

# Encoding the Ethiopian Script for $\LaTeX$

Berhanu Beyene      Manfred Kudlek  
Olaf Kummer      Jochen Metzinger

Universität Hamburg, FB Informatik, AB TGI  
Vogt-Kölln-Straße 30, D-22527 Hamburg

5th May 1997

## 1 Introduction

Parallel to the development of information and communication technology, text processing systems like  $\TeX$  gained attention. The  $\TeX$  text processing system produces the finest quality of technical and scientific manuscripts, books, documents etc. In addition,  $\LaTeX$  (a macro that enriches the development of  $\TeX$ ) and `babel` (a special package to support non-Latin scripts) have contributed to boost text processing technology in different scripts.

For this reason we developed the `ethiop` package [5] for Ethiopian script supported by the `babel` package which is a collection of `METAFONT` and  $\TeX$  /  $\LaTeX$  macros. The package helps to typeset the character sets of the different languages in Ethiopia and Eritrea.

We also consider the origin of the Ethiopian languages and script, the `ethiop` package, the encoding system and the applications. 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.

### 1.1 The origins of the Ethiopian languages and script

Linguists suppose that the original home of the Afro-Asiatic cluster of languages was somewhere in northeastern Africa, possibly in the area between the Nile River and the Red Sea. Some assert that during the first millennium B.C. and possibly even earlier, various Semitic-speaking groups from Southwest Arabia began to cross the Red Sea and settle along the coast and in the nearby highlands. These migrants brought with them their Semitic languages (Sabaeans and perhaps others) and script (Old Epigraphic South Arabic) and monumental stone architecture (based on information from [2, 4]).

There exist 86 different languages (4 of them extinct), classified in four major language families with around 200 dialects. Three language families belong to the Afro-Asiatic super-language family: Semitic (represented by the branches Ethio-Semitic and Arabic), Cushitic, and Omotic. Speakers of these language families amount to about 98% of the population. The fourth family is East

South Arabian	Ethiopian	South Arabian	Ethiopian	South Arabian	Ethiopian
’	ከ	’	አ	ś	ṣ
b	በ	b	በ	t	ተ
g	ገ	g	ገ	t̄	ፈ
d	ደ	d	ደ		
ḍ	ዘ	z	ዘ	1	፩
h	ሃ	h	ሀ	5	፮
w	ወ	w	ወ	10	፱
z	ዘ	z	ዘ	50	፷
ḥ	ሐ	ḥ	ሐ	100	፻
ḫ	ኀ	ḫ	ኀ	1000	፻፺፮
t	ተ	q	ቀ	N	፻፺፮
z̄	ፈ	r	ረ		
y	የ	š	ሠ	:	:

Table 1: Comparison of the South Arabian and the Ethiopian script

Sudanic, which belongs to the Nilo Saharan super language family. It is represented by Gumuz, Suri, Komman, Berta, Kunama and some others.

The Ethiopian script has its origin in the South Semitic alphabet which has been used for Thamudene, Dedanite, Liḥyanite, Ṣafaitic, Minaean, Sabaeen, Ḥimyaritic, Qatabanic, and Ḥaḍramautic. These are Semitic languages once spoken in the Arabian peninsula. The alphabet was without vocalization. The South Semitic characters are known from stone monuments only, whereas the Ethiopian characters are known from stone inscriptions as well as from parchment and paper documents. The writing direction of South Semitic was from right to left or the reverse (Bustrofedon).

In the 2<sup>nd</sup> half of the 1<sup>st</sup> millennium B.C. immigrants from South Arabia established a kingdom at Aksum. Later they changed the original South Arabic script into the Ge’ez (ግዕዝ) script, which was vocalised about 350 A.D. To trace the origins and development of the Ethiopian script a comparison between the South Arabian and the Ethiopian script is shown in Table 1.

At present the Ethiopian script is used as an official writing system for a number of languages in two states, Ethiopia (አ.ጎ.ዮ.ጵ.ዶ) and Eritrea (ኢ.ር.ጎ.ረ.). For some major languages statistical data are presented (summarised from [8]). Note that Amarinya, Guraginya, Afarinya, Hadiyyinya, Kambaatinya, Orominya (Oromiffa), Sidaminya, Somalinya and some others are the expressions in Amharic for the different languages, whereas the names in the second column of Table 2 are from [8].

