

DON OSBORN

AFRICAN LANGUAGES IN A DIGITAL AGE

Challenges and opportunities for
indigenous language computing

This page intentionally left blank

DON OSBORN

AFRICAN LANGUAGES IN A DIGITAL AGE

Challenges and opportunities for
indigenous language computing

International Development Research Centre
Ottawa • Cairo • Dakar • Montevideo • Nairobi • New Delhi • Singapore



This book is an output of the IDRC-funded African Network for Localization,
www.africanlocalization.net

Published by HSRC Press
Private Bag X9182, Cape Town, 8000, South Africa
www.hsrcpress.ac.za
and
International Development Research Centre (IDRC)
PO Box 8500, Ottawa, ON, Canada K1G 3H9

First published 2010

ISBN (soft cover) 978-0-7969-2249-6
ISBN (PDF) 978-0-7969-2300-4
ISBN (epub) 978-0-7969-2301-1
eISBN (IDRC) 978-1-55250-473-4

© 2010 Human Sciences Research Council

The views expressed in this publication are those of the authors. They do not necessarily reflect the views or policies of the Human Sciences Research Council ('the Council') or indicate that the Council endorses the views of the authors. In quoting from this publication, readers are advised to attribute the source of the information to the individual author concerned and not to the Council.

Copyedited by Robyn Arnold
Typeset by Simon van Gend
Cover design by Hothouse South Africa
Printed by Logo Print, Cape Town

Distributed in Africa by Blue Weaver
Tel: +27 (0) 21 701 4477; Fax: +27 (0) 21 701 7302
www.oneworldbooks.com

Distributed in Europe and the United Kingdom by Eurospan Distribution Services (EDS)
Tel: +44 (0) 20 7240 0856; Fax: +44 (0) 20 7379 0609
www.eurospanbookstore.com

Distributed in North America by Independent Publishers Group (IPG)
Call toll-free: (800) 888 4741; Fax: +1 (312) 337 5985
www.ipgbook.com

Contents

	List of tables and figures	viii
	Foreword	ix
	Preface	xii
	Acronyms and abbreviations	xv
1	Introduction	1
2	Background	5
	Importance of African languages and implications for ICT	5
	What is localisation?	7
	Overlapping regional contexts: localisation where?	12
	Who localises?	14
	What is the current state of localisation across the African region?	15
3	Introducing 'localisation ecology'	17
	An ecological perspective on the environment for localisation	17
	The PLETES model	20
	Dynamic complexes within localisation ecology	25
	Relevance to questions of ICT and localisation	29
4	Linguistic context	31
	Languages, dialects and linguistic geography	31
	Sociolinguistics and language change	34
	Oral and literate traditions	35
	Language and language in education policies	38
	Basic literacy, pluriliteracy and user skills	40
	Terminology and accommodation of ICT concepts	41
5	Technical context I: physical access	43
	Physical and soft access	44
	Basic infrastructure	45
	Computer hardware and operating systems	46
	Connectivity and ICT policy	47
6	Technical context II: internationalisation	49
	The facilitating technical environment	49
	Handling complex scripts: from ASCII to Unicode	50
	The 'last mile' of internationalisation	54
	Internationalisation and localisation	55

7	African-language text, encoding and fonts	57
	Non-Latin scripts and ICT	57
	Typology of Latin-based African orthographies	58
	Evolution of African-language text use in ICT	59
	Fonts	61
	Languages without writing systems	63
8	Keyboards and input systems	65
	Keyboards	65
	Keyboards for Africa	68
	Alternative input methods	70
9	Defining languages in ICT: tags and locales	73
	Languages and the ISO 639 standards	73
	Locale data	75
10	Internet	79
	E-mail	79
	Internationalisation and the web	80
	Web content in and about African languages	80
	Internationalised domain names	83
11	Software localisation	85
	Applications and operating systems	85
	Trends in proprietary software	85
	Trends in free and open-source software	86
	Software localisation in Africa	87
	Web interfaces	88
12	Mobile technology and other specialised applications	91
	Mobile technology	91
	Audio dimensions: voice, text-to-speech and speech recognition	92
	Geographic information systems	93
	Computer-assisted translation	94
13	Achieving sustainable localisation	97
	Needs by kind of localisation and localiser	97
	Understanding the needs of localisers	99
	Analysis of needs from a pan-African perspective	102
	Facilitating communication about localisation	103

