

Cahiers **GUT** *enberg*

☞ T_EX IN AFRICA
☞ Jörg KNAPPEN

Cahiers GUTenberg, n° 10-11 (1991), p. 15-24.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1991__10-11_15_0>

© Association GUTenberg, 1991, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.

TEX and Africa

Jörg KNAPPEN

*Institut für Kernphysik, Universität Mainz, Postfach 39 80, D-W 6500 Mainz
e-mail knappen@vkpmzd.kph.uni-mainz.de*

Abstract. At the present time, TEX is not usable for typesetting many african languages. They use special letters which don't occur in the standard fonts (and aren't included in the ec-scheme). The letters used in the major languages of africa can put into *one* font. A font encoding scheme (fc) and some METAFONT code is prepared. There is work in progress on hausa TEX (done by Gos Ekhaguere, Ibadan).

Résumé. *Nous n'étions pas en mesure, jusqu'à l'heure actuelle, d'utiliser TEX pour la composition de textes rédigés dans plusieurs langues africaines. Ces langues utilisent des caractères spéciaux, qui ne sont pas disponibles dans les fontes standards. Les symboles utilisés dans les langues africaines importantes sont regroupés dans une seule fonte, dessinée avec METAFONT, pour laquelle nous avons préparé un schéma d'encodage. Gos Ekhaguere (d'Ibadan, Nigeria) poursuit actuellement ce travail et utilise hausa TEX.*

Key words: african languages, special letters, font encoding, METAFONT.

1. Introduction

Whenever the latin alphabet was adopted to another language some adjustments were made. This was done by shifting the meaning of seldom used letters or by introducing new letters.

The main way of creating new letters is the addition of a diacritical mark, for example an accent, a cedilla or an extra stroke like in polish ł. The other way is creating new shapes from the existing ones and borrowing characters from other alphabets. Famous examples are the icelandic characters ð and þ, the latter borrowed from runes.

Although TEX can handle accented letters by the use of control sequences, it is advantageous to have them ready in a font. The ec font encoding scheme, proposed at the TEX conference of Cork 1990, satisfies the need of european TEX users well.

For african languages the situation is quite different. Many of them use special letters for phonemes unknown to european languages. These letters are sometimes borrowed from the international phonetic alphabet (IPA), sometimes derived by modification of standard latin letters. Accents are used to mark the different tones. (Many african languages are tonal languages like chinese.)

Unfortunately, the extensions of the latin alphabet are not made in a consistent way. For each language exist an own solution or – even worse – more than one solution. The process of standardisation has just begun and is far away from reaching its end point.

In africa several hundred languages are spoken. It seems to be impossible to scan all these languages and their writing system. In this work I confined myself to the so-called critical languages as defined by the US departement of education in 1986. Each of these languages has at least 1 mio. speakers.

2. How to Extend T_EX to African Languages

The first thing to do is the creation of fonts which contain the zusatz letters used in african languages. It turns out that *one* 256-character font will do the job for all major african languages. It also supports french and portugese.

The second step is to create appropriate format files, including the new fonts and setting up the control sequences for the accents and special letters. Hyphenation pattern for the different languages are needed.

Then make national style files for T_EX/L^AT_EX. The national style should provide some prescription how to use standard ascii input. This can be done by making some character(s) active, like in `german.sty`. Alternatively, you could access the special letters as ligatures. This approach would lead to different fonts (different by their ligtable) for different languages. Therefore it was discarded. For L^AT_EX, all the names (like references) should be set into the national language.

If existing, special typographical conventions should be included.

The last step: After finishing the work, make it accessible to the public and look how it spreads. For this, it is helpful, if there are simple introductions adapted for the different languages (and written in these languages).

3. Work in Progress

I have prepared a font containing that characters which have entirely new shapes. I plundered nearly every source code available to do this and added the missing shapes. For some letters, variant glyphs are provided. At this moment, the shapes are not completely refined.

Italics are under development now. The letters “long f (f)” and “round v (v)” provide a new kind of problem. The usual italic shape denotes a different character in some african languages. The solution is to redesign the italic f and v (and as a consequence of this also the w and y).

The fcmr font is the First aFrican Computer Modern Roman font. It includes computer modern roman in the lower 128 code positions and 36 african zusatz characters in the upper half. The latter already have their fc code positions, but some of the lower half will be moved.

The next step will be an afc (Almost aFrican Computer) font. This will have all characters positioned according to the EC/FC scheme, but will not be complete. Some accented letters might be absent (and are probable so).

After completion, the name will change to fc, which will be the final form.

Gos Ekhaguere from the Departement of Mathematics, Ibadan, Nigeria, is working on hausa T_EX. This is the first project in Africa known to me.

4. Acknowledgements

I want to thank Prof. N. Cyffer from the Institut für Ethnologie und Afrikastudien, Mainz, for leading my interest to the critical languages and for help in finding good sources.

A. Characters Required for African Languages

Here is a list of african special letter including the languages in which they are used. The first part of the list contains entirely new shapes, the second part deals with letters carrying diacritical marks.