## 1.2 The origins of the babel and ethiop package

The ethiop package has been developed as a project at the University of Hamburg since 1995. To develop the package, fortunately, we did not have to start from scratch, because there were some Ethiopian fonts available already.

The fonts are based on EthT<sub>E</sub>X which was originally distributed by Abass B.

Language family	Languages		No. of speakers in 1000 (year 91–95)
Semitic	Ge'ez	ግዕዝ	only in church
	Amharic	አማርኛ	20000
	Gurage	ጉራግኛ	1850
	Tigrinya	ትግርኛ	6050
Cushitic	Afar	አፋርኛ	750
	Agew	አገዌኛ	1000
	Hadiyya	ሀዲያኛ	1000
	Kambaata	ክምባትኛ	1000
	Oromo	ኦሮምኛ	14000
	Sidamo	ሲዳሞኛ	1500
	Somali	ሶማልኛ	2050
Omotic	Gamo/Gofa/Dawro	ጋምኛ	780
	Wolaytta	ወላይትኛ	2000
East Sudanic	Gumuz	ጉሙዚኛ	50
	Suri	ሱሪኛ	20

Table 2: Languages and number of speakers

Alameneh. The genuine EthT<sub>E</sub>X package can be found on CTAN in the directory `language/ethiopia/ethtex/` or in `http://www.neosoft.com/ethiosys/` [7]. We changed the existing letters a little to make them more robust at low resolutions. 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 T<sub>E</sub>X ligatures to the fonts in order to support our input transcriptions.

When we chose our transcription scheme, it became obvious that we would have to activate some of the input characters. 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 [6] as framework for the implementation. The T<sub>E</sub>X macros are derived from the file `language.skeleton` provided with that package, which also allows a more well-rounded support of multiple languages.

The `babel` package is a collection of T<sub>E</sub>X macros that simplify the typesetting of multi-lingual documents. It is now an integral part of the L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> 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 texts.

Since the original EthT<sub>E</sub>X package used a special editor that is not available for all platforms, it was not easily portable. Moreover, a special version of L<sup>A</sup>T<sub>E</sub>X 2.09 was generated, thereby preventing an upgrade to the now current L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>. By rewriting the language support completely and by embedding the Ethiopian fonts into the framework of the `babel` package we produced a more robust and portable system that can be used with future L<sup>A</sup>T<sub>E</sub>X versions.

## 2 The Encoding System

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

The main features of the Ethiopian script are the syllabary (consonants with vowels) (**ረደል**), numerals (**ዐሐአዝ**) and punctuation marks (**ሥርዐተ ነጥብ**). A syllable consists of a consonant and a vowel, making one character. As mentioned before the Ethiopian script is not only different from Latin script, but it also consists of more than 400 characters. The order of the syllabary is simply adopted from the modern Ethiopian syllabary layout i.e. from *ha* to *pa* (**ሀ** to **ጥ**) in seven columns and the extras (i.e. the so-called dikallas) in five columns. The dikallas represent labialized consonants (consonants with *w*), like *k<sup>w</sup>* in contrast to *k*.

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 Ethiopian text within an ASCII environment. However, the SERA encoding was not realizable as an input encoding for T<sub>E</sub>X.

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 E. Hammerschmidt [10], Guidi [9], Littmann/Hoefner [16], Leslau [15], SERA [21], and that mostly used by the University of Addis Ababa [20]. All the adopted systems have their peculiar features. In most of the cases almost all consonant representations are 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 and to serve the purpose, we have summarised the variations in the application of vowels in Table 3.

The representation of vowels, namely at the first (**ግዕዝ**), fourth (**ረብዕ**), fifth (**ሃምስ**), sixth (**ሳድስ**) positions and those in the last five columns for dikallas (**ዲቃላዎች**) vary.

