Aethiopica 3 (2000)

International Journal of Ethiopian and
Eritrean Studies

Notes on Encoding Ethiopic for LaTeX

BERHANU BEYENE, MANFRED KUDLEK, OLAF KUMMER

1. Introduction

Ethiopic is a script consisting of more than 450 syllabic characters. It is used to write the Ethiopian and Eritrean languages. Although the name of the script has been given variously as "Ethiopic", "Ge'ez", "Sabaean", Ethiopian script", "Feedel", "Hoheyat" etc, we prefer to use *Ethiopic*, since it is used as a standard name by Unicode[1] and the Ethiopian Computer Standard Association (ECoSA)[2] of which the governmental sole Standard Authority (ESA) is a member. Moreover, many research societies like Scientific Society, EthCITA, EthioSystem use the name Ethiopic.

Parallel to the development of information and communication technology, electronic document processing systems in general and TeX/LaTeX in particular have attracted authors, publishers and typesetters. This is due to the fact that the TeX/LaTeX text processing system produces the finest quality of technical and scientific manuscripts, books, documents etc. In addition, the TeX/LaTeX special package to support also non-latin scripts, called babel provides a multilanguage switch system which has contributed to boost text processing technology with different scripts.

As a computer has no native tongue, it accepts and processes every character supplied by the user as long as it is predefined to match the characters "0" and "1". Therefore all characters have to be encoded with "0" and "1". The best known and most widely applied character encoding is the ASCII code (American Standard Code for Information Interchange). The original 7-bit ASCII encoding system consists of only 128 characters, whereas the present encoding systems have 256 characters, 64 of which are reserved for numerals, special commands and characters. The remaining 192 characters are assigned for latin scripts and some special characters. It is therefore quite necessary to consider the encoding

[1]   UNICODE.
[2]   ECOSA.

mechanism for non-latin scripts consisting of more than 256 characters, as in the case of Ethiopic. Thus encoding Ethiopic also needs special consideration.

This article highlights the encoding of Ethiopic for LaTeX and the package ethiop developed at the University of Hamburg. The package ethiop is a collection of METAFONT and TeX/LaTeX macros and helps to typeset the character sets of the different languages in Ethiopia and Eritrea. Besides, some sample texts are shown in Appendix A. For those interested in, a short installation instruction of the system is also given in Appendix B.

2. Document Preparation with TeX/LaTeX

Computer aided document preparation systems started timidly in the 1960s, with the appearance of TROFF or RUNOFF formatters as an extension to program editing tools. The first systems were low-level formatters, handling text input through the editor used for writing programs. Essentially was the construction of lines of equal length, and the production of justified pages on the basis of a *ribbon* of text. A few commands could be introduced into the text to control page layout. The language in which commands were expressed, showing little formalism, was at a low level, similar to an assembly language[3].

With the rapid growth in information and communication technology, and more particularly in line with developments in programming languages, the methods of electronic document preparation system have advanced to an interactive processing system, usually called WYSIWYG (What You See Is What You Get). The interactive editing systems (like Microsoft Word for Windows), with sophisticated typographical and user friendly features, have dominated the document preparation world, at least in great parts of business and administration. However, scientific and technical documents dominantly remained to be processed by the formatting systems (such as TeX/LaTeX), due to their immense advantages, mainly high quality of output and freedom of authors to choose their formatting style.

2.1. TeX/LaTeX and METAFONT

TeX is a platform independent (available for most computers, ranging from PC to Mac, from UNIX to VMS and Atari systems) typesetting system developed by DONALD E. KNUTH[4] in early 1970's. LaTeX is a macro package developed in 1986 by L. LAMPORT, that has made TeX a powerful system with predefined
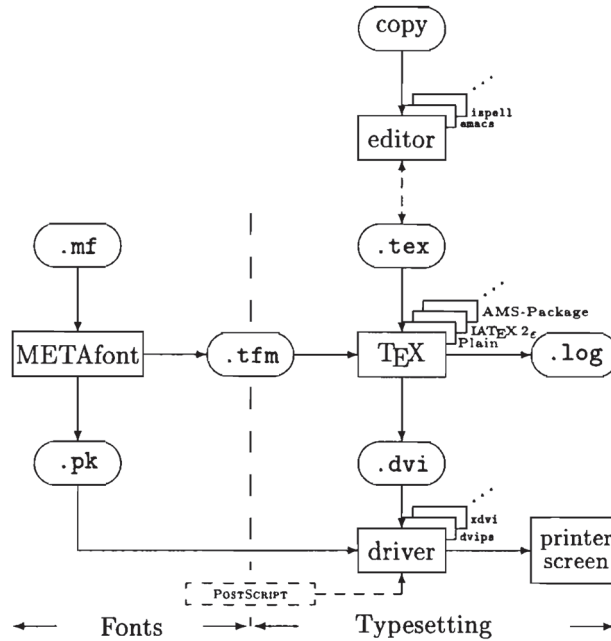
---

[3] ANDRE.
[4] KNUTH (1986A).

Figure 1: Components of a TeX System

layout to produce the highest quality output. METAFONT is another program developed by DONALD KNUTH to produce bitmap fonts for use by TeX, its viewers, printer drivers, and related programs. METAFONT utilises scaling, rotation, reflection, skewing and shifting, and other complex transformations and mathematical calculations. The output of the METAFONT are so-called *generic fonts* (GF) which can be compressed to equivalent *packed* (PK) fonts so that they could be used by the TeX/LaTeX system. TeX/LaTeX uses the compiled files called TeX Font Metric (TMF), to describe the dimensions, ligatures and kerns of a font. Figure 1 (taken from wots.tex by Kees van der Laan[5]) shows how TeX and LaTeX 2ε work together.

