New Review of Hypermedia and Multimedia,
Vol. 11, No. 2, December 2005, 247–255

Taylor & Francis
Taylor & Francis Group

Technical Note

Review of script displays of African languages by current software

QUINTIN GEE*
Learning Technologies Group, Electronics and Computer Science, University of Southampton, SO17 1BJ, Southampton, UK

All recorded African languages that have a writing system have orthographies which use the Roman or Arabic scripts, with a few exceptions. While Unicode successfully handles the encoding of both these scripts, current software, in particular Web browsers, take little account of users wishing to *operate* in a minority script. Their use for displaying African languages has been limited by the availability of facilities and the desire to communicate with the 'world' through major languages such as English and French. There is a need for more use of the indigenous languages to strengthen their language communities and the use of the local scripts in enhancing the learning, teaching, and general use of their own languages by their speaking communities.

1. Introduction

All African languages use the Roman script written left to right, or the Arabic script, written right to left; however, mixtures of both styles rarely occur (Bendor-Samuel 1996). There are some exceptional scripts used in Africa in addition to the two mentioned, which are Bamun, which has been used to serve a population of 215 000 in the Cameroon (SIL a 2005, Battestini 1994) and the Ethiopic script used for Amharic (17 m speakers), Ge'ez (solely for religious purposes), Tigrinya (3 m) and Oromo (9 m) (although this was Romanized in 1991) (SIL 2005b).

Until the advent of the worldwide Web, it could reasonably be assumed that the first target of minority language users was to operate a word processor in their own language/script. It is clear (for example in Mafu 2004) that browser use is on at least equal footing with word processing, and has in many cases overtaken it. Output to publishing devices such as files and printers comes a distant third, and poses little problem, as we shall see later in this paper.

If you want to use a word processor or browser, you currently need to use one that operates in a popular and widespread language such as English, Spanish, or Chinese. The availability of software tailored for use by speakers of other than the 'popular' languages and non-standard scripts has been minimal for a long time, although the process of change is beginning to accelerate.

*Email: qg2@ecs.soton.ac.uk

New Review of Hypermedia and Multimedia
ISSN 1361-4568 print/ISSN 1740-7842 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13614560500350818

There are two elements that concern us with word processing in minority languages. First is the ability to handle the extra characters that appear in the orthography. Second is the ability of a vernacular speaker to operate the word processor in their own language.

2. Orthographies

Unicode (Unicode 2005) handles Roman and Arabic scripts well and also orthographies based on them, whether by direct representation (e.g. the š in Venda is already included as 0161) or by composition of a diacritic with an existing character (e.g. ọ in Yoruba from 05BD and the letter o). Unicode also encompasses Ethiopic script (Unicode Ethiopic script 2005).

The typing of Arabic also causes no problems, and the script has a long history of adaptation to represent languages for which it was not originally designed with suitable representation of 'alien' phonemes (Kaye 1996).

A problem might occur when characters are being used that do not occur in conventional orthographies. It appears that the groups dedicated to creating new orthographies are mindful of this. Whenever it is a government institution such as the South African National Language Service (SANLS 2002a), the Kenyan Ministry of Culture, or the Nigerian Ministry of Culture, there is usually a clear policy laid down that precludes the addition of new characters, pressing for the use of existing letters, with the use of existing diacritics where necessary. Sometimes, the sheer number of languages poses considerable difficulties, as it does in Nigeria: 'the linguistic problem is also important, as most indigenous languages do not have developed orthographies' (Oluge 1987).

One of the largest originators of new orthographies worldwide would seem to be Wycliffe Bible Translators (WBT 2004) and its associated organization, the Summer Institute of Linguistics (SIL 2005c). These organizations are dedicated to identifying and formalizing unrecorded languages, with a view to commencing Bible translation in them, where necessary. Prior to commencing such a publication, much work needs to be done to develop a new orthography, to have it adopted and to build an ownership of it by the community of mother-tongue speakers. This includes literacy materials of all sorts and for all levels of learner. In doing this, these organizations are guided by liaison with national governments and universities as well as the target users.

