Typesetting African languages

AN INVESTIGATION BY CONRAD TAYLOR

Typesetting African Languages

by Conrad Taylor

This investigation, experiments and report were originally inspired and prompted by the author's interest in the publishing aspirations of a South London voluntary organisation, but this research was not 'commissioned' (or paid for) by them. I took the project on as a private one. I hope that this has worked out to everyone's advantage, as I became fascinated by the subject – and took my investigations further than anyone would reasonably have anticipated (or have been prepared to paid for).

Despite that, the effort is very incomplete. It was particularly difficult to track down clear reference sources for the character sets used for writing African languages. Linguists seem largely uninterested in this problem. The least ambiguous sources are language tutorials and dictionaries, but they are hard to find. As a result, many important languages (e.g. Buganda, Kongo, Mossi, Mandinka, Ndebele, Shona) remain undocumented here. However, the principles remain the same.

I do not deal at all with Arabic here (it is a particularly difficult typesetting problem, but lucrative enough to have attracted the attention of software developers; the solutions are well documented elsewhere). I also decided not to include the Austronesian languages of Madagascar, so the focus here is on continental Black Africa.

Today, Black Africa is in urgent need of better means for transmitting vital information about health issues, agricultural techniques and other means of improving life in the African countryside and the rapidly-growing cities. I hope that more linguists, scholars, writers, designers and software engineers will contribute their skills to ensure that more effective means can be developed to propagate this information effectively in the many languages of Africa.

It is my belief and hope that computers should be our salvation in this work, and not be a part of the problem, though the issue of 'intellectual property' in font technology does require some careful thinking – and creative, generous solutions – to ensure that Africans are not being charged more than they can pay for the right to communicate in print in their own languages.

Contents

African languages and writing systems

A large quilt of small patches

Africa is a continent of many indigenous languages: over 2,000 – more than are found in any other continent. The widely-accepted classification scheme of Joseph Greenberg divides these into four main language families – Afro-Asiatic, Nilo-Saharan, Niger-Congo and Khoisan – the approximate geographical distribution of which is shown on the map below.

Around 1,350 African languages are members of the Niger-Congo family, which predominates in sub-Saharan Africa. The Bantu sub-family of 400

Fig. 1: Language families

languages was first identified as having a common origin by Wilhelm Bleek in 1862, based on the Kongo/Luba word *bantu* for 'people', the equivalent word for which in other languages of the same family is quite similar (*banto*, *abantu*, *abandu*, *baat*, *bato*, *vanhu* etc.). Starting from a common homeland between the Niger and Zaïre river basins, the Bantu peoples have spread out to occupy East and Southern Africa.

In contrast, a total of only 300,000 people speak languages of the small and shrinking Khoisan family, which formerly would have been in widespread use in Southern Africa among the hunter-gatherer peoples. Some Khoisan languages have recently become extinct. However, the distinctive 'click' consonants of these languages have influenced their Bantu neighbours such as Xhosa and Zulu.

The languages of the Afro-Asiatic family predominate in North Africa. The most prominent of these, Arabic, was imported from the Arabian peninsula during the Muslim conquests of the seventh and eighth centuries, but there are several highly significant indigenous Afro-Asiatic languages, especially in the Horn of Africa and the Sahel.

The Nilo-Saharan family is probably the least 'tidy' classification, comprising 200 languages. In particular there has been dispute about how to classify Songhai, a language spoken around the Niger Bend.

(As for the Austronesian languages of Madagascar, they were brought by colonisation from South East Asia, and are not considered in this paper.)

Only about 5% of indigenous African languages have more than a million speakers, and only six are used by more than ten million people:

Language name	Population (approx)	Family	Where spoken
Swahili	30 million	Niger-Congo	Tanzania, Kenya, Uganda (lingua franca for 25m)
Hausa	25 million	Afro-Asiatic	North Nigeria, Niger
Yoruba	20 million	Niger-Congo	Nigeria, Benin
Amharic	14 million	Afro-Asiatic	Ethiopia (official language)
Igbo (Ibo)	13 million	Niger-Congo	Nigeria
Fula (Fulfulde)	13 million (various dialects)	Niger-Congo	Several West African countries
Oromo (Galla)	11 million	Afro-Asiatic	Ethiopia, Kenya

These six account for less than 20% of the entire population of Africa, and the percentage would be far lower if mother-tongue speakers only were to be counted. This is a striking contrast with South Asia (India, Pakistan, Bangladesh, Nepal and Sri Lanka), where 17 languages have more than ten

million mother-tongue speakers, and between them account for about 900 million people – some 70% of the combined populations of those countries.

In consequence of this linguistic fragmentation, and of long-distance trade and colonisation, Africa has been described as a continent of lingua francas, where Arabic and English, French and Portuguese have provided the basis for much communication. Several African languages also function as lingua francas, especially Swahili ('coastal language') which developed from Sabaki dialects in East Africa but was massively influenced by Arabic and other languages spoken by trading partners; it is now the official language of Tanzania and is the most widely spoken single language in Uganda and Kenya, usually as a speaker's second language.

European colonisation led to the evolution of several important Creole languages, such as Krio in Sierra Leone. Other indigenous languages became creolised to a degree, and were promoted as colonial powers required a common language for their locally-recruited armies and administrations. This was particularly so in the Belgian Congo, where Lingala became the language of the army.

European colonial presence has also, of course, determined the setting within which most indigenous African languages have acquired their writing systems.

'Written African'

The traditional histories and story-telling, poetry and liturgies of almost all African societies have been oral, not written down.[1] This may seem ironic, when we consider that five thousand years ago the Egyptians were among the first people to create a writing system, which was a mixture of pictogram and alphabet.

Ancient Egypt's writing system did have some influence on later writing systems, such as the modified hieroglyphic system used in the Kushite empire of Meroë, but then it died out, and its inscriptions remained a mystery until they were decoded in the early nineteenth century by Jean-François Champollion.

Pure alphabets were more successful. The first fully alphabetic script was devised around 1700 BC in northern Palestine and Syria, with 22 signs for consonants. This gave rise to a number of different alphabetic systems, for instance Hebrew and Arabic, Sabaean, and the script of the Phoenicians – which was also transferred to the North African Phoenician settlement of Carthage. The Phoenician alphabet was taken as a model by the Greeks, who added vowels; this alphabet in turn inspired Etruscan and Roman alphabets, and so led to the development of all the 'roman' alphabets in use today.

1 For example, a West African equivalent of the Homeric tradition is the Malian epic of the magician-king Sundiatta, retold for centuries by Malian griots (minstrels).

A small number of African alphabetic systems made their own separate development from these early beginnings:

- **Tifinagh** is an ancient alphabetic writing system still used today to write Tamashek, the language of the Tuareg Berbers. It consists of consonants only, usually written right to left, in rather square letters made up of straight lines and dots (see Fig. 2 below). It seems to be an ancient Libyan script derived from Carthaginian Phoenician writing and dates from about 300 BC; rock-carved examples have been found across North Africa and in the Canary Islands. Interestingly, Tifinagh is used for rather domestic purposes; within the Tuareg communities, the 'official' and written language is Arabic, in which most men but only some women are literate.

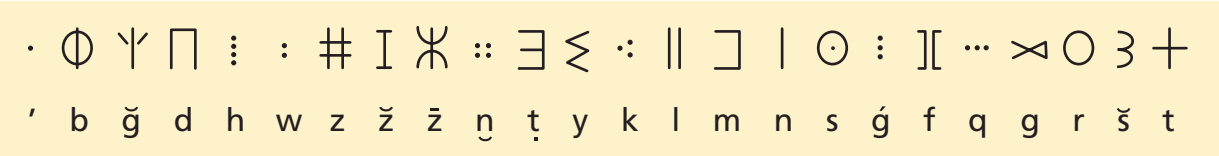


Fig. 2: Tifinagh script

- **Coptic** or **Old Nubian** script is a modified form of the Greek alphabet, which was used to write the Coptic language, descended from ancient Egyptian. Coptic became extinct as a living language around 1600 AD but continued in use in the liturgy of the Egyptian (Monophysite) Christian church. One language which continues to use the Coptic alphabet today is 'Nile Nubian' or Dongolawi, a Nilo-Saharan language spoken in Egypt and Sudan by about a million people, which has a written literature dating back to the 8th century. To support the Nubian language, four extra consonants were added to the script.

- **Ge'ez** or 'Ethiopic' script is the unique alphabet of the Horn of Africa, developed from the old Sabaean script from the south of the Arabian peninsula for Ge'ez, the old language of Ethiopia which survived in liturgical use. This script was also used by the Ethiopian Jews, the *falashas*, to write their scriptures. Today it is used to write three Semitic languages of Africa: Amharic, which was promoted as the national language of Ethiopia by Emperor Tewodros II in the 19th century; also Tigrinya, the major language of Eritrea, and the related language Tigré used in the north of Ethiopia. This script was originally a system of consonants only, but has turned into a syllabary by adding an extension to each consonant to indicate the following vowel sound.

Alphabets with a mission

Legend has it that the Greek theologian St. Cyril (827–869 AD), assisted by his brother and fellow-missionary St. Methodius, modified the Greek alphabet so that the Gospel could be brought to the heathen Slavs in their own language – which had sounds for which Greek didn't have letters. Thus

was devised the 'Cyrillic' alphabet which is used for Bulgarian, Russian and Serbian today. Similar missionary processes developed the Roman script so that it could be used to write Irish, Saxon and other tongues; over time, this adaptation of the roman alphabet also led to the addition of new letters such as *y* and *j* and *w*, the ligatured letterforms *ß* and *æ* and *œ,* and various accents to distinguish between a much wider range of vowel sounds than were found in Latin itself.

Essentially, that is also how most African languages have acquired their writing systems. Just as 5th-century monks adapted the alphabet to bring the Good News to the Angles and Saxons, latter-day missionaries devised further modifications to the latin script to print Bibles in Yoruba and Igbo, Gikuyu and Swahili. And this, broadly, is the origin of most of the Africa writing systems the typesetting of which is being considered in this paper. In fact, in studying this history I came time and time again across accounts of how standardisation of spelling systems was slowed down by rivalry between Catholic and Protestant promoters or alternative systems.

These details need not concern us, fascinating though they doubtless are. However, I would like to make three points to counteract the impression that the bringing of writing systems to Africa was entirely a missionary endeavour:

- In West Africa in the region of the Niger Bend and Lake Chad, societies were involved in sophisticated trading networks across the Sahara, centuries before Europeans anchored their ships off the Gold Coast. For these societies, literacy was first encountered in the form of written Arabic. Hausa is an example of a language which was written in Arabic letters from about the 16th century, but which latterly has converted to a latin script with some special consonants added; Swahili, used along the East African trade routes, was also written in Arabic script in the early 18th century.
- In the post-colonial period, some African governments established national commissions to reform and standardise the writing systems and promote their use. An example of such an enterprise is the Ghana Bureau of Languages.
- Some of these writing systems were standardised very recently indeed. For example, there was a great deal of controversy in Somalia about how the language should be written, and it was a stated objective of the 1969 revolution to settle the question. One favoured contender was the unique Osmanian alphabet, named after its inventor, Osman Yusuf. However, the military government of Siad Barre decreed in 1972 that a simple latin alphabet would be employed, without accents, and with long vowels signified simply by writing the vowel twice. This decree was followed up by an effective literacy campaign (civil servants were given a three-month deadline to learn how to spell!) and by these forceful means Somalia's modern writing system was established.

Fonts for typesetting African languages: the issues

Modern typesetting is done using standard personal computers, with software of various degrees of sophistication, plus type fonts which contain the repertoire of characters we need.