14	Summary, recommendations and conclusion	107
	Major themes	107
	Strategic perspectives	109
	Conferences and workshops	111
	Training and public education on localisation	113
	Information resources and networking	115
	Languages, policy and planning	116
	Basic localisation, and ICT policies and programmes	118
	Africa and ICT standards for localisation	120
	Advanced applications, tools and research	123
	Conclusion	127
	Notes	130
	References	139
	Index	146

Tables

TABLE 2.1:	Dimensions of localisation	8
TABLE 7.1:	Approaches to using Latin-based orthographies with extended characters and/or diacritics (category 3 and 4 orthographies) in ICT	61
TABLE 7.2:	Some legacy 8-bit fonts for extended Latin scripts in Africa	62
TABLE 9.1:	iso 639 categories for identifying language (current and planned)	74
TABLE 9.2:	African languages filed in CLDR 1.6.1	76
TABLE 11.1:	OpenOffice localisation projects	89
TABLE 13.1:	E-mail forums on African languages and ICT	103

Figures

FIGURE 3.1:	Model of language management	21
FIGURE 3.2:	Three basic factors in localisation ecology	22
FIGURE 3.3:	The PLETES model	23
FIGURE 3.4:	The three key factors of localisation in the PLETES model	26
FIGURE 3.5:	Applied linguistics, translation in localisation, and social uses of ICT	26
FIGURE 3.6:	Comparison of the main concerns of language policy and ICT policy	27
FIGURE 3.7:	Digital divide projects: from basic to more complex dynamics, without language	28
FIGURE 3.8:	Localisation, localisation follow-through and localisation follow-up	29

Foreword: Language, money and the information society

AT THE BEGINNING OF the 21st century, national languages and cultures play a much more important role in international affairs and relations among peoples and governments than some 20th-century analysts and researchers had predicted. Among the potentially devastating effects of globalisation, linguistic unification – not to mention Anglicisation – of societies and cultures has very often been referred to as its most dangerous negative impact. So dangerous, in fact, that global summits have been held on cultural and linguistic diversity, and monumental efforts have been made to prevent cultural homogenisation.

However, global tensions since September 2001 have reawakened decision-makers and global institutions to the need to understand and to master the language of others so as to better understand them and better protect ourselves.

Information and communication technologies (ICTs) facilitate this interaction as tools that use languages or as language processing and representation tools. While humanity's main languages are now well served by ICTs, there are still thousands of languages in the world in which one cannot send an email or read a website. Some languages do not yet have standardised characters, while others have two or three groups of characters: one group uses the local alphabet; another group uses the alphabet of a formerly dominant foreign language; and the third group often uses the Latin alphabet.

When ICTs are not available in a given local language, the opportunity to produce and disseminate local content (educational, administrative or tourism content) on the Internet is reduced. As a result, the chances that the culture conveyed by this language will be shared and made accessible to its speakers, researchers and linguists who would like to study it are also decreased. Worse yet, given the widespread use of ICTs (mobile phones, computers, multimedia and digital audio-visual aids, etc.), the de facto language imposed on users (be it English, French, Spanish, Arabic or other) ends up gaining the upper hand and replacing the local language for ICT and other purposes.

This phenomenon is not unique to ICTs. In a recent conference on translation, one of the speakers attributed the predominance of a particular foreign language in his government's correspondence and invitations to tender to the language preference of administrative representatives. This resulted in favouring Anglophone companies when invitations to tender were drafted in English and Francophone companies when they were drafted in French. The impact of a particular trend therefore extends beyond its own linguistic dimension to become political, economic and social in nature.

In the Information Society, in addition to being a means of communication, language has a socio-economic role similar to that of money in industrial society. While money is used to acquire material goods, language is used to acquire knowledge and intangible goods.*

This book is the result of several years of observation, analysis, consultation and synthesis of the adaptation of ICTs to local languages in Africa. The goal of the *Pan Africa Localization* project led by Don Osborn was to closely track the progress of ICTs in African languages and clearly identify the priorities that the *Pan African Network for Localization* (ANLOC) will pursue in its work plan. This book is a revised version of the project's final report. By collecting and compiling all the data presented in this book, Don has helped establish ANLOC's research network and has provided an accurate picture of ICT localisation in Africa.