As graphically shown in Figure 1, the document preparation with TeX/LaTeX complies with the principle of formatting, i.e. the user edits the input file and the formatting commands. Finally, with the interplay of the TeX/LaTeX and

---

[5] TOBIN.

METAFONT systems, the file is compiled. The successfully compiled file, a device independent (DVI) file, can either be previewed or printed on paper. Thus the document processing in such a system goes through three distinct phases, i.e. editing the input, processing it, and previewing or printing the output. There exist also some quasi interactive (WYSIWYG) systems performing the editing and previewing.

2.2. The babel package

The babel package has been developed by JOHANNES BRAAMS[6] to support multilingual text processing in the LaTeX system. More vividly, the babel package is a collection of LaTeX macros that simplify the typesetting of multi-lingual documents. It is now an integral part of the LaTeX 2_ε system.

In the first versions, babel dealt mostly with languages that use the Latin script, but in later versions it was extended to Greek and Cyrillic. In addition to typesetting special characters, babel supports the replacement of built-in words like 'Chapter' or 'Figure' by language specific expressions.

2.3. The ethiop package

Since 1995 ethiop is developed by our research group at the University of Hamburg. This package supports Ethiopic for LaTeX. A similar product had been developed by ABASS B. ALAMNEHE known by the name EthTeX[7]. The ethiop group has based its work on EthTeX. Furthermore, the group has changed the existing letters slightly to make them more robust.

The main reason to develop ethiop was that the EthTeX package uses a special editor that is not available for all platforms and was not easily portable. Moreover, a special version of LaTeX 2.09 was required.

When we chose our transcription of Ethiopic characters, as near as possible to the scientific transliteration, it became obvious that we would have to activate some of the input characters, i.e. to treat them in a special way like 'accents'. Although we wrote our own set of macros to handle the activation, we soon decided not to introduce yet another incompatible mechanism for this task. Instead we used the babel package by JOHANNES L. BRAAMS[8] as framework for the implementation. The TeX macros are derived from the file `language.skeleton` provided with that package, which allows to add further languages to the babel package.

[6] BRAAMS.
[7] ETHTEX.
[8] BRAAMS.

The ethiop group rewrote the fonts completely and embedded them into the framework of the babel package. By this the ethiop produced a more robust and portable system that can be used with future LaTeX versions. The major part of the work was spent on adding new characters that were not present in the original fonts. Still more important is the addition of TeX ligatures to the fonts in order to support our input transcriptions.

3. Encoding Ethiopic

Encoding is defined as a system of assigning numeric values to characters [9], so that the characters can be used by a computer. Since the birth of computers, users strive to develop a universally applicable encoding system. This effort has come out with standards like ASCII, Unicode, etc. The encoding effort has broadened to include new scripts, also using new techniques for that purpose. A group of researchers interested in encoding Ethiopic for Unicode[10] succeeded with a first approval of recognition by the Unicode Standard Committee.

Until major standard organisations and software producing firms incorporate Ethiopic to their systems, electronic publishing with Ethiopic has explicitly to be set up by the users.

Because our transcription method does not provide support for a direct entry of Ethiopic characters, the input text has to be coded in some machine readable representation. One convenient method is to enter Ethiopic text via Latin letters. Hence we had to represent the Ethiopic character sets by Latin characters.

3.1. The Encoding System

The main features of the Ethiopic script are the syllabary (consonants with vowels) (ፊደል), numerals (አኃዝ), and punctuation marks (ሥርዓተ ነጥብ). A syllable consists of a consonant and a vowel, making one character. As mentioned before, the Ethiopic script is not only different from the Latin script, but it also consists of more than 450 characters.

The order of the syllabary is simply adopted from the modern Ethiopic syllabary layout, i.e. from *ha* to *pa* (ሀ to ፐ) in seven columns and the extras (i.e. the so-called diqalla) in five columns. The diqalla represent labialized consonants (consonants with ^{w}), like k^w in contrast to k.

---

[9] NADIN.
[10] YACOB.

Author/Group	1	2	3	4	5	6	7
	ጎዝ	ከዕብ	ሳልስ	ራብዕ	ኃምስ	ሳድስ	ሳብዕ
1. Guidi	a	u	i	ā	ē	−/ī	o
2. Littmann/Höfner	a/ä	u	i	ā	ē	ǝ/e	o
3. Leslau	a/ä	u	i	ā	ē	ǝ	o
4. Hammerschmidt	a	u	i	ā	ē	−/e	o
5. SERA	e	u	i	a	E	−	o
6. Aethiopica	ä	u	i	a	e	−/ǝ	o
7. ethiop	a	u	i	A	E	−/e	o

Author/Group	1	2	3	4	5
	ዲቃላዎች				
1. Guidi	ua	ui	uā	uē	uí/é
2. Littmann/Höfner	ᵘa/ᵘä	ᵘi	ᵘā	ᵘē	ᵘǝ/ᵘe
3. Leslau	ᵘä	ᵘi	ᵘa	ᵘe	ᵘǝ
4. Hammerschmidt	ʷa	ʷi	ʷā	ʷē	ʷ−/ʷe
5. SERA	We	Wi	Wa	WE	W
6. Aethiopica	ʷä	ʷi	ʷa	ʷe	ʷ−/ʷǝ
7. ethiop	ua	ui	uA	uE	ue

Table 1: Transcription and transliteration systems by some scholars

This encoding is based on scientific transcription techniques and is closely related to other encoding standards. One of these standards is SERA (System for Ethiopic Representation in ASCII), which is mainly intended for the recording and transmission of Ethiopic text within an ASCII environment. However, the SERA encoding was not realizable as an input encoding for TeX.

