

African Languages and Information and Communication Technology: Localising the Future?



Donald Osborne

1. Introduction

The dual processes of internationalisation and localisation of information and communication technologies (ICT)¹ are still felt incompletely and unevenly in Africa. Situated on the disfavored side of what is commonly referred to as the “digital divide,”² Africa has not benefited directly from internationalisation of ICT as much as it might, largely due to low connectivity and lack of access, but also to the fact that some aspects of internationalisation are not widely available or well understood. Africa also has not yet seen much activity in the matter of localisation for its languages and cultures even though this is vitally important for the future use and utility of ICT in the region.

As a continent with perhaps 2000 languages and various special script requirements, Africa stands to benefit greatly from internationalisation and localisation of ICT,³ but it lacks the means and strategies to take advantage of them. Moreover, the majority of discussions about “bridging” the digital divide in Africa and how to make ICT work for its people pay little attention to these processes and make only passing mention of the language dimension. This, even though a majority of the population either does not speak or does not master the originally European tongues that both serve as the official languages in their countries and are dominant in computing and the internet.⁴

The future development of Africa, and indeed all world regions, depends increasingly on the appropriation and appropriate application of ICT. These in turn cannot be fully achieved without taking full advantage of internationalisation in order to progressively implement localisation of the technology in order to respond to its cultural and linguistic, as well as socio-economic, needs and realities. Localisation emerges as a key to successful adoption and use of ICT and thus also as an indispensable consideration as Africa looks to the future.

Localisation, of course, has several definitions relating to the adaptation of computer applications and/or the content of computing to the linguistic and cultural realities of a particular country, region, or national community. The sense used here focuses on practical linguistic aspects and addresses both software and content. Nevertheless, localisation is understood in the context of emerging discussions of “ethnocomputing” (Tedre et al 2002) and is captured in Daniel Yacob’s (2004) broad defini-

tion: “the transfer of cultural consciousness into a computer system, making the computer a natural extension of the society it serves.”

This paper therefore focuses on the status of African language⁵ localisation of ICT and outlines elements of a strategy to facilitate this process. It underlines the importance of two overlapping areas of action – content and tools – and also of prerequisite factors like wider use of Unicode, to successful localisation in, for and by Africa.

II. African Languages and ICT: On the Periphery of the Information Society

A. Importance of the Issue

There are several reasons why African language use in ICT is of interest.

- 1) So long as a language is spoken and used in other spheres of activity, it is assumed that it is worth providing the opportunity for it to be used in various ways by its speakers in computer and internet technologies. This thinking could be both framed in terms of “linguistic rights” and taken as a practical observation that opportunities for expression and learning are favored by providing means to work with a powerful technology in one’s first language.
- 2) Such an opportunity becomes a critical concern because although African languages are themselves important vehicles for communication and the creation of knowledge, there is little organized educational activity beyond small-scale programs for adult basic literacy and a limited amount of first language instruction at the primary school level in some countries. ICT in African languages could be important in a range of activities favoring dissemination of information, enhancement of skills, and generation of knowledge.
- 3) Reliance almost exclusively on English, French, and Portuguese for the transmission of new knowledge (see Enguehard and Mbodj 2003) puts people who are not skilled in these languages, and arguably the entire societies of which they are a part, at a disadvantage.
- 4) The issue certainly goes further, as it is legitimate to ask what sort of future there is for languages that are not used actively in ICT.

¹ ICT is used here mainly in the narrower sense of computers and the internet, what has also been called simply “information technology” (IT) and also the “new” ICTs. In wider usage, ICT can also cover radio, telephones, and other electronic media, the boundaries among which are becoming ever less clear in the wake of rapidly changing technology.

² The uneven distribution of computers and internet access in favor of certain regions and groups. Various statistics on Africa like low numbers of telephone connections, lack of electrification and high illiteracy that have long characterized the continent are now joined, unsurprisingly, by low indicators for connectivity and access to ICT.

³ Jacques Ngangala Balade Tongamba (2001) sees ICT as presenting an opportunity “without precedent” for Africa, in part because of how it can facilitate use of African languages.

⁴ Clinton Robinson (1996) cites figures as high as 90 percent in some countries of sub-Saharan Africa who do not speak the official language of their country (English, French or Portuguese).

⁵ African languages are defined here as the modern languages indigenous to the continent plus major historically indigenized languages (such as Arabic, Malagasy and Afrikaans).

Lack of usable multilingual capacities on computers and the internet in Africa – whether because of old operating systems, lack of necessary fonts and input methods, or system managers insufficiently trained in aspects of multilingual computing – effectively limits many people's access to full use of the technology.⁶ This includes the capacity to display and share text with extended characters or non-Latin scripts – precisely the basic problem that Unicode is intended to resolve – as well as more sophisticated adaptations and uses of the technology, from software localisation to manipulations and transformations of content.

Localisation of the technology therefore involves the entire orientation of the technology with regard to language capacities. In Africa, however, it emerges as both a casualty of language problems (illiteracy, language and education policies that in principle or through incomplete application disfavor African languages) and as part of the solution to those same problems (by opening new possibilities for use of the languages and dealing with multilingual issues).

In the following, the current state of African languages and ICT is briefly surveyed and factors affecting their level of use are examined.

B. Brief Overview of the Current State of African Languages and ICT

The relative level of use of African languages in computing and on the internet is hard to gauge but important to attempt to characterize. To begin with, it seems clear that African languages are not yet widely represented in computing applications or internet content. We can deduce this for instance from the very small amount of software localised even for major African languages and the infrequency and character of such web content as one does find in African languages.

