanimity with regard to symbols used in transliteration of Sanskrit text. Our intent has been to provide all the symbols that have been used

over the past 200 years. We were fortunate to be able to obtain facsimili of type specimens of a Telugu font cut in 1802 from a museum in London. We also looked at various publications produced since then.

Complexity of Script English text is composed by laying letter after letter in a linear manner. Composition of Telugu is very different. Composition is done by laying syllable after syllable. But composing a syllable (see page 1 for the definition of a syllable) is very complex. The total number of possible syllables is very large. So we cannot store the images of all the syllables. Syllables need to be composed on the fly. Composing a syllable means juggling symbols, stacking symbols, juxtaposing symbols. Composing a syllable requires several character lookahead. The first step in composing a syllable is to modify the base consonant (first consonant) by the vowel which appears after the intervening consonants (consonant conjuncts). This problem of building up a syllable brings us to the next item.

Standards In case of Telugu, we should distinguish between the coding system for text storage/transmission (we will call this Information Interchange code IIC) and the coding system for the font that is used for rendering (printing/ displaying) text. The font contains graphical elements that can be used to compose syllables. Many of the symbols in the font may not look like any letter in the alphabet. IIC contains codes for consonants, vowels, accents, digits and punctuation marks. Each syllable is stored as a sequence of constituent consonants, vowels and accents. In Telugu there are 16 vowels, 36 consonants and 11 accents. There is a standard for Telugu information interchange. In fact it is the same for most of the languages used in India since they all have the same phonetic structure. But there is no standard for the font. It is very difficult to define a standard for assigning codes to elements of a font. It is difficult to decide what should be in the font. The structure of each font depends on the font designer. A font designer (for Telugu) would use various criteria i.e. number of symbols permitted in a font, adherence to tradition, choice of alphabet, complexity of composition, features available in the font development system, features available in the composition system etc. to determine what primitive strokes or forms should be included in the font.

In English, information interchange code would correspond directly to encoding used in the font. This makes life a lot easy. It also explains why it is a very difficult task to adapt a text composition system developed primarily for English to Telugu.

Letterforms Most of the existing Telugu letterforms are heavy along base line. It makes it very hard on the eyes of the reader. Many letterforms lack in consistency. This may be due to the fact that Telugu fonts are typically designed by metal-smiths/calligraphers not very well skilled in modern principles of type design and do not have access to modern technology. We have designed our letter forms from scratch. No attempt was made to imitate an existing font. In fact we have not found any font that we thought was good. We are fortunate in being able to use a very sophisticated font development system called METAFONT.

Parametrization One of the major strengths of METAFONT is its ability to generate fonts for a variety of devices in a variety of styles and sizes. In order to realize this facility we need to identify key parameters of the font. Then we can play with these

parameters to obtain various fonts. We have tried to define various parameters for Telugu script.

Font Metric Data In case of Telugu we need to provide certain measurements about location of some key points in a symbol. These measurements need to be stored along with the font. The measurements are used by the text composition system to align/position the symbols during syllable construction. Font metric files generally do not have any provision for storing optional font metric data. To overcome this problem we write the measurements in the form of \TeX macros to the log file. After font generation the macros are extracted and placed in a separate file (called offsets file). The measurements (offsets) are generated in relative units (*em* units). This means that only one offset file is needed (regardless of size/style).

Speed Syllable building is a very CPU intensive job. Each syllable is pieced together from constituent parts. Typesetting Telugu text is a slow process.

Hyphenation \TeX hyphenation algorithms are not applicable to Telugu text. We use the rule that a syllable should not be broken across lines. Since a syllable is built in a box and boxes are not broken up, this rule is satisfied. But then we are not giving \TeX much freedom to do line breaking. We insert a discretionary hyphen between two syllables. Our experience indicates that this works fine.

