Ethiopian Language Support for the **Babel** Package Version 0.6

Berhanu Beyene Manfred Kudlek
Olaf Kummer Jochen Metzinger

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

18th March 1998

Abstract

The Ethiopian script differs considerably from the Latin script. Most important, it consists of more than 350 different letters. A new transcription method is presented that can be used for the LaTeX typesetting system. It was implemented on the basis of the multilingual typesetting package babel.

In addition to a guide to the concrete usage of our system, we provide technical details of the implementation and sketch the reasons for our design decisions. Linguistic and historical information on the Ethiopian script is also included.

አኅጽሮተ ጽሁፍ

የኢትዮጵያ ጽሁፍ ከላቲን ጽሁፍ እጅግ የተለየ ነው። ጎልቶ የሚታየው ከ 350 በላይ ፊደላትን የያዘ መሆኑ ነው። እነዚህን ፊደላት LaTeX በተባለ የጽህፈት አጣጣል ማቀነባበሪያ ዘዴ የኢትዮጵያን ጽሁፍ በላቲን ቋንቋ ለመጻፍ የሚያስችል አዲስ ያአጻጻፍ ዘየ (Transcription) ተዘጋጅቶ ቀርቧል።

የኛን የአጻጻፍ ዘዴ ለመጠቀም ከሚያስችል መመሪያ ጋር፤ የአጠቃቀሙን ቴክኒካል ዝርዝርና ለምን ይሄን የአጻጻፍ መንገድ ለመጠቅም እንደወሰን አሳይትናል። ስለ ኢትዮጵያ ቋንቋዎች አጻጻፍ ዘዴና የቋንቋ ታሪካዊ መረጃዎች በዚህ ጥናት ላይ አክለንበታል።

1 Introduction

The ethiop package [3] is a collection of fonts and TeX macros that enable you to typeset the characters of the languages of Ethiopia.

1.1 The origins of the Ethiopian script

The Ethiopian script has its origin in the South Semitic alphabet which has been used for Thamudene, Dedanite, Liḫyanite, Ṣafaitic, Minaean, Sabaean, Ḥimyaritic, Qatabanic, and Ḥaḍramautic. These are Semitic languages once spoken in the present state of Yemen. The alphabet was without vocalization. The South Semitic characters are known from stone monuments only, whereas the present Ethiopian characters originate from paper documents. The writing direction of South Semitic was from right to left or alternating (Bustrofedon).

In the 2^{nd} half of the 1^{st} milliennium BC Semitic groups (Ge'ez, ግዕዝ) from South Arabia established a kingdom at Aksum. After 350 AD a vocalization was introduced.

Today the Ethiopian script is an official writing system in two states, Ethiopia (ኢትዮጵያ) and Eritrea (ኤርትራ). There exist 86 languages (4 of them extinct) from 4 language families with around 200 dialects in Ethiopia and Eritrea. The writing system is used for a number of languages in these states, see Tab. 1 for the most important languages (figures taken from [6]). 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 Tab. 1 are from [6].

1.2 The origins of this package

The current Ethiopian LaTeX environment started as a project assignment at the university of Hamburg in 1995. Luckily, we did not need to start from scratch, because there were some Ethiopian fonts already.

The fonts are based on EthTeX which was originally distributed by Abass B. Alameneh. The genuine EthTeX package can be found on CTAN in the directory `language/ethiopia/ethtex/` or in [5]. We changed the existing letters a little to make them more robust at low resolutions, but the most effort went into 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 transcription.

When we chose our transcription scheme, it became apparent 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 [4] as framework for the implementation of the input transcription. The TeX macros are derived from the file `language.skeleton` provided with that package, which also allows a more well-rounded support of multiple languages.

Since the original EthTeX package used a special editor that is not available for all platforms, it was not easily portable. Moreover, a special version of LaTeX 2.09 was generated, thereby preventing an upgrade to the now current

Language family	Languages		No. of speakers in 1000 (year 91–95)
Semitic	Ge'ez	ግዕዝ	only in church
	Amharic	አማርኛ	20 000
	Gurage	ጉራግኛ	1 850
	Tigre	ትግሬ	?
	Tigrinya	ትግርኛ	6 050
Cushitic	Afar	አፋርኛ	750
	Hadiyya	ሀዲይኛ	1 000
	Kambaata	ከምባትኛ	1 000
	Oromo	ኦሮምኛ	14 000
	Sidamo	ሲዳምኛ	1 500
	Somali	ሶማልኛ	2 050
Omotic	Gamo/Gofa/Dawro	ጋምኛ	780
	Wolaytta	ወላይትኛ	2 000
East Sudanic			?