This situation obviously arises from the underlying sociolinguistic, language policy, and educational contexts, though in this paper these will not be explored in depth. However, it is worth noting that computers and the internet, like formal educational systems a century earlier, have been introduced and disseminated as more or less monolingual media relying on one or another European language. This is a reflection of both the dominance of the languages inherited from colonization in software and internet content, and the use of these languages by those people in Africa most likely to have access to the technology.

A quick overview of web content, use in e-mail, use in non-internet applications, and localisation of software and web interfaces, helps to elucidate the situation.

1) The Web

African languages are represented on the web, but not prominently as media of communication. The actual level of use is emerging as a topic of discussion.

It is easy to get the impression that African language content is rare. To begin with, simple surveys of web content by lan-

guage that relied on search engines unsurprisingly did not find enough in African languages – even the ones most widely spoken – to rank them as high as some minority European languages with relatively few speakers.⁷

More focused or systematic surveys have come up with more interesting results. For instance, an informal survey done in Tanzania in 2001 as part of a larger report for the Swedish International Development Agency estimated that ten percent of websites with a Tanzanian focus had at least some Swahili content (Miller Esselaar Associates, 2001), but most of the sites did not have majority content in the language. A more extensive study by Diki-Kidiri and Edema (2003) found a significant number of sites that treat African languages in one way or another, but these generally have minimal content in the languages themselves. In effect, a large proportion consists of sites about African languages, including online dictionaries and instructional pages.

On the other hand, Van der Veken and de Schryver (2003), using a different search methodology and statistical extrapolation, suggest that there may actually be more African language web content than we realize.

A useful schema to attempt to understand this evolving situation is that proposed by Ballantyne (2002) for analyzing content in terms of two parameters: “expression,” or whether the content is of local or international origin, and “application,” or whether the audience is local or international. Much web content relating to Africa, even concerning African languages, and whatever the origin, has an external and largely non-African audience, and so would logically tend to use languages understood internationally. Also, much of the web content with intended local application originates from outside of Africa, where production of content in languages other than English, French or Portuguese is not an easy option. What is at issue here is in part the extent to which content originating outside of Africa is adapting to use of African languages, but more importantly the evolution of content of local origin and local application that can naturally and efficiently use those languages.

2) E-mail and E-mail Lists

E-mail has long been a significant use of the limited internet connectivity in Africa.⁸ By its nature it is harder to track the contents but there is other information that can be used to get an indication of the use of African languages for this purpose. For instance, at one point there were two web-based e-mail services that provided for composition in several African languages: Africast.com and Mailafrica.net (though the latter no longer seems to function). In addition, recent years have seen the setting up of a number of e-mail fora in which much or most of the traffic is in one or another African language. For instance there are several Hausa and Swahili lists in which these, probably the most widely spoken indigenous tongues on the continent, are the primary languages of communication,⁹ and Van der Veken and de Schryver (2003) found fora in Hausa, Somali, and Lingala.

⁶ This is a broader definition of access than that commonly used in discussions of the digital divide. However, there are some sources that offer multiple definitions of access. For instance, Telecommons (2000) discusses “physical access” to ICT infrastructure and applications, and “soft access,” which we define as software and applications which are designed to enable rural African users to utilize ICTs for their own needs and uses once the physical access has been established.” The organization Bridges.org goes further to define twelve dimensions of what it calls “real access,” of which “relevant content” mentions language (see <http://www.bridges.org/digitaldivide/realaccess.html>).

⁷ A simple survey of websites by language done in 2000 by Vilaweb, the website of a Barcelona newspaper (Pastore 2000), listed no African languages among the 31 ranked and showed many more pages for languages such as Basque and Slovenian than for any language in Africa surveyed three years later by Diki-Kidiri and Edema (2003). A follow-up to the Vilaweb survey which ranked the top 48 languages on the web found Afrikaans 42nd after the abovementioned languages, and Swahili last following, among others, Frisian and Faeroese (Mas 2003).

⁸ There was even a “web-page by e-mail” service hosted for several years by Kabissa.org, in recognition of the fact that many people in Africa could not access the web but did have limited e-mail access.

⁹ These include several that use the free Yahoogroups service: “Kiswahili” <http://groups.yahoo.com/group/Kiswahili/>; and for Hausa language: “Finafinan_Hausa” groups.yahoo.com/group/Finafinan_Hausa/, “HausaDaHausawa” groups.yahoo.com/group/hausadahausawa/, “Marubuta” groups.yahoo.com/group/Marubuta/, “Matasa” groups.yahoo.com/group/Matasa/. The scholarly-oriented H-Swahili list also has traffic in the Swahili language: <http://www.h-net.msu.edu/~swahili>

3) Non-Internet use of African languages in computing

It is harder still to attempt to quantify the degree to which African languages are used in the content of computer applications in Africa, for instance on word processors for the production of printed documents. Among users in this category have been specialists in African languages and linguists in Africa and beyond. Certainly the publication of books and news in African languages is computerized using specialized publication software, but use on public, office and personal computers is less visible.

As for popular usage, a glimpse of African language use on computers in a Senegalese telecenter is given in a brief article (Elder 2002) that mentions use of Pulaar and Wolof. Also in Senegal, a local non-governmental organization, ANAFA,¹⁰ has been doing computer training (including basic literacy) in national languages. Beyond such anecdotal evidence however, there are apparently no surveys of such non-internet use.