When called upon to typeset an 'unusual' language, the first issue that arises is: do we have all the letterforms that this language requires – and does the computer system[2] have the means to assemble them in a manner acceptable to the users of that language? From this standpoint, I believe it is useful to grade African languages into five grades of difficulty:

- **LEVEL 1** — these languages use only characters shared with the English language, and also use no accents in conjunction with letters. Thus they are extremely easy to typeset by computer.
- **LEVEL 2** — these languages do not have any specially constructed letterforms. They do use some accents over vowels, but in a way that is standard to common European languages such as French, Spanish or Portuguese. This means that they can be typeset using standard fonts and software – presenting only a slight learning difficulty, in that the operator has to learn how to access special characters such as ô or é.
- **LEVEL 3** — The next step up in difficulty is those languages which use 'ordinary' letterforms but in some non-standard combinations – such as a dot under a vowel, or an acute accent over a consonant. These languages cannot be set with standard applications and fonts. There are two possible approaches: one is to use special typesetting software based on 'graphic decomposition' which allows compound letterforms to be assembled from their constituent elements; the other is to use standard publishing software, but with specially created fonts in which the combinations exist in ready-assembled form.
- **LEVEL 4** — These are the languages which clearly require a number of special letterforms that do not exist in the standard fonts oriented towards Western European language typesetting, for example the 'hooked consonants' of Hausa. Here, a special font is definitely required, but no other modification of the system is needed.
- **LEVEL 5** — The most problematic languages have a non-latin character set which is so large in its required repertoire that a single standard font cannot contain them all – or perhaps they have unusual behaviours, such as requiring different forms of letter depending on where

2 For now, I use the general term 'computer system' so as to treat the system as a whole, without yet distinguishing the separate contributions made by the operating system software, publishing application software, etc.

they occur in a word. This level of problem requires more than just a special font: some other modifications will be needed, such as special software or operating system extensions. As we are not considering Arabic typesetting in this paper, the only script system which poses this level of difficulty for us is the Ethiopic script system of Amharic, Tigré and Tigrinya, for which a satisfactory solution is available if desired.

This five-level classification scheme is a useful way to assess how difficult it would be from a technical point of view to start publishing in a particular language. Thus, according to my investigations so far, I find that *Swahili* and *Somali* are at Level One, *Tswana* is at Level Two, *Igbo* and *Yoruba* and *Nyanja* are at Level Three, *Twi* and *Krio* and *Hausa* are at Level Four and *Amharic* is at Level Five.

What is a font? And what's in it?

With some special exceptions, a font in a modern computer system is a software resource, installed in a special relationship with the computer's operating system[3] so that once in place, it allows the letterforms stored in the font to be used in a wide variety of programs on the computer such as a word-processor, DTP program or illustration program.

Internally, a modern computer font consists of a range of letterforms, each of which is described mathematically as one or more closed paths made up of straight lines and geometric curves; and each letterform is located within a rectangular framework which determines the space around it. This can be seen in the screen-shot image below.

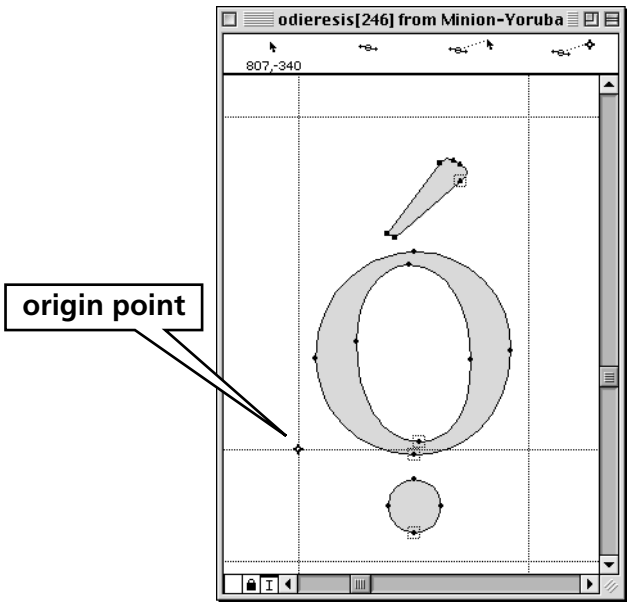


Fig. 3: A compound character

This is a screen capture of the editing window of a font editing program, Fontographer, which is often used to create new typefaces or modify existing ones. In this case, a new combined letterform is being created for use in typesetting Yoruba, and is being stored in the existing slot for the o-dieresis character (ö) – which Yoruba does not need.

Observe how the shape of the letter is defined by mathematical curves that run between digitization points on the letter's contour. Also note how the letter sits in relation to its origin point and the 'bounding box' which surrounds it.

(Fontographer does not show the letters shaded in grey; this shading has been added afterwards in the interests of clarity of presentation.)

3 The operating system is the most basic layer of software which a computer requires to operate, and which provided central services to all other software. Examples of operating systems: MS–DOS, Windows, Unix, Mac OS.

Additional font data

In addition to the character data, a font will also contain tables of values for various purposes. **Hinting** data provides guidance about how best to convert outline font data to pixels for the best possible display on a screen or printer, and **kerning** data provides fine adjustment to inter-letter space for letter pairs which do not fit well together naturally.

Fonts are usually provided in matched sets known as 'families'. In a well-constructed font family, each font is encoded with details of its 'family membership' so that if the operator issues a simple request to switch from the normal font to **bold** or *italic*, the correct alternate font is substituted.

The two principal formats in which type fonts can be purchased for either a Windows or Macintosh computer are **PostScript Type One** and **TrueType**. A brief explanation of the difference is given in Appendix A; it is not an important distinction in terms of language support.

Standard-repertoire character sets

Within a computer system, each type character is assigned a numeric code to identify it. In most computer systems, a single byte's worth of data is used to store this code, and because a byte is a binary number with eight binary 'places', this gives a maximum of 256 characters to which a unique code can in theory be assigned.

Ninety-three characters are numerically encoded identically on all systems, a standard which was established in **ASCII** – the American Standard Code for Information Interchange. This standard makes communication of textual data possible between different programs running on the same computer, and also between different computers, as in email applications. This very limited characters set is satisfactory for basic English communication, and is illustrated below:

Fig. 4: The ASCII standard character set

This characters set is common to all computer systems – with some old, rare exceptions. Note that as an American-defined standard it includes no accented characters, nor the pound sterling sign.