Table 1: Languages and number of speakers

LaTeX 2_ε. By rewriting the language support completely and by embedding the Ethiopian fonts into the framework of the babel package we got a more robust and portable system that will be usable with future LaTeX versions.

Our transcription method does not provide support for the direct entry of Ethiopian characters. Instead a natural encoding has been developed that allows us to enter Ethiopian text via Latin letters. This encoding is based on scientific transcription techniques and is closely related to other encoding standards. One of these standards is SERA, which is mainly intended for the recording and transmission of Ethiopian text within an ACSII environment. However, the SERA encoding was not realizable as an input encoding for TeX. Also important is Unicode, which provides a unified framework for *all* languages by encoding characters with 16 bits instead of 8.

2 Installation

Installing the ethiop package is not overly difficult, especially if you adhere to the following instructions.

1. Check the prerequisites for this package.
    - Make sure you have installed TeX and METAFONT.
    - Make sure that the files `cmbase.mf` and `romand.mf` from the Computer Modern fonts are accessible to METAFONT.
    - Make sure that you have installed LaTeX 2_ε with a release date of 1996/12/01 or later. LaTeX 2_ε can be found on CTAN in the directory `macros/latex/`.

- Make sure that you have installed the babel package with a release date of 1997/01/23 (version 3.6h) or later. Babel can be found on CTAN in the directory `macros/latex/packages/babel/`.

2. If a previous version of this package is installed, remove all the files, especially the font files that were generated by METAFONT after the first installation.

3. Depending on how you obtained this package it might be necessary to unpack/uncompress an archive. Now the files mentioned in `MANIFEST` should be present.

4. Generate the TEX files from their `docstrip` source. To do this, run

   ```
   tex ethiop.ins
   ```

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

5. The files `ethiop.ins` and `ethiop.dtx` can be removed, or you can run

   ```
   latex ethiop.dtx
   ```

   to generate the source code documentation. *This is not required.*

6. Move the files to their destination.

The files	are	and are needed by
`*.tfm`	font metric files	TEX
`*.fd`	font definitions	TEX
`*.sty`	style files	TEX
`*.ldf`	babel language definitions	TEX
`*.tex`	TEX sources	TEX
`*.mf`	METAFONT sources	METAFONT

   The exact location where the files belong depends on your installation. As a first approximation, install them near other files with the same extension.

   For some installations it might be possible or even necessary to place all the files in the directory where the user documents will be placed. *This is not recommended.*

7. Remove the font files that might be left over from a previous version of ethiop, that is all files `eth*.gf` and `eth*.pk`.

8. 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` should be identical to this text.

It is highly probable that after doing all of the above, you have a working Ethiopian language package set up. But maybe you ran into trouble during the installation. In this case try the following:

- If the run on `ethiop.ins` fails, the files might have been corrupted during transmission or one of the necessary files might not be accessible to TEX.
- If LATEX complains about a missing input file, check whether the files `*.fd` and `*.sty` are accessible and readable.
- If LATEX complains about a missing font, check the placement of the files `*.tfm`.
- If LATEX issues warning messages, you might have an old version of LATEX or babel.
- If LATEX issues strange errors, you might have an old version of LATEX or babel.
- If LATEX issues strange errors, the files might have been corrupted during transmission. Conversions of CR, LF, and so on might cause this problem.
- If the previewer or the printer driver complains about missing fonts and does not automatically call METAFONT to generate these font, either adapt your installation or generate the fonts by hand. Depending on your installation you must run something like

  ```
  mf '\mode=localfont; input etha10'
  ```

  for each of the fonts.
- If METAFONT is called and complains about missing source files, check whether these (`*.mf`) are placed correctly.
- If METAFONT is called and complains about missing source files, check whether the Computer Modern fonts have been properly installed.
- If METAFONT is called and complains about strange paths, you are probably generating the font at a lower resolution than 200 dpi. There is little you can do except ignoring the errors or telling METAFONT to do so. Please report such errors only if they occur at resolutions above 200 dpi.
- If the previewer or the printer driver complains about missing characters, check whether you have deleted all files from previous versions of this font.
- If the previewer or the printer driver complains about a checksum error, check whether you have deleted all files from previous versions of this font.