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

### 2.1 Syllabary

The Ethiopian character set (syllabary) consists of 41 consonants (  $\check{c}$  occurring in two variants of characters ) followed by a vowel in the first 7 orders (assumed to be basic), and 5 other orders (dikallas) for labialized consonants. The Ethiopian script thus theoretically consists of  $42 \times (7 + 5)$  letters. But not all positions are actually filled (see Table 4).

Note that the first column of Table 4 refers to the transcription used by Hammerschmidt (and some other authors) while the second column denotes our encoding which the user should use for typing the text.

The reason for putting **ጸ** and **ፉ** under *ua* is that these characters are derived from **አ** and **የ**, respectively.



		a	u	i	ā	ē	e	o	wa	wi	wā	wē	we
		a	u	i	A	E	e	o	ua	ui	uA	uE	ue
		U	I					0	uI	<i>if preferred</i>			
h	h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ			ሸ		
l	l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ			ሸ		
h	.h	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ			ሸ		
m	m	መ	ሙ	ሚ	ሚ	ሚ	ም	ሞ	ሙ	ሙ	ሚ	ሚ	ም
ś	's	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ			ሢ		
r	r	ረ	ሩ	ሪ	ራ	ራ	ር	ሮ			ሪ		
s	s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ			ሲ		
š	˘s	ሸ	ሹ	ሺ	ሻ	ሼ	ሽ	ሾ			ሺ		
q	q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ	ቁ	ቁ	ቂ	ቂ	ቁ
q	.q	ቈ	቉	ቊ	ቋ	ቌ	ቍ	቎	ቁ	ቁ	ቂ	ቂ	ቁ
b	b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ	ቡ	ቡ	ቢ	ቢ	ቡ
v	v	ቨ	ቩ	ቪ	ቫ	ቬ	ቭ	ቮ	ቡ	ቡ	ቢ	ቢ	ቡ
t	t	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ			ቲ		
č	˘c	ቸ	ቹ	ቺ	ቻ	ቼ	ች	ቾ			ቺ		
h	_h	ሳ	ሴ	ስ	ሶ	ሷ	ሸ	ሹ	ሳ	ሳ	ስ	ስ	ሳ
n	n	ነ	ኑ	ኒ	ና	ኔ	ኖ	ኘ			ኒ		
ñ	˘n	ኸ	ኹ	ኺ	ኻ	ኼ	ኽ	ኾ			ኺ		
,	,	አ	ሉ	ሊ	ላ	ሌ	ል	ሎ	አ				
k	k	ከ	ከ	ከ	ከ	ከ	ከ	ከ	አ	አ	አ	አ	አ
k	_k	ኸ	ኸ	ኸ	ኸ	ኸ	ኸ	ኸ	አ	አ	አ	አ	አ
w	_w	ወ	ወ	ወ	ወ	ወ	ወ	ወ	አ	አ	አ	አ	አ
‘	‘	ዐ	ዐ	ዐ	ዐ	ዐ	ዐ	ዐ					
z	z	ዘ	ዘ	ዘ	ዘ	ዘ	ዘ	ዘ			ዘ		
ž	˘z	ዠ	ዡ	ዢ	ዣ	ዤ	ዥ	ዦ			ዢ		
y	y	የ	የ	የ	የ	የ	የ	የ	የ				
d	d	ደ	ደ	ደ	ደ	ደ	ደ	ደ			ደ		
d	.d	ደ	ደ	ደ	ደ	ደ	ደ	ደ			ደ		
ḡ	˘g	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ			ጀ		
g	g	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ
g	.g	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ገ
t	.t	ጠ	ጠ	ጠ	ጠ	ጠ	ጠ	ጠ			ጠ		
č	˘C	ጠ	ጠ	ጠ	ጠ	ጠ	ጠ	ጠ			ጠ		
	_C	ጠ	ጠ	ጠ	ጠ	ጠ	ጠ	ጠ			ጠ		
p	.p	ጸ	ጸ	ጸ	ጸ	ጸ	ጸ	ጸ			ጸ		
s	.s	ጸ	ጸ	ጸ	ጸ	ጸ	ጸ	ጸ			ጸ		
ç	.c	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ			ፈ		
f	f	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ	ፈ
p	p	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ	ፐ
q	'q	ቆ	ቆ	ቆ	ቆ	ቆ	ቆ	ቆ	ፐ	ፐ	ፐ	ፐ	ፐ
k	'k	ኸ	ኸ	ኸ	ኸ	ኸ	ኸ	ኸ	ፐ	ፐ	ፐ	ፐ	ፐ
h	'h	ኸ	ኸ	ኸ	ኸ	ኸ	ኸ	ኸ	ፐ	ፐ	ፐ	ፐ	ፐ
ḡ	'g	ገ	ገ	ገ	ገ	ገ	ገ	ገ	ፐ	ፐ	ፐ	ፐ	ፐ