It is understandable that new characters are introduced with apprehension, and the policy on orthography is similar to that used by government institutions mentioned above (Lojenga 2001). Policy varies, of course, across the world, but generally the emphasis is on using a local script where it is known and politically acceptable, and enhancing that to take into account the elements of the target language.

Turning to Africa, despite the UNESCO 1978 conference suggesting an 'African Reference Alphabet', this has not been taken up by any African

governments that we are aware of. However, more recent endeavours by UNESCO have centred on the theme of making governments aware of the impending loss of their minority languages, which can be a valuable cultural unifier, and the need for multilingual resources to be made available, thus creating ownership of the scripts (Robinson and Gadelii 2004) as well as the language. While this is all well in theory, many African governments have rather more pressing needs than conservation of minority languages. Thus, it may be more a question of political will and lack of language planning. In some cases, governments are distinctly hostile. Mafu (2004) reports that the 'Tanzanian central government continues to be unenthusiastic about promoting indigenous languages (which amount to about 130 altogether)'.

3. Origination of ethnic textual material

Several African-language word processors are available on the market, or have been created as research projects (Gee 1990, 1991). These are essential to get the minority-language speakers to take ownership of the source materials so that they can originate them themselves.

One difficulty that occurred during the preparation of the Bemba word processor was the lack of technical terms for many 'typesetting' concepts. While the word for 'character' was easy to come by, even though it was a neologism, the term for 'left justify' caused some problems. There was no single word for left or right, and we had to make do with a phrase that, loosely translated, means 'move to the left-hand side'. This crossed half the screen on a pull-down menu! After 3 months' use of this word processor, the end users were questioned about improvements that could be made, and they suggested a two-word idiom that could replace the cumbersome phrase that had been implemented. They were happy with this evaluation, and to be consulted and felt that they 'owned' the product.

The reason that more developments of this nature have not taken place is largely due to the requirements for ongoing support, whether it is of errors and improvements in the word processor, or in responding to genuine queries on its use. There is room for local academic institutions to undertake such a service. Anderson (2004) notes that her Script Encoding Initiative at Berkeley uses, and continues to seek, volunteers because 'While the business interests have been actively behind much of the character encoding. . . . advocates for the lesser-known scripts have not had a similarly strong presence among the Unicode Consortium membership'. We suspect that major firms are not willing to invest in 'small' markets. For example, the word-processing programs for Yoruba and kiSwahili cited herein were created and paid for by academics and small commercial firms.

Some work has been done on the most populous language in Africa, Yoruba (19 m speakers). Paradigm's Lingua claims to work in Yoruba (figure 1), but we have yet to witness a demonstration of this (Paradigm 2003); likewise for a kiSwahili word processor, although the spell checker appears to work only under Unix (Jambo Open Office 2005).