If that does not help, have a look at our web page located at `http://www.informatik.uni-hamburg.de/TGI/mitarbeiter/wimis/kummer/ethiop_eng.html` to read about possible updates and bug reports. If this does not provide any clue and your friendly TEXnician is unavailable, we will try to help you, if time permits. Send a mail to `ethiop@informatik.uni-hamburg.de`, hopefully we will get back to you.

Please do **not** report bugs to Johannes L. Braams (the maintainer of babel) or to the LATEX team. They are **not** responsible for our bugs and they are **very** busy already.

3 Usage

An important fact about the Ethiopian script is that it uses more than 350 different characters. Hence the text entry itself is a big problem. But we must also deal with other topics like punctuation and spacing that arise when using this package.

3.1 Including the package

The babel language support is selected by adding the line

```
\usepackage[english]{babel}
```

to the preamble of a document. Of course a different language might also be selected. But since the ethiop package is not yet a part of the babel package, you must select it with

```
\usepackage{ethiop}
```

in the preamble of a document. There is no need to include babel explicitly unless we want to use two or more languages. But if babel is explicitly included, is *must* be included *before* the `\usepackage{ethiop}`. Putting the reference to babel after the inclusion of ethiop will not work. Now we can use the ordinary language switching mechanism of babel to take us from one language to the other. For example

```
\selectlanguage{english}'adis 'ababA
\selectlanguage{ethiop}'adis 'ababA
\selectlanguage{english}'adis 'ababA.
```

will give you: 'adis 'ababA አዲስ አበባ 'adis 'ababA.

3.2 Typing the text

In Tab. 2 you can see the Ethiopian characters that are accessible using this package.

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*.

We will now explain how the characters are entered. Every character represents a syllable that consists of a consonant followed by a vowel. If possible, every character is encoded in a way that matches its pronounciation or its scientific transcription as closely as possible.

As an example we choose the letter ደ 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 ድ 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 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. ቸ has *ča* as its transcription, hence we will enter it as `^ca` which is as close to the proper transcription as we can get.

		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				O		uI		*if preferred*	
h	h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ			ሇ		
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	ᎀ	ᎁ	ሟ	ᎂ	ᎃ
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
q̣	.q	ቐ	ቑ	ቒ	ቓ	ቔ	ቕ	ቖ	ቘ	ቚ	ቛ	ቜ	ቝ
b	b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ	ᎄ	ᎅ	ቧ	ᎆ	ᎇ
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		
ẖ	_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		
’	'	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					
‘	`	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	ዯ				
d	d	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ			ዷ		
ḍ	.d	ዸ	ዹ	ዺ	ዻ	ዼ	ዽ	ዾ			ዿ		
ğ	^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	ⶓ	ⶔ	ጟ	ⶕ	ⶖ
ṭ	.t	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ			ጧ		
č̣	^C	ጨ	ጩ	ጪ	ጫ	ጬ	ጭ	ጮ			ጯ		
p̣	.p	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ			ጷ		
ṣ	.s	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ			ጿ		
ç	.c	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ					
f	f	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ	ᎈ	ᎉ	ፏ	ᎊ	ᎋ
p	p	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ	ᎌ	ᎍ	ፗ	ᎎ	ᎏ
q́	'q	ⷀ	ⷁ	ⷂ	ⷃ	ⷄ	ⷅ	ⷆ					
ḱ	'k	ⷈ	ⷉ	ⷊ	ⷋ	ⷌ	ⷍ	ⷎ					
h́	'h	ⷐ	ⷑ	ⷒ	ⷓ	ⷔ	ⷕ	ⷖ					
ǵ	'g	ⷘ	ⷙ	ⷚ	ⷛ	ⷜ	ⷝ	ⷞ					

~mA	ፙ
~ri	ፘ
~fi	ፚ

Table 2: The Ethiopian characters

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

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 diphtong, i.e. a combination of two vowels. A diphtong occurs e.g. in ዷ which is best transcribed *dwā*. We will code it as `duA`, thereby slightly deviating from the standard transcription. But this is unavoidable, because if we enter `dwA`, we will get an ambiguity with `d wA` which we want to result in ድዋ. The SERA encoding, which is developed by Daniel Yacob, Yitna Firdyiwek, and Yonas Fisseha, suggests `dWA`, which has been considered as an alternative to the present encoding. It did not show any significant advantages, however.