4) Software and web-interface localisation

Localisation of software and web-interfaces for African languages is getting increasing attention. There have been efforts for localising software on smaller scales for several years,¹¹ but these are becoming more numerous and the level of activity is increasing.

The recent agreements concluded by Microsoft Corporation to localise for Swahili (Anyanzwa 2004), Wolof (APS 2004), and Hausa, Yoruba, and Igbo (*This Day* 2004) indicate the importance that the company is attaching to the issue. The announcement early last year (Microsoft Corporation 2004) concerning its increased work on localising its software, including for Africa, follows on preparation that apparently goes back some years.¹²

Localisation of open-source software is also emerging as an important factor. A localisation project for South African languages, Translate.org.za,¹³ has received a fair amount of attention, announcing completion of software in Zulu, Sepedi, and Afrikaans last year. In late 2004 another open source localisation project, in Uganda, released a web browser in Luganda (Otter 2004), and a Tanzanian-based project, Kilinux,¹⁴ completed a Swahili-language word processor. Other projects are in the works though information on them is sometimes not readily available.¹⁵

There are also some African projects that have produced software for composition in African languages but without language localised commands, including several based in or focusing on Nigeria - K_nyin,¹⁶ Afára,¹⁷ ALT-I,¹⁸ and Paradigm.¹⁹

As for web-interfaces, the popular search engine Google has had a program for localised versions that already has several

African language versions translated by volunteers. Another example is a "V-webmail" e-mail interface which was recently localised for Swahili.²⁰ There may be more of this sort of language localisation going on than is apparent,²¹ but one concern with easier localisation such as that offered by Google is the quality of the translations done by volunteers.

C. Factors Affecting the Level of Use of African Languages in ICT

Despite the examples cited in the previous section, African language use in ICT appears to be relatively marginal in Africa. Why is that? First of all, the factors that define the digital divide also tend to minimize the potential for African language use in ICT. Connectivity is centered on cities and towns where official languages – the same languages that are dominant on the internet – may be more widely spoken. In addition, only people with means and education, who are also more likely to have facility in use of the official languages, can access computers and internet connections. The digital divide therefore is arguably more localised than bridged, being replicated on national and local levels along the lines of older and deeper social, economic, and linguistic divides.²² In effect there are a number of interrelated factors that disfavor languages that are not associated with wealth and power, even before one considers possible issues specific to particular languages.

The other reasons why African languages are not more used in ICT can be grouped into two categories: motivation (the will to use them) and structural factors relating to the languages that affect access to and use of ICT (the means to do so).

1) Motivation

With regard to the motivation, it might be said that while those people who have access to computers in Africa tend to be people who are educated in and socialised to some degree to use the official languages and thus less likely to actively seek to use their first languages in ICT, those people who use their first languages but not the official language tend not to be in a position to do much in this area even if they wanted to. Of course the issues surrounding use of the vernacular in Africa are complicated by factors such as status and attitudes towards what is indigenous (including language) vis-à-vis languages that are seen as providing more economic opportunity. It becomes easy then for all to assume that ICT is for European languages. This is all the more so since for many internet users, the medium is primarily a way to access information from other parts of the world.

Another factor in the case of web content has to do with the intended audience(s) to which web sites are addressed. Website designers who are concerned primarily with addressing content

¹⁰ L'Association nationale pour l'alphabétisation et la formation des adultes. See <http://anafa.ouvaton.org/>

¹¹ For example, a Somali language word processor, "Hikaadiye," is at least six years old – see <http://www.somitek.com/>. In the late 1990s a word processor for Oromo was marketed by a company called Oromosoft.

¹² Some correspondence on the A12n-collaboration list in 2002 revealed that Microsoft had already been gathering data on African language transcriptions for its research and development.

¹³ Its website, <http://translate.org.za/>, has information on the releases.

¹⁴ See <http://www.kilinux.org/>.

¹⁵ For instance, the author is aware of individuals considering localisation for Fon and for Sonrai, and of new projects for Lingala and three languages of Burkina Faso.

¹⁶ See <http://www.konyin.com/>.

¹⁷ See <http://www.pin.itgo.com/afara/>.

¹⁸ African Languages Technology Initiative. See <http://www.alt-i.org/> and also Egbokhare (2003).

¹⁹ See <http://www.paradigmint.net/lingua.htm>.

²⁰ See <http://webmail.variomedea.de>. One needs to have an account to fully access the service, but it is possible to see Swahili among the language options.

²¹ Localisation extends to other ICTs such as mobile telephone technology. For instance, one project is localizing cellphone interfaces in Afrikaans, Sesotho, Swahili, Xhosa and Zulu (Shanglee 2004), and another is focusing on the Ethiopic script (IRIN News 2004).

²² Another perspective is offered by Kenneth Keniston (2004) who describes four digital divides of which one is linguistic and cultural. The other three are socioeconomic within countries, digital between North and South, and the gap between the technical elite and everyone else. A roundtable on the digital divide at UCLA considered "a whole range of digital disparity gaps" among which language issues figure prominently (Afnan-Manns and Dorr 2003).

about Africa to a wide international audience would not generally choose an African language to convey information. In the case of localisation of applications, another factor is the perceived market for the software.