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890 ?!.,;:-'"*
@$%&+=^~_/|\  <([{}])>
```

In early computing systems, the range of characters that a computer could process was limited, because of the eight binary digits or 'bits' in each byte of data, one was reserved for use in checking the integrity of communicated data. However, the development of more sophisticated error-checking schemes which did not rely on reserving a 'parity bit' in each byte means

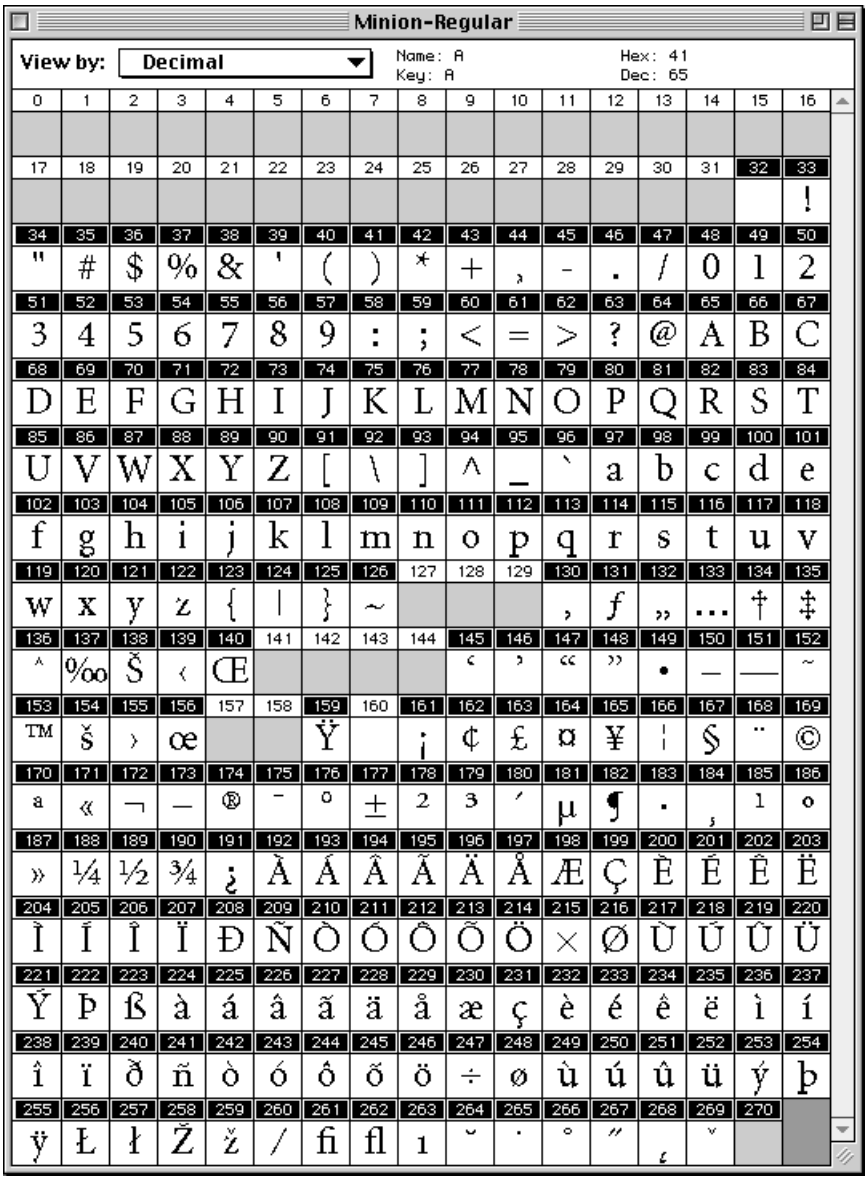


Fig. 5:
Windows font encoding

To compile this reference, an Adobe font for Macintosh (Minion) was converted to Windows encoding within a font editing program. A comparison with the original Mac encoding on the next page is most revealing.

Note that there appear to be more 'slots' in the Windows font than the 256 which we would expect a byte's worth of data to provide for, but in practice there are only as many characters as a byte can reference. All 32 of the initial ASCII slots (0–31) are reserved for their original purpose such as control codes. (All the unoccupied slots are shaded grey here.)

that a modern computer character set can contain about 240+ characters, which is the case in the fonts used for word processing or desktop publishing on Windows or Macintosh computers. However, this extended character space was implemented differently on different operating systems, as is illustrated in Figures 5 and 6.

Fig. 5 at the top of this page shows the standard character encoding scheme used by Adobe Systems for the fonts it supplies for use on the Windows operating system. Note that the first 31 slots are reserved for control codes. 32 is the standard word-space, and the range from 33 to 126 constitutes the standard ASCII character set. Character 127 is the 'delete' control code, also reserved by ASCII. A range of extended punctuation marks, symbols, accented vowels and other special characters required by some European languages are deployed in most of the remaining upper slots.

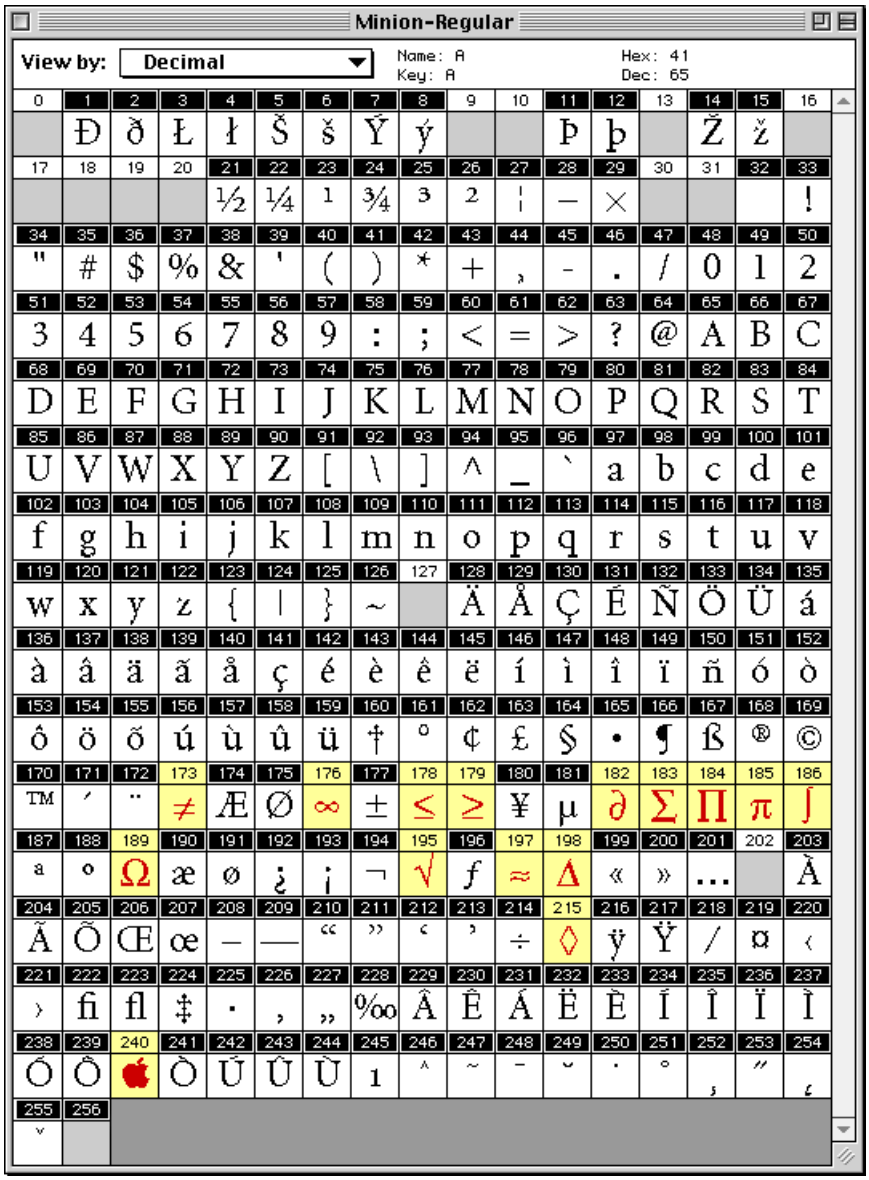


Fig. 6: Mac-encoded font

In a standard Macintosh font encoding, some built-fraction characters, and letters required for Icelandic and East European languages, are moved into the slots reserved under Windows for control characters. Without special operating system extensions, these characters (plus those in the range 245–255) are rendered inaccessible to the user.

Characters in the range 128–244 are quite easy to access on a Macintosh, due to easy-to-remember key combinations (such as option–e + e for **é** or option–a for **å**).

The characters marked in colour, required mostly for mathematical expressions, are not actually part of each Mac font, but are borrowed from the standard Symbol font instead.

Contrast this arrangment with Fig. 6 above, which shows the equivalent extended encoding for the Apple Macintosh operating system, widely used in the graphic arts. The standard ASCII characters all occupy the same slots as in the Windows-encoded font – as of course they must. However, the 'extended' characters are deployed using different encodings.

In practice, this can and does lead to file translation errors when files are passed from a Windows computer to a Macintosh computer or vice-versa – for example, Windows text with typographically appropriate single quotes 'like this' would look ìlike thisî when transferred to a Macintosh, and the common Windows bullet character [•] transforms to a sigma [Σ] on Mac.[4]

4 However, there are file translation utilities which correct for this encoding mis-match, and some desktop publishing programs likewise re-encode the characters while importing a text file to preserve the original intended appearance.

African typesetting using standard fonts

If the reader examines the character tables in Figures 4 and 5 above in conjunction with the tables of character use by some African languages in Appendix B, it becomes clear that many African languages pose no special typesetting problems because all of the characters required are provided for in standard fonts. To refer the reader back to the five-step classification introduced on page 6…

- **LEVEL ONE** languages which have no accents or special letters (so can be typeset as easily as English) include Oromo, Swahili, Somali and Zulu.
- **LEVEL TWO** languages do require the use of some accented vowels, but when the operator has figured out how to access these from within the standard fonts there will be no problem to typeset them.

In addition, two of the important *lingua francas* of Africa, French and Portuguese, are 'Level Two' languages for the purpose of this discussion. French is widely used in e.g. Algeria, Mali, Niger, Chad, Senegal, Cameroon, Guinea (*Konakry*), Côte d'Ivoire, Togo, Central African Republic, Gabon, Congo (*Brazzaville*), Rwanda, Burundi and the Democratic Republic of Congo (*formerly Zaïre*). Portuguese is widely used in Angola, Mozambique, Guinea-Bissau and the Cape Verde islands.

Level 1 & 2 languages and the Internet

Because all of the characters required to display Level One and Level Two languages are in standard computer fonts, there is no difficulty using these languages in email messages or on Web pages. The Level Two languages do however pose something a minor problem, because of the variation between different 'standards' for how these extended character sets are numerically encoded. These difficulties have been resolved for the Web, and to a less uniform degree for email users:

- **HTML encoding:** to make sure that an accented character displays as intended in all Web browsers whether on Windows, Unix, Macintosh or other systems, it is re-encoded as a special 'character entity' within the text of a Web page. For example, **Lomé** would be encoded behind the scenes as `Lomé` … The `é` fragment is displayed on a Windows Web browser as character 233 and on a Mac as character 142 but as **é** in both cases.
- **Email:** there are two re-encoding methods used to transfer these characters in the body of standard email messages. The older system is called *Quoted-Printable* and uses an equals sign as an escape character followed by a two-digit hexadecimal code. Some more recent email programs use HTML encoding.

Using TEX to typeset African languages

An inexpensive shareware-based typesetting system popular in academic circles can handle a broader range of African languages than standard DTP systems. But as experiments have shown, it is not that easy to use…

Introducing TEX

In 1977, Professor Donald Knuth of Stanford University began to investigate the use of standard computers for typesetting complicated publications. In particular, as a mathematician he was concerned about difficulties in typesetting mathematical books, journals and papers, where equations are a big problem. He devised a typesetting system called *tau epsilon chi*, which are the three letters at the root of the Greek word *techne* (for art, or craft), from which we get the word 'technology'. This is often typeset as TEX and pronounced 'tek'.

Knuth placed TEX into the public domain, together with the METAFONT system which he devised to make the computer typefaces which are used by TEX typesetting systems. Hundreds of programmers, usually based at universities, have likewise contributed their efforts to developing the TEX typesetting system; and through this collaboration, TEX has been converted to run on a wide range of computers – from multi-user mainframes to personal microcomputers. The software and fonts can be dowloaded for free, or for modest shareware fees, from a network of Internet servers devoted to the project (the CTAN archives).

Glueing accents to characters

One of the problems we have already discussed in typesetting African languages is that diacritical marks are often required to be combined with letters in ways that are not usual in European languages. This is a problem for standard word-processing programs and DTP programs, because they simply place each character to the right of the preceding one, and so they need to have access to 'ready-composed' common combinations of letters and diacritical marks. This means that the cedilla of **ç** cannot be placed under an **s** or the acute accent of **é** be placed over an **m**.

TEX is different because it builds up compound accented characters from a base character, plus floating accents. This also means that the fonts specially designed for use with TEX are very differently organised from those shown in figures 5 and 6 above. This can be understood better by examining the character encoding for Computer Modern, a font designed by Knuth himself, a PostScript equivalent of which is shown in Fig. 7 overleaf.

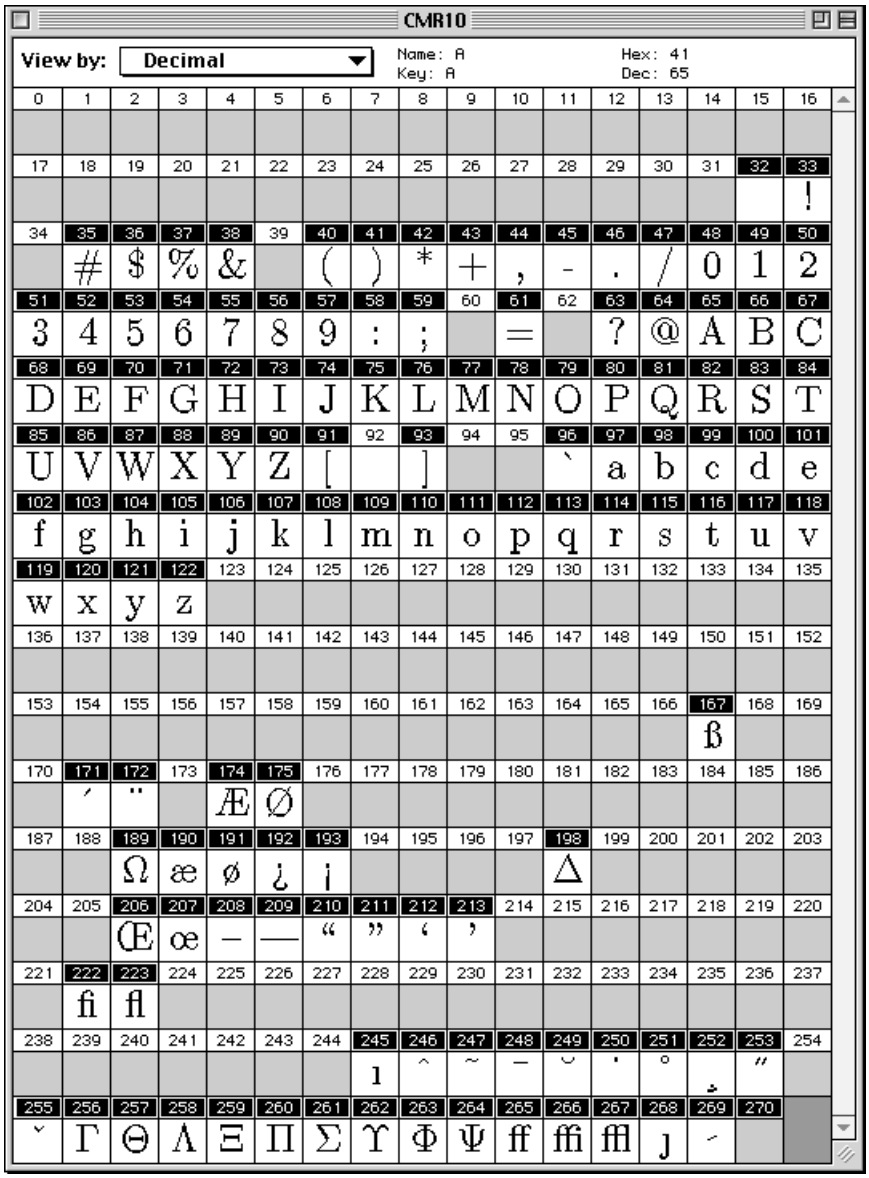


Fig. 6: Computer Modern

Donald Knuth's font for use with the TEX typesetting system is encoded very differently from the fonts for use with standard text composition programs. The array of pre-composed European accented characters found in standard fonts is simply missing here.

Instead, TEX relies on picking up floating accents from slots 96, 171 and 172, 246–253, 255 and 259 and composing them together with a base character. The full stop may also be repositioned as an underdot character.

This approach allows a much wider range of accented characters to be set with TEX than with standard systems. Note the provision of dotless i and j (at 245 and 268) to facilitate this form of character composition.

TEX in practice

The biggest shock about TEX for someone who has only recently been introduced to text processing on a computer is that TEX is not interactive, and the view you get while preparing a document is definitely not 'What You See Is What You Get' (WYSIWYG). It is a **code-driven system**, and the pages are composed in a **batch process**.

(This in part explains the success of TEX within the academic community: free from the need to support an interactive editing view of the document, programmers can deliver parsimonious implementations of TEX which use very little memory and processor power, and which can be used successfully even from a basic terminal on a time-sharing multi-user computer.)

To explain how TEX works, we shall take the example of typesetting a short section from modern Yoruba literature, which was done as an experiment

Fig. 7:
A Yoruba typset sample
from A. Ìṣọlá – "Ó le kú"
(Ibadan, OUP, 1974)

> Ọ̀rọ̀ tí Àjàní sọ wọ Àṣàkẹ́ létí. Ó ní òun rí i pé òdodo ni àlàyé tí Àjàní ṣe. Inú Àjàní dùn. Bí ó tilẹ́ jẹ́ pé ohun tí ó lè gbe Àjàní l'ó ńsọ fún Àṣàkẹ́, síbẹ̀ ọ̀nà tí ó gbá gbé ọ̀rọ̀ náà kalẹ̀ wọ ni létí púpọ́. Bí Àṣàkẹ́ bá le mu ìmọ̀ràn yì lò, àti máa lọ s'ọ́dọ́ Àjàní kò níí ṣòro mọ̀. Kẹ̀kẹ̀ bẹ́ẹ̀ imù ẹlẹ́dẹ̀ á wọgbà. Àṣàkẹ́ ní oun á bẹ̀rẹ̀ sí í máa ṣ'àlàyé ọ̀rọ̀ fun bàbá òun, ṣùgbọ́n pé díẹ̀díẹ̀ ni òun yíó máa ṣe é o. Ijọ́ tí a bá gùn kọ́ ni a ńkan ọ̀run. Nwọ́n fi ipade sí kejì ní yunifásítì ni yàrà Àjàní.

using a Macintosh implementation of TEX (CMacTEX 3.2). The final typeset version of the file is illustrated in Fig. 7 above.

(I selected Yoruba for this experiment because I determined that, if it were deemed acceptable to use the simple underdot character which is sometimes used for the letters **ọ**, **ẹ** and **ṣ**,[5] rather than the vertical stroke which is the alternative, then the standard Computer Modern TEX font would have all of the components required to compose the text. I also chose it because Yoruba is a significant, popular language – and a significant challenge.)

Preparing the typesetting file

A TEX typesetting file is an ordinary computer text file, using the standard ASCII characters illustrated in Fig. 3 on page 8 above. The typesetting file can be prepared with any simple text-editing program, on any computer, and is afterwards processed through the TEX software. As a simple text file, the typesetting file can also be transferred easily to another kind of computer – for instance by email – where it can be used to create identical output.

The TEX file contains a mixture of plain text content, and TEX 'command words' which are preceded by a forward-slash character (\). This idea of a mixture of text content and formatting codes will be familiar to anyone who has worked with HTML coding to make Web pages.

For this exercise, I used the BBEdit 4.5 text editing program for Macintosh. This is popular with Mac programmers and Web page creators, and is also a good choice for Mac TEX-ers, because there are extensions available for BBEdit which help by 'syntax colouring' command words so that they are easier to distinguish during the editing process.[6]

5 You may wonder how I managed to insert those characters in this text. The answer is that the FrameMaker software I have used to prepare this report has means to move characters around from their original typeset position. However, it is a slow an difficult process and does not offer a solution for typesetting African languages using standard fonts.

6 On a Windows system, WordPad or TextPad might be used for a similar purpose, though Word can also be used provided that the file is saved as plain ASCII text and given a **.tex** file extension.

Unfortunately, the complexity of the accents in Yorùbá meant that the TEX typesetting file became quite difficult to read – as can be seen below. (Standard TEX codewords are coloured blue and comments green, author-defined codewords are red, and line numbers have been added which were not in the original file.)

Defining personal macro codewords…