For the consonant series ጨ ጩ ጪ ጫ ጬ ጭ ጮ ጯ there exists a variant form that looks like ጨ ጩ ጪ ጫ ጬ ጭ ጮ ጯ and denotes the same syllables. The variant series can be activated by issuing the command `\ethvariantCtrue` once. The original letterforms can then be restored by typing `\ethvariantCfalse`. This option will only be needed in comparative studies, usually the standard series should be preferred.

In the lower right corner of Tab. 2 you can see an inlay with the three characters ፙ, ፘ, and ፚ. These characters are probably the remnants of three complete series of seven syllables that were once used.

We added these three in order to completely cover the character set of the proposed Unicode standard for code positions 1200h to 137Fh. The encoding of these characters is not fixed yet and may change at any time in the future. At the moment the encoding is based on the characters' appearance, but we are aware that the pronounciation is different today.

3.3 Punctuation

Although the punctuation characters look different from the punctuation of the Latin script, they have essentially the same meaning. We made some compromises between visual similarity and similar interpretation when we chose the encoding of the punctuation characters. In Tab. 3–5 we have collected the appropriate inputs for each of the characters.

The punctuation characters match their SERA equivalents closely.

3.4 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 LaTeX documents.

But in former times word breaks used to be denoted by the character ፡ and even today this method is used for handwriting. To get the appropriate effect look at the following LaTeX source

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

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

Table 3: The Ethiopian punctuation characters

input	<	<<	>	>>	'	''	`	``
output	‹	«	›	»	’	”	‘	“

Table 4: The Ethiopian quotation characters

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

Table 5: Special characters

```
'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 preceeding 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.

3.5 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 justification. Nevertheless, the lack of hyphenation in the Ethiopian language 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` enviroment 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 changable, e.g. because it is a quote from some other source, or the author insists on that very phrase. In that case you an 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.

3.6 Numbers

Since today Arabic numbers are more frequently used than the original Ethiopian numbers, the `ethiop` package outputs the Arabic numbers when the letters `0` up to `9` occur in the source code.

But Ethiopian numbers 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 999 999 to their Ethiopian equivalents: `\ethnum{999999}` gives ፺፼፱፼፺፻፱፻፺፱. We can see that the program knows quite a lot about the representation of large numbers.

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	፼

Table 6: The Ethiopian numbers

In LaTeX we must output the contents of a counter from time to time. This can be accomplished using the macro `\ethiop`. Saying `\ethiop{subsection}` will cause TeX to print ፮, since this is subsection 6.

3.7 Math mode

One of TeX's most important features is its math mode. We can use math within Ethiopian text, but by default all letters in mathematical formulas will be taken from the ordinary TeX fonts.

Ethiopian letters can be used in a formula, although this will require slightly more work. When the macro `\ethmath` appears in math mode while the Ethiopian language is selected, its single argument will appear in the proper size typeset with the Ethiopian fonts.

For example

```
$$b+\frac{d}{f^g}\iff
  \ethmath{ba}+\frac{\ethmath{da}}{\ethmath{fa}^{\ethmath{ga}}}$$
```

will result in

$$b+\frac{d}{f^g}\iff \text{በ}+\frac{\text{ደ}}{\text{ፈ}^{\text{ገ}}}$$

where you should note the varying fonts in the first and second subformula. Actually all the work is done by the macro `\ethmath` that outputs the Ethiopian characters.

3.8 Ethiopian dates

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. The calendar system is implemented in our package, so it is possible to type `\today` and get መጋቢት 9 1990. (This is the date on which this document has been translated. Compare it to the date on the title page!)

3.9 Two examples

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