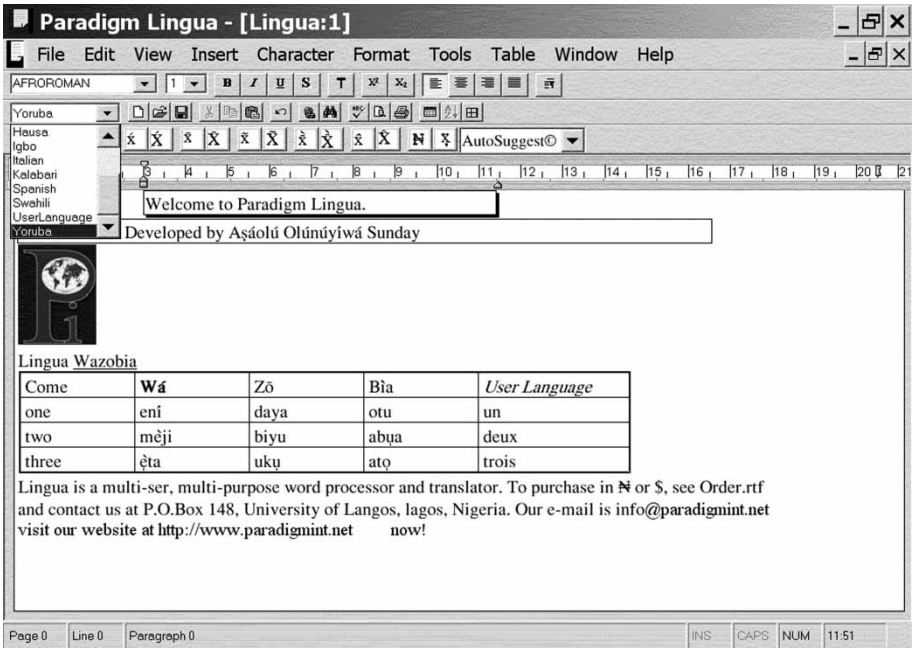

Figure 1. Lingua 2.4 in Yoruba (from Paradigm 2005).

4. Printing of textual material

Laser jet, ink jet, bubble jet, and most other modern computer printers can efficiently handle the required fonts, diacritics, and other symbols.

However, the storage and transmission of material, for example to a printing establishment is another matter. This is due to the fact that the printing organization may not have equipment and software that are compatible with the originator's coding system.

The text stream may be 'mangled' by the transmission process, which can take into account content and may view Unicode bytes as spurious or unrecognizable characters, and replace them with something else. Anyone who has tried emailing accented characters will have experienced this. With the more unusual combinations and with minority scripts, the current versions of mailing applications are inadequate.

There are standards in place to ameliorate this problem, Unicode for one, but compatibility must always be checked with the destination files before this can be said to work reliably.

5. Screen display of material

From the display of minority languages and their scripts, we assume that most modern computer screens have the facility to display the required fonts. However, it must not be assumed that the end user will necessarily have a WYSIWYG screen available, and may be working with a character-only

screen. Character-only screens come in a bewildering variety—some may be able to display accented characters, some may not. Some can manage Arabic, Hebrew, or Chinese character set, but not necessarily mixtures of these. The use of these screens in any case will eventually be diminished.

Even with modern screens, the ability of operating systems to handle the different encoding sets for Unicode is sporadically implemented, although right-to-left working and two-dimensional working can be handled by many applications.

Turning now to the delivery of minority-language material across the Web, most modern browsers are Unicode-enabled, which means they should, in theory, have no trouble in displaying to the end users the orthographies that we have been discussing, and this is where the greatest benefit for minority languages lies. However, the reality to date is rather mixed (see Unicode I18N Tests, for example.) It may be necessary to set the encoding system for a particular script using a menu function.

However, as we have said above, the ability to *operate* in the minority language is the challenge. Several African-language browsers are available on the market, such as for Afrikaans using Opera (Opera 2005), Firefox (Mozilla Organization 2005), and MS Internet Explorer (Microsoft 2005), and for isiXhosa using Firefox (Bailey 2005). These are not difficult to provide, and require little support since, for the most part, they are display mechanisms only. However, it is confusing to the user that an error message is still displayed in English when using a vernacular browser.