```
%%%%%%%%%%%%%%% macros for Yoruba font characters  %%%%%%%%%%%%%%

\def\Ed%%  upper-case E with dot below
    {E\kern-.4em\lower.45ex\hbox{.}\kern.2em}

\def\ed%%  lower-case e with dot below
    {e\kern-.35em\lower.45ex\hbox{.}\kern.1em}

\def\Od%%  upper-case O with dot below
    {O\kern-.54em\lower.45ex\hbox{.}\kern.24em}

\def\od%%  lower-case o with dot below
    {o\kern-.38em\lower.45ex\hbox{.}\kern.13em}

\def\sd%%  lower-case s with dot below
    {s\kern-.35em\lower.45ex\hbox{.}\kern.1em}

\def\Aj%%  character Ajani in story
    {\`Aj\`an\'\i{ }}

\def\As%%character Asake in story
    {\`A\sd \`ak\'\ed{ }}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

Setting margins…

```
\raggedbottom
\baselineskip=16pt
\leftskip=100pt
\rightskip=30pt
\parindent=0pt
\hsize=4.5in
\vsize=7in
\voffset=.75in
\magnification = \magstep 1

```

Entering the text, and codes for special characters

```
\`\Od r\`\od{} t\'\i{} \Aj s\od{} w\od{} \As
l\'et\'\i. \'O n\'\i{} \`oun r\'\i{} i p\'e \`ododo ni \`al\`ay\'e t\'\i{}
\Aj \sd e. In\'u \Aj d\`un. B\'\i{} \'o til\'\ed{} j\'\ed{} p\'e
ohun t\'\i{} \'o l\`e gbe \Aj l'\'o \'ns\od{} f\'un \`A\sd \`ak\'\ed{},
s\'\i b\`\ed{} \`\od n\`a t\'\i{} \'o gb\'a gb\'e \`\od r\`\od{} n\'a\`a
kal\`\ed{} w\od{} ni l\'et\'\i{} p\'up\'\od{}. B\'\i{} \As b\'a le mu
\`\i m\`\od r\`an y\`\i{} l\`o, \`ati m\'aa l\od{} s'\'\od d\'\od{}
\Aj k\`o n\'\i \'\i{} \sd \`oro m\`\od{}. K\`\ed k\`\ed{} b\'\ed \`\ed{}
im\`u \ed l\'\ed d\`\ed{} \'a w\od gb\`a. \As n\'\i{} oun \'a
b\`\ed r\`\ed{} s\'\i{} \'\i{} m\'aa \sd '\`al\`ay\'e
\`\od r\`\od{} fun b\`ab\'a \`oun, \sd \`ugb\'\od n p\'e
d\'\i \`\ed d\'\i \`\ed{} ni \`oun y\'\i \'o m\'aa \sd e \'e o.
Ij\'\od{} t\'\i{} a b\'a g\`un k\'\od{} ni a \'nkan \`\od run.
Nw\'\od n fi ipade s\'\i{} kej\`\i{} n\'\i{} yunif\'as\'\i t\`\i{}
ni y\`ar\`a \`Aj\`an\'\i{}.\par
```

End of file.

```
\bye
```

Typesetting the file

Once the typesetting file had been prepared, the **tex** program was started, and run against the prepared file. (In practice, this had to be done several times because there were invalid commands in the file to which the **tex** program objected, and these had to be tracked down and fixed.)

The result of a successful typesetting operation is a device independent DVI file, which the **tex** program creates on the basis of character image data from the METAFONT font files and the sophisticated batch-processing algorithms for composing pages. The DVI file is a series of page images, in which type characters are represented as black and white pixels at a predefined resolution – which by default was 300 dots per inch. I previewed this file without wasting paper by using the **dvipreview** program included in the package.

At this stage, I could have sent the file to a laser printer, but I wanted to see if a more versatile image of the page could be generated using the PostScript language. The **dvi2ps** conversion program produced a PostScript file from the DVI data, and because I had installed a PostScript Type One version of the Computer Modern font in addition to the METAFONT version, the file was created with resolution-independent characters instead of the fixed-resolution 300 dpi ones.

I discovered that by processing the PostScript file using **Acrobat Distiller** software from Adobe Systems, I could create a Portable Document Format (PDF) file of the TEX- typeset document. Such a file could be placed on a Web site, to be viewed by users of Adobe's free Acrobat Reader software. I was also able to use Acrobat software to make an Encapsulated PostScript image file, which allowed me to place the TEX typesetting as an image on an ordinary DTP page (which is how it made its way onto page 14).

Some notes and conclusions

- Not every language is as troublesome to typeset in TEX as Yorùbá. In Igbo, for instance, the underdotted i, o and u characters never have a superimposed accent, and this makes them easy to typeset using the basic TEX codeword **\d** – which will place an underdot under any desired character. But it proved impossible to precede a character with two simple accent-placement codewords: **\d \`o** does not produce **ọ̀**. Thus, for Yorùbá, I had to resort to a 'box placement' strategy using the command string `{o\kern-.38em\lower.45ex\hbox{.}\kern.13em}` to generate an underdotted o. This could then be preceded by the **\`** grave-accent-placement command-word.

- One can make life easier by defining one's own codewords in TEX, and I used this strategy here, so that the long string of codes just described was aliased to a simple author-defined codeword, **\od** (see lines 12–13 in code on previous page).

- Of course, a typesetting strategy which relies on this basic implementation of TEX using standard METAFONT resources cannot solve the problems of typesetting languages which have extra, special letters that cannot be made up of existing components. Twi (Akan) and Hausa are examples of such languages, which I therefore refer to as 'level four' in my scheme of difficulties.
- However, searches of the Cornell University Africana Web resources indicate that some programmers *have* developed TEX-compatible METAFONT fonts for African typesetting – notably Jörg Knappen at the University of Mainz in Germany. His **fc** font package, which is free shareware, is said to support Akan (Twi), Bambara, Bamileke, Bassa, Bemba, Ciokwe, Dinka, Dholuo, Efik, Ewe-Fon, Fulani, Gã, Gbaya, Hausa, Igbo, Kanuri, Kikuyu, Kikongo, Kpelle, Krio, Luba, Mende, More, Nhala, Njanja, Oromo, Rundi, Kinya Rwanda, Sango, Serer, Shona, Somali, Songhai, two systems of Sotho, Swahili, Tiv, Yao, Yoruba, Xhosa and Zulu.[7]
- It is clear that TEX is a powerful and impressive typesetting system when used to typeset materials with a very simple columnar structure such as a textbook. Indeed, its enthusiastic supporters point out that TEX has much better automated algorithms for producing pleasing hyphenation and space distribution in the kind of formal justified-text setting used in such publications. However, it would be very difficult, if not impossible, to use a TEX system for the graphical and exciting layouts required for newsletters, posters or publicity leaflets.

Therefore, while I have come away from this experiment impressed at the capabilities of the TEX system, I find that I cannot recommend it at all for the kind of typesetting and publication-design task which a voluntary organisation would require. That needs an easy-to-use WYSIWYG system which integrates the entry, placement and formatting of text and graphics in an interactive editing view, and which shows the real font characters on screen as you type – not scary formatting codes.

7 We have not been able to verify this, nor see samples of output. Of course, some of the languages in this list are not at all problematic to typeset; others definitely are. Apparently the **fc** fonts were used to typeset an important Hausa–English dictionary, for instance.

Obtaining modified fonts for African languages

An examination of the character charts in Appendix B shows that a number of African languages require special characters and diacritical marks in combination with characters, and standard fonts do not support these needs. (Indeed, an illustration program had to be used to prepare those charts, because they could not be typeset by normal means.)

Using a font editing program

One way to get the letterforms and combinations required would be to take an existing font, open the font data package using an editing program such as Macromedia Fontographer (as seen in figures 2, 4 and 5 above), delete unwanted characters from the font, and replace them with the characters required by the language in question. In many cases this editing could be achieved by copy-and-paste methods that would pose few technical or aesthetic challenges, for example to place a circumflex over a w character to create the ŵ required for Nyanja. The edited font would then be saved as a new font[8] and could then be installed and used in the normal way.

A few points should be made about this process:

- **Levels of difficulty —** Making custom assemblies of existing letterforms with existing accents is very easy. Creating new letterforms – as would be required by e.g. Krio or Hausa – is more difficult, especially as one would wish these to fit in smoothly with the rest of the letters. The spacing arrangements between newly-created and existing letterforms would also need to be checked and adjusted.

- **Legality** — A font is a software program, and when you 'buy' one you in fact merely license the right to use it on a number of designated computers and printers (see the font license from the particular font vendor for particulars of each licene). One should be very careful to ensure that modifying a font and re-saving it does not constitute a breach of the licencing agreements.

  In general we know that *rearranging* or *modifying* a font to which one has a right of usage is not taken by most font vendors to be a breach of the licence agreement, so long as the modified font is used only by the original licensee; but to give that modified font to another party would be a clear breach of contract. Having said that, we are not qualified to give a detailed legal opinion on this matter and more authoritative advice might need to be sought.

8 This could be either in PostScript Type One or TrueType format, in either Macintosh or Windows encoding, and the Fontographer software is available for both Windows and Macintosh. See Appendix A.

Purchasing a specially-engineered font or a 'superfont' set

One way to avoid legal problems would be to find a legitimate vendor who could sell a valid licence to a font with the range of characters required to typeset the language or languages in question. The problem is that support for African languages has not been a priority for the established vendors of quality fonts such as Adobe, Agfa, Monotype, Heidelberg etc. Either they are unaware of the problem, or they do not want to act upon it because the market for such fonts would be too small to justify the effort.

(Software companies also consider developing economies to be a poor risk because unauthorised copying or 'piracy' is, understandably, most common in these markets.)

Dalton-Maag: a font-house willing to customise

A conversation with Bruno Maag of the specialist type design company Dalton-Maag indicates that they *would* be prepared to make a custom version of any of their own existing type designs, in order to prepare it for use in African-language typesetting.

Fig. 6

The Lexia and Pan type families from Dalton-Maag could be customised.