5 Implementation

The system has two components. First is the font developed using METAFONT. These symbols are used for composing Telugu text. The second component is used for Telugu text composition. This is written in \TeX macro language. Telugu is a phonetic language. The user enters Telugu as he/she would speak (i.e. phonetically). But the script is not linear. Unlike English, the keys on the keyboard do not correspond with the symbols in the font. We need a mechanism that interprets the phonetic input and composes aksharas (or syllables). This requires considerable amount of processing. This is what is done by the second component. Phonetic transcription is interpreted and syllables are formed. As mentioned earlier, there are two very different approaches to inputting Telugu (and other Indian languages). One is phonetic (as ours) and the other is called Graphical approach. In this approach, the user picks individual symbols from the font and composes the syllable. For Telugu, the Graphical approach would be tedious and very difficult. We have tried to make the second component applicable for other Indian languages as well. Since all Indian languages have similar phonetic structure, this component can be easily adapted to other Indian languages.

Next we provide an outline of our implementation. Subsection 1 deals with font development and Subsection 2 deals with text composition.

5.1 Font Development

The font is developed using METAFONT program. The font has over 230 symbols. In addition to symbols used to compose Telugu text, the font includes Telugu as well as Hindu-Arabic digits, punctuation symbols, non-alphabetic characters from ASCII, accents and symbols needed for transliteration of Sanskrit. One may wonder why worry about

symbols that are available in Computer Modern Roman. The answer is that we want the text composed to look consistent and harmonious. Since Telugu characters have a rounded appearance, we decided that we should use circular pens. In olden days Telugu used be written on palm leaves using a metal stylus with a round nib.

Following are the steps in developing the font:

1. Identify all the symbols that are needed to compose syllables.
2. Define a grid framework for drawing the symbols. In this grid we define a baseline, ascender height, descender depth, top shoulder, bottom shoulder and a few other lines.
3. Assign codes to each symbol. All code references are made symbolically and codes are defined in one separate file. It would be very easy to change code assignment. We have used the same code as ASCII for the symbols that also occur in ASCII.
4. Make rough sketches of all these symbols and other elements of the font on a graph paper using a consistent scale.
5. Enumerate the parameters of the font. This is a delicate job. The parameters are modified to obtain various styles/sizes of the font.
6. Define font dimensions (see METAFONT book).
7. Identify all the strokes/shapes that occur more than once. Write macros to draw these strokes/shapes.
8. Write a METAFONT program for each symbol in the font taking care to use macros whenever possible and refer to parameters.

First we would look at the sketch and identify control points. Specify coordinates in relative terms. That is, the coordinates are defined in terms of font parameters and box dimensions (width, height and depth). Box dimensions are defined in terms of parameters, u (unit width) and uh (unit height). Parameters are either unit free (e.g. ratios, scaling factors etc.) or are defined in terms of other parameter(s) and/or u and/or uh . u and uh are defined interms of the design size. One exception to this rule is `pen.width` (diameter of the circular pen). This is specified as an absolute quantity. In order to generate the font we need to provide the design size, pen size/shape, slant and optionally, the magnification factor.

9. METAFONT produces dimensions of the bounding box of each symbol in the font and these dimensions are used by text composition software. For Telugu we have some special requirements. When a consonant is modified by a vowel, vowel modifier is attached to the consonant base. The point of contact varies from consonant to consonant. The offset to this point should be known to the syllable building mechanism. This offset (in em units) is generated as a \TeX macro and written onto the log file. Similarly, when various symbols are stacked on the top/bottom of a C+V, the syllable building mechanism needs to be aware of the offset by which to shift the symbols. This offset is also generated as a \TeX macro and written onto the log file. All the macros are extracted from the logfile and copied into another file that is read by syllable building software. The offsets are very sensitive to changes

to the font. Because of the way we handle offsets, we are free to change the font and new offsets are automatically recorded. As stated earlier, there is no standard for the assignment of codes in the Telugu font. We have decided on a particular encoding scheme. But the codes are not hardwired into the syllable building mechanism. The codes are generated as T_EX macros and written onto the log file. If somebody wants to change the codes, all that he/she needs to do is alter the entries in the code file and generate the font.