There is a South African team named translate.org.za, which is a non-profit organization dedicated to producing free and open-source software. The Translate Project started in 2001 with the vision of providing free software translated into the 11 official languages of South Africa. Free

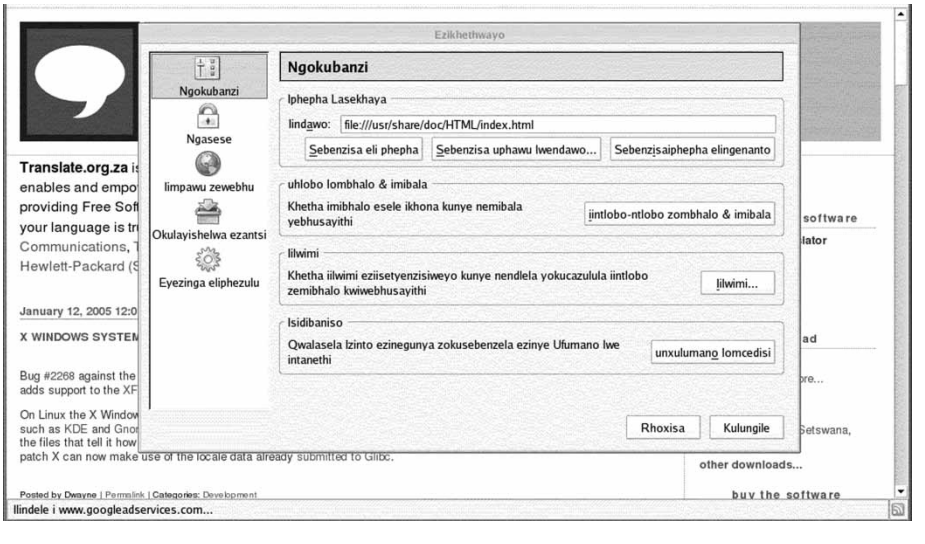

Figure 2. Firefox in isiXhosa (from translate.org.za 2005).

software in each language is both empowering and enabling (Bailey 2001). Progress has been sporadic but it is being made on several fronts. As a cooperative venture, it relies on volunteer work in translating the menus and dialogues into the target languages, and as we have pointed out, this can be problematic if the technical terms do not exist in the language. This is why much reliance is placed on the sterling work done by the terminological services (SANLS 2002b).

6. The need for these components

We take South Africa as an example, with its 11 official languages. Of these, English is spoken by 45% as a lingua franca, and Afrikaans by 62% likewise. The number of speakers of the remaining languages varies from 8.5 m in Zulu, to 100 000 for Xitsonga and Tshivenda. Clearly, on a world scale, the 10 listed in table 1 can be considered minority languages.

We can see that these languages are well under-represented at the moment as regards being able to operate in them using application software. Work is continuing in providing spell checkers in all these languages, and the South African National Language Service (SANLS 2002b) seems to have produced acceptable add-ons for MS Word which will be of great use for text origination.

It should also be noted that Microsoft Corporation has adopted a methodology that allows for a standard implementation of their office products to be supplemented by a Multilingual User interface pack that allows a wide range of languages to be installed as user interfaces (Microsoft 2005). Unfortunately, this list does not include any African languages as yet. However, the well-established Centre for Text Technology at the North-West University (Sentrum vir Tekstegnologie 2005) is the vendor of spelling checkers for five South African languages—Afrikaans, isiXhosa, isiZulu, Sesotho sa Leboa, Setswana—to Microsoft (Prof Gerhard van Huyssteen, private communication 22 April 2005).

Table 1. Current resources available in some African languages.

As at Jan 2005	Browser	Word processor	Spell checker	Web pages
Afrikaans	Yes	Yes	Yes	Yes
isiZulu	Yes		Yes	Few
seSotho	Yes		Yes	
siSwati			Yes	Few
isiXhosa	Yes		Yes	
Setswana			Yes	Few
isiNdebele			Yes	
Sepedi			Yes	
Tshivenda			Yes	
Xitsonga			Yes	