There is also a question about how much interest there is among foreign sponsors of ICT projects on or for the continent in considering African languages. In general there seems to be a tendency to rely on official languages and not pay much attention to indigenous ones. One foreign development expert based in Senegal who had years of African experience asked this author two years ago quite sincerely why someone who could read French would want to use another (i.e., an African) language.²³ This mindset is hardly unique; an American, also in Senegal who was connected with the U.S. "Digital Freedom Initiative" (DFI) there dismissed any consideration of providing for any indigenous Senegalese language use of the systems that that project is introducing there – for several reasons, including the multiplicity of languages, complexities in handling their scripts, and lack of literacy in them. He further offered the opinion that African language use in ICT is fifteen years off, if it ever happens.²⁴

However, even where foreign-funded initiatives would be more favorably inclined to multilingual ICT there is little incentive to actively initiate efforts for African languages. This author's communication with the BusyInternet center in Accra, for instance, yielded generally positive but ultimately unenthusiastic appraisals of the potential for providing basic Ghanaian language computing capacities to their systems. In the absence of a vocal demand for such service, it is easy to assume that it does not matter.

2) Structural factors

There are several structural factors limiting African language use. Some of these relate to standardization of orthography, which in some cases is subject to change or individual experimentation,²⁵ and in more than a few cases varies for the same language across borders.²⁶ A significant number of less widely spoken languages apparently do not have any established orthographies.

Another factor is that of the special characters or non-Latin scripts used in many orthographies that required specialized 8-bit fonts but now can use Unicode fonts.²⁷ However, there is still some debate on the provisions in the Unicode standard for certain diacritical characters used in some languages, which currently pose some inconveniences (see Tassé 2003). On the whole, though, the problem is that the use of Unicode is still not widely understood among computer technicians and systems administrators on the continent.

This in turn relates to a lack of intersection of language policies and ICT policies in most African countries. In fact, it appears that there is little collaboration between linguists and ICT technicians in Africa outside of programs like those of RIFAL (Réseau international francophone d'aménagement linguistique). Similarly, in development agencies there is generally a lack of knowledge about African languages and linguistics or about basic technical options to facilitate computing in multiple

languages (especially in the case where characters or scripts beyond standard ASCII are used). The opinion of the individual with DFI mentioned above concerning the prospects for African languages in ICT is merely a more extreme example of lack of knowledge of the realities and possibilities.

Another structural factor that is as essential as it is obvious, relates to lack of resources to advance work in these categories, even where there is the will and know-how to implement multilingual ICT projects in Africa. Often though, the know-how is insufficient especially outside of technical circles that are most often based uniquely in the capital city, so plans for training to build skills of local level technicians is necessary.

III. Elements of a Strategy for African Language Localisation of ICT

A strategy to favor African language use in ICT needs to concentrate on two main concerns: favoring dissemination of content and providing tools. In other words, any concentration on what the technology can do with African languages to meet people's needs and aspirations should focus on what it can deliver and how it can best facilitate more production in these languages. Also, these two concerns are essential in expanding access to ICT. These will be discussed below.

In addition, and as a prerequisite (or corequisite), structural factors including those mentioned in the previous section, that together affect the "environment" for localisation, need to be addressed in order to facilitate content production and providing tools. Some of these are beyond the scope of this paper but bear reiterating: basic literacy (which also can be understood as enhancing user profiles), language and education policies (what priorities do the governments have in terms of learning in and other use of indigenous tongues, and how do these priorities mesh with ICT policy), and availability, quality and cost of connections outside of the capital cities (some innovative use of the technology for rural Africa might require significant bandwidth). The others will be outlined below before discussing content and tools.

These structural factors include: supporting existing efforts and improving coordination and collaboration among people active in the field; standardization of orthographies and spelling conventions; and dissemination and adoption of Unicode continentwide.

A. Structural or "Environmental" Factors

A trio of factors needs explicit attention as prerequisites or at least corequisites of successful efforts to develop content and tools:

- 1) Supporting ongoing efforts and improving coordination and collaboration among the actors in the field (software developers, content managers, etc.)
- 2) Standardizing orthographies, rules of transcription, and locales

²³ This and similar comments heard by the author in Africa and also regarding indigenous languages in the U.S. echo Keniston's (1999) observations concerning India: "It can be argued that, given the fusion of language, wealth and power in India, there is simply no market (and perhaps no need) for software in any language other than English. Asked about localisation to Indian languages, international software firms sometimes reply, 'But everyone speaks English in India,' by which of course they mean that the present market consists of people who speak English."

²⁴ Such argumentation in the end resembles the vicious circle of rationalization that hobbles foreign assistance to literacy efforts: it is sometimes argued on the one hand that printing materials in African languages is pointless since few people know how to read these languages, and on the other hand that it makes little sense to conduct literacy training in these languages since there is so little to read in them. The author has encountered such arguments in Niger.

²⁵ In Niger, for instance, the most recent revision of orthographies dates to 1999. In the case of the Igbo language of southeastern Nigeria, a recent dictionary (Echueiro 1998) departed from established practice by substituting a dieresis for the subdot on vowels and /c/ for /ch/ - a decision that aroused some controversy.

²⁶ Significant effort has gone into trying to harmonize transcriptions of cross-border languages, including international expert meetings facilitated by UNESCO several decades ago (see <http://www.bisharat.net/Documents/>). Nevertheless there are frequently small differences.

²⁷ An overview of African orthographies and ICT usage is surveyed by the author elsewhere (Osborn 2001).

- 3) Adopting and adapting Unicode continentwide, especially for those countries whose languages use non-Latin scripts or extended characters in the Latin alphabet

These factors are interrelated and together they can create a facilitating environment for localisation.