As far as the encoding is concerned, we have also referred to different sources adopted by different groups and individual scholars. These are the transliteration and transcription codings by HAMMERSCHMIDT[11], GUIDI[12], LITTMANN/ HOEFNER[13], LESLAU[14], SERA[15], in AETHIOPICA, and that one mostly used by the University of Addis Ababa[16]. All the adopted systems have their peculiar features. In most of the cases almost all consonant representations are

[11] HAMMERSCHMIDT.
[12] GUIDI.
[13] LITTMANN – HÖFNER.
[14] LESLAU (1987).
[15] YACOB.
[16] WRIGHT.

similar. The main difference is in the representation of the respective vowels. To comment the details of the strength and shortcomings of these different representations is beyond the scope of this paper. For the sake of comparison we have summarised the variations in the application of vowels in Table 1.

The representation of vowels, namely at the first (ገብዝ), fourth (ራብዕ), fifth (ኃምስ), the sixth (ሳድስ) positions, and those in the last five columns for diqalla (ዲቃላዎች) vary. These different representations could be classified into:

- In the first column or order (ገብዝ) the variation could be summarised in three types: a, a/ä and e.

- The fourth column or order (ራብዕ) shows variations between: a, a/ā, A .

- The fifth column or order (ኃምስ) shows variations between: E, e, and e/ē .

- The sixth column or order (ሳድስ) shows variations between: –/e, ī, e/ə, and ə.

- The variation in the diqalla is also similar to the representation of the vowels in the basic orders, but it uses some consonant + (^{w}(vowel), ^{w}(vowel), W(vowel), or two (i.e. (u+ vowel)) vowels together).

All the characters of Ethiopic for ethiop are developed using METAFONT. METAFONT is a system for the design of characters (and fonts) suited to raster-based devices that print or display text.

3.2. Syllabary

Ethiopic consists of 42 consonants (č occurring in two variants) followed by a vowel in the first 7 orders (assumed to be basic), and 5 other orders (diqalla) for labialized consonants. Ethiopic thus theoretically consists of 42 × (7 + 5) letters. But not all positions are actually filled (see Table 4).

In Tables 2 and 3 you can see the Ethiopic characters that are accessible using this package.

Characters with .d are only used for *Oromo*, all with 'q, 'k, 'h, 'g and mua, mui, muE, mue only for *Gurage (Chaha)*, all with .q only for *Tigrinya*, and all with fu, pu only for *Wolaytta*.

Note that the first column of Tables 2 and 3 refers to the transcription used by many authors while the second column denotes our encoding which the user should use for typing the text.

The reason for putting ኸ and ፇ under *ua* is that these characters are derived from the first column (ገብዝ), ኸ and ፀ, respectively.

		a	u	i	ā	ē	e	o	^{w}a	^{w}i	^{w}ā	^{w}ē	^{w}e
		a	u	i	A	E	e	o	ua	ui	uA	uE	ue
			U	I				O		uI	*if preferred*		
h	h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ			ኊ		
l	l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ			ሏ		
ḥ	.h	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ			ሗ		
m	m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ	ሟ	ሟ	ሟ	ሟ	ሟ
ś	's	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ			ሧ		
r	r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ			ሯ		
s	s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ			ሷ		
š	^s	ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ			ሿ		
q	q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ	ቈ	ቊ	ቋ	ቌ	ቍ
q̇	.q	ቐ	ቑ	ቒ	ቓ	ቔ	ቕ	ቖ	ቘ	ቚ	ቛ	ቜ	ቝ
b	b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ			ቧ		
v	v	ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ			ቯ		
t	t	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ			ቷ		
č	^c	ቸ	ቹ	ቺ	ቻ	ቼ	ች	ቾ			ቿ		
ḫ	_h	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ	ኈ	ኊ	ኋ	ኌ	ኍ
n	n	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ			ኗ		
ñ	~n	ኘ	ኙ	ኚ	ኛ	ኜ	ኝ	ኞ			ኟ		
ʾ	'	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ	ኧ				
k	k	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ	ኰ	ኲ	ኳ	ኴ	ኵ
ḵ	_k	ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ	ዀ	ዂ	ዃ	ዄ	ዅ
w	w	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ					
ʿ	'	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ					
z	z	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ			ዟ		
ž	^z	ዠ	ዡ	ዢ	ዣ	ዤ	ዥ	ዦ			ዧ		
y	y	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ					
d	d	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ			ዷ		
ḍ	.d	ዸ	ዹ	ዺ	ዻ	ዼ	ዽ	ዾ			ዿ		
ǧ	^g	ጀ	ጁ	ጂ	ጃ	ጄ	ጅ	ጆ			ጇ		
g	g	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ	ጐ	ጒ	ጓ	ጔ	ጕ
g̣	.g	ጘ	ጙ	ጚ	ጛ	ጜ	ጝ	ጞ			ጟ		