This publication will thus be useful for decision-makers intending to develop a language policy, developers working on language processing, researchers in the area of languages and information technologies, donor agencies that fund projects to support local languages, and ICT users wanting to use these technologies in their local language.

By publishing this book and supporting ANLOC's work, we are contributing to the implementation of the World Summit on the Information Society's plan of action and its Tunis Agenda. The decision-makers who gathered in Geneva in 2003 and Tunis in 2005 signed a declaration in which they committed themselves to:

- ◆ encourage the development of content and to put in place technical conditions to facilitate the presence and use of all world languages on the Internet;
- ◆ in the context of the Information Society, provide content that is relevant to the cultures and languages of individuals by providing access to traditional and digital media services;
- ◆ nurture the local capacity for the creation and distribution of software in local languages, as well as content that is relevant to different segments of population, including non-literate, persons with disabilities, disadvantaged and vulnerable groups, especially in developing and transition countries.

The Tunis Agenda is very clear in this regard. The signatories committed to '*working earnestly towards multilingualization of the Internet, as part of a multilateral, transparent and democratic process, involving governments and all stakeholders, in their respective roles.*' They also supported '*local content development, translation and adaptation, digital archives, and diverse forms of digital and traditional media*'.†

Despite all of the efforts to respect these commitments and to promote multilingualism on the Internet, we have to admit that there is still a long way to go before all world languages appear on the World Wide Web. Few international

or regional mechanisms have been implemented, whereas volunteer efforts, small industry initiatives, and research projects such as ANLOC have sometimes had a significant impact on the lives of citizens.

But all these efforts are not enough if policies do not follow and are not appropriately implemented. For several years, IDRC has been funding a research network on Asian languages, PAN Localization, which has played an important role in ICTs and Asian languages. The African project, ANLOC, is producing dictionaries, terminology and regional language settings for software. It is also supporting the professional training of software translators in African languages (in collaboration with the *Localisation Research Centre* in Limerick, Ireland), as well as software translations and the development of software translation management tools that comply with industry standards and even define new innovative practices using global and African knowledge to speed up the development of ICTs in African languages.

The results of this enormous effort should subsequently guide national policies, which would guarantee and regulate the supply and demand of ICTs in local languages so that computers delivered to African schools would be equipped with local language keyboards and software, as well as with keyboards and software in an international language. It will take a great deal of time and energy, but it is feasible and worth the effort. ANLOC and its collaborators will succeed.

Adel El Zaïm

Senior Program Specialist

Regional Office for the Middle East and North Africa

International Development Research Centre

Preface

IN ITS BROADEST SENSE, the term ‘information and communication technology’ (ICT) covers a range of technologies, including broadcast and telephony, which in their audio aspects can be readily used with any spoken language. The focus in this book is on newer ICTs built mainly around use of text, namely computing, the internet and text on mobile devices, as well as a range of advanced human language technologies such as machine translation, speech recognition and text-to-speech.

These newer ICTs – digital computing and the internet – do lend themselves to adaptation in diverse human languages but require effort and resources to achieve that end. Indeed, ICT is increasingly being put to use in processing, analysing, reading, transcribing, and translating an ever widening range of languages. What does this mean for a world region where, on the one hand, there are a great number of indigenous languages that for the most part are not well resourced, often relatively few speakers, and are assigned low status or even denigrated, and on the other hand, ICT penetration is low and the means to increase its use are limited?

That is the broad picture in Africa today. It presents a number of challenges to efforts to use ICT more effectively for development and education. Yet at the same time, the potential to enhance initiatives in African languages is increasing and is already being explored.