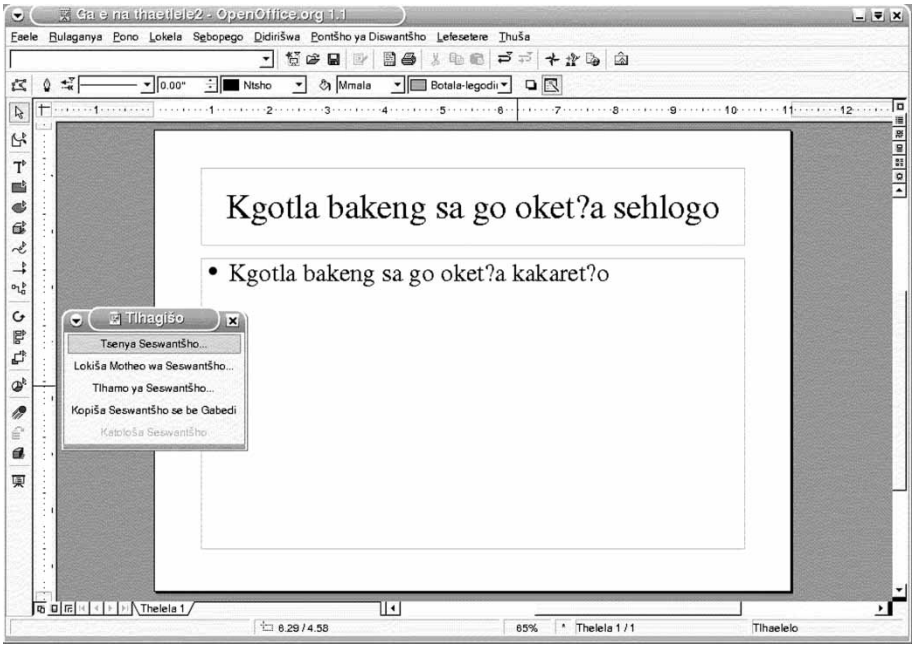

Figure 3. Open Office in North Sotho (from WaZoBiaSOFT).

7. Conclusion

We have shown that the display and printing of minority languages is not a difficult problem. However, the access to facilities for generating and making materials available in Web page or printed format remains primitive and generally unsung.

Current Web browsers and applications are only just beginning to take account of scripts that are serving other than the majority of users. Their use for displaying African languages has been limited by the availability of people with knowledge of the language/script and the computer competencies to write suitable interfaces or adaptations to existing software or for the creation of new products.

There is a need for more use of the indigenous languages to perpetuate their language communities and the use of their scripts in enhancing the learning, teaching, and general use of their own languages by their speaking communities. Unfortunately, in many of the countries that in theory could benefit from this, few people in the key groups are literate, own a computer, are computer-literate, or even live in an area that has electricity. Mafu (2004) notes that this is a big problem in Tanzania. Most people have to go to a cybercafé to access the Internet. How can we help them break this 'vicious cycle?'

Although, within Africa, the larger language groups (Yoruba, Swahili, Hausa) are not among the most endangered minority languages, the continent as a whole represents a sizeable population whose linguistic

needs are largely overlooked by most development in the current software market.

A start could be made by the ownership and creation of Web pages that display the minority script. In addition, we feel there is a need for organizations with the expertise in software adaptation to work closely with appropriate language informants in some of the target languages to generate momentum in this area, as is being done at a volunteer level by translate.org.za. A Nigerian view is that first 'the educational sector should encourage and embrace indigenous language software' (Asaolu 2003).

References

D. Anderson, "The Script Encoding Initiative", *SIGNA*, 6, April 2004, pp. 1–12.

O.S. Asaolu, "Language software development: principles, processes and prospects for Nigerian education", in *An Introduction to Information and Communication Technologies in Education*, O. Owhotu, Ed., Lagos: University of Lagos, 2003.

D. Bailey, *Open Source Software Translation Project*, 2001. Available online at: translate.org.za (accessed April 2005).

D. Bailey, *Xhosa Translate-athon Completes Firefox Translation*, January 24, 2005. Available online at: translate.org.za (accessed April 2005).

S. Battestini, *African Writing and Text*, Ottawa: Legas, 2000. ISBN 1894508068.

J. Bendor-Samuel, "African languages", in *The World's Writing Systems*, P.T. Daniels and W. Bright, Eds., Oxford: Oxford University Press, 1996, pp. 689–691.