Table 4: The Ethiopian characters

Arabic	1	2	3	4	5	6	7	8	9		
Ethiopian	፩	፪	፫	፬	፭	፮	፯	፰	፱		

  

Arabic	10	20	30	40	50	60	70	80	90	100	10000
Ethiopian	፲	፳	፴	፵	፶	፷	፸	፹	፺	፻	፳፻

Table 5: The Ethiopian figures

input	:=	:-	::	,	;	?	'?	!	'!	...
output	፩	፪	፫	፬	፭	፮	፯	፰	፱	...

Table 6: The Ethiopian punctuation characters

input	<	<<	>	>>	,	''	'	''
output	<	፪	>	፬	,	''	'	''

Table 7: The Ethiopian quotation characters

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

Table 8: Special characters

## 2.4 Addition of new characters

A number of Ethiopian languages contain long vowels as phonemes (in most cases  $\bar{a}$ ,  $\bar{e}$ ,  $\bar{i}$ ,  $\bar{o}$ ,  $\bar{u}$ ). E.g. in Oromiffa, *oromoo* is written ጸሮጦ።, pronounced as 'oromo disregarding the long  $\bar{o}$ . Actually, *A* and *E* (4<sup>th</sup> and 5<sup>th</sup> order) can be interpreted as long vowels  $\bar{a}$  and  $\bar{e}$ . But if *a* (1<sup>st</sup> order) has the meaning  $\bar{a}$  ( $\epsilon$ ) and *E* means *e*, then some problems arise.

Some other languages contain also tones, or even both, long vowels and tones. Another feature are doubled (geminated) consonants (ጥብቅ). E.g. [20] the *m* in ስምንት is doubled (*semment*) but not written, whereas in ስምም (*se-mem*) it is doubled in written form but not in pronunciation, and in መኩንንና (*mak<sup>w</sup> annenennā*) both occur.

Several solutions are possible. The first we propose follows the system of Indian scripts with separate symbols after basic syllables. This can be done either by one symbol *l* denoting the length of any vowel, e.g. ሀl, ከl, ጸl, ጎl, ሰl, ጸሮጦl.

Or individual symbols denoting the length of each vowel can be used: *l* for  $\bar{a}$ , *l* for  $\bar{e}$ , *l* for  $\bar{i}$ , *l* for  $\bar{o}$ , and *l* for  $\bar{u}$ , (or, if needed, *l* for long  $\epsilon$  and *l* for  $\bar{a}$ ). E.g. ሀl, ከl, ጸl, ጎl, ሰl, ጥl, ጸሮጦl.

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

If  $A$  and  $E$  are used for  $\bar{a}$  and  $\bar{e}$  then those for  $\bar{i}$ ,  $\bar{o}$ ,  $\bar{u}$  may be used only. We have designed these symbols in the style of Ethiopian characters, actually from the corresponding characters  $q$  or  $t$ :  $\Phi$ ,  $\Psi$ ,  $\Theta$ ,  $\Upsilon$ ,  $\Xi$ ,  $\Gamma$ ,  $\Sigma$ ,  $\Lambda$ ,  $\Omega$ .

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 Ethiopian characters.

Following the usage of linguistic articles in Ethiopia, the following diacritical symbols above basic characters may be introduced:  $\acute{u}$  for gemination  $\Psi\Omega\Phi$ ,  $\bar{u}$  for long vowels, and  $\bar{u}$  for their combination, e.g.  $\acute{u}\bar{u}\bar{u}$  and  $\bar{u}\bar{u}\bar{u}$ .