Lexia

ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒ&
abcdefghijklmnopqrstu
vwxyzæœfiflß.,!?
$¢£€0123456789%‰

ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒ&
abcdefghijklmnopqrstu
vwxyzæœfiflß.,!?
$¢£€0123456789%‰

ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒ&
abcdefghijklmnopqrstu
vwxyzæœfiflß.,!?
$¢£€0123456789%‰

Pan

ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒ&
abcdefghijklmnopqrstu
vwxyzæœfiflß.,!?
$¢£€0123456789%‰
ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒffiffl ss.,!?
$¢£€0123456789%‰

ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒ&
abcdefghijklmnopqrstu
vwxyzæœfiflß.,!?
$¢£€0123456789%‰

ABCDEFGHIJKLMNOPQRSTU
VWXYZÆŒ&
abcdefghijklmnopqrstu
vwxyzæœfiflß.,!?
$¢£€0123456789%‰

Dalton-Maag (see `www.daltonmaag.com`) is a small company with offices in Brixton, most of whose work is in producing custom fonts as part of corporate identity projects. The example best known to the public are the fonts produced for the National Westminster Bank, as used in their leaflets and promotional literature, and even in the interface of their cash machines. However, Dalton-Maag have latterly put more energy into designing their own fonts, for public licencing. Three font families are currently available, of which Lexia and Pan are the two most suitable for general-purpose typesetting, and therefore for modification for African languages.

A modified font for African typesetting sourced from Dalton-Maag would be of high technical and aesthetic quality, and Bruno Maag –who is a world-class expert in font engineering – has a number of ideas about how to make such a font easy to use as well. However, this might be quite an expensive option. On top of their standard font licencing fees, Dalton-Maag typically charge £500 a day for font customisation work – though Bruno says he is open to negotiation.

Summer Institute of Linguistics

The Summer Institute of Linguistics (see `www.sil.org`) is a US-based organisation, which publishes an encyclopaedia of linguistics and culture called *The Ethnologue*. It is my understanding that the origins of the Institute are in Christian missionary work – which, as has been noted above, has often been a driving force in the promotion of literacy and the development of writing systems.

The Institute offers for sale four 'extended Latin' font families with the purpose of assisting the typesetting of a wide range of languages. The fonts have more characters in them than a keyboard can typically accommodate, but a utility is provided so that one can create a custom set of the characters to suit the job in hand, and assign easy-to-use key sequences for inputting the characters.

The best way to illustrate the range of supported characters is to show the sample graphics for the 'Doulos' font – somewhat like Times Roman – from the Summer Institute of Linguistics Web site…

Fig. 7 – part A: 'Doulos' font characters from A to R

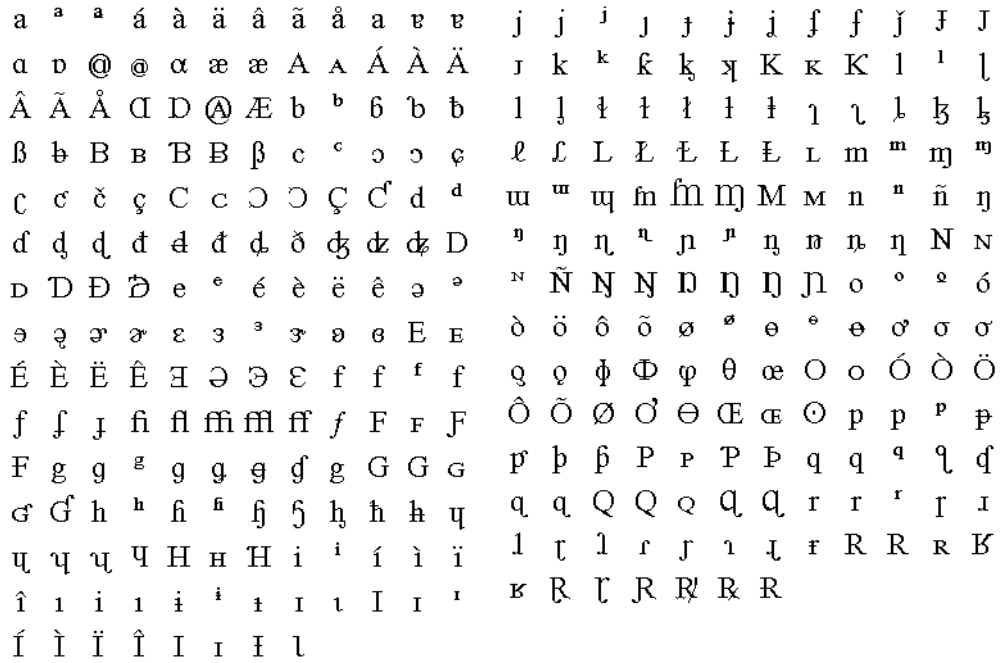


Fig. 7 – part B: ‘Doulos’ S to Z, numerals and punctuation

s ˢ š ʂ ʂ ʂ ʃ ʃ ʃ ʃ S s
Š ʃ t t ᵗ ʈ ʈ ᵗ ʈ ʈ ʈ f
ʧ ʦ ʧ T ᵀ Ŧ T T u ᵘ ú ù
ü û ʉ ᵘ ʊ ʊ ʊ ʊ ʊ µ U
ʊ Ú Ù Ü Û Ʊ Ʊ Ʊ Ʊ Ʊ Ʊ v
ᵛ v ʌ ʋ ᵛ ᵛ ᵛ ʏ ʏ ʏ ʏ ʏ
ʏ ʏ ʊ V v Λ Ʋ Ʋ Ʋ w ʷ ʷ
ʋ ʌ ɯ ɯ ɯ W w x ˣ ˣ x χ
X x y y ʸ ý ÿ ʎ ʏ ʏ λ λ
Y ʏ ʏ Ý Ÿ ʏ Y z ᶻ ʒ ʒ ʒ
ʒ ʒ ž ʒ ʒ ʒ ʒ ʒ ʒ ʒ Z z
Ʒ Ž Ʒ Σ Ʒ Ʒ

0 ⁰ ⁰ ⁰ 0 1 ¹ ¹ ¹ 1 2 ²
² ² 2 ² 3 ³ ³ ³ 3 4 ⁴ ⁴
⁴ 4 5 ⁵ ⁵ ⁵ 5 6 ⁶ ⁶ ⁶ 6
7 ⁷ ⁷ ⁷ 7 8 ⁸ ⁸ ⁸ 8 9 ⁹
⁹ ⁹ 9 ½ ⅓ ¼ ⅛ ⅔ ¾ ⅜ ⅝ ⅞
? ¿ ? ʔ ʔ ? ʔ ʔ ʔ ʔ ʔ ʔ
ʔ ʔ ʔ ʔ ʔ ʔ ! ¡ ! ! ! ‼
" " . · . … : ⁝ ⁝ , ‚ ‚
; ; ; ‘ ‘ ‘ ’ ’ ’ ’ “ “
“ ” ” „ ( ) [ ] { } ‹ ›
‹ › « » & @ ¶ § $ ¢ £ ¥
% % % ‰ © ® ℗ ℠ ™ # ⁜ *
† ‡ ˙ ⁚ ⁝

Fig. 7 – part C: ‘Doulos’ diacritics sample; plus pi characters

ʒ ạ ā Ā ʒ ạ ā Ā ä Ä ȧ Ȧ
ạ ạ ã Ã ã Ã à À ǎ Ǎ ạ ǎ
Á ǎ Á ǎ Ǎ ǎ Ǎ ạ à À à À
à À ȧ Ȧ ȧ Ȧ ạ ä Ä ä Ä ǎ
Ǎ â Â ʒ ạ ȧ Ȧ ʒ ạ ȧ Ȧ â
Â â Â ǎ Ǎ ǎ Ǎ à À ǎ Ǎ ã
Ã ạ ǎ Ǎ ǎ Ǎ ạ ǎ Ǎ ǎ Ǎ ǎ
Ǎ à À ǎ Ǎ ǎ Ǎ ʒ ạ â Â â
Â ʒ ạ ã Ǎ ạ ạ ǎ Ǎ ʒ ạ ȧ
Ȧ ȧ Ȧ ʒ ạ ȧ Ȧ ʒ ạ ȧ Ȧ ǎ
ǎ Ǎ ǎ Ǎ ǎ Ǎ ǎ Ǎ ǎ Ǎ ǎ Ã
ȧ Ȧ ạ ȧ Ȧ ç ç ç ç ạ ạ ạ
ạ ḍ ʒ ḍ ʒ ḍ ạ ạ ʒ ḍ ḍ ḍ
ạ ạ ạ ạ ạ ạ

· • ● ● ❧ ○ ◑ ◐ ◐ ◑ ◑ ◐
■ ■ □ □ ◆ ◆ ◇ ◇ ◇ ◊ ❖ ★
☆ ▶ ◀ ▼ ▲ ▶ ◀ ◗ ◖ ◗ ◗ ◗
◖ ⇨ ⇦ ↓ ↑ → ← ↓ ↑ ↖ ↘ ↗
↙ ↕ ↕ ↔ ⇌ ⇋ ⇒ ⇐ ⇓ ⇑ ⇕ ⇔
☞ ☜ ☛ ☚ ✂ ✂ ☎ ✉ ☺ ♂ ♀ ✈
✍ √ ✓ ~ ¤ ␣ ° ˍ ▮ ¶ ˍ ˈ
| | ǂ ¦ ‖ | | ▌ — — ▬ —
— — — — ▬ ⌐ ¬
\ \ / / + + − - ± ±
∓ × ÷ √ = ≠ ≡ ~ ≈ ∞ ∨ ˅
∧ ^ < ≺ ˂ > ≻ ≽ ˃ ≤ ≥ ∫
⊕ ⊕ ⊗ ‖ ∫ ⎫ ⎰ ⎨ ⎩ | ⎭ ⎬
⎭ ⎺ ⎻ ⎼ ⎯ ⎽ ⎽ ⎽ ∪ ∩ ⊂ ⊃
⊂ ⊃ ⊂ ⊂ ⊃ ⊃ ⊆ ⊆ ⊇ ⊇ ° °
∅ Δ ∂ Σ Ω Π π α β Γ σ µ
τ Φ θ δ ø ε η ς ξ ζ ω ℵ
² . ∫ ⎭ ∩ ☑ ☐ ℞ ✍ ♦ ♥ ♣
🔈 ♭ ♮ ‖

The Ethiopic script system

The only major writing system in Africa, apart from Arabic, which does not use the Roman script at all is the ancient Ethiopic script.

A number of closely related African Semitic languages which are spoken by a total of about 18–20 million people in Eritrea and Ethiopia – **Amharic**, **Tigré** and **Tigrinya** – use variants of an ancient writing system derived from the South Semitic *Sabaean* script. The advanced kingdom and culture of Saba‘ (Sheba) in what is now Yemen had already had a long interplay of influence with the Horn of Africa, and indeed legend has it that Bilqis, Queen of Sheba (‘Makeda’, as she is called in Ethiopian tradition) bore to King Solomon of the Jews a son, Menelek, who is claimed as the founder of the royal Ethiopian dynasty.

Around the 4th century AD there was a high degree of contact, cultural interchange and settlement between Saba‘ and the emerging kingdom of Axum in the north of modern Ethiopia. Both a Sabaean script and Greek had been in use in the area from about the fifth century BC, and in the 4th century AD these were joined by a purely Ethiopic script similar to Sabaean, known as **Ge’ez**, which came to predominate from this date onwards.

The Ge’ez language is the common ancestor of modern Tigré and Tigrinya, and it became the language of the Ethiopian Orthodox Christian church. Thus, although Ge’ez ceased to be spoken around the 9th or 10th centuries, it was retained as the liturgical language of that church, and the height of classical Ge’ez literature was between 13th and 17th centuries.

Slightly different sets of letters are used to write southerly Amharic on the one hand, and northerly Tigré and Tigrinya on the other. Oromo, the other major language of Ethiopia, used to be written in a form of Ge’ez script, but is now more commonly written with a Roman alphabet.

Like Sabaean, the original Ge’ez script consisted purely of consonants, of which it had 26. As the same script was adopted for use with the related language of Amharic, it gained more consonants, now having a total of 33. This process was one of gradual evolution. However, in what was probably a conscious act of reform, the consonants were later conceived as having seven ‘orders’ – depending on the vowel-sound pronounced after each – and each letter acquired a vocalisation marker attached to it. Thus the modern Ge’ez/Amharic scripts have hundreds of distinct compound glyphs, and function as a kind of ‘syllabary’.

The large number of glyphs obviously can cause a problem for using a computer to typeset the languages that use this script.