1) Supporting existing efforts

Support for efforts to localise ICT in African languages needs to be considered broadly. Not only are means needed, but also information, networking and collaboration that goes beyond narrow technical issues. Special attention is needed to involve linguists, policymakers, and educators.

Support for efforts to develop mono- or multilingual web content in African languages first needs to take account of who is developing the content for whom, perhaps using Ballantyne's (2002) schema mentioned above. The support itself can take various forms such as means to finance or encourage such work in Africa, and ways to facilitate communication among people doing it wherever they are.

A workshop on African languages and the internet at the 2002 African preparatory conference in Bamako for the World Summit on the Information Society (WSIS), for instance, proposed establishment of a "Highway of African Multilingual Information" (HAMI) fund to support creation and maintenance of African language web content.²⁸ Although this proposal does not appear to have been acted on, the idea of donor support for African language content is one that should be seriously considered in any long-term ICT strategy on the continent. The mechanics of such a program could range from something as simple as annual prizes for websites developed in Africa with African language content, to something more complicated such as training, or other approaches.

Support for software localisation in Africa similarly would benefit from material and technical support. At the time of writing, a workshop for African localisation developers is in the planning stages.

Any effort to support African language localisation also needs ways to facilitate ongoing communication among people working on separate projects, including Africans and others in Africa, Africans in the diaspora,²⁹ and others abroad who are motivated to help. There are already several electronic fora for discussion of issues related to African language in ICT that seem to have demonstrated the potential of this medium for fostering exchange of information and even collaboration on small projects.³⁰ The possibility of organizing conferences or meetings should also be considered.

2) Standardizing orthographies, transcription, and locales

The standardization of orthographies and rules of transcription is a process that needs accelerated attention in the many languages for which these are still not fixed. Such standardization is essential for virtually the whole range of computer applications from simple wordprocessing and navigation of the internet to use of advanced processes such as machine translation and text-to-speech. This is a process that relies mainly on lin-

guists, but also needs to involve coordination with technical experts.

Such standardization may also incorporate efforts to harmonize transcriptions for Africa's many cross-border languages. This has been a concern going back at least forty years but needs to be actively considered in current processes of localisation.³¹

Technical localisation for languages and cultures also needs attention to locales. These similarly need to be addressed in a systematic way by experts in language and ICT.

3) Adopting and adapting Unicode continentwide

The promise of Unicode (ISO-10646) for African languages that are written either with non-Latin scripts or the Latin alphabet with additional "extended characters" is that text can be shared readily and that different languages and scripts can be combined in the same documents. Although as mentioned above, there are some questions about how the standard meets some character needs, the main problem with Unicode is that despite the fact that it is being adopted as the industry standard, it is not yet widely understood by technicians in Africa. One recent example is a telecenter project in Mali where the director indicated that telecenter managers did not know much about Unicode, and that the font they had for national language use is a legacy 8-bit font (ironically one of the same legacy fonts used for documents that the RIFAL project is assisting Malian authorities to convert to Unicode fonts).

Wider training in and discussion of use of the Unicode standard therefore seems to be needed, mainly in those countries (most of them in West and Central Africa) whose languages are written with non-Latin or extended characters. However, since it is intended as a universal standard, and in the interests of African unity, Unicode's use should be promoted throughout the continent.

B. Content

Content here is considered broadly, including web content and also e-mail and traffic on discussion fora. It is one of the two main strategic concerns of localisation in Africa.

Beyond supporting existing web content creation efforts there is a need to consider other creative approaches to increasing the amount of material on the web that is accessible in African languages. These could include for instance:

- 1) Putting extant African language texts on the web
- 2) Increasing the amount of current high-quality Africa-related web content that is translated into appropriate African languages
- 3) Facilitating use of African languages on internet fora
- 4) Exploring the potential of audio content

Creation of web content directed to Africa in its diverse languages is a way to make the technology more relevant to the people of the continent. Such creation can be accomplished in several ways, and it is instructive to review these and how ways can be found to speed up the process.

First of all, development of original text-based web content from scratch requires much time and resources. Furthermore, one cannot develop significant amounts of content in African

²⁸ See <http://www.bisharat.net/Documents/Bamako2002-workshop.htm>. Reference to this was also made at a later WSIS preparatory meeting in Accra in February 2005 (Diakité 2005).

²⁹ The African diaspora's role in African language web content production seems to be a significant if overlooked factor (this author has touched on it elsewhere – Osborn 2004). Use of internet among diaspora communities in other ways is already a noted factor (Ajibewa and Akinrinade 2003).

³⁰ These include several in English or French: "Unicode-Afrique" <http://fr.groups.yahoo.com/group/Unicode-Afrique/>; "A12N-collaboration" <http://lists.kabissa.org/mailman/listinfo/a12n-collaboration>; "A12n-forum" <http://lists.kabissa.org/mailman/listinfo/a12n-forum>; "A12n-entraide" <http://lists.kabissa.org/mailman/listinfo/a12n-entraide>. There are in addition several message boards for selected languages and countries of West Africa accessible at <http://www.quicktopic.com/share?s=QSp0>.