10. Assign space on the left and/or on the right of symbols. Since the symbols are pieced together to form syllables, we need to be very careful about placing white space on the sides.
11. Prepare ligature tables.

By varying the font size, the pen size/shape and the slant we can obtain several fonts. All font file names have “tel” prefix. This prefix is followed by a number indicating the font size (in points). Font size is defined as the distance between the top edge of the top shoulder and the bottom edge of the bottom shoulder. If the size in a font filename is followed by ‘s’ then it indicates a slanted font. If the size in a font filename is followed by ‘b’ then it indicates a bold font.

5.2 Text Composition

The programs for syllable building are written in T_EX macro language. There are two modules. One is used to interpret user input. This module implements the transliteration scheme proposed in Section 2 (see page 8). If a different transliteration scheme or input system is to be used then this module will need to be modified. The second module identifies the components of a syllable and builds the syllable. Every syllable is output as a horizontal box. Each syllable box may have many constituent boxes. The main activity in this part involves box manipulation. Boxes are stacked, shifted, attached to obtain the complete syllable.

5.2.1 Transliteration Module(tlxlite.tex)

This module reads one or two characters and identifies the corresponding letter (vowel or consonant or accent) in Telugu and invokes corresponding macro in the Syllable module. The logic is:

1. Read a character
2. If more than one transliteration codes can begin with this character then lookahead one character.
3. Identify the letter (vowel/consonant/accnt)
4. Invoke the corresponding macro.

The principal mechanism for implementing transliteration is the concept of *active* characters. The main advantage of this technique is that it does not require any extra effort on the part of the user. All the characters that can begin a transliteration code are declared as active characters. There is a macro for each active character. These macros take one of the following two forms:

1. Invoke the macro corresponding to the letter (If the character cannot begin another transliteration code).
2. Else look at the next character to identify the letter.

To input a non-alphabetic character, one need only enter the corresponding symbol (subject to \TeX input rules). The user can choose between Hindu-Arabic digits and Telugu digits. It is easy to switch between these two kinds of digits. Two macros are provided to make this switch. To start with, Hindu-Arabic digits are provided. The macro used to switch to Telugu digits does the following:

1. Declare all the digits as active characters.
2. In the macro for a digit, output the corresponding Telugu digit.

The macro for switching to Hindu-Arabic digits declares all digits to be non-active characters. The transliteration module does not bother about non-alphanumeric characters because \TeX would automatically handle those symbols (just as in case of English text). All non-alphanumeric characters and Hindu-Arabic digits in Telugu font appear at the same positions as in ASCII.

5.2.2 Syllable Module(`tsyllable.tex`)

In this module we try to build a syllable from constituent consonant(s) and the vowel. The focus of macros in this module is on composing a syllable from constituent letters. There are two subsections in this section. First we give a brief introduction to Syllables in Telugu script and then we outline our implementation.

Introduction

We may recall that a syllable (or Akshara) can take one of the two forms.

1. a stand-alone vowel(V)
2. Consonant + Consonant + ... + Consonant + Vowel($C_1 + C_2 + \dots + C_n + V$)

C_1 is considered the base consonant. C_2 to C_n are called consonant conjuncts. Consonant conjuncts are optional. The Vowel modifies the base consonant. The vowel modifier may be absent. This is indicated by a special code which appears in place of the vowel modifier. For the purpose of implementation we will treat this code as another vowel modifier. The form of Base Consonant + Vowel (henceforth referred to as $C+V$) varies depending on both the consonant and the vowel. The vowel modifiers can appear to the right, on the top or at the bottom of the base consonant. In many cases a completely different symbol is needed. After Consonant + Vowel is formed, consonant conjuncts are attached in the order they appear in the syllable. Consonant Conjuncts are placed to the right or below $C+V$. Consonant Conjuncts have a different form than either base consonant or $C+V$'s.

An Akshara can optionally have accent(s). Accents could appear at the top or on the side or at the bottom of a syllable.