In order to better understand the overall situation and find ways to support such localisation efforts, the International Development Research Centre (IDRC) sponsored the PanAfrican Localization (PAL) project. One of the objectives of the PAL project was to conduct the research that forms the basis for this book. The research sought to assess localisation in two overlapping regions – Africa and the Arabic-speaking countries, focusing on sub-Saharan Africa and predominately Arabic-speaking North Africa – while acknowledging the fundamental linguistic and cultural connections of the latter with the Arabic Middle East. It was concerned with the localisation of ICT in languages particular to Africa and in Arabic, which we collectively refer to as African languages except when there is a reason to treat Arabic separately.

The present work advances an understanding of the current status of localisation with respect to ICT, as well as the need for localisation in African languages and the potential to do so. Because of the nature of the topic and the range of factors involved, the research was extensive and challenging. Nevertheless, a volume such as this will remain incomplete as a result of the rapid changes in technology and its adaptation. The geographic scope involved is enormous, given that Africa is

the second-largest continent and is home to about a third of the world's languages. Despite regional variations, the continent is generally disadvantaged with regard to ICT and resources for researching, adapting and extending the technologies to the whole population.

In principle, ICT should be capable of accommodating people in any language and serving as a tool for development in its fundamental and most comprehensive sense of revealing potentialities. In the context in which basic needs are often not met, health crises persist, literacy in any language is low and many languages do not have a set orthography, it may appear to be a luxury to consider localising ICT in any form. To consider doing so, however, is an expression of hope, an affirmation of the value and relevance of Africa's linguistic and intellectual heritage, and a practical attempt to use new tools to help find new solutions to old problems, perhaps in the very languages and idioms most familiar to the disadvantaged.

This book is therefore the start of an initiative in a new direction. In addition to the printed volume, an online version is being published with links to an extensive web-based resource in wiki form (permitting ongoing online input by diverse experts in language, localisation and ICT) on languages, countries, writing systems or scripts, organisations and localisation resources or tools. The wiki may be accessed at <http://www.panafril1on.org/>.

The overall objective of the research was to identify issues, concerns, priorities and lines of work with regard to localisation in Africa and, more broadly, the meeting between ICT and African languages. Within that context, the book also discusses current and potential areas of focus in localising in African languages.

The book is organised into several thematic chapters on language, ICT and localisation. In order to help make sense of the processes of localisation, the concept of 'localisation ecology' is proposed as a way of accounting for various factors that may impact on current and potential future localisation efforts, and a model is suggested with which to organise that line of thought. Details of actors and activities, which tend to change frequently, are dealt with in the five parts of the abovementioned wiki that serves in part as a companion to this book: languages, countries, writing systems or scripts, organisations and tools.

The localisation of ICT is currently a popular topic internationally, but it is neither a fad nor a passing fancy. Nevertheless, the observations of Peter Senge, who researched the 'fad cycle' in business management, are worth noting. He observed that new ideas often go through a fairly predictable cycle starting with the initial interest, during which there is considerable activity and many people become involved, followed by an inevitable slackening of interest, during which the initial enthusiasm wanes and most people move on to other things (Senge 2006). Senge further noted that the difference between an idea that has an enduring effect or becomes institutionalised in a sustainable way on the one hand, and a passing fad

that has little long-term effect on the other, is the degree to which the idea is solidly supported by or linked to theory. Apart from some work on the importance of first languages (L1s) as media for communication and learning, there has been very little research that articulates at a theoretical level the importance and utility of using Africa's indigenous languages for all levels of computing and communication via the internet. In addition to reviewing activities and proposing practical measures, this book therefore also seeks to define localisation in the African context and demonstrate its importance in the long term.

Acknowledgements

This book was originally a document produced as part of the PanAfrican Localization project funded by International Development Research Centre (IDRC). The author is grateful to IDRC and in particular Laurent Elder, who initiated contact with the author on the concept for the project, and Adel El Zaim, who guided the effort over three years from initiation to completion.