Table 2: The Ethiopic characters (continued on next page)

		a	u	i	ā	ē	e	o	^{w}a	^{w}i	^{w}ā	^{w}ē	^{w}e
		a	u	i	A	E	e	o	ua	ui	uA	uE	ue
			U	I				O		uI		*if preferred*	
ṭ	.t	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ			ጧ		
č	^C	ጨ	ጩ	ጪ	ጫ	ጬ	ጭ	ጮ			ጯ		
p	.p	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ			ጷ		
ṣ	.s	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ			ጿ		
c	.c	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ					
f	f	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ	ፚ				
p	p	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ					
q́	'q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ					
k	'k	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ					
h	'h	ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ					
ǵ	'g	ጘ	ጙ	ጚ	ጛ	ጜ	ጝ	ጞ					

		~mA	ᎃ
		~ri	ᎄ
		~fi	ᎅ

Table 3: The Ethiopic characters (continued)

3.3. Numbers

Today Arabic figures are more frequently used than the original Ethiopian figures. The ethiop package outputs the Arabic figures when the letters 0 up to 9 occur in the source code.

But Ethiopic figures can be typeset, too, by using the command \ethnum. If we enter \ethnum{1} we get ፩ as the result. But the macro \ethnum can do more than that. In fact it can convert all numbers up to 999999 to their Ethiopic equivalents: \ethnum{999999} gives ፱፻፺፱፼፱፻፺፱.

It should be noted that there exists some confusion in literature about the meaning of ፼ , some sources assigning to it the value 1000, and not 10000. If there is some uncertainty the user may also use the Amharic words ሺ for 1000, and ሚልዮን for 1000000.

Arabic	1	2	3	4	5	6	7	8	9		
Ethiopic	፩	፪	፫	፬	፭	፮	፯	፰	፱		
Arabic	10	20	30	40	50	60	70	80	90	100	10000
Ethiopic	፲	፳	፴	፵	፶	፷	፸	፹	፺	፻	፼

Table 4: The Ethiopic figures

3.4. Punctuation

Effort is made to code as much as possible all punctuations relevant to the Ethiopic writing system. Currently we have found 24 different punctuation symbols in use. At the time of encoding the application of ¿ is not clear, however we have included it.

Although the punctuation characters look different from the punctuation symbols of the Latin script, they have essentially the same meaning. We made some compromises between visual similarity and similiar interpretation when we chose the encoding of the punctuation characters. In Tables 5–7 we have collected the appropriate inputs for each of the characters.

input	:=	:-	::	,	;	?	'?	!	'!	...
output	፥	፦	፡፡	፣	፤	?	¿	!	¡	...

Table 5: The Ethiopic punctuation characters

input	<	<<	>	>>	'	''	'	''
output	‹	«	›	»	'	"	'	"

Table 6: The Ethiopic quotation characters

input	(	)	[	]	\$
output	(	)	[	]	$

Table 7: Special characters

The punctuation characters match their SERA equivalents closely.

3.5. Addition of new characters

A number of Ethiopian languages contain long vowels as phonemes (in most cases *ā, ē, ī, ō, ū*). E.g. in Oromo, *oromoo* is written ኦሮሞ, pronounced as *'oromo* disregarding the long *ō*. Actually, A and E (4th and 5th order) can be interpreted as long vowels *ā* and *ē*. But if *a* (1st order) has the meaning *ä* (ε) and E means *e*, then some problems arise.

Some other languages also contain tones, or even both, long vowels and tones. Another feature are doubled (geminated) consonants (ተበቅ). E.g.[17] the *m* in the word ስምንት is doubled (*semment*) but not written, whereas in ስምም (*semem*) it is doubled in written form but not in pronunciation. In መኩንኝ (*mak^w annenennā*) both occur.

Several solutions are possible. One of our proposals follows the system of Indian scripts with separate symbols after basic syllables. This can be done either by one symbol ǀ denoting the length of any vowel, e.g. ባǀ, ካǀ, ዳǀ, ጋǀ, ሳǀ, እርጕǀ.

Or individual symbols denoting the length of each vowel can be used: ǀ for *ā*, ι for *ē*, ι for *ī*, ٢ for *ō*, and ⊦ for *ū*, (or, if needed, ǀ for long *ɛ* and ɹ for *ā*). E.g. ባǀ, ካι, ዳι, ጋ٢, ሳ⊦, ጶɹ, እርጕ٢.

To solve the gemination problem, another suffix has to be introduced denoting the fact that the consonant in question is without vowel, e.g. ۱ put after consonants with *e* as in ብ۱.

If *A* and *E* are used for *ā* and *ē* then those for *ī*, *ō*, *ū* may be used only. We have designed these symbols in the style of Ethiopic characters, actually from the corresponding characters *q* or *t*: ቀ, ተ, ቄ, ቴ, ቂ, ቲ, ቆ, ቶ, ቄ, ቱ.

The other possibility is the use of diacritical symbols. Unfortunately, such symbols either increase the distance between lines, or the basic characters, and these are more or less all, have to be decreased in size, thus multiplying the set of Ethiopic characters.

Following the usage of linguistic articles in Ethiopia, the following diacritical symbols above basic characters may be introduced : ṅ for gemination ተብቅ, ṅ for long vowels, and n̈ for their combination, e.g. ስ̃ምንት and መኩ̈ንኝ̈.