ḃ hooktop b

used in many languages including Hausa, Fulful, Kpelle. In Zulu and Xhosa now obsolete. Uppercase: Ḃ (Hausa, Fulful, Kpelle), B (Xhosa, Zulu)

ḋ d with tail

Ewe. Uppercase: Ḍ

ḏ hooktop d

Fulful and Hausa. Uppercase: Ḑ

ɛ open e

very common, e.g. Basa (Kru), Dinka, Ewe, Fulful, Gã, Kpelle, Mandekan (Bambara), Mende, Ngala (Lingala). Uppercase: E

ə inverted e

Kanuri. Uppercase: E̅

f long f

Ewe. The usual f is used, too. Uppercase: F̄

ɣ ipa gamma

Dinka, Ewe, Kpelle. Uppercase: Y

ħ crossed h

Maltese. Uppercase: H

k hooktop k

Hausa. Uppercase: K

ɲ enj

Songhai. Uppercase: N

ɳ eng

Basa(Kru), Bemba, Dinka, Efik, Fulful, Ewe, Gã, Kpelle, Mende. Uppercase: D, variant: □

ɔ open o

very common, e.g. Basa (Kru), Dinka, Ewe, Fulful, Gã, Kpelle, Mandekan (Bambara), Mende, Ngala (Lingala). Uppercase: O

ʃ esh

Gã. Uppercase: Ʒ. I've seen also Σ, but have no reference for this.

t crossed t

Sami. Although not african, I included this, because it is missing in the Cork scheme. Uppercase: T.

v round v

Ewe. Must be distinguished from u,v,y. Uppercase: U

w raised small w

Efik. Occurs only after certain consonants.

y hooktop y

Fulful. Must be distinguished from normal y. Uppercase: Y. Maybe Y is acceptable, too.

Accented letters are also in frequent use in african writing. In some languages, accents are used to distinguish phonemes, in other languages they are tonal marks, and even double use is possible. Since tonal marking is mostly optional, tonal marks should be created with floating accents. The most common tonal marks are acute, grave, and circumflex, efik uses also háček and the universal accent ('). The other letters where the accent is obligatory should be included in an african font.

The list of accented letters is

Acute ń, ñ: Efik

Grave è, ò: Mandekan, Sotho (South African Orthografie)

Circumflex ê, ô: Sotho (Basutoland Orthografie)

Diaresis ë, ö: Sotho (B)

Tilde ã, ê, ē, ī, õ, ò, û: Kpelle

 ĩ, ü: Kikuyu

 ñ: Efik, Oromo

Háček ě, õ: Sotho (SA)

 š: Oromo, Sotho (SA)

Breve ŵ: Gã

Dot above è, ò: Basa(Kru)

 ñ: Bamileke, Basa(Kru)

 ġ, ž: Maltese

Macron	ē, ō: Sotho (SA)
Dot below	ẹ, ɣ: Yoruba ọ: Igbo, Yoruba ị, ỳ: Igbo
Line below	ẹ, ọ: Basa(Kru), Kikuyu ṅ: Luganda (Kinya Rwanda)

B. The fc Character Encoding Scheme

The TUG conference of Cork 1990 has proposed a 256 character font encoding scheme well suited for european languages. This scheme does not fit for the various african languages with latin writing. So I want to propose a scheme suited for the so called critical languages of africa. It should be named FC or FCM for aFrican Computer.

The following rules are used

- The lower 128 codes are identical to the Cork scheme
- A glyph also occurring in the Cork scheme is placed on the same code point as in the Cork scheme
- Each letter from the 128-char cm-font is saved (Thus ị, L, Ø, etc. will work)
- The uppercase/ lowercase mechanism holds

The following languages are covered: Akan, Bamileke, Basa (Kru), Bemba, Ciokwe, Dinka, Efik, Ewe-Fon, Fulani (Fulful), Gã, Gbaya, Hausa, Igbo, Kanuri, Kikuyu, Kikongo, Kpelle, Krio, Luba, Mandekan (Bambara), Mende, More, Ngala, Nyanja, Oromo, Rundi, Kinya Rwanda, Sango, Shona, Somali, Songhai, Sotho (two different writing systems), Suaheli, Tiv, Yao, Yoruba, Xhosa and Zulu.

Also covered are: Maltese and Sami (European languages not covered by the Cork scheme).

I tried to consult the most recent dictionaries. A good part of the mentioned languages has not yet a standardised writing system.

I considered accents which are only tonal marks and optional in writing not to be a part of a letter. These should be created by using floating

accents. Even double accenting is possible, e.g. Open e with tilde and acute (Ě). Accented letters, where the accent is a part of the letter, are included.