³¹ A series of expert meetings sponsored by UNESCO, of which one in Bamako in 1966 is considered seminal, dealt with these issues and have had an important impact on the transcriptions we see today. Documents from many of these meetings are available at <http://www.bisharat.net/Documents>. A South African based organization, the Center for Advanced Studies of African Societies (CASAS), is currently working on this issue.

languages very quickly, as there is a lack of available human and monetary resources to apply to the task and in some cases the orthographies are not settled or require specialized scripts. In any event, the dynamic here is by its nature slow: the proportions of the languages on the internet change, but gradually. Some expect that these proportions will continue to approach an approximation of the current percentages of speakers of languages in the world, but that assumes relatively comparable written traditions and available resources – conditions that don't hold for much of Africa.

1) Putting extant African language texts on the web

One quick way of increasing quality African language content is to put published and other extant texts on the web. This could include texts published in African languages (often with parallel text in a European language)³² and monographs with African language content. Such literature, currently accessible only to limited audiences would thus be accessible to a new generation of readers. In addition, a range of other, unpublished materials could also be made available on the web.

In addition to obvious cultural importance, the presence of such material online could encourage development of other original content in African languages and help set a high standard for presentation in them.

2) Translating Africa-related content

The content on the web is vast and continually changing, so no one would propose translating it or any major part thereof into any language, however important. However, some content relevant to culture, development, education, health, and agricultural and natural resource extension in Africa might be very appropriate and useful to have translated into various African languages. The focus could be on relatively static content on basic themes and be approached as an investment, much like publishing books.

This process would be facilitated by advances in machine translation for African languages (see below, III.C.4).

3) Facilitating use of African languages on internet fora

The existence of e-mail fora in which African languages are used has been noted above (II.B.2). In cases where the languages in their official orthography use characters not supported by existing web-based groups or distribution list archives, efforts should be made to use existing technology to make them Unicode aware. Web-based groups could also facilitate input of extended characters, for instance, by the use of click-on buttons or pop-up keyboards.

But more than the technical solutions, the use of the languages in various kinds of announcements and mailings by projects and people with the means to do so will demonstrate the potential for African language use in this medium, and develop models for their use by others.

As the growing movement to publish weblogs becomes more common in Africa, software for these should also be Unicode compliant to facilitate diverse language use.

4) Exploring the potential of audio technology

African cultures are often referred to as “oral cultures”, and even where the written forms of its languages are being used, the importance of speech and orality is still high. It seems therefore

worthwhile to explore the ways in which the audio capacities of computers and the internet can be better exploited to create relevant content.

Improvements in audio technology can in theory allow a vast quantity of spoken content in any language to be quickly added to or communicated on the internet. Audio files will never replace text (part of the problem is their size), but they make it possible to create new forms of content with various combinations of text, image and audio (based, if so desired, mainly on the audio). Audio e-mail, for instance, is a marginal technology in the Northern countries but might be very interesting to develop for users in multilingual societies of Africa and other parts of the global South.

In addition, text-to-speech and speech-to-text technologies could also be important. Could localising ICT in Africa open up new paths for evolution of oral culture as well as for developing written culture?

C. Tools

Tools that facilitate the use of African languages in ICT and the production of material such as web content in them is a second pillar of a strategy to enhance use of these languages. Several specific elements are important, including fonts, keyboard layouts, software localisation, and machine translation. All of these rely on standardized orthographies and benefit from use of Unicode.

1) Fonts

Although many African languages do not have special script or character requirements and thus can be typed in a range of readily available fonts, many others do require fonts with modified Latin letters (extended characters) or non-Latin alphabets. There are not many such fonts and so a basic unmet need is for more quality Unicode fonts to facilitate use of these languages. This, of course, is both to be able to view text on the web and to be able to compose text for diverse purposes. Various means to promote such font development need to be examined – for instance if it will work best on proprietary or open-source levels – and implemented.

It would be ideal for Africa if the major Latin fonts used internationally were to include all Latin character ranges and if other fonts for Africa could include, as a matter of course, all Latin ranges plus all Arabic ranges, Ethiopic/Ge'ez, Tifinagh, N'ko, and yet to be encoded scripts like Vai.

2) Keyboard layouts

Along with fonts there is a need both for means to input the necessary characters and for standardization of layouts. When extended Latin characters or diacritics (sometimes used to mark tone or to indicate alternate sounds) are needed to compose text in a language, keyboard layouts can be designed to facilitate input. This can be done through creation of specialized keyboard layouts for use with existing software, or as part of software localisation.

Interfaces for input of special characters can be done in a number of ways such as using programs like Tavultesoft's Keyman program, Microsoft's keyboard layout utility, or simply by assigning keys within a wordprocessor program.³³ These are not particularly hard and in fact there is an increasing number of these available for various languages and countries or regions.³⁴

³² For example a series of books in the “Classiques africains” collection published over the last three or four decades has numerous titles with parallel French and African language text.

³³ Three main approaches include: key combinations (with the Alt or Ctrl keys), deadkeys (where one key typed before another yields a certain character or diacritic after striking the second key), or substitution (simply reassigning a key to another character).

³⁴ See for instance the Tavultesoft site <http://www.tavultesoft.com> or the keyboard projects links at <http://www.bisharat.net/A12N/Projects>.

Non-Latin scripts such as the Ethiopic/Ge'ez syllabary used for Amharic and Tigrinya may pose a challenge. In some cases an alternative approach using graphics tablets with keyboard assignment or even handwriting recognition may be useful. Indeed, the potential to use traditional keyboards in tandem with graphics tablets is an option that should be explored in contexts where multiple scripts are used.