Acronyms and abbreviations

ACALAN – African Academy of Languages (Academie Africaine des Langues)
ACCT – Agency of Cultural and Technical Cooperation (Agence de Cooperation Culturelle et Technique), now the oif
ANLOC – African Network for Localization project
ANSI – American National Standards Institute
ASCII – American Standard Code for Information Interchange
CAT – Computer-assisted translation
CLDR – Common Locale Data Repository
DOS – Disk Operating System
EIWC – European (or europhone) languages of wider communication
FOSS – Free/open-source software
FUNREDES – Networks and Development Foundation (Fundación Redes y Desarrollo)
GIF – Graphics interchange format
GIS – Geographic information system
GRASS – Geographic Resources Analysis Support System
HTML – Hypertext markup language
ICANN – Internet Corporation for Assigned Names and Numbers
ICT – Information and communication technology
ICT4D – ICT for development
ICT4E – ICT for education
IDN – Internationalised domain names
IDRC – International Development Research Centre
ISO – International Organization for Standardization
ISP – Internet service provider
L1 – First language
L2 – Second or additional language
LED – Light-emitting diode
LIP – Language Interface Packs
LWC – Language of wider communication
MSKLC – Microsoft's Keyboard Layout Creator
MT – Machine translation
NICI – National information communications infrastructure
OLPC – One Laptop per Child
PAL – PanAfrican Localization project
PDF – Portable document format
PLETES – Politics, languages, economics, technology, education, sociocultural model

RFC – Request for comment
 RIFAL – International Francophone Network for Language Management (Réseau
 International Francophone d'Aménagement Linguistique)
 SAT-3/WASC – South Atlantic 3/West Africa Submarine Cable
 SMS – Short message service
 STT – Speech-to-text
 TM – Translation memory
 TTS – Text-to-speech
 UCLA – University of California, Los Angeles
 UCS – Universal Character Set (another way of referring to Unicode/iso 10646)
 UNDP – United Nations Development Programme
 UNECA – United Nations Economic Commission for Africa
 UNESCO – United Nations Educational, Scientific and Cultural Organisation
 USA – United States of America
 USAID – United States Agency for International Development
 USB – Universal Serial Bus
 USINFO – American Information Web
 UTF – Unicode transformation format
 V-MAIL – Voice e-mail
 VOIP – Voice over Internet Protocol
 VSAT – Very Small Aperture Terminal two-way satellite ground station

Introduction

GIVEN THE GROWING NUMBERS of computers and the penetration of the internet around the world, the localisation of ICT and the content it carries into the many languages people speak is becoming an increasingly important area for discussion and action. Localisation, simply put, includes the translation and cultural adaptation of user interfaces and software applications, as well as the creation of internet content in diverse languages and the translation of content from other languages. Defined in this way, it can be appreciated that localisation is essential in:

- ◆ Making the new ICTs' more accessible to the populations of poorer countries, for whom ICT is supposed to offer new possibilities for advancing development;
- ◆ Increasing the relevance of ICT to their lives, needs and aspirations;
- ◆ Ultimately, bridging the 'digital divide'.

Africa, which is recognised today both as a continent struggling with aspects of its own development and one where the use of ICT lags behind most of the rest of the world, is beginning to see some attention to localisation. This is gradual, with projects limited to certain regions, sometimes the result of personal initiatives, but generally without much in the way of organisation, resources or long-term planning. In addressing this situation, this book and the PanAfrican Localization PAL project are motivated by the intention of assisting the region in maximising the potential of ICT for development (ICT4D) by identifying ways of supporting effective and sustainable localisation.

This book therefore seeks to explore the following four sets of questions:

- ◆ Why is localisation important? What are the barriers to greater use of African languages in computing and the internet? How do these affect the potential for localisation?
- ◆ What is actually being done to increase localisation? By whom, for which languages, and in which countries are such efforts being made? What challenges are being encountered, and what solutions are being found?
- ◆ What future trends can be anticipated? Which areas should receive priority?
- ◆ How do these relate to one another, and how should they be addressed in localisation work?

To accomplish this, it is necessary to consider the situation in Africa, as well as among African-language speakers abroad, with respect to information on the languages spoken, the body of speakers, language and educational policies, and basic information on current ICT situations, policies, plans and initiatives. Existing localisation initiatives should also be considered, including where these are taking place and who is driving them. Two broad areas – language and technology, as well as their relationships with the social and cultural context in which they are used – represent the fundamental preoccupations of localisation, but other factors such as economics, politics and education also have to be taken into account.