For each consonant, the conjunct form is different from the base form. The base consonant takes a different form depending on the vowel modifier. The vowel modifies the

base consonant. The way the consonants are modified by the vowel is not governed by any uniformly applicable rules. This lack of consistency contributes to the complexity of the Telugu script. Some consonant conjuncts are stacked beneath the base consonant+vowel, whereas others appear on the side.

The logic to process a syllable is :

1. Identify the base consonant (C_1)
2. Scan for the next vowel while saving the intervening consonant conjuncts ($C_2 \dots C_n$).
3. Once the vowel modifier is identified, compose C_1+V . This is the messy part. Though many of the consonants are modified uniformly by a given vowel, there are too many exceptions.
4. Generate consonant conjuncts.
5. Apply accents.
6. Add space following a syllable.
7. If preceded by a box then insert a discretionary hyphen.
8. Release the syllable box.

Implementation

Next we will discuss the implementation of the syllable building mechanism.

The main strategy to build a box containing the syllable involves piecing together boxes containing constituent symbols from the font. Boxes are shifted, lowered, raised and stacked to obtain the right appearance for the syllable. In summary, syllable building is an exercise in box juggling or box manipulation. But all the shifts and kerns must be carefully controlled. All the box dimensions must be accurately specified and we cannot have any part of the symbol spill over the bounding box. This calls for extreme care in specifying box dimensions in the METAFONT programs. Else we will have malformed syllables. In this module we have an extensive collection of macros to shift/lower/raise boxes and to recompute box dimensions affected by shifts/kerns. After processing, `\syl` box contains the syllable. The consonant conjunct tokens are built up in the token list `\cctok`. The name of the base consonant is saved in `\cbtok`. In this module various letters of Telugu are identified by unique names. These names are local to this module. The names we use for various letters are significant. They cannot be changed arbitrarily. The names are used in defining macro names. For example, the macro to process consonant conjunct x is `\tlccx`. The macro to process $C(x)+V(y)$ is `\tlcvxy`. It is unlikely that the user will invoke the macros defined here directly.

When a vowel x is to be processed, the macro `\tlvox` is invoked. When a consonant x is to be processed, the macro `\tlcbx` is invoked. These macros are invoked from the transliteration module.

The consonant processing macro determines if the consonant is the base consonant or a conjunct; if base consonant, the syllable processing is initiated. If conjunct, the macro to form the conjunct is added to the token list `\cctok`.

The vowel processing macro determines if the vowel is a stand-alone vowel or a modifier. If it is a stand-alone vowel, it just outputs the character. If it is a modifier, then the following actions are taken :

- Some C+V combinations need special treatment. The macro to process consonant x and vowel y is `\tlcvxy`. Otherwise the macro to process the vowel modifier(`\tlvmy`) is called.
- Process consonant conjuncts.

How does the syllable module know which (C+V) combinations need special treatment? Vowels are partitioned into groups such that if a consonant needs special handling when combined with a vowel in a group then the consonant also needs special treatment when combined with other vowels in the same group. The groups are {అ}, {ఇ,ఈ}, {ఉ,ఊ}, {ఋ,ౠ}, {ఌ} and others. We assign a prime number to each group. We also assign a number to each consonant which is the product of the prime numbers associated with the vowel groups that need special treatment when combined with this particular consonant. If a consonant needs no special treatment then it gets a distinct prime number. Given a consonant(C) and a vowel(V) if the number for V divides the number for C evenly then this C+V combination needs special care.

This module contains macros to process accents. There is a macro for each accent. Processing accents involves retrieving the last syllable box and placing the accent. The placement could be at the top or at the bottom or on the side depending on the particular accent. A syllable can have multiple accents. It is the responsibility of the user to ensure that a proper combination of accents is used.

6 Type Specimens

Specimen of tla (tel10)

“పరులను మొసగించి హనివేచు తలపెట్టినవారు తామే వెడివోతారు”.

Specimen of tlanx (tel10nx)

“పరులను మొసగించి హనిచేచు తలపెట్టినవారు తామే వెడివోతారు”.