An important issue, however, is the standardization of layouts for either individual languages or groups of languages in a country or countries. The object is to provide users with predictable input systems and software localisers with information to guide their work. This stops short of discussing physical keyboards designed to better accommodate multilingual needs in Africa, but these are a logical extension of discussion about (virtual) keyboard layouts.

3) Software and web-interface localisation

As indicated above (II.B.4), localisation of software in African languages by open source and proprietary projects is beginning to evolve. In effect, the terrain of localisation of applications is diversifying as it evolves. Proprietary firms – both multinational (notably Microsoft) and local – are in the field as is a growing open source software movement. In addition, one notes what could be characterized as two different levels of software localisation: full translation of the user interface and software that still relies on a language of wider communication for commands, etc., but facilitates use of specific sets of African languages through making available appropriate fonts and dictionaries.

These efforts should be encouraged and collaboration among the local ones facilitated (per III.A.1, above). One notes for instance that the Translate.org.za project in South Africa has offered assistance to similar efforts in other countries – this kind of mutual assistance can go a long way to building capacity in various parts of the continent to carry out localisation projects.

In the long run, such software localisation provides tools that can be readily used for web content development and may also have a very positive impact on a range of other production in the languages concerned. One might expect that localised software would contribute to a boon in local content.

Beyond promoting collaboration among localisation developers, there are needs to develop continental strategies for training, prioritize types of software, and do outreach in countries where there is no localisation. For each language, moreover, there is also the need to involve appropriate language experts so that the terminology and orthographies conform to operant standards.

4) Machine Translation

Although the technology of computer translation is in some ways still rough, it is being steadily improved. Beyond figuring in any strategy to increase the quantity of materials in African languages available on the web as discussed above, this technology can also be a key part of efforts to develop materials in African languages for publication in print. The object would be the creation of software for translation between the principal official languages (English, French, Portuguese) and African languages, as well as software for translation among the dialects of certain African languages.

In any event, the ability to read any text on the internet in the mother tongue (even if the translation is not perfect) could have immense implications, as would the potential for speeding translation of reference and other materials into less widely spoken languages.

As with software localisation, there is a need to plan for training, but beyond that also the opportunity to consider the development of the specialty of machine translation in Africa (including long-term development of a professional association under the International Association for Machine Translation on a par with those in North America, Europe and Asia).

IV. Conclusion

African language localisation is an essential part of any effort to address the digital divide on the continent and use ICT to meet the development needs of its population. This fact is now beginning to get more attention, but there is a need to formulate and elaborate plans to support aspects of localisation. In the current situation of African languages and ICT one sees some useful elements to build on, including of course Unicode, as well as some challenges. At this time, a strategy for localisation of ICT in Africa should focus on facilitating the creation of web content and the development of software tools, and such a strategy should address several elements as outlined above. In the longer run there are other issues to be addressed and probably new dynamics that will come into play. Indeed, by actively taking steps now, the environment for future localisation will be enhanced.

Bibliography

- Afnan-Manns, Sheila, and Aimée Dorr. 2003. "Re-evaluating the Bridge! An Expanded Framework for Crossing the Digital Divide through Connectivity, Capability, and Content. A Report on the Digital Divide's Multiple Dimensions: Indicators for Measuring Success." Los Angeles: The Pacific Bell/UCLA Initiative for 21st Century Literacies at the UCLA Graduate School of Education & Information Studies. http://www.newliteracies.gseis.ucla.edu/publications/reval_bridge.pdf
- Agence de Presse Sénégalaise (APS). 2004. "Microsoft s'engage à traduire ses logiciels en Wolof (Officiel)." APS 25-10-04. http://www.aps.sn/artfiche.php?page=id_article=3220
- Ajibewa, Aderemi, and Sola Akinrinade. 2003. "Globalisation, Migration and the New African Diasporas: Towards a Framework of Understanding." International Workshop on Migration and Poverty in West Africa, March 13-14, 2003, University of Sussex. <http://www.sussex.ac.uk/migration/research/transrede/workshop/IWMP15.pdf>
- Anyanzwa, James. 2004. "Microsoft Launches Kiswahili Edition." East African Standard, Oct. 29, 2004. <http://allafrica.com/stories/200410280925.html>
- "Africa on the Internet." 2002. (Africa Action) <http://www.africaaction.org/inet.htm>
- Ballantyne, Peter. 2002. "Collecting and Propagating Local Development Content: Synthesis and Conclusions." Research Report No. 7. IICD in association with the Tanzania Commission for Science and Technology; funded by DFID. http://www.icconnect-online.org/base/ic_show_news?sc=107&cid=1878
- Davies, Mark. 2003. Personal correspondence.
- Diakité, Salam. 2005. "African languages and content development in the cyberspace." PowerPoint presentation for the World Summit on the Information Society, African Regional Preparatory Conference, Accra, Ghana, 2-4 February 2005. <http://www.wsisaccra2005.gov.gh/updates/ppt/African%20languages%20and%20content%20development%20in%20the%20cyberspace.ppt>