Tones may be denoted by the following diacritical symbols above characters:  $\acute{u}$  for high and  $\grave{u}$  for low tone. These may be combined with long and geminated consonants, e.g.  $\acute{u}$ ,  $\grave{u}$ ,  $\acute{u}$  (high-low),  $\grave{u}$  (low-high),  $\acute{u}$ ,  $\grave{u}$ ,  $\acute{u}$ ,  $\grave{u}$ ,  $\acute{u}$ , and  $\grave{u}$ .

### 3 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 Ethiopian writing system with Ethiopian character sets. Besides some applications like inserting a date and writing mathematical notations and formulas are included.

#### 3.1 Typing the text

In Table 4 you can see the Ethiopian 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  $\eth$ , or a pure consonant, especially in combination with other consonants. (Example:  $\Psi\Omega\Upsilon\Gamma$  pronounced as  $t\eth mh\eth r$ ). 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  $\eth$  which is pronounced  $da$  and entered as  $\eth a$ . You will find the character in the character table in row  $\eth$  and column  $a$ .

The letter  $\eth$  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  $\eth e$  or  $\eth$  at the users choice.

If accented characters are used in the scientific transcription of an Ethiopian syllable, they are entered without the accent, but prefixed with an appropriate special character.  $\eth$  has  $\check{a}$  as its transcription, hence we will enter it as  $\hat{c}a$  which is as close to the proper transcription as we can get.

Long vowels are usually indicated by a bar,  $d\bar{e}$  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  $\eth E$  as our transcription of the aforementioned syllable  $\eth$ .

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  $\eth i$  or  $\eth I$  and get  $\eth$  in both cases.

Note that the consonant series `_C` is a variant form of the series `^C` and that the letters `^Ca` and `_CE` are identical. We offer both forms at the users' choice.

Some consonants may be followed by a diphtong, i.e. a combination of two vowels. A diphtong occurs e.g. in ቋ 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 `dWA` which we want to result in ቋፆ. The SERA encoding, which was developed by Daniel Jacob, Yitena Firdiywok, Yonas Fessiha 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 *Orominya (Oromiffa)*, all with `'q`, `'k`, `'h`, `'g` and `mua`, `mui`, `muE`, `mue` only for *Guraginya (Chaha)*, all with `.q` only for *Tigrinya*, and all with `fu`, `pu` only for *Wolayttinya*.

### 3.2 Spaces

When the Ethiopian script 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 L<sup>A</sup>T<sub>E</sub>X 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:
'abAs: 'alamenahe: 'abAs: 'alamenahe::
```

which results in

```
አባስ፡ አለምነህ፡ አባስ፡ አለምነህ፡ አባስ፡ አለምነህ፡ አባስ፡ አለምነህ፡ አባስ፡ አለምነህ፡ አባስ፡
አለምነህ፡ አባስ፡ አለምነህ፡ አባስ፡ አለምነህ፡
```

or

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

which results in