Specimen of tlany (tel10ny)

“పరులను మొసగించి హనివేచు తలపెట్టినవారు తామే వెడివోతారు”.

Specimen of tlab (tel10b)

“పరులను మొసగించి హనివేచు తలపెట్టినవారు తామే వెడివోతారు”.

Specimen of tlas (tel10s)

“పరులను మొసగించి హనిచేచు తలపెట్టినవారు తామే వెడివోతారు”.

Specimen of tlb (tel11)

“పరులను మొసగించి హనివేచు తలపెట్టినవారు తామే వెడివోతారు”.

Specimen of tlbnx (tel11nx)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlbnx (tel11ny)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlbb (tel11b)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlbs (tel11s)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlc (tel12)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlcnx (tel12nx)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlcny (tel12ny)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlcb (tel12b)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlcs (tel12s)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tld (tel15)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tldb (tel15b)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlds (tel15s)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tle (tel18)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tleb (tel18b)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tles (tel18s)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlf (tel20)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlf (tel20b)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlfs (tel20s)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlg (tel25)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlg (tel25b)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlgs (tel25s)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlh (tel30)

“పరులను మొసగించి హానిచేయ తలపెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlhs (tel30s)

“పరులను మొసగించి హానిచేయ తల పెట్టినవారు తామే చెడిపోతారు”.

Specimen of tli (tel35)

“పరులను మొసగించి హానిచేయ తల పెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlis (tel35s)

“పరులను మొసగించి హానిచేయ తల పెట్టినవారు తామే చెడిపోతారు”.

Specimen of tlj (tel40)

“పరులను మొసగించి హానిచేయ తల పెట్టినవారు తామే ”

Specimen of tljs (tel40s)

“పరులను మొసగించి హానిచేయ తల పెట్టినవారు తామే ”

Specimen of tlk (tel55)

“పరులను మొసగించి ”

Specimen of tlks (tel55s)

“పరులను మొసగించి ”

Specimen of tll (tel72)

“పరులను ”

Specimen of tlm (tel100)

“పరులను ”

Specimen of tln (tel172)

పరుల

Specimen of tlspa (telspa)

“పరులను ”

Specimen of tls pb (telspb)

“పరులను ”

Specimen of tls pc (telspc)

“వదలిను”

7 Summary

In this project, we have undertaken to produce a typesetting system for a very complex script. We have used two top quality programs (T_EX and METAFONT) developed by Prof. Donald E. Knuth of Stanford University. There are two parts to the project. One is the development of fonts using METAFONT and the other is writing programs (i.e. macros) in T_EX. METAFONT is a very difficult language to program in. T_EX is not really a good programming language. In order to realize the full potential of these programs we need to do extensive ground work. Parameters need to be defined, we have to write macros for frequently occurring shapes and also write a program for each symbol of the font. This is an iterative process.

All these are time consuming and highly skilled tasks. When developing the font, we had to ensure that our font would blend harmoniously with Roman fonts when used in a document that contains both Roman and Telugu fonts. This is a very delicate job because we are dealing with two radically different languages. Printers producing Telugu use hundreds of different types (metal blocks) with the image of a symbol cut on the face. But METAFONT and T_EX allow no more than 256 symbols in a font. By identifying the independent forms and exploiting the capabilities of T_EX we have managed to keep the font size well within this limit.

The quality of the font we have developed is satisfactory but will need to be refined as we gain experience with the system. Telugu input to T_EX/ L^AT_EX is presented in the form of Roman transliteration. Our macros interpret this english text, identify syllables and compose the syllables. All these tasks are very CPU intensive and hence typesetting Telugu takes considerably more time than typesetting English text.

Kannada is another language spoken in South India. Kannada script is very similar to Telugu script. Our typesetting system can be easily adapted to Kannada. Parts of our system can be used for extending T_EX to other South Asian & Southeast Asian languages. Our system consists of well defined and independent modules. This confers a lot of flexibility on the typesetting system.