```
\subsubsection*{`amala~nAytun mArq}

pAduwA bametbAl web ya'i.tAliyA katamA si~nor bAptisetA
yatasa~nu 'and tegu_h `sarAta~nA yenofu nabar:: 'enih sawe
hulat qon^go sEto^c le^g^c nabaruA^caw::
```

and gives the following result:

ዐመለኛይቱን ማርቅ

ፓዱዋ በምትባል ውብ የኢጣሊያ ከተማ ሲኞር ባፕቲስታ የተሰኙ አንድ ትጉኅ ሠራተኛ ይኖሩ ነበር። እኒህ ሰው ሁለት ቆንጆ ሴቶች ልጆች ነበሯቸው። ኳዷሒሊና የተባለችው ውበቷና ጠባኞ የተሟላ ነው። ፔፊሷ ውበቷ እንደ እኅቷ አሟልቶ ስላልሰጣት በዐመለኛነቷ ትታወቃለች።

ፔፊሷ ላባቷ የበኩር ልጅ በመሆኗ በዚህ አለብላቢት ምላሷ ሳቢያ ምናልባት ባል አጥታ ቁሞ-ቀር እንዳትሆን አባቷ ስጋት አድሮባቸው ነበር።

የፈሩት አልቀረም ለቆንጆዋ ከቬርና እስከ ቍዳኅላ ሲጎርፍ ለዐመልኛዋ አጠገቧ የሚደርስ 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 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.

3.10 Using ethiop with ArabTeX

Starting with version 3.6i, babel is now compatible with ArabTeX. But still there are some problems with the many active characters which are used by ethiop, so that the two packages do not work right away.

But it is possible to use ArabTeX with ethiop by including the special style `etharab.sty` after ArabTeX has been loaded. Some internal macros of ArabTeX are redefined, so this is not guaranteed to work with every version of ArabTeX, but it has been successfully used with version 3.06g3 of ArabTeX.

```
\documentclass{article}
\usepackage{arabtex}
\usepackage[english]{babel}
\usepackage{ethiop}
\usepackage{etharab}
\begin{document}
  \selectlanguage{english}
```

```
  The Arabic (<al.har_t>) and the Ethiopian script
  (\selectlanguage{ethiop}sel.tAnE\selectlanguage{english})
  may occur within one sentence.
\end{document}
```

We do not provide the output of the example, because this document is intended to be translatable even in the absence of ArabTEX. In fact, everything works as usual, Arabic text can be inserted using the `arabtext` environment or using `<>` pairs. However, these commands must not be used in arguments to other commands or in command definitions. If that is desired the complete commands or command definitions must be enclosed in a `noethiop` environment.

```
\begin{noethiop}
\section{<al.har_t> -- cultivation}
\end{noethiop}
```

provides an example. If Ethiopian characters are needed, too, then a little trick is in order.