C. The fc Font Encoding Table

glyph	octal code	description
Ɓ	200	Capital letter hooktop B
Ɗ	201	Capital letter hooktop D
Ɛ	202	Capital letter open E
Ǝ	203	Capital letter reversed E
Ƒ	204	Capital letter long F
Ě	205	Capital letter E with háček
Ƴ	206	Capital letter ipa Gamma
Ɔ	207	Capital letter double barred H
Ƒ	210	Capital letter hooktop K
Ƒ	211	Capital letter Enj
Ɔ	212	Capital letter open O (reversed C)
Ƒ	213	Capital letter N with acute
Ƒ	214	Capital letter Esh
Ƒ	215	Capital letter Eng
Ƒ	216	Capital letter round V
Ƒ	217	Capital letter hooktop Y
Ĝ	220	Capital letter G with dot above
Ĝ	221	Capital letter M with acute
Š	222	Capital letter S with háček
Ń	223	Capital letter N with dot above
Ń	224	Capital letter N with line below
Š	225	Capital letter S with dot below
Ŵ	226	Capital letter W with breve
Ƨ	227	Capital letter crossed T
Ě	230	Capital letter E with dot above
Ě	231	Capital letter E with dot below
İ	232	Capital letter I with tilde
Ō	233	Capital letter O with macron
Ƨ	234	ligature t-esh
Ƨ	235	ligature fj

glyph	octal code	description
đ	236	Small letter crossed d
˘	237	double grave accent
ḃ	240	Small letter hooktop b
ḋ	241	Small letter hooktop d
ε	242	Small letter open e
ə	243	Small letter inverted e
f	244	Small letter long f
ě	245	Small letter e with háček
ɣ	246	Small letter ipa gamma
ḥ	247	Small letter crossed h
ḵ	250	Small letter hooktop k
ɲ	251	Small letter enj
ɔ	252	Small letter open o
ń	253	Small letter n with acute
ʃ	254	Small letter esh
ŋ	255	Small letter eng
ʋ	256	Small letter round v
ɣ	257	Small letter hooktop y
ḡ	260	Small letter g with dot above
ḿ	261	Small letter m with acute
š	262	Small letter s with háček
ň	263	Small letter n with dot above
ṅ	264	Small letter n with line below
ṣ	265	Small letter s with dot below
ṽ	266	Small letter w with breve
t	267	Small letter crossed t
è	270	Small letter e with dot above
ẹ	271	Small letter e with dot below
ĩ	272	Small letter i tilde
ō	273	Small letter o with macron
w	274	small raised w
!	275	inverted exclamation mark
?	276	inverted question mark
˘	277	universal accent

glyph	octal code	description
À	300	Capital letter A with acute
Ì	301	Capital letter I with dot below
Â	302	Capital letter A with circumflex
Ã	303	Capital letter A with tilde
Ë	304	Capital letter Open E with tilde
Õ	305	Capital letter Open O with tilde
Æ	306	Capital ligature AE
Ç	307	Capital letter C with cedilla
È	310	Capital letter E with grave
É	311	Capital letter E with acute
Ê	312	Capital letter E with circumflex
Ë	313	Capital letter E with diaeresis
Ē	314	Capital letter E with line below
Ĕ	315	Capital letter E with macron
Ë	316	Capital letter E with tilde
Ï	317	Capital letter I with diaeresis
Ð	320	Capital letter crossed D (Edh)
Ñ	321	Capital letter N with tilde
Ò	322	Capital letter O with grave
Ó	323	Capital letter O with dot above
Ô	324	Capital letter O with circumflex
Õ	325	Capital letter O with tilde
Ö	326	Capital letter O with diaeresis
Œ	327	Capital ligature OE
Ø	330	Capital letter crossed O
Ɔ	331	Capital letter O with dot below
Ɔ	332	Capital letter O with line below
Û	333	Capital letter U with circumflex
Ů	334	Capital letter O with háček
Ū	335	Capital letter U with dot below
Û	336	Capital letter U with tilde
·	337	cross piece for polish L and l

glyph	octal code	description
á	340	Small letter a with acute
ì	341	Small letter i with dot below
â	342	Small letter a with circumflex
ã	343	Small letter a with tilde
ẽ	344	Small letter open e with tilde
õ	345	Small letter open o with tilde
æ	346	Small ligature ae
ç	347	Small letter c with cedilla
è	350	Small letter e with grave
é	351	Small letter e with acute
ê	352	Small letter e with circumflex
ë	353	Small letter e with diaeresis
e	354	Small letter e with line below
ē	355	Small letter e with macron
ē	356	Small letter e with tilde
ï	357	Small letter i with diaeresis
đ	360	Small letter d with tail (note: not edh!)
ñ	361	Small letter n with tilde
ò	362	Small letter o with grave
ó	363	Small letter o with dot above
ô	364	Small letter o with circumflex
õ	365	Small letter o with tilde
ö	366	Small letter o with diaeresis
œ	367	Small ligature oe
ø	370	Small letter crossed o
ø	371	Small letter o with dot below
o	372	Small letter o with line below
û	373	Small letter u with circumflex
ö	374	Small letter o with háček
u	375	Small letter u with dot below
ũ	376	Small letter u with tilde
ß	377	Small letter eszet