It is one of the premises of this work that a broader view of apparent and expressed needs can serve to contextualise such information and more fully inform programmes to assist localisers and ICT4D projects. Understanding the basic information is largely a matter of drawing on existing research on languages and ICT in Africa, which provides the context for discussing and planning localisation.

Systematically uncovering what is being done towards localisation is more difficult, as such activities are often not publicised and thus tend to remain out of view to those (sometimes even in the same country) who might be interested in knowing about them. Attention is needed to identify trends and potentialities – that is, where localisation might be headed and what that means for the future. ICT is a new and unavoidable fact of life for Africa, no less than for other regions, although the particular issues and needs may differ from area to area. Africa is one of the most multilingual regions of the world, so the meeting of technology and language would seem to be of particular consequence for African development, even though that fact is not yet receiving the attention it deserves.

This in turn relates to the visions one may formulate about the current and potential direction of technological change, since the evolution of ICT is constant and rapid, and the object of localising and utilising it for development in Africa cannot remain limited to catching up with practice and applications in other world regions.

Beyond that, and returning to the basic realities one encounters in Africa, one cannot separate the tasks and objectives of localisation from the broader development and education efforts, policy contexts and socioeconomic dynamics at play on the continent. This is especially true as one considers, on the one hand, the sustainability of localisation and associated long-term planning, and on the other, the role that the localisation of ICT could play in addressing larger problems of development.

In order to achieve the goals of this book, therefore, one must always bear in mind the main components of localisation – namely, language, technology and their sociocultural contexts – as well as the relationships among those and several other factors that affect the possibilities for localisation and its actual implementation, and that are in turn affected by the process and achievement of localisation.

For that set of relationships, this book will introduce the concept of 'localisation ecology' to account for the key factors, facilitate discussion of their interaction, and call attention to how planning and implementing localisation can and should consider these.

Another issue to take into account in documenting a rapidly changing field is that the information will soon become outdated. As Esselink (1998) put it in his first book on localisation: 'Writing about software localisation is like fighting against time.' In this case, we are seeking as full a picture of the current situation as possible – in effect, a composite of snapshots – with the understanding that much of the information will soon be out of date. The advantage of this approach, despite the inherent problem of documenting rapidly changing technology, is that it is useful to understand connections in the system at any particular time, for the purposes of comparative study and for understanding the localisation ecology. The decision to publish this work in print form, despite the rapidly evolving nature of the subject matter, was based on the consideration that the print form could reach audiences in different ways than a purely electronic publication would have. At the same time, given the advantages of being able to update the information as changes occur, the companion website can extend the usefulness and relevance of the printed volume.

A further task is therefore to determine the sorts of information resources and practical skills that are needed to assist and facilitate localisation work on the ground in Africa. It is hoped that these findings will contribute to the evolution of the localisation resource website that complements this book (<http://www.panaf-rilion.org/>).

This book is organised into several chapters. Following a background discussion on localisation in chapter 2, localisation ecology is presented and modelled in chapter 3. The next three chapters consider the linguistic context (chapter 4) and technical context (chapters 5 and 6) of localisation in Africa. Chapters 7 to 9 discuss aspects of enabling systems. The following four chapters deal with aspects of localisation. These are followed by a summary, recommendations and the conclusion.

This page intentionally left blank

Background

THIS CHAPTER INTRODUCES THE importance of African language use in ICT, responds to the question of why it is important to localise in African languages, defines localisation as used in this book, discusses the regional context of the research, and outlines the approach of the book towards localisation in Africa.

Importance of African languages and implications for ICT

As the information revolution worldwide becomes increasingly multilingual, and as the presence of the new ICTs in Africa extends to larger areas beyond the capital cities, there is a growing need to accommodate the use of diverse African languages and greater potential to tap the linguistic wealth of the continent for development and education. There are two aspects to this issue: European languages cannot meet all of Africa's needs, and African languages have much to contribute.

It is generally agreed that the availability of software and content in the languages most familiar to users is an essential element in the adoption and optimal use of computers and the internet. One might add that in the context in which people may speak several languages – as is common in Africa – the option of using different languages is also empowering.