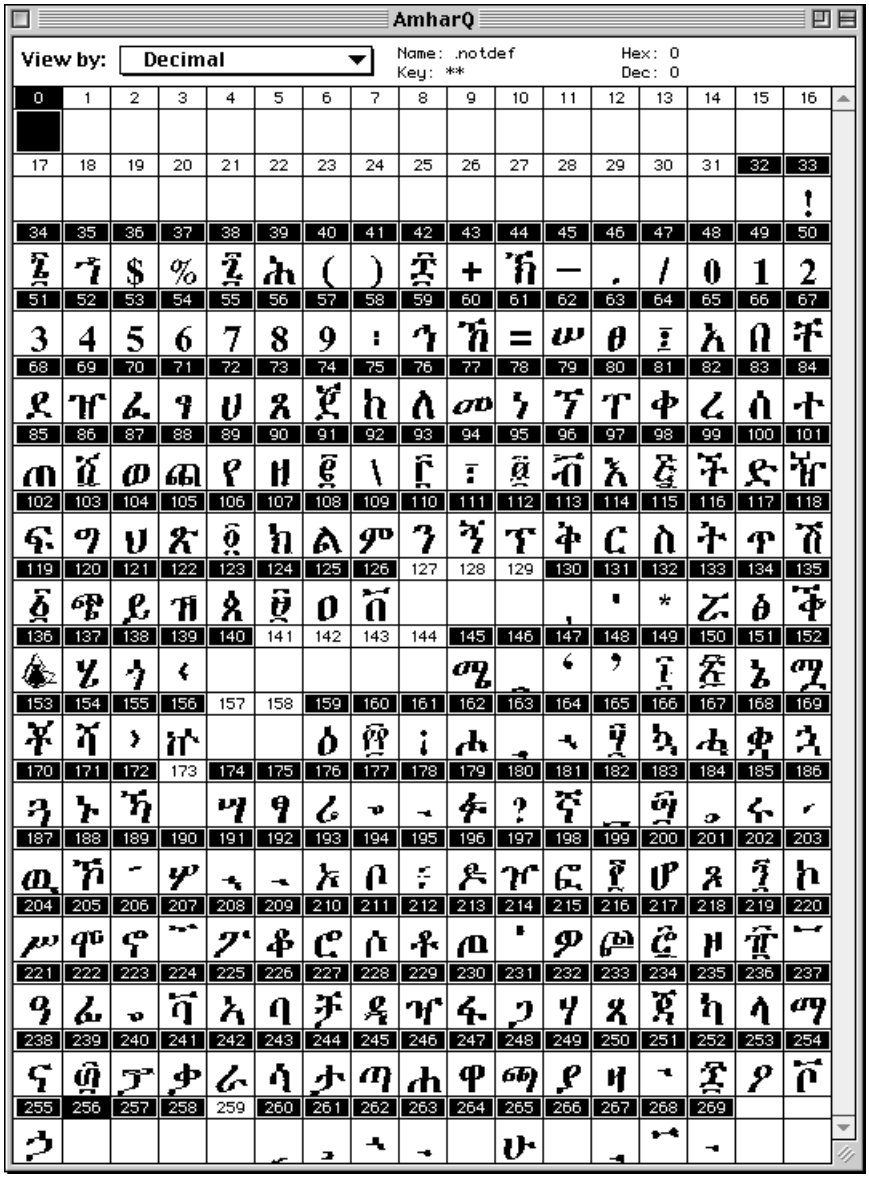


Fig. 8: an Amharic font: AmharQ.ttf

This shareware TrueType font with Windows encoding is available free from a number of Web sites.

One can get some idea of the appearance of an Ethiopic script by looking at a Fontographer character-repertoire view of a free shareware TrueType font for Windows called **AmharQ**; see above. The script has a distinct classical form strongly influenced by the manuscript tradition, for which a broad-edged reed pen was used with the edge held nearly horizontal. Unlike its Sabaean ancestor, modern Ethiopic script is written from left to right.

A complete solution?

A number of free shareware Ethiopic fonts are available in TrueType format for use with Windows. However, I believe the most complete Windows solution is offered by the enterprising EthiO Systems Company of Houston, Texas (`www.neosoft.com`). I cannot think of a better way of describing their product offering than by reproducing some of their Web pages on the following pages of this paper…

Software & Document | Ethiopian Computing | Ethiopicing Internet | Authorized Distributors | Home

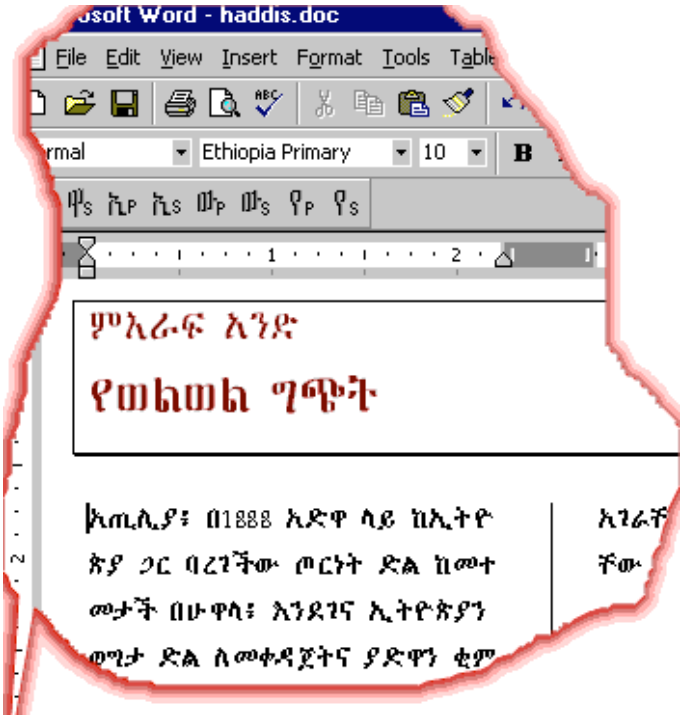
ዋሽራ 3.0

Introduction

WashRa is designed to provide services where Ethiopian script and language users can take advantage of popular Windows 95 and Windows 3.1 software including word processing, spreadsheets, database, presentation, multimedia, Internet, and many other programs. Users can exchange mails, write or read USENET news, design illustrations, and publish documents with Ethiopian text.

WashRa 3.0 introduces new features:

- A new user interface, where users can have access to all services at once with a single mouse click. Setting application mode, configuring the keyboard layout, or reading on-line help is easier than before.
- WashRa 3.0 comes with Enhanced KWK keyboard in additon to the standard one. Now, users of MS Word and WordPerfect can type all characters in Ethiopian script with out switching between the *primary* and *secondary fonts*.
- Four fonts designed to meet your classical, modern, and Internet based publication needs.
- Support for latest Win95 application programs including Office97, CorelDRAW 7, Netscape Communicator .01, PageMaker 6.5, Adobe Acrobat, MS Internet Explorer,...
- The online help has been revised and new features are added; more specifically, the English section has major overhaul.



Using WashRa with Word97

WashRa from EthiO Systems: page 1

Supported Application Programs

WashRa 3.0 supports several Windows application programs.

- **Word Processing:**
  - WordPerfect, MS Word, AmiPro, WordPro, MS Write,...
- **Spreadsheets:**
  - Lotus 1-2-3, Excel, Quattro Pro,...
- **Database:**
  - MS Access, Approach,...
- **Presentation:**
  - Freelance, PowerPoint, Presentation
- **Illustration:**
  - CorelDRAW, Adobe Illustrator, PhotoShop, Paintbrush. PhotoWorks,...
- **Publishers:**
  - PageMaker, Corel Ventura, QuarXpress,...
- **Internet:**
  - Netscape Navigator, Netscape Communicator, MS Internet Explorer, Eudora, Adobe Acrobat,...
- **Multimedia:**
  - Director, Authorware,...

Comments, please write to ethiosys@neosoft.com

WashRa from EthiO Systems: page 1 part 2

Software & Document | Ethiopian Computing | Ethiopicing Internet | Authorized Distributors | Home

Enhanced KWK Keyboard

The **EKWK** virtual keyboard layout provides standard and application dependent keyboard services, in which the former is used on all supported programs and the later with **MS Word** and **WordPerfect**.

KWK Keyboard

- A **basic key** is used to enter character from the ***1st*** order if it is in the primary set of the font as shown below.
- A combination of **basic** and **qualifier keys** is used to enter characters from ***2nd*** to ***7th*** order.
- Qualifier keys are "**u**", "**i**", "**a**", "**y**", "**e**", "**o**", and "/".

Basic Key	Qualifiers Keys						
		u	i	a	y	e	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
...	...	...	...	...	...	...	...

A glimpse of KWK keyboard table

When the user type a key from the 1st order, **KWK** will display the matching character, but if the user follows the ***1st*** order key with one of the qualifier keys as shown above, KWK will map the two sequence of keys to the proper character and display it.

Enhanced KWK Keyboard

The EKWK keyboard is essentially the same as KWK, but has more option to the MS Word and WordPerfect users. Now, users can enter all characters in the Ethiopian script with out switching fonts between *primary* and *secondary*. This is done using the "**Alt**" key.

- A basic key is used to enter character from the ***1st*** order if it is in the primary set of the font, but "**Alt + a basic key**" if it is in the secondary set.
- A combination of basic and qualifier keys is used to enter characters from ***2nd*** to ***7th*** order.
- Qualifier keys are "**u**", "**i**", "**a**", "**y**", "**e**", "**o**", and "/".

WashRa from EthiO Systems: page 2 part 1

Basic Key	Qualifiers Keys						
		u	i	a	y	e	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
Alt + h	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
Alt + s	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
...	...	...	...	...	...	...	...

A glimpse of EKWK keyboard table

When the user type a key from the 1st order, EKWK will display the matching character from the *primary set* , but if the user follows the 1st order key with one of the qualifier keys as shown above, KWK will map the two sequence of keys to the proper character and display it. However, if the user type "**Alt + Basic key**" EKWK will display the matching character from the *secondary set* .

Next Section

Comments, please write to ethiosys@neosoft.com

WashRa from EthiO Systems: page 2 part 2

Software & Document | Ethiopian Computing | Ethiopicing Internet | Authorized Distributors | Home

ፊደል
ከመሳል

Ethiopian Script Fonts

Ethiopian Script Fonts

WashRa provides four fonts ***Washra*** , ***Ethiopia*** , ***Wookianos*** , *and* ***YebSe*** . They can be used for classical and modern publishing, Web page and artistic illustration. They are not designed based on decomposition method which lend itself to distortion of the script. They come in TrueType form, but they are also available in PostScript Type-1 format. A glimpse of each font:

WashRa

«ጥፈት የማያውቁ መምህራን የተጣፈውን
ከመተርጐም በቀር የጸሐፍትን ስሕተትና
ግድፈት ማረምና ማቅናት እንዳይችሉ፤
ሰዋሰው የማያውቅ ጸሐፊም መልክአ ፊደ
ሉን፤ ሀረጉን፤ ቅጠሉን አጣጣሉን ብቻ
ከማሳመር በቀር ንባቡንና አገባቡን አጥርቶ
ማወቅ ፊደሉን መጠንቀቅ አይችልም።»

አለቃ ኪዳነ ፡ ወልድ ፡ ክፍሌ

Wookianos

«ጥፈት የሚያውቁ ሠምህራን የተጣፈውን ከመተር
ጐም በቀር የጸሐፍትን ስሕተትና ግድፈት ማረምና
ማቅናት እንዳይችሉ፤ ሰዋሰው የሚያውቅ ጸሐፊም
መልክአ ፊደሉን ፤ ሀረጉን፤ ቅጠሉን አጣጣሉን ብቻ
ከማሳመር በቀር ንባቡንና አገባቡን አጥርቶ ማወቅ
ፊደሉን መጠንቀቅ አይችልም።»

አለቃ ኪዳነ ፡ ወልድ ፡ ክፍሌ

Ethiopia

WashRa from EthiO Systems: page 3 part 1

«ጥፊት የማያውቁ መምህራን የተጣፈውን ከመተር ጐም በቀር የጸሐፍትን ስሕተትና ግድፈት ማረምና ማቅናት እንዳይችሉ፥ ሰዋሰው የማያውቅ ጸሓፊም መልክአ ፊደሉን ፥ ሀረጉን፥ ቅጠሉን አጣጣሉን ብቻ ከማሳመር በቀር ንባቡንና አገባቡን አጥርቶ ማወቅ ፊደሉን መጠንቀቅ አይችልም።»

አለቃ ኪዳነ ፡ ወልድ ፡ ክፍሌ

YebSe

<<ጥፊት የማያውቁ መምህራን የተጣፈውን ከመተር ጐም በቀር የጸሐፍትን ስሕተትና ግድፈት ማረምና ማቅናት እንዳይችሉ፣ ሰዋሰው የማያውቅ ጸሓፊም መልክአ ፊደሉን ፣ ሀረጉን፣ ቅጠሉን አጣጣሉን ብቻ ከማሳመር በቀር ንባቡንና አገባቡን አጥርቶ ማወቅ ፊደሉን መጠንቀቅ አይችልም። >>

አለቃ ኪዳነ ፡ ወልድ ፡ ክፍሌ

Next Section

WashRa from EthiO Systems: page 3 part 2

Software & Document | Ethiopian Computing | Ethiopicing Internet | Authorized Distributors | Home

ኢትዮ ሲስተምስ

Documentation

Reference Manual

A reference manual written in Amharic provides a detail and step by step instructions on installation, configuration, keyboard layout, using WashRa applications programs, Internet, Haddis Character Code, and many more.