Tones may be denoted by the following diacritical symbols above characters : ṅ for high and ṇ for low tone. These may be combined with long vowels and geminated consonants, e.g. ṅ́, ṅ̀, ṅ̂ (high-low), ṅ̌ (low-high), ቀ̃, ṅ̋, ṅ̋, ṅ̈, ꞓ̂, and ṅ̃.

So far, none of these proposals have been integrated into ethiop, but the corresponding characters have been designed using METAFONT.

4. Applications

Using the ethiop package is not really that difficult once you get used to it. The package consists of what we need to process a text in Ethiopic writing system with Ethiopic character sets. Besides some applications like inserting a date

---

[17] WRIGHT.

and writing mathematical notations and formulas are included.

4.1. Typing the text

In Tables 2 and 3 you can see the Ethiopic characters that are accessible using this package.

We will now explain how the characters are entered. Every character represents a syllable that consists of a consonant followed by a vowel. The characters of the sixth order either phonetically represent a syllable with ə, or a pure consonant, especially in combination with other consonants. (Example : ትምህርት pronounced as təmhərt). Wherever possible, every character is encoded in a way that matches its pronunciation or its scientific transcription as closely as possible.

As an example we choose the letter ዳ which is pronounced *da* and entered as da. You will find the character in the character table in row d and column a.

The letter ደ might represent the syllable *de*, but it might also denote the consonant *d* without an accompanying vowel. To reflect this, it is possible to enter either de or d at the user's choice.

If accented characters are used in the scientific transcription of an Ethiopic syllable, they are entered without the accent, but prefixed with an appropriate special character. ጨ has *ča* as its transcription, hence we will enter it as ^ca which is as close to the proper transcription as we can get.

Long vowels are usually indicated by a bar, *dē* can serve as an example. But since long vowels are fairly frequent, we did not want to use two letters for their representation. Instead we have employed the uppercase letters for this purpose, which leads us to dE as our transcription of the aforementioned syllable ዲ.

When a vowel does not have both a short and a long form, like the *i* in *di*, there is no need to insist on the proper case for the vowel. Hence we might enter either di or dI and get ዲ in both cases.

Some consonants may be followed by a diphthong, i.e. a combination of two vowels. A diphthong occurs e.g. in ዷ which is best transcribed *dwā*. We have coded it as duA, thereby slightly deviating from the standard transcription. But this is unavoidable, because if we enter dwA, we get an ambiguity with d wA which we want to result in ድዋ. The SERA encoding, which was developed by DANIEL YACOB, YITNA FIRDYIWEK, YONAS FISSEHA and others, suggests dWA, which has been considered as an alternative to the present encoding. It did not show any significant advantages, however.

All characters with `.d` are only used for *Oromo*, all with `'q`, `'k`, `'h`, `'g` and `mua`, `mui`, `muE`, `mue` only for *Gurage (Chaha)*, all with `.q` only for *Tigrinya*, and all with `fu`, `pu` only for *Wolaytta*.

4.2. Spaces

When Ethiopic is printed today, an interword gap is signalled by a white space, in the same way as it is done for the Latin script. This kind of spacing can be used simply as in ordinary LaTeX documents.

But formerly word breaks used to be denoted by the character ፡ and even today this method is used for handwriting. One can compare the following two examples, i.e. with ፡ and without ፡ representing the old and the handwriting, and the modern writing systems.

```
'abAs :   'alamenahe:   'abAs :'alamenahe:
'abAs:'alamenahe:'abAs:'alamenahe:
'abAs:'alamenahe:'abAs:'alamenahe::                    results in
```

አባስ፡አላምነሁ፡አባስ፡አላምነሁ፡አባስ፡አላምነሁ፡አባስ፡አላምነሁ፡አባስ፡አላምነሁ፡አባስ፡
አላምነሁ።

or

```
'abAs 'alamenahe 'abAs 'alamenahe

'abAs 'alamenahe 'abAs 'alamenahe::                    gives
```

አባስ አላምነሁ አባስ አላምነሁ

አባስ አላምነሁ አባስ አላምነሁ።

in the output. As can be seen, line breaks are allowed after ፡ even if there is no space character in the source. Space characters immediately following or preceeding a ፡ in the input are ignored as we can see from the first line. Therefore new lines in the input will not cause any problem either. An empty line in TEX indicates a new paragraph as can be seen in the second example.

4.3. Line breaking

The ordinary spaces as well as the white space surrounding an ፡ can be stretched a little, so that it is possible to achieve proper adjustment. Nevertheless, the lack of hyphenation in the Ethiopic makes itself felt from time to time, when TEX cannot find suitable breakpoints for a paragraph.

There are a few standard solutions to this problem, the easiest is to use a `sloppypar` environment which allows TeX to stretch the interword spaces more than usual. But this does not work when some words are simply too long. In that case one may want to rewrite the sentence that causes the bad break, maybe only by changing a few words.

However, the text might not be easily changeable, e.g. because it is a quote from some other source, or the author insists on that very phrase. In that case you can insert a \- into a suitable breakpoint where the text will be split between two lines. Unlike the usual TeX behaviour, no hyphen will be added at the breakpoint. This kind of line breaking is especially well suited when the character ፡ is used for interword spaces.