```
አባስ አለምነህ አባስ አለምነህ አባስ አለምነህ አባስ አለምነህ አባስ አለምነህ
አባስ አለምነህ አባስ አለምነህ አባስ አለምነህ
```

in the output. As you can see, line breaks are allowed after the `:` even if there is no space character in the source. Space characters immediately following or preceding a `:` in the input are ignored as we can see from the first line. Therefore newlines in the input will not cause any problem either.

### 3.3 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 Ethiopian script 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 break points is impossible. In this case, some explicit `\hspace` must be added in a suitable position or a `raggedright` layout must be selected during the whole document or part thereof.

### 3.4 Ethiopian dates

The calendar system is implemented in our package, so it is possible to type `\today` and get `ጥ.ዓ.ዓ. 27 1989`. (This is the date on which this document has been translated. The Ethiopian calendar is based on the Julian calendar with twelve months of 30 days and one month of 5 days. Every fourth year is a leap year, which means that the last month will have 6 days. Compare the date on the title page!)

A long calculation is required to convert the Gregorian date provided by the `TeX` primitives `\year`, `\month`, and `\day` to the Ethiopian date. 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.

### 3.5 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 explicitly select 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 might be used for other languages, too, e.g. to provide a unified input mechanism for all Latin characters.

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 L<sup>A</sup>T<sub>E</sub>X to get a more readable version.

### 3.6 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, although 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 T<sub>E</sub>X dimensions while Ethiopian text is being typeset. We can circumvent this problem by using a `,` instead of the `.` when entering numbers for T<sub>E</sub>X. Note, that we can use the `.` without problems when we have temporarily switched to a language other than Ethiopian, e.g. English.

We could expect that the activation of `^` and `_` spoils T<sub>E</sub>X's math mode, but this is not the case. In fact math mode behaves just like before, with subscripts and superscripts in their proper position.

The complex calendrical calculation require the allocation of a large number of counters to hold the intermediate results.

In connection with other counter intensive packages L<sup>A</sup>T<sub>E</sub>X may run out of counters when using the `ethiop` package. Maybe we can get rid of two or three of the counters in the next version, but it will still remain a problem.

This problem is actually due to the lack of temporary counters in L<sup>A</sup>T<sub>E</sub>X, which are not provided, even though this is done for all the other types of registers.

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 Ethiopian, 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, Ethiopian is always selected as the main language 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.

## 4 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 Ethiopian script and text processing system.

### 4.1 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 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 in the Ethiopian script.
- Conventions for the encoding of other vowels, or short and long vowels must be looked for.

### 4.2 Thanks

We would like to thank Abass B. Alameneh, Johannes L. Braams, Donald E. Knuth, and Leslie Lamport for their efforts. The programs provided by them made our package possible.

## References

- [1] Alamneh, Abass: EthTEX.  
`ftp://ftp.dante.de/tex-archive/languages/ethiopia/ethtex/`
- [2] Abate, Yohannis; Wubneh, Mulatu: Ethiopia. Boulder, Colorado, 1989.
- [3] Bender, M.L.; Bowen, J.D.; Cooper, R.L.; Ferguson, C.A.: Language in Ethiopia. Oxford University Press, London 1997.
- [4] Bender, M. Lionel (ed.): The Non-Semitic Languages of Ethiopia. Michigan State University, East Lansing, 1976.

- [5] 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/>
- [6] Braams, Johannes L.: The babel package.  
<ftp://ftp.dante.de/tex-archive/macros/latex/packages/babel/>
- [7] EthiO Systems: EthT<sub>E</sub>X.  
<http://www.neosoft.com/~ethiosys/ethtex/ethtex.html>
- [8] Summer Institute of Linguistics: Ethnologue. Dallas, Texas, 1996.  
<http://www.sil.org/ethnologue/>
- [9] Guidi, Ignazio: Vocabolario Amaharico-Italiano. Casa Editrice Italiana, Roma, 1901.
- [10] Hammerschmidt, Ernst: Äthiopische Handschriften vom Ṭānāsee 1. Franz Steiner Verlag, Wiesbaden, 1973.
- [11] Knuth, Donald E.: The T<sub>E</sub>Xbook. Addison Wesley, Reading, Massachusetts, 1986.
- [12] Knuth, Donald E.: The METAFONTbook. Addison Wesley, Reading, Massachusetts, 1986.
- [13] Lamport, Leslie: L<sup>A</sup>T<sub>E</sub>X – A Document Preparation System. Addison Wesley, Reading, Massachusetts, 1986.
- [14] Leslau, Wolf: Ethiopians Speak, Studies in Cultural Background, Vol. 2. University of California Press, Berkeley 1965.
- [15] Leslau, Wolf: Comparative Dictionary of Ge'ez (Classical Ethiopic). Otto Harrassowitz, Wiesbaden, 1987.
- [16] Littmann, Enno; Höfner, Maria: Wörterbuch der Tigrē-Sprache. Franz Steiner Verlag, Wiesbaden, 1962.
- [17] The Unicode Technical Report #1 – Draft Proposal on Ethiopian Script. Houston, TX, August 1993.
- [18] አክሊል ብርሀን ወልደ ኪርኮስ፥ ስለ ግዕዝና አማርኛ ቋንቋ ታሪክ። ንግድ ማትሚያ ብት፣ አዲስ አበባ 1958።
- [19] ተክለ ወልድ ደስታ፥ የአማርኛ መዝገበ ቃላት። አርቲስቲክ ማትሚያ ቤት፣ አዲስ አበባ 1970።
- [20] Wright, Stephen: The Transliteration of Amharic. Journal of Ethiopian Studies, Vol. II No. 1, Addis Ababa, 1964.
- [21] Yaqob, 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 L<sup>A</sup>T<sub>E</sub>X 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^g^c nabaruA^caw::
...
```

and gives the following result.