```
\def\temptext{%
  \selectlanguage{ethiop}sel.tAnE\selectlanguage{english}}
\begin{noethiop}
\section{<al.har_t> -- cultivation -- \temptext}
\end{noethiop}
```

But there should be really, really few occasions for such ugly code.

4 Advanced topics

Although the usage of the ethiop package is not really difficult once one gets used to it, there are a few points to be aware of. We will highlight the internal structure of the package first, to make it more plausible why some problems just cannot be easily patched away.

4.1 Implementation notes

Because we have to deal with so many characters, we placed them in two separate fonts. The two codetables are shown in Tab. 7 and Tab. 8, together with the recommended input strings. (Remember that there might be different ways to achieve the same result.)

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

	0	1	2	3	4	5	6	7
0	mua ᎀ	muI ᎁ	muA ሟ	muE ᎂ	mue ᎃ	qua ቈ	quI ቊ	quA ቋ
8	quE ቌ	que ቍ	bua ᎄ	buI ᎅ	buA ቧ	buE ᎆ	bue ᎇ	kua ኰ
16	kuI ኲ	kuA ኳ	kuE ኴ	kue ኵ	gua ጐ	guI ጒ	guA ጓ	guE ጔ
24	gue ጕ	fua ᎈ	fuI ᎉ	fuA ፏ	fuE ᎊ	fue ᎋ		
32		! !	'' ”	'! ¡	$ $	'? ¿		' ’
40	((	))	:= ÷		, ፣	- -	. .	/ /
48	0 0	1 1	2 2	3 3	4 4	5 5	6 6	7 7
56	8 8	9 9	: ፡	; ፤	< ‹		> ›	? ?
64		'A ዓ				'E ዔ		
72		'I ዒ						'O ዖ
80						'U ዑ		
88				[[	`` “	]]		
96	` ‘	'a ዐ	be ብ		de ድ	'e ዕ	fe ፍ	ge ግ
104	he ህ			ke ክ	le ል	me ም	ne ን	
112	pe ፕ	qe ቅ	re ር	se ስ	te ት		ve ቭ	we ው
120		ye ይ	ze ዝ		| ፥		:: ።	
128	ha ሀ	hU ሁ	hI ሂ	hA ሃ	hE ሄ	hO ሆ	la ለ	lU ሉ
136	lI ሊ	lA ላ	lE ሌ	lO ሎ	ma መ	mU ሙ	mI ሚ	mA ማ
144	mE ሜ	mO ሞ	ra ረ	rU ሩ	rI ሪ	rA ራ	rE ሬ	rO ሮ
152	sa ሰ	sU ሱ	sI ሲ	sA ሳ	sE ሴ	sO ሶ	qa ቀ	qU ቁ
160	qI ቂ	qA ቃ	qE ቄ	qO ቆ	ba በ	bU ቡ	bI ቢ	bA ባ
168	bE ቤ	bO ቦ	va ቨ	vU ቩ	vI ቪ	vA ቫ	vE ቬ	vO ቮ
176	ta ተ	tU ቱ	tI ቲ	tA ታ	tE ቴ	tO ቶ	na ነ	nU ኑ
184	nI ኒ	nA ና	nE ኔ	nO ኖ	ka ከ	kU ኩ	kI ኪ	kA ካ
192	kE ኬ	kO ኮ	wa ወ	wU ዉ	wI ዊ	wA ዋ	wE ዌ	wO ዎ
200	za ዘ	zU ዙ	zI ዚ	zA ዛ	zE ዜ	zO ዞ	ya የ	yU ዩ
208	yI ዪ	yA ያ	yE ዬ	yO ዮ	da ደ	dU ዱ	dI ዲ	dA ዳ
216	dE ዴ	dO ዶ	ga ገ	gU ጉ	gI ጊ	gA ጋ	gE ጌ	gO ጎ
224	fa ፈ	fU ፉ	fI ፊ	fA ፋ	fE ፌ	fO ፎ	pa ፐ	pU ፑ
232	pI ፒ	pA ፓ	pE ፔ	pO ፖ	pua ᎌ	puI ᎍ	puA ፗ	puE ᎎ
240	pue ᎏ						huA ኋ	vuA ቯ
248	tuA ቷ	nuA ኗ	zuA ዟ	yua ዯ	duA ዷ	luA ሏ	ruA ሯ	suA ሷ

Table 7: The Ethiopian codetable ETHA

	0	1	2	3	4	5	6	7
0	.qua ቈ		.quI ቊ	.quA ቋ	.quE ቌ	.que ቍ		
8	_hua ኈ		_huI ኊ	_huA ኋ	_huE ኌ	_hue ኍ		
16	_kua ኰ		_kuI ኲ	_kuA ኳ	_kuE ኴ	_kue ኵ		
24	.gua ጐ		.guI ጒ	.guA ጓ	.guE ጔ	.gue ጕ		
32								
40	~mA ፙ	~ri ፘ	~fi ፚ					
48		፩	፪	፫	፬	፭	፮	፯
56	፰	፱	፲	፳	፴	፵	፶	፷
64	፸	፹	፺	፻	፼			
72	.ha ሐ	.hU ሑ	.hI ሒ	.hA ሓ	.hE ሔ	.he ሕ	.hO ሖ	.huA ሗ
80	'sa ሠ	'sU ሡ	'sI ሢ	'sA ሣ	'sE ሤ	'se ሥ	'sO ሦ	'suA ሧ
88	^sa ሸ	^sU ሹ	^sI ሺ	^sA ሻ	^sE ሼ	^se ሽ	^sO ሾ	^suA ሿ
96	.qa ቀ	.qU ቁ	.qI ቂ	.qA ቃ	.qE ቄ	.qe ቅ	.qO ቆ	
104	^ca ቸ	^cU ቹ	^cI ቺ	^cA ቻ	^cE ቼ	^ce ች	^cO ቾ	^cuA ቿ
112	_ha ኀ	_hU ኁ	_hI ኂ	_hA ኃ	_hE ኄ	_he ኅ	_hO ኆ	
120	~na ኘ	~nU ኙ	~nI ኚ	~nA ኛ	~nE ኜ	~ne ኝ	~nO ኞ	~nuA ኟ
128	'a አ	'U ኡ	'I ኢ	'A ኣ	'E ኤ	'e እ	'O ኦ	'ua ኧ