English and Amharic On-line Help

Besides the reference manual, WashRa 3.0 comes with an on-line help document both in English and Amharic. The on-line help includes, but not limited to:

Introduction to WashRa 3.0,
Using WashRa with with Windows applications programs,
Entering character in Ethiopian script---KWK keyboard layout,
Tutorial on exchanging mails written based on Ethiopian script and building database,
Troubleshooting,

Price

Price

WashRa 3.0 $115.00 plus shipping and handling.
Owners of WashRa 2.0 can upgrade to version 3.0 for $75.00 plus shipping and handling.

Shipping and handling:

US, $4.00
Overseas, $12.00

Payment Method

Major credit cards, money order (cashier check), or wire transfer; and check (only US).
If you prefer wire transfer, please contact us for more information.

Comments, please write to ethiosys@neosoft.com

WashRa from EthiO Systems: documentation & ordering

Appendix A: Type technology notes

PostScript font format

The current phase in computerised publishing was initiated in California in the mid 1980s through close co-operation between Apple Computer and Adobe Systems. Apple's Macintosh was the first affordable computer which could display pages on screen as they would appear on the printer, but the first fonts for the Mac were crude bit-maps, appropriate only for their ImageWriter dot-matrix printer.

Meanwhile, Adobe had created the PostScript page description language, by means of which an image of a page could be sent to a printing device, irrespective of whether that device was a medium-resolution office laser printer or a high-resolution graphic arts imagesetter; they also devised the **Type One** scaleable outline fonts to work inside a PostScript workflow.

In 1985, Apple licensed the PostScript technology for its first laser printer, together with a number of Adobe's PostScript Type One fonts. Linotype also licensed the system for its laser imagesetters. Provided with this technical base, Macintosh DTP programs such as PageMaker and QuarkXPress quickly swept away previous methods of typesetting machines.

Origins of TrueType

However, a strong faction within Apple felt that the company was unduly reliant on Adobe's technology – for which Adobe charged hefty licensing fees – and quietly planned an alternative outline font format based on a different kind of geometric curves.[9] This became the **TrueType** font format. Apple entered into an agreement with Microsoft whereby both companies would support the rendering of TrueType fonts by their operating systems, and Microsoft would provide a page description language – a PostScript alternative – to work with the new font format.

When Apple and Microsoft went public about their TrueType initiative, some observers thought that Adobe's Type One format was doomed, especially as the Apple and Microsoft operating systems were upgraded to output high-quality font data both to the screen and to low-cost printers such as ink-jets. However, within the graphic arts industry, publishers had come to rely on PostScript workflows to publish their magazines and adverts and *make money* – and it was difficult to use TrueType fonts within such workflows. Adobe also improved their position by creating Adobe Type Manager (ATM), a system extension for Windows and Macintosh which renders PostScript Type One font data nicely to screen and to low-cost

9 PostScript Type One fonts have their outlines defined in cubic-equation curves; TrueType outlines are quadratic curves.

printers, and which is bundled with desktop publishing and other graphic arts programs. Therefore, in practice, both TrueType and PostScript Type One font formats have survived; the former are used mostly in business communication and the home, and the latter in professional publishing.

Does it matter whether one uses TrueType or PostScript fonts for publishing projects? Recent technical developments in PostScript interpreter software for imagesetters mean that it should not – but former bad experiences with trying to use TrueType fonts in a graphic arts workflow mean that most graphics professionals are still unwilling to use TrueType fonts in projects which will be sent out for professional imagesetting and printing. They may be wrong, and probably now are, but printers are known to be difficult to separate from their technical prejudices.

Beyond eight bits: the role of Unicode

The Unicode Consortium is a project which aims to give each character in each of the world's languages a unique reference code of its own, as a better way of allowing computer systems to reference large character sets. The idea is to use not just one byte's worth of data to reference a character, but two; with sixteen binary digits, this creates a 'reference space' to refer to more than 64,000 characters.

Windows NT4.0 was the first widely available computer operating system to use Unicode encoding for referencing characters, and Unicode is also supported by the Windows 2000 operating system and the Microsoft Office 2000 application suite. Adobe InDesign, Adobe Illustrator and QuarkXPress 5.0 are three document preparation systems likely to be early supporters of Unicode.

Unfortunately, not all of the characters required by African languages appear to have been yet indexed by the Unicode consortium and it is not clear what impact Unicode will have on the develop of African typesetting.

The OpenType project

Collaboration between Adobe Systems and Microsoft Typography has gone into developing a 'next-generation' font format which would be able to contain a larger number of characters. The OpenType format may use either PostScript's Bézier curves or TrueType's quadratics to describe the character outlines, and will have extensive sets of tables to control the relationships between letters.

Once publishing applications are developed which support the feature, OpenType's **glyph substitution** features may be of great interest for typesetting African languages because this will allow a sequence of key-presses to be 'collapsed' or 'fused' into the presentation of a single composite glyph. Thus one might press the key sequence **** + **`** + **o** and be presented with **ọ̀** on screen.

Appendix B: Character sets for African languages

The charts following this page are intended to show the Roman letterforms and accents required to typeset a variety of African languages. (However, it should be noted that it was hard to find trustworthy sources of information, and more diligent research is required to improve upon these findings.)

The language charts are presented in alphabetical order, without page numbers.

- **Baule**
- **Chewa, Chichewa** or **Nyanja**
- **Edo** or **Bini**
- **Fulfulde** or **Pular**
- **Hausa**
- **Kikuyu**
- **Krio**
- **Igbo**
- **Oromo** or **Galla**
- **Somali**
- **Swahili**
- **Tswana**
- **Twi, Akan, Fante** or **Ashanti**
- **Wolof**
- **Xhosa**
- **Yoruba**
- **Zulu**

Baule

Baule is a member of the **Kwa** sub-group of the **Niger-Congo** family of languages. It is spoken by some 1·5 million people in Côte d'Ivoire, and half a million people in Ghana.

Consonants

Bb Dd Ff Gg Kk
Ll Mm Nn Pp Ss
Tt Vv Ww Yy Zz

Vowels

Aa Ee Ɛɛ Ii Oo
Ɔɔ Uu

- Baule is not a difficult language to typeset by computer, provided one has access to a special font with the correct letterforms. It should be possible to map these consonants to existing keys on the keyboard which are not required for other purposes.

Chichewa or Nyanja

The language variously called Chewa, Chichewa or Nyanja is a member of the **Bantu** sub-group of **Benue-Congo** languages, with a tradition of origin in the Zaïre basin. It is spoken by 3.8 million people in south-east Africa, notably in Malaŵi, and in the south east of Zambia, where it is the second most common language after Bemba.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Ŵŵ Xx Yy Zz

Vowels

Aa Ee Ii Oo Uu

Further notes

- I have marked the consonants **Q**, **V** and **X** in lighter grey above, as I am not certain that they are used in Nyanja.
- The only letterform required by Nyanja that is not in the standard Macintosh or Windows fonts is **W with superscript circumflex**. It may be of interest to note that this is a letterform also required by Welsh – for which it is possible to obtain some modified fonts.
- There are two solutions for typesetting Nyanja. If a modified font (as for Welsh) is available, any DTP or word-processing program can be used to typeset it. Alternatively, one could use a typesetting system which can place a floating accent on top of any arbitrarily chosen character (the 'composed character' approach). Any implementation of TEX could do this with ease, but complex document layouts such as leaflets and newsletters are hard to do in TEX.

Edo or Bini

Edo (or Bini) is a member of the **Niger–Congo** family of languages, spoken in Nigeria on the West bank of the Niger south of the confluence with the River Benue, and in Benin. Estimates of the number of speakers vary widely, up to 2·5 million.

Consonant range and sequence

b d f g gb gh h k kh kp l m mw n p
r rh rr s t v vb w y z

Consonantal letterforms

Bb Dd Ff Gg Hh Kk
Ll Mm Nn Pp Rr Ss
Tt Vv Ww Yy Zz

Vowels

Aa [Àà Áá] Ee [Èè Éé] Ẹẹ [Ẹ̀ẹ̀ Ẹ́ẹ́]
Ii [Ìì Íí] Oo [Òò Óó] Ọọ [Ọ̀ọ̀ Ọ́ọ́]
Uu [Ùù Úú]

Further notes

- In addition to the seven vowels shown above, Edo also has five nasalised wowels, but these are signalled simply by adding an n – e.g. an.
- Edo has tonal accents (signalled with acute and grave diacriticals), but the practice is to use these only in the hundred or so words where the lack of an accent would result in ambiguity.
- There are two possible approaches to typesetting Edo. Any implementation of TEX could do the job, but complex document layouts such as leaflets and newsletters are hard to do in TEX. For use with standard DTP or word-processing programs, a font with an extended character set would be preferred.

Fulfulde or **Pular**

Fulfulde (Pular, Pullaar, Pulle) is the language of the Fulani or Fulbe people, who are widely dispersed throughout West Africa in a zone from Senegal to Cameroon. Related to Wolof and Serer, Fulfulde is a member of the West Atlantic sub-group of the Niger-Congo family and may have as many as 15 million speakers. It is a national language in Guinea, Mali and Niger. Early adopters of a pastoral lifestyle, the Fulani have also played an important role in the dispersion of Islam in West Africa.

Consonants

Bb Ɓ Cc Dd Ɗɗ Ff
Gg Hh Jj Kk Ll Mm
Nn Ɲɲ Pp Qq Rr Ss
Tt Ww Xx Yy Ƴƴ Zz

Vowels

Aa Ee Ii Oo Uu

Further notes

- **V** is not required for Fulfulde, but modified **B**, **D**, **N** and **Y** letterforms are required for extra consonants.
- Hausa is not a difficult language to typeset by computer, provided one has access to a special font with the correct letterforms. It should be possible to map these consonants to existing keys on the keyboard which are not required for other purposes.

Hausa

Hausa is by far the most widely spoken member of the Chadic sub-group of Afro-Asiatic languages, and the only one to have a written literature. Arabic script ('ajami') was introduced in the 16th century, but now a modified Roman alphabet is used. About 25 million people speak Hausa as their mother tongue, in south Niger and northern Nigeria, and several million more speak Hausa as a second language. In Nigeria, the Hausa-speaking Muslim community is politically influential.

Consonants

Bb Ɓ Cc Dd Ɗɗ Ff
Gg Hh Jj Kk Ƙƙ Ll
Mm Nn Rr Ss Tt Ww
Yy Zz

Vowels

Aa Ee Ii Oo Uu

Further notes

- **Q**, **V** and **X** are not required for Hausa, but modified **B**, **D** and **K** letterforms are required for three glottalised consonants. (A fourth, **TS**, can be written with existing letterforms.) **P** is occasionally met as a non-standard representation of **F** – which in Hausa has a pronunciation closer to **P**.
- Hausa has both long and short vowels. As an aid to pronunciation in learning Hausa, a macron is sometimes used over a vowel (ō) to show when it is long. However, long and short vowels are not distinguished thus in everyday written Hausa. Similarly, Hausa is a partially tonal language, with three tones: low, falling and high. A low tone may be indicated by a grave mark over a vowel (ò) and a falling tone with a circumflex (ô), but these are not used in everyday written Hausa.
- In summary, Hausa is not a difficult language to typeset by computer, provided one has access to a special font with the correct letterforms. It should be possible to map these consonants to existing keys on the keyboard which are not required for other purposes.