But maybe even the insertion of breakpoints is impossible. In this case, some explicit `\hspace`, i.e. inserting some white space, must be added in a suitable position or a ragged right layout must be selected during the whole document or part thereof.

4.4. Ethiopian dates

The calendar system is implemented in our package, so it is possible to type `\today` and get ሕምሌ 13 1992. (This is the date on which this document has been translated).

The simple algorithm converting Gregorian dates into Ethiopian dates can be translated into a program using the TeX primitives `\year`, `\month`, and `\day`. The implementation in TeX is straightforward, but hardly readable, because TeX's expressiveness for formulas is very weak. As calendar routines are provided for all the other babel language definitions, it was obvious that this problem had to be addressed, too.

4.5. Implementation notes

Because we have to deal with so many characters, we placed them in two separate font tables. In the first font we preferred to place characters that result from a TeX ligature (which must not be confused with a ligature from ordinary printing) in the positions 0–31 and 128–255. These characters usually cannot be entered from a keyboard and hence it is safe to assume that they resulted from a ligature.

Only the characters in the primary font are accessible by entering ordinary characters and forming ligatures. For the other characters it is necessary to select explicitly the secondary font within the TeX code. But this requires the execution of TeX macros, hence the activation of some characters was required.

The activated characters can inspect the following characters and request the necessary font change. In fact this method of enlarging the number of available characters is quite general and is also used for other languages, and to provide a unified input mechanism for all Latin characters with accents.

For a complete documentation of the input parsing mechanism we refer the interested reader to the commented source code in `ethiop.dtx`, which can also be typeset by LaTeX to get a more readable version.

4.6. Common Pitfalls

After discussing some of the internals of the ethiop package, we are now prepared to examine some of the problems that result from our implementation.

First of all, the characters ~, ^, ', _, and . are made active. This is unavoidable since we chose a transcription as near as possible to the scientific transliteration. But there are some drawbacks.

- We cannot use ^^ for entering special characters. Usually this is done in package files only, so we do not get into real trouble, since babel activates the characters only at the beginning of the document.
- We cannot use a . in numbers and TeX dimensions while Ethiopic text is being typeset. We can circumvent this problem by using a , instead of the . when entering numbers for TeX. Note, that we can use the . without problems when we have temporarily switched to a language other than using Ethiopic, e.g. English.

In the newest version of babel some problems with activation might occur.

Since our package is not yet an integral part of the babel system, we have some other difficulties to overcome.

- If the babel package is included explicitly in the header of a document, e.g. to typeset both English and Ethiopic, it must be included *before* the ethiop package. A failure to do so results in weird error and warning messages. But by putting the `\usepackage{ethiop}` in the second place, Ethiopic is always selected as the main script of the text. If this is not intended, the main language must be explicitly reset by a `\selectlanguage` command after the `\begin{document}`.
- We do not know what the future will bring. While the present transliteration system *might* work with future versions of babel, there is no guarantee that it will. So you should keep your old version of babel until you are sure it works with the ethiop package or until an updated version of ethiop is issued.

Typing errors will usually not result in an error message. Instead, a black rectangle will appear in the output, if some illegal character is encountered.

- Because it is allowed to enter consonants without a trailing vowel, there are plenty of typos that simply result in the wrong letters being printed.

So look at your finished document carefully.

5. Concluding Remarks

Much work has been done during the development of this package, but much work had been done before and there are still open points. As the preceding efforts and developments have given us insight, we believe our work might also be helpful for those who can continue researching in the development of the Ethiopic script and text processing system. Although ethiop was first initiated as an academic interest, the result has become a vital system, and could enrich document preparation with Ethiopic. It is also more viable that it could facilitate further research activities in different or similar fields.

5.1. What remains to be done?

- Bugs must be detected. We know that there are some, but we do not know which. Please report bugs to ethiop@informatik.uni-hamburg.de, we will be happy to make some updates.
- More languages should be added. To do this, we have to know the phonemic inventory of such languages, especially which additional phonemes have to be encoded. We need examples of usage for as many of such languages of Ethiopia and Eritrea as possible. Suggestions for these and other improvements can be sent to ethiop@informatik.uni-hamburg.de.
- More characters must be added. We have provided almost all characters of Unicode, but there might be some more.
- We have to find out more about the conventions for typesetting Ethiopic.
- Conventions for the encoding of other vowels, or short and long vowels must be looked for.

5.2. Acknowledgements

References

AKLILU BERHAN = አክሊሉ ብርሀን ወልደ ቄርቆስ፥ ስለ ገዐዝና አማርኛ ቋንቋ ታሪከ። ንግድ ማተሚያ ቤት፤ አዲስ አበባ 1958።

ALAMNEHE = ALAMNEHE, ABASS: EthTₑX.
    ftp://ftp.dante.de/tex-archive/languages/ethiopia/ethtex/

ABATE – WUBNEH = ABATE, YOHANNIS – WUBNEH, MULATU: *Ethiopia.* Boulder (Colorado 1989).

ANDRE = JACQUES ANDRE: *Structured Documents: Winter School on Document Structures;* Cambridge Univ. Press (Cambridge 1989).

BENDER – BOWEN – COOPER = BENDER, M.L. – BOWEN, J.D. – COOPER, R.L.; FERGUSON, C.A.: *Language in Ethiopia.* Oxford University Press, (London 1997).

BENDER = BENDER, M. LIONEL (ed.): *The Non-Semitic Languages of Ethiopia.* Michigan State University, (East Lansing 1976).