Accommodating the languages most familiar to people is a consideration of primary importance in any efforts to use ICT for development. This should come as no surprise, as education and communication are generally easier in the first language (L1) than in languages that people acquire later. Furthermore, at a community or societal level, L1s are considered a central and indispensable aspect of social and cultural systems.²

ICT was originally introduced to Africa and Arabic-speaking regions in English and French, as well as in Portuguese and Spanish in certain sub-Saharan countries. The same languages, of European origin, were used in colonising these regions and have served as official languages since their independence, especially south of the Sahara. Such languages will be referred to in this book as European

(or europhone) languages of wider communication (ELWCs).³ One of the problems of relying on ELWCs is that a large majority of people on the continent either do not speak these languages or do not speak them well.⁴ Even if they did, having computer access and internet content only in ELWCs would be a limitation to populations that also speak other languages.

In any event, the use of ICT in Africa's indigenous languages should not be considered merely as a means of compensating for people's lack of knowledge of ELWCs, nor as a second-best or interim solution for such people until knowledge of ELWCs increases and improves.⁵ It is also a question of fairness with respect to access, which is a long-term practical issue, since it is difficult to imagine that Africans, any more than the populations of any other region, would universally be comfortable or efficient in using ELWCs in ICT to the exclusion of their LIS.

Using ICT in Africa's indigenous languages is a solution that also opens up new possibilities for more effective use of the technology by the most highly educated, thus complementing and expanding upon the potential offered by applications in ELWCs.

Challenges

At the same time, the sheer number and diversity of languages on the continent – over 2 000 languages according to *Ethnologue* (Gordon 2005), which is about a third of all living languages in the world – poses a challenge for localisation efforts and indeed for educational programmes to support them. The fact that many of the languages that have been considered to be separate also fall into clusters of very closely related and interintelligible languages shows that Africa's linguistic complexity has multiple dimensions.

Initiatives that aim to expand the use of ICT in Africa for development, education or other purposes are beginning to recognise the need to respond to these sociolinguistic realities. Such efforts are benefiting from advances in internationalisation of the technology, the greater use of Unicode (ISO 10646)⁶ for handling diverse scripts and extended characters, and the availability of utilities for creating keyboard layouts.

However, there are still a number of hurdles. Some are technical, relating for instance to the use of extended character sets and Unicode on older computer systems and European keyboards. Some hurdles relate to economic factors, such as the costs of translating content. Others are social in nature, relating to education levels, as well as the sometimes negative attitudes towards African languages among foreign development and education experts and even some native language speakers.⁷ In some countries, government language and education policies disfavour African languages, which in turn has an impact on ICT usage.

What is localisation?

The term 'localisation' is used in various contexts relating to ICT, but the definitions revolve around the adaptation of user interfaces and digital information to the local modes of communication, culture and standards. Daniel Jacob (2004) offers a broad interpretation that defines the object of localisation as: 'the transfer of cultural consciousness into a computer system, making the computer a natural extension of the society it serves'. In practical terms, the key consideration in localisation is invariably language.

The concerns related to localisation were arguably inherent or latent in computer technology itself from the very beginning. In other words, it was inevitable that computing would eventually enable the handling of human language, that questions would then arise about the choice of languages, and that the possible use of additional languages would be raised by users from diverse linguistic backgrounds. As computers became more readily able to convey images, sounds and styles of presentation, it was also inevitable that issues of cultural appropriateness would follow.

In practice, localisation is both a technical set of approaches and techniques for adapting software and content to particular languages and cultures and, more broadly, an enterprise activity that incorporates those technical dimensions, linguistic information planning and organisation necessary to make it happen. Altogether, localisation aims to facilitate the use of target languages in ICT and can further be understood as an active component of wider efforts to adapt science and technology to diverse societies and cultures.

Localisation as a technical task

Computer systems, and ICT in general, involve two levels of consideration: hardware and bits (binary encoding). Together these define the technical possibilities for localisation.

At its simplest, the hardware aspect of ICT can be understood as involving devices and connections. The devices – computers as well as increasingly powerful handheld devices – can operate independently for certain purposes, including the storage and manipulation of data such as text, spreadsheets and other files. They also can connect to a network linked to other devices – the internet (or an intranet) – for the retrieval and exchange of information such as e-mail, webpages and streaming media. Localisation relates to both the independent and networked aspects of ICT.