Q.H. Gee, "A word processor in Nyanja", *Comput. Forum*, 1990, 1(1), pp. 11–17.

Q.H. Gee, "A Bemba word processor", in *Proceedings of 1st National Computer Conference*, November 1991, Maputo.

Jambo Open Office, "Tuxpaint, kiSwahili Unix spell checker", 2005. Available online at: www.cis.yale.edu/swahili/ (accessed April 2005).

A.S. Kaye, "Adaptations of Arabic script", in *The Worlds Writing Systems*, P.T. Daniels and W. Bright, Eds., Oxford: Oxford University Press, 1996, pp. 743–762.

C.K. Lojenga, *Develop an Orthography*, Summer Institute of Linguistics, 2001. Available online at: http://www.sil.org/lingualinks/literacy/developanorthography/contents.htm (accessed April 2005).

S. Mafu, "From the oral tradition to the Information Era: the case of Tanzania", *Int. J. Multicult. Societ.*, 6, pp. 99–124, 2004.

Microsoft Corporation, *Office 2003 MUI Pack Languages and Corresponding Locale Identifiers*. Available online at: office.microsoft.com/en-gb/assistance/HA011402211033.aspx (accessed April 2005).

Mozilla Organization, *Download Firefox Language*, 2005. Available online at: www.mozilla.org/products/firefox/all.html. (accessed April 2005).

B. Oluge, "National language and national development", in *Proceedings of the Congress of the Language Association of Nigeria*, 1987.

Opera. Available online at: www.opera.com/download/languagefiles/ (accessed April 2005).

Paradigm, *Lingua 2.4 for Yoruba*, 2003. Available online at: www.paradigmint.net (accessed March 2005).

C. Robinson and K. Gadelii, Writing unwritten languages, 2004, UNESCO. Available online at: portal.unesco.org/education/en/ev.php-URL_ID = 28300&URL_DO = DO_TOPIC&URL_SECTION = 201.html (accessed April 2005).

SANLS, *Towards a Human Language Technologies (HLT) Strategy for South Africa*, South African National Language Service, 2002a. Available online at: www.dac.gov.za/about_us/cd_nat_language/language_planning/HLT_brochure/english.htm (accessed April 2005).

SANLS, *Spell checkers for African Languages*, South African National Language Service, 2002b. Available online at: www.dac.gov.za/about_us/cd_nat_language/home.htm (accessed April 2005).

Sentrum vir Tekstegnologie, North-West University, Potchefstroom. Available online at: www.nwu.ac.za/ctext (accessed April 2005).

SIL, *Ethnologue 14 and Bibliography Information on BAMUN*, Summer Institute of Linguistics, 2005a. Available online at: www.ethnologue.com/show_language.asp?code = BAX (accessed April 2005).
SIL, *Ethnologue 14 and bibliography information on AMHARIC*, Summer Institute of Linguistics, 2005b. Available online at: www.ethnologue.com/show_language.asp?code = AMH (accessed April 2005).
SIL, *SIL International: Partners in Language Development*, Summer Institute of Linguistics, 2005c. Available online at: www.sil.org (accessed April 2005).
Unicode Code Charts, Unicode Consortium. Available online at: www.unicode.org/Public/4.1.0/charts/CodeCharts.pdf (accessed April 2005).
Unicode Ethiopic Script, Unicode Consortium. Available online at: www.unicode.org/charts/PDF/U1200.pdf (accessed April 2005).
Unicode I18N Tests: Inline bidi Markup 1, Unicode Consortium. Available online at: www.w3.org/International/tests/sec-inline-bidi-1 (accessed April 2005).
WBT, *Current state of translation*, 2004, Wycliffe Bible Translators. Available online at: www.wycliffe.org/annualreport.html (accessed April 2005).