BEYENE – KUDLEK – KUMMER = BEYENE, BERHANU – KUDLEK, MANFRED – KUMMER, OLAF; Metzinger, Jochen: *The* ethiop *package.* Fachbereich Informatik, Universität Hamburg, 1997.
    ftp://ftp.dante.de/tex-archive/languages/ethiopia/ethiop/

BRAAMS = BRAAMS, JOHANNES L.: *The* babel *package.*
    ftp://ftp.dante.de/tex-archive/macros/latex/packages/babel/

ECOSA = http://ecosa.ethiopianonline.net

ETHTEX = EthiO Systems: EthTₑX.
    http://www.neosoft.com/~ethiosys/ethtex/ethtex.html

ETNOLOGUE = SUMMER INSTITUTE OF LINGUISTICS: *Ethnologue.* (Dallas, Texas 1996).
    http://www.sil.org/ethnologue/

GUIDI = GUIDI, IGNAZIO: *Vocabolario Amharico-Italiano.* Casa Editrice Italiana, (Roma 1901).

HAMMERSCHMIDT = HAMMERSCHMIDT, ERNST: *Äthiopische Handschriften vom Ṭānāsee 1, Verzeichnis der orientalischen Handschriften in Deutschland XX 1.* Franz Steiner Verlag, (Wiesbaden 1973).

KNUTH (1986A) = KNUTH, DONALD E.: *The* TₑX *book.* Addison Wesley, (Reading, Massachusetts 1986).

KNUTH (1986B) = KNUTH, DONALD E.: *The* METAFONT *book.* Addison Wesley, (Reading, Massachusetts 1986).

LAMPORT = LAMPORT, LESLIE: LATₑX – *A Document Preparation System.* Addison Wesley, (Reading, Massachusetts 1986).

LESLAU (1965) = LESLAU, WOLF: *Ethiopians Speak*, Studies in Cultural Background, Vol. 2. University of California Press, (Berkeley 1965).

LESLAU (1987) = LESLAU, WOLF: *Comparative Dictionary of Ge'ez (Classical Ethiopic)*. Otto Harrassowitz, (Wiesbaden 1987).

LITTMANN – HÖFNER = LITTMANN, ENNO – HÖFNER, MARIA: *Wörterbuch der Tigrē-Sprache*. Franz Steiner Verlag, (Wiesbaden 1962).

NADIN = NADIN, J.: *Developing International Software: for Windows95 and WindowsNT*; Microsoft Press 1997.

SERA = *The Unicode Technical Report #1 – Draft Proposal on Ethiopian Script*. (Houston, TX, August 1993).

TAKLA WALD = ተክለ ወልድ ደስታ፥ የአማርኛ መዝገበ ቃላት። አርቲስቲክ ማተሚያ ቤት፣ አዲስ አበባ 1970።

TOBIN = TOBIN, GEOFFREY: METAFONT *for Beginners*. Baskerville - The Annals of the UK TeX users' Group Vol. 3. No.1.

UNICODE = `http://www.unicode.org/`

WRIGHT = WRIGHT, STEPHEN: *The Transliteration of Amharic*. Journal of Ethiopian Studies, Vol. II No. 1, (Addis Ababa 1964).

YACOB = YACOB, DANIEL: *SERA FAQ*.
`http://www.cs.indiana.edu/hyplan/dmulholl/fidel/sera.html`

A. Examples

We provide the first sample text to illustrate the appearance of our font. The LaTeX source of the text begins with

```
pAduwA bametbAl web ya'i.tAliyA katamA si~nor bAptisetA
yatasa~nu 'and tegu_h 'sarAta~nA yenoru nabar:: 'enih sawe
hulat qon^go sEto^c le^go^c nabaruA^caw::
kuAduA.helinA yatabAla^cew webatuAnA
.tabAyua yatamuAlA naw:: puafEluA
webatuA 'enda 'e_hetuA 'amuAlto selAlesa.tAt ba'amala~nAnatuA
tetAwaqAla^c::

puafEluA lAbAtuA yabakuir le^g bamahonuA bazih 'alabelAbit
melAsuA sAbiyA menAlbAt bAl 'a.tetA qumo-qar 'enedAthon 'abAtuA
segAt 'aderobA^caw nabar::
```

```
yafarut 'alqaram laqon^gowA kavEreno 'eska quidA_hualA sigorf
la'amal~nAwA 'a.tagabuA yamiders saw .tafA::
```

and gives the following result:

ፓዱዋ በምትባል ውብ የኢጣሊያ ከተማ ሲኞር ባፕቲስታ የተሰኘ አንድ ትጉግ
ሠራተኛ ይኖሩ ነበር።። እኒህ ሰው ሁለት ቆንጆ ቤቶች ልጆች ነበሯቸው።። �托ዲሕሊና
የተባለቸው ውብቲና ጠባየ የተሟላ ነው።። ፔፉኒ ውብቲ አንደ አጓቲ አሚልቶ ስላልሰጣት
በመጠኛነቲ ትታወቃለች።።

ፔፉኒ ላባቲ የበኩC ልጅ በመሆኒ በዚህ አለብላቢት ምላጒ ሳቢያ ምናልባት ባል
አጥታ ቄሞ-ቀር አንዳትሆን አባቲ ስጋት አድርባቸው ነበር።።

የፉሩት አልቀረም ለቆንጀዋ ከሺCና አስከ ቀኑዳጐላ ሲጎርፍ ለ0መልኛዋ አጠገጒ
የሟድርስ ሰው ጠፉ።።

The next example will clarify the usage of bold and slanted Ethiopian fonts.
Italic characters are mapped to slanted characters. The font selection works
just as in ordinary LaTeX with NFSS. The source text

```
'adis 'ababA \textbf{'adis 'ababA}
\textsl{'adis 'ababA \textbf{'adis 'ababA}}
```

gives us አዲስ አበባ **አዲስ አበባ** አዲስ አበባ **አዲስ አበባ** as the output.