136	_ka ኸ	_kU ኹ	_kI ኺ	_kA ኻ	_kE ኼ	_ke ኽ	_kO ኾ	
144	^za ዠ	^zU ዡ	^zI ዢ	^zA ዣ	^zE ዤ	^ze ዥ	^zO ዦ	^zuA ዧ
152	.da ዸ	.dU ዹ	.dI ዺ	.dA ዻ	.dE ዼ	.de ዽ	.dO ዾ	.duA ዿ
160	^ga ጀ	^gU ጁ	^gI ጂ	^gA ጃ	^gE ጄ	^ge ጅ	^gO ጆ	^guA ጇ
168	.ga ጘ	.gU ጙ	.gI ጚ	.gA ጛ	.gE ጜ	.ge ጝ	.gO ጞ	
176	.ta ጠ	.tU ጡ	.tI ጢ	.tA ጣ	.tE ጤ	.te ጥ	.tO ጦ	.tuA ጧ
184	^Ca ጨ	^CU ጩ	^CI ጪ	^CA ጫ	^CE ጬ	^Ce ጭ	^CO ጮ	^CuA ጯ
192	ጨ	ጩ	ጪ	ጫ	ጬ	ጭ	ጮ	ጯ
200	.pa ጰ	.pU ጱ	.pI ጲ	.pA ጳ	.pE ጴ	.pe ጵ	.pO ጶ	.puA ጷ
208	.sa ጸ	.sU ጹ	.sI ጺ	.sA ጻ	.sE ጼ	.se ጽ	.sO ጾ	.suA ጿ
216	.ca ፀ	.cU ፁ	.cI ፂ	.cA ፃ	.cE ፄ	.ce ፅ	.cO ፆ	
224	'qa ⷀ	'qu ⷁ	'qi ⷂ	'qA ⷃ	'qE ⷄ	'qe ⷅ	'qo ⷆ	
232	'ka ⷈ	'ku ⷉ	'ki ⷊ	'kA ⷋ	'kE ⷌ	'ke ⷍ	'ko ⷎ	
240	'ha ⷐ	'hu ⷑ	'hi ⷒ	'hA ⷓ	'hE ⷔ	'he ⷕ	'ho ⷖ	
248	'ga ⷘ	'gu ⷙ	'gi ⷚ	'gA ⷛ	'gE ⷜ	'ge ⷝ	'go ⷞ	

Table 8: The Ethiopian codetable ETHB

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 LaTeX to get a more readable version.

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.

4.2 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, 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 Ethiopian 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 Ethiopian, e.g. English.

We could expect that the activation of `^` and `_` spoils TeX'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 calendarical calculation require the allocation of a large number of counters to hold the intermediate results.

- In connection with other counter intensive packages LaTeX 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 LaTeX, 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 this package *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 Copyright and Liability Notice

6 Changes

Version 0.2 was the first version to be publicly released.

6.1 Version 0.3

- The letter ፚ was added.
- Some bugs regarding subscripts and superscripts in ordinary math mode were fixed.
- The command `\ethmath` was added to allow Ethiopian characters in math mode.
- Fonts in 5 point size were added to allow Ethiopian characters in subscripts and superscripts.

6.2 Version 0.4

- The punctuation characters ፡ and ። where added.
- The syllables ፘ, ፙ, and ፚ were introduced. Now all characters of the proposed Ethiopic Unicode standard for code positions 1200h to 137Fh are included.
- `\ethvariantCtrue` and `\ethvariantCfalse` were introduced. At the same time the shorthand `_C` was removed. The variant forms of the series `^C` can only be reached via the aforementioned commands from now on.

6.3 Version 0.5

- Problems with activating `'` in math mode have been solved.

6.4 Version 0.6

- `etharab.sty` has been added to allow cooperation with ArabTEX.
- As suggested by Paul Seelig from the Debian team, the license and copyright terms were made explicit.

7 Closing 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.

7.1 To Do

What remains to be done?