- Diki-Kidiri, Marcel, and Atibakwa Baboya Edema. 2003. "Les langues africaines sur la Toile." *Cahiers du Rifal*, 23: 5-32.
- Echeruo, Michael J.C. 1998 *Igbo-English dictionary : a comprehensive dictionary of the Igbo language, with an English-Igbo index*. New Haven: Yale University Press.
- Elder, Laurent. 2002 "What a Difference IT Makes ..." (an Acacia project site visit) http://web.idrc.ca/en/ev-9481-201-1-DO_TOPIC.html
- Egbokhare, Francis. 2003. "Interview with Francis Egbokhare, African Languages Technology initiative (ALTI) of Nigeria." *iConnect Africa* Vol. 1, Issue 5, July 2003. <http://www.uneca.org/aisi/IConnectAfrica/v1n5.htm>
- Enguehard, Chantal, and Chérif Mbodj. 2003. "Flore: un site coopératif pour recueillir et diffuser les noms des plantes dans les langues africaines." *Cahiers du Rifal*, 23: 46-54.
- Internet World Stats: Africa. 2004. <http://internetworldstats.com/africa.htm>
- IRIN News. 2004. "Ethiopia: Old alphabet adapted for modern use in technology." (11 Nov. 2004). http://www.irinnews.org/report.asp?ReportID=44097&SelectRegion=Horn_of_Africa&SelectCountry=ETHIOPIA
- Keniston, Kenneth. 2004. "Introduction: The Four Digital Divides." In K. Keniston and D. Kumar, eds. *IT Experience in India: Bridging the Digital Divide*. Delhi: Sage Publishers. http://web.mit.edu/~kken/Public/PAPERS/Intro_Sage.html
- Keniston, Kenneth. 1999. "Language, Power, and Software." In Charles Ess, ed. *Cultural Attitudes Towards Technology and Communication*. New York: SUNY Press. <http://web.mit.edu/~kken/Public/papers1/Language%20Power%20Software.htm>
- Mas, Jordi. 2003. "La salut del català a Internet." *Softcatalà*. <http://www.softcatala.org/articles/article26.htm>
- McLean, Pamela. 2004. Personal correspondence.
- Microsoft Corporation. 2004. "Microsoft Enables Millions More to Experience Personal Computing Through Local Language Program." PressPass - Information for Journalists, March 16, 2004. <http://www.microsoft.com/presspass/press/2004/mar04/03-16LLPPR.asp>
- Miller Esselaar and Associates. 2001. "A Country ICT Survey for Tanzania: Final Report.", November 2001, Prepared for Swedish International Development Cooperation Agency <http://www.milless.co.za/downloads/Sida%20report%20-%20Tanzania.pdf>
- Ngangala Balade Tongamba, Jacques. 2001. "Technologie de l'information et langues africaines : Lieux des enjeux existentiels." *Cahiers du Rifal*, 22: 3-9.
- NUA Internet Surveys. 2002. "How Many Are Online?" http://www.nua.ie/surveys/how_many_online/
- Osborn, Don. 2004. "African Languages and ICT: Literacy, Access and the Future." Paper presented at the Annual Conference of African Linguistics, Harvard University, Cambridge, MA - April 2-4, 2004
- Osborn, Don. 2001. "The Knotty Problem of Using African Languages for E-Mail and Internet." *Balancing Act's News Update*, No. 69. http://www.balancingact-africa.com/news/back/balancing-act_69.html
- Otter, Alistair. 2004. "Uganda gets indigenous language browser." *Tectonic*, Sept. 15, 2004. <http://www.textonic.co.za/view.php?id=342>
- Pastore, Michael. 2000. "Web pages by language." ClickZ Stats, July 5, 2000. http://www.clickz.com/stats/big_picture/demographics/article.php/5901_408521
- Robinson, Clinton D. W. 1996. *Language Use in Rural Development: An African perspective*. Contributions to the Sociology of Language, 70. Berlin: Mouton de Gruyter.
- Shanglee, Russell. 2004. "Localisation in African Languages: Translators Face Linguistic Challenges as They Localize Modern Technology." *Multilingual Computing and Technology* 61 (Vol. 15, Issue 1): 37-38.
- Sonaike, S. Adefemi. 2004. "The Internet and the Dilemma of Africa's Development." *Gazette: The International Journal for Communication Studies* 66(1): 41-61.
- Tassé, Etienne. 2003. "The First Steps of African Languages on the Internet." *iConnect Africa* Vol. 1, Issue 5, July 2003. <http://www.uneca.org/aisi/IConnectAfrica/v1n5.htm>
- Tedre, Matti, Piet Kommers, and Erkki Sutinen. 2002. "Ethnocomputing. A Multicultural View on Computer Science." Paper presented at the IEEE conference ICALT 2002, Russia, 9-12 September 2002. http://lutf.ieee.org/icalt2002/proceedings/t501_icalt105_End.pdf
- TeleCommons. 2000. "Rural Access to Information and Communication Technologies: The Challenge of Africa." <http://www.telecommons.com/reports.cfm?itemid=122>
- This Day. 2004. "Microsoft Endorses Due Process in IT." *This Day* (Lagos), Dec. 15, 2004. <http://allafrica.com/stories/200412160060.html>
- Van der Veken, Anneleen, and Gilles-Maurice de Schryver. 2003. "Les langues africaines sur la Toile: Etude des cas haoussa, somali, lingala et isixhosa." *Cahiers du Rifal*, 23: 33-45.
- Yacob, Daniel. 2004. "Localize or be Localized: An Assessment of Localisation Frameworks." Paper presented at the International Symposium on ICT: Education and Application in Developing Countries, Addis Ababa, 19-21 October 2004.

Donald Z. Osborn is the founder and director of Bisharat, Ltd. With a background in environment, agricultural development, and African languages, he has developed an expertise in African language computing and localisation with an eye towards applications for sustainable development. He can be reached at dzo@bisharat.net