Besides we can also apply different type sizes as shown below.

```
{\tiny 'adis 'ababA} {\small 'adis 'ababA} 'adis 'ababA
{\large 'adis 'ababA} {\huge 'adis 'ababA}
```

gives us:

አዲስ አበባ አዲስ አበባ አዲስ አበባ አዲስ አበባ አዲስ አበባ

as output.

In the following examples we illustrate the use of characters indicating vowel
length in some Ethiopian languages.

Amharic : የጦግን ብልጠት፦ የላምን የዋሀነት፦ የአንበሳን ጉልበት የሟለካ ሰው
አልተገኘገም።።

Tigrinya : ዋI አንታ በሎI፡ጉ-IአጉIአ ከIበCታ።። or ዋI አንታ በሎየ፡ጉ-ተአጉ-ተአ
ከየበCታ።።

Oromo : በሩ፦ ቄቢውወን ደቻ በከከ ሰዲትቲ ቆIድኄI ከIኡIስ ኂ ደንደIኘ ዱIላ።።
or በየሩ፦ ቄቢውወን ደቻ በከከ ሰዲትቲ ቆየድኄI ከIኡተስ ኂ ደንደIኘ ዱIላ።።

We can also apply math mode:

```
$$\frac{\ethmath{ba}+\ethmath{da}^2}{\ethmath{^ca}}\qquad
  (b+b)^2-(\ethmath{ba}+\ethmath{ba})^2\qquad
  \ethmath{ba}+\frac{\ethmath{ba}}%
  {\ethmath{ba}\sp{\ethmath{ba}\sp{\ethmath{ba}}}}$$
```

$$\frac{በ + ደ^2}{ቻ} \qquad (b+b)^2 - (በ + በ)^2 \qquad በ + \frac{በ}{በ^{በ^በ}}$$

In Table 9 we show a financial budget as an example of the combination of LaTeX and ethiop.

ርእስ	የወር ወጪ አይነት		የወር ወጪ (በብር)
ምግብ	ጤፍ	1 ኩንታል	250
	ቅቤት	10	120
	መጠጥ	10020 ሊትC	80
	ቅጠላቅጠል	20 ኪሎ	50
	ጥራጥሬ	15 ኪሎ	45
ትራንስፖCት	ነዳጅ	40 ሊትC	100
	የአውቶቡስ ቲኬት	በቀን 2 ብC	45
መዝናኛ	ለስፖCት ክለብ	በወC 50 ብC	50
	ሽCሽC	በወC 50 ብC	50
ጠቅላላ ድምC			795

B. Installation

Installing the ethiop package is not difficult.

1. Check the prerequisites for this package. Make sure:

   - that you have installed TeX and METAFONT.
   - that the files cmbase.mf and romand.mf from the Computer Modern fonts are accessible to METAFONT.
   - that you have installed LaTeX 2_ε with a release date of 1996/12/01 or later. It can be found on CTAN (e.g. at ftp.dante.de) in the directory tex-archive/macros/latex/.
   - that you have installed the babel package with a release date of 1997/01/23 (version 3.6h) or later. It is also on CTAN in the directory macros/latex/packages/babel/.

Berhanu Beyene, Manfred Kudlek, Olaf Kummer

- to get the ethiop package, which is located on CTAN in the directory `languages/ethiopia/ethiop/`. Now the files mentioned in `MANIFEST` should be present.

2. Generate the TeX files from their `docstrip` source by running

   `tex ethiop.ins`

   in the directory where the files `ethiop.ins` and `ethiop.dtx` reside.

3. Verify the installation by generating this documentation file from its source `ethiodoc.tex`. Run LaTeX twice to get the references right.

   ```
   latex ethiodoc.tex
   latex ethiodoc.tex
   ```

   The resulting file `ethiodoc.dvi` is the user level documentation of our package.

Summary

Encoding Ethiopic for computer application is increasingly important for linguists, librarians, information and communication technologists. This paper focusses on encoding Ethiopic for the LaTeX document preparation system, and on the methods and principles of the ethiop package which supports Ethiopic for LaTeX. ethiop is a package developed by a research group at the University of Hamburg, that integrates Ethiopic to the TeX/LaTeX system.

We hope that this package opens up a new venue to scientific and mathematical document processing, using the rich and well developed TeX/LaTeX system. Moreover, it could render a better data exchange mechanism by virtue of its platform independence and plain ASCII nature. [18]

For questions and remarks please contact

Berhanu Beyene, Manfred Kudlek, Olaf Kummer
Universität Hamburg, FB Informatik, AB TGI
Vogt-Kölln-Straße 30, D-22527 Hamburg

---

[18] Different from the software normally used in AETHIOPICA the layout of this article has been produced with LaTeX and the special ethiop package.