Igbo

Igbo (or Ibo) is a member of the **Niger–Congo** family of languages, variously classified with the **Bantu** or **Kwa** language sub-groups. It is one of the chief literary and cultural languages of southern Nigeria, and is spoken by about 12 million people. In the past there have been rival writing systems for Igbo sponsored by Catholic and Protestant missionaries; the system now used was set out in 1961 by S. E. Onwu.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Xx Yy Zz

Vowels

Aa Ee Ii Ị ị Oo Ọ ọ
Uu Ụ ụ

Further notes

- At the time of writing it is not clear whether all of the standard Latin consonants shown are actually required for Igbo, as the information was obtained from a reference source listing only the language's *extended* Latin font requirements.
- There are two possible approaches to typesetting Igbo. Any implementation of TEX could typeset Igbo with ease, using the **\d** control-word to position the underdots; however, complex document layouts such as leaflets and newsletters are hard to do in TEX. For use with standard DTP or word-processing programs, a font with an extended character set would be preferred.

Kikuyu

Kikuyu is an easterly member of the **Bantu** sub-family of **Niger-Congo** languages, spoken in Kenya by about five million people between Nairobi and Mt. Kenya. The Kikuyu were very active in the Kenyan independence struggle and the language is politically influential.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Yy

Vowels

Aa Ee Ii Ĩĩ Oo Uu Ũũ

Further notes

- There is no **X** or **Z** in Kikuyu. The letters **F**, **L**, **P** and **V** are also unused in Kikuyu, except when spelling words of foreign origin that require them.
- There are two possible approaches to typesetting Kikuyu. Any implementation of TEX could do the job, but complex document layouts such as leaflets and newsletters are hard to do in TEX. For use with standard DTP or word-processing programs, a font with an extended character set would be preferred.

Kikuyu

Kikuyu is an easterly member of the **Bantu** sub-family of **Niger-Congo** languages, spoken in Kenya by about five million people between Nairobi and Mt. Kenya. The Kikuyu were very active in the Kenyan independence struggle and the language is politically influential.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Yy

Vowels

Aa Ee Ii Ĩĩ Oo Uu Ũũ

Further notes

- There is no **X** or **Z** in Kikuyu. The letters **F**, **L**, **P** and **V** are also unused in Kikuyu, except when spelling words of foreign origin that require them.
- There are two possible approaches to typesetting Kikuyu. Any implementation of TEX could do the job, but complex document layouts such as leaflets and newsletters are hard to do in TEX. For use with standard DTP or word-processing programs, a font with an extended character set would be preferred.

Krio

Krio is an English-facing creole language, spoken and written by approximately 350,000 people in Sierra Leone. Most of the vocabulary is recognisably derived from English.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Xx Yy Zz

Vowels

Aa Ee Ɛɛ Ii Oo
Ɔɔ Uu

Further notes

- Three tones can be distinguished in Krio – low, high and falling – and these are sometimes marked in reference books with grave (ò), acute (é) and circumflex (ô) accents over the vowels. But these accents are not employed in everyday usage.
- Krio is not a difficult language to typeset by computer, provided one has access to a special font with the correct letterforms. It should be possible to map these consonants to existing keys on the keyboard which are not required for other purposes.

Oromo or **Galla**

Of all the members of the Cushitic sub-group of the Afro-Asiatic language family, Oromo has the most native speakers – about 11 million people, mostly in Ethiopia, and some in Kenya. In the past it has been written in Ethiopic script, but it was not officially favoured as a written language until 1970. However, there has long been a rich oral poetic tradition.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Ww
Xx Yy

Vowels

Aa Ee Ii Oo Uu

Further notes

- It may be noted that Oromo does not use **V** or **Z**. There is a total of 25 recognised consonants, five of which are written as digraphs (e.g. **ch**). The glottal stop or 'qoqsa' is written with an apostrophe.
- Oromo is not a difficult language to typeset by computer. All the required characters are already in the basic character set provided.

Somali

Somali is a **Cushitic** language within the **Afro-Asiatic** family of languages. It is spoken by some 6 million people in Somalia, in parts of Ethiopia and Kenya, and by substantial refugee communities abroad. The use of this simple orthography for Somali based on roman letters was made official by the government of Siad Barre in 1972, setting aside the unique 'Osmanian' script proposed by Osman Yusuf; a fairly successful literacy campaign followed. Some Somali-English children's picture books have been published in Britain.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn
Qq Rr Ss Tt Ww
Xx Yy

Vowels

Aa Ee Ii Oo Uu

Further notes

- It may be noted that Somali does not use **P**, **V** or **Z**. An apostrophe is used to indicate the glottal stop.
- Somali is not a difficult language to typeset by computer. All the required characters are already in the basic character set provided with all computers.

Swahili

Swahili (kiSwahili, 'coastal language') developed in Zanzibar and on the East African coast, based on a **Bantu** language structure with extensive borrowings of vocabulary from Arabic and Indian traders. Swahili was first written in 1728 in Arabic script, but later changed to Roman letters; the first Swahili newspaper *Habari ya Mwezi* was published at Magila in 1895. It has grown to become one of the principal languages of East Africa, spoken by more than 30 million people; it is the official language in Tanzania, and recognised as a secondary language in Kenya and Uganda.

Consonants

Bb Cc Dd Gg Hh
Jj Kk Ll Mm Nn Pp
Rr Ss Tt Vv Ww
Yy Zz

Vowels

Aa Ee Ii Oo Uu

Further notes

- Swahili does not need **F**, **Q** or **X**, but sometimes uses **R** to spell words of a European language which use the letter, such as 'regulation'. (However, as in a similar confusion among North East Asians, many Swahili speakers cannot distinguish between the European **R** and Bantu **L** sounds.)
- Swahili is not a difficult language to typeset by computer. All the required characters are already in the basic character set provided.

Tswana

Tswana is a southern member of the **Bantu** subgroup of the **Niger–Congo** language family, related to Sotho and Venda. It is spoken by about 3·3 million people in south east Africa, especially in Botswana where it is the principal language.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Xx Yy Zz

Vowels

Aa Ee Êê Ii Oo Ôô
Uu

Further notes

- At the time of writing it is not clear whether all of the standard Latin consonants shown are actually required for Tswana, as the information was obtained from a reference source which listed only the language's *extended* Latin font requirements.
- Tswana is extremely easy to typeset with existing DTP or word-processing programs for Windows or Macintosh computers. The accented vowels **Ê** and **Ô** – shown in green above – are part of the standard font encoding for these computers.

Twi

Twi – also known as Akan, Fante or Ashanti – is a member of the **Kwa** group of West African **Niger–Congo** languages, and is spoken by 6–7 million people in Ghana and Côte d'Ivoire. The orthographic system shown below was developed by the Ghana Bureau of Languages.

Consonant range and sequence

p t k ky b d g gy f s h hy m n ṇ ṇg ñ ñy ññy
ñỹ ñg r w w̆ tw dw ḍw gu hw ṇw ñw̆ ñũ ṇh l v

Consonantal letterforms

Bb Dd Ḍḍ Ff Gg Hh Kk
Ll Mm Nn Ññ Ṇṇ Pp Rr
Ss Tt Vv Ww W̆w̆ Yy Ỹỹ

Vowels

Aa Ạạ Ãạ̃ Ee Ẹẹ Ẽẽ Ɛɛ
Ɛ̃ɛ̃ Ii Ĩĩ Ɔɔ Ɔ̃ɔ̃ Oo Ọọ Õõ
Uu Ũũ

Further notes

- Like some other West African languages, Twi has a relativistic system of three tones ('tone terracing'), but no tone markers are used in the writing system.
- The *tilde* mark (~) indicates nasalisation of a vowel or consonant. In common use, nasalised vowels are usually not marked, but all possible combinations are shown above. Those letterforms marked in green can be achieved with the standard Mac/Windows character set (ñ+õ).
- Some of the letterforms shown could be achieved as composed characters using a typesetting system such as TEX – especially if a round 'underdot' can be substituted for the vertical understroke glyph. However, it is clear that a special font would be needed anyway for the two 'open' vowels.

Wolof

Wolof is a member of the **West Atlantic** sub-group of the **Niger–Congo** language family. It is spoken by about 2·6 million people in Senegal. (The Senegalese scholar Cheikh Anta Diop has claimed controversially that Wolof is closely related to ancient Egyptian.)

Consonants

Bb Cc Dd Ff Gg Ḥḥ
Hh Jj Kk Ll Mm Nn
Pp Qq Rr Ss Tt Ṭṭ
Vv Ww Xx Yy Zz

Vowels

Aa Ââ Ăă Ee Èè Éé
Ii Oo Uu

Further notes

- At the time of writing it is not clear whether all of the standard Latin consonants shown are actually required for Wolof, as the information was obtained from a reference source which listed only the language's *extended* Latin font requirements.
- There are two solutions for typesetting Wolof. If a modified font is available, any word-processing or DTP program can be used to typeset it. Alternatively, one could use a typesetting system which can place an accent above or below any arbitrarily chosen character (the 'composed character' approach). Any implementation of TEX could do this with ease, but complex document layouts such as leaflets and newsletters are hard to do in TEX.

Xhosa

Xhosa is a member of the **Bantu** sub-group of the **Niger–Congo** family of languages, and is spoken by about 6–7 million people in the north and north-east of the Republic of South Africa. The Xhosa people merged lineages with some neighbouring Khoe peoples, and as a consequence Xhosa has absorbed some of the Khoisan 'click'-consonant sounds.

Consonant range and sequence

b bh d f g h hl dl j k kh kr l m n ng ny p ph r
rh s sh t th tsh ts ty dy v w y z + clicks c q x

Consonantal letterforms

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Xx Yy Zz

Vowels

Aa Ee Ii Oo Uu

Further notes

- Because it has been possible to 'recycle' three latin consonants into click-sound representations, Xhosa can be typeset with completely standard DTP or word-processing software.

Yoruba

Yoruba, spoken by 20 million people in southern Nigeria and Benin, is one of the principal **Bantu** languages in the **Benue–Congo** subgroup of the **Niger–Congo** family of languages.

Consonant range and sequence

b m f t d n s l r s j y k g p gb w h

Consonantal letterforms

Bb Dd Ff Gg Hh Kk Ll
Mm Nn Pp Rr Ss Ṣṣ Tt
Vv Ww Yy

Vowels

Aa [Àà Áá] Ee [Èè Éé] Ẹẹ [Ẹ̀ẹ̀ Ẹ́ẹ́]
Ii [Ìì Íí] Oo [Òò Óó] Ọọ [Ọ̀ọ̀ Ọ́ọ́]
Uu [Ùù Úú] [M̀m̀ Ḿḿ] [Ǹǹ Ńń]

Further notes

- Yoruba is a tonal language, and acute or grave accents are used over vowels to indicate tone. When the letters **M** and **N** are used as nasalised vowels, they too may require the placement of tonal accents. The vowel-accent combinations already provided as combined glyphs in the Windows and Macintosh standard character sets are shown in green.
- The precise form of the 'underdot' character varies in use. Sometimes a circular dot is used, sometimes a vertical stroke, and sometimes the vertical stroke intersects with the profile of the character above.
- All the letterforms shown could be built up as composed characters using the control codes of a type-setting system such as TEX – especially if a round 'underdot' can be used, accessed with the **\d** code. However, it is difficult to create complex layouts for leaflets and newsletters with TEX, so a special font that supplies all of the combinations of letters and accents as prebuilt letterforms may be preferred.

Zulu

Zulu is one of the most southerly members of the Bantu sub-group of the Niger–Congo family of languages, and is spoken by about 8–9 million people in the east of the Republic of South Africa (KwaZulu-Natal Province). In common with some other extreme-south Bantu languages such as Xhosa, Zulu has absorbed 'click'-consonant sounds from neighbouring Khoisan languages.

Consonants

Bb Cc Dd Ff Gg Hh
Jj Kk Ll Mm Nn Pp
Qq Rr Ss Tt Vv Ww
Xx Yy Zz

Vowels

Aa Ee Ii Oo Uu

Further notes

- The Zulu language has a very rich range of consonants. Some of the 'un-clicked' Zulu consonants are written as latin digraphs or trigraphs: e.g. **sh**, **tsh**, **kh**, **ng**.
- The three 'click' consonants in Zulu are produced by an implosive separation of the tongue from various parts of the palette – from just behind the teeth, the middle of the hard palette, and the soft palette at the back of the mouth. The letters **C**, **Q** and **X** are used to represent the simple form of these clicks, which can also be aspirated (**ch**, **qh**, **xh**), nasalised (**nc**, **nq**, **nx**) or voiced (**pc**, **gq**, **gx**).
- Because it has been possible to 'recycle' three latin consonants into click-sound representations, Zulu can be typeset with completely standard DTP or word-processing software.