ፓዱዋ በምትባል ውብ የኢጣሊያ ከተማ ሲኞር ባፕቲስታ የተሰኙ እንደ ትጉጎ ሠራተኛ ይኖሩ ነበር። እኒህ ሰው ሁለት ቆንጆ ሴቶች ልጆች ነበሯቸው። ኳዲሶሊና የተባሉት ውበቷና ጠባኛ የተሟላ ነው። ፔጶሊ ውበቷ እንደ እጎቷ እሟልቶ ስላልሰጣት በዐመለኛነቷ ትታወቃለች።

ፔጶሊ ላባቷ የበኩር ልጅ በመሆኗ በዚህ አለብላቢት ምላሷ ሳቢያ ምናልባት ባል እጥቃ ቁም-ቀር እንዳትሆን አባቷ ስጋት እድርባቸው ነበር።

የፈሩት አልቀረም ለቆንጆዋ ከቬርኖ እስከ ቊዳጎላ ሲጎርፍ ለዐመልኛዋ አጠገቧ የሚድርስ ሰው ጠፋ።

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 L<sup>A</sup>T<sub>E</sub>X 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 አዲስ አበባ አዲስ አበባ as the output.

Besides we can also apply different type the 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 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 : የጦጣጣን ብልጠት፣ የላምጣን የዋህነት፣ የአንበሳጣን ጉልበት የሚለካ ሰው አልተገኘም።

Tigrinya : ዋገ እንታ በሎ፣ገገገገገገ ኮበርታ። or ዋገ እንታ በሎ፣ገገገገገገገ ኮበርታ።

Oromiffa : ቦፋ፣ ቁቤውውን ደቻ በከከ ሰዲገትቲ ቆፋይይ ከከገገ ኒ ደንደገኘ ዱገላ።

or ቦፋ፣ ቁቤውውን ደቻ በከከ ሰዲገትቲ ቆፋይይ ከከገገ ኒ ደንደገኘ ዱገላ።

We can also apply math mode:

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

$$\frac{n + x^2}{\pi} \quad (b + b)^2 - (n + n)^2 \quad n + \frac{n}{n^n}$$

In Table 9 we show a financial budget as an example of the combination of L<sup>A</sup>T<sub>E</sub>X and ethiop.

ርእስ	የወር ወጪ አይነት	የወር ወጪ (ብር)
ምግብ	ጤፍ	1 ኩንታል
	ቅባት	10
	መጠጥ	10020 ሊትር
	ቅጠላቅጠል	20 ኪሎ
	ጥራጥሬ	15 ኪሎ
ትራንስፖርት	ነጻጅ	40 ሊትር
	የአውቶቡስ ቲኪት	በቀን 2 ብር
መዝናኛ	ለስፖርት ክለብ	በወር 50 ብር
	ሽርሽር	በወር 50 ብር
ጠቅላላ ድምር		795

## B Installation

Installing the ethiop package is not overly difficult. Just make sure that

1. Check the prerequisites for this package. Make sure:
  - that you have installed T<sub>E</sub>X and METAFONT.
  - that the files `cmbase.mf` and `romand.mf` from the Computer Modern fonts are accessible to METAFONT.
  - that you have installed L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> with a release date of 1996/12/01 or later. It can be found on CTAN in the directory `macros/latex/`.
  - that you have installed the `babel` package with a release date of 1997/01/23 (version 3.6h) or later. It can be found on CTAN in the directory `macros/latex/packages/babel/`.
  - to get the ethiop package, which can be found on CTAN in the directory `languages/ethiopia/ethiop/`. Now the files mentioned in MANIFEST should be present.

2. Generate the T<sub>E</sub>X 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 L<sup>A</sup>T<sub>E</sub>X 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.