- Bugs must be removed. 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 must be added. To do this, we must get to know the names for 'bibliography', 'chapter', 'index' and so on for as many of the languages of Ethiopia as possible. Suggestions for these and other non-technical improvements can also be sent to `ethiop@informatik.uni-hamburg.de`.
- It must be possible to typeset whole books in the Ethiopian script. Although this is possible at the moment, there are several problems that occur with the different LATEX structures like `\part`, `\section`, or even `\item`.
- We need to find out more about the conventions for typesetting in the Ethiopian script.

7.2 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.

Daniel Yacob provided encouraging and constructive feedback on our project.

While we are on the subject, here is a list containing some of the names that appeared in this text. You can already guess it, they are written in the Ethiopian script. But which name is which?

ብርሃኑ በየነ	ዮሃንስ ብራምስ	ማንፍሬድ ኩድሌክ
አባስ አለምነህ	ኦላፍ ኩመር	ለስሊ ለምፖርት
ዮኸን ምጽንገር	ዶናልድ ክኑት	ደንኤል ያዕቆብ

References

[1] Alamneh, Abass: EthTEX.
`ftp://ftp.dante.de/tex-archive/languages/ethiopia/ethtex/`

[2] Bender, M.L.; Bowen, J.D.; Cooper, R.L.; Ferguson, C.A.: Language in Ethiopia. Oxford University Press, London 1997.

[3] 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/`

[4] Braams, Johannes L.: The babel package.
`ftp://ftp.dante.de/tex-archive/macros/latex/packages/babel/`

[5] EthiO Systems: EthTEX.
`http://www.neosoft.com/~ethiosys/ethtex/ethtex.html`

[6] Summer Institute of Linguistics: Ethnologue. Dallas, Texas, 1996.
`http://www.sil.org/ethnologue/`

[7] Guidi, Ignazio: Vocabolario Amaharico-Italiano. Casa Editrice Italiana, Roma, 1901.

[8] Hammerschmidt, Ernst: Äthiopische Handschriften vom Ṭānāsee 1. Franz Steiner Verlag, Wiesbaden, 1973.

[9] Knuth, Donald E.: The TEXbook. Addison Wesley, Reading, Massachusetts, 1986.

[10] Knuth, Donald E.: The METAFONTbook. Addison Wesley, Reading, Massachusetts, 1986.

[11] Lamport, Leslie: LATEX – A Document Preparation System. Addison Wesley, Reading, Massachusetts, 1986.

[12] Leslau, Wolf: Ethiopians Speak, Studies in Cultural Background, Vol. 2. University of California Press, Berkeley 1965.

[13] Leslau, Wolf: Comparative Dictionary of Ge'ez (Classical Ethiopic). Otto Harrassowitz, Wiesbaden, 1987.

[14] Littmann, Enno; Höfner, Maria: Wörterbuch der Tigrē-Sprache. Franz Steiner Verlag, Wiesbaden, 1962.

[15] The Unicode Technical Report #1 – Draft Proposal on Ethiopian Script. Houston, TX, August 1993.

[16] አክሊለ ብርሀን ወልደ ኪርኮስ፥ ስለ ግዕዝና አማርኛ ቋንቋ ታሪክ። ንግድ ማትሚያ ብት፣ አዲስ አበባ 1958።

[17] ተክለ ወልድ ደስታ፥ የአማርኛ መዝገበ ቃላት። አርቲስቲክ ማተሚያ ቤት፣ አዲስ አበባ 1970።

[18] Yaqob, Daniel: SERA FAQ. `http://www.cs.indiana.edu/hyplan/dmulholl/fidel/sera.html`