In order for one to make use of the hardware, the bits that are used at the most basic level, both to encode and manipulate information, and to write the

software for facilitating that encoding, are organised in forms that permit human interface with the devices and networks, and with the storage and transmission of information. In other words, two aspects are involved: the interface for accessing and using the technology, and the information content, including documents and data. Table 2.1 illustrates and cross-indexes these two levels – the two fundamental categories of hardware and the two fundamental ways in which the technology is used. In effect, by considering the two in a matrix, it is easier to understand the aspects of ICT that are involved in localisation.

From this analysis, we can identify three separate but overlapping concerns, which are listed as follows and then further discussed:

- ◆ Equipping systems deployed in various localities – or actualising their existing capacities – to handle local language needs. This facilitates the production of documents and the display of multilingual web content (related to item 1 in Table 2.1).
- ◆ Production of web content – original and translated – for diverse audiences in languages and formats that they can understand (related to items 2 and 4 in Table 2.1).
- ◆ Localisation of user interfaces on individual devices and the internet (related to items 1 and 3 in Table 2.1).

The PAL project focused on all three of these concerns, but the localisation of interfaces (particularly software) is pivotal, as it is the logical extension of efforts to equip systems to handle local language needs and it has the potential to facilitate the production of localised content.

Equipping systems

Equipping systems relates mainly to actualising the potential of computer systems to handle local languages in various ways, notably non-ASCII (American Standard Code for Information Interchange) text.⁸ The main issues are fonts, input and display.

TABLE 2.1 *Dimensions of localisation*

	Interface/access (how we interact with the technology)	Information storage, communication, retrieval (what we use the technology for)
Computer (individual piece of hardware)	1. Operating system, software for various purposes, keyboard, display	2. Documents and files of various sorts, created by user(s)
Network (connections among computers: the internet and intranets)	3. The above (under item 1) plus specialised software resident on servers such as search engines, databases	4. Web content, remote storage, ability to link individual computers in real time

Many African languages are written with an extended Latin script, while a number of others, such as Arabic, use non-Latin scripts. For these languages – unlike languages whose orthographies use essentially the same character set as Western European languages (as is the case with many languages in southern and East Africa) – the advent of Unicode represents a new era of possibilities. The necessary basics are an adequate selection of complete fonts and provision for standardised and user-friendly input, which relates mainly to keyboard layouts but will ultimately also include speech recognition software. (Keyboards and input methods are discussed in more detail in chapter 8.) The first step of localisation for these languages is in effect this ‘last mile’ of internationalisation, which in turn refers to the process of improving computers and systems to accommodate the diverse language needs of the world.

Fonts are effectively the first issue, since without fonts that include the necessary extended characters or non-Latin scripts, software applications will not fully or correctly display text in a number of languages. This means Unicode fonts in which all characters are encoded according to the Unicode standard, since legacy 8-bit fonts – while they may be able to display the characters and diacritics used in whichever writing system they were designed for, and be useful for someone producing documents meant only for personal use – are not readable on systems on which those fonts are not installed. Basically, 8-bit fonts are not intercompatible, because each uses the limited number of codepoints for characters in a different way, while Unicode in principle provides a single code-point for each character in every writing system (as discussed in more depth in chapter 6).

For the input of text in languages that use non-ASCII characters, specialised keyboard layouts are necessary, and these may be created for languages or groups of languages for which localised software does not yet exist. Apart from the capacity to handle text, the capacity of systems to permit users to create and use multimedia that does not rely solely on text is another important, although sometimes overlooked, consideration.

At the same time, it is recognised that there are many older computers in existence in Africa – often the result of donations of used equipment – that cannot handle Unicode and may be limited in other respects (as discussed in chapter 5).

Content

Content is usually taken to mean web content – the information conveyed on the pages of sites on the World Wide Web. More broadly, we may understand it to include information stored as documents or data on computers or conveyed over the internet by other means, such as e-mail. The latter is of interest in measuring the use of diverse languages and the demand for capability to do so.

The production and display of information via the web that is relevant and accessible to users is facilitated by the considerations related to equipping systems, as discussed in the previous subsection. The choice of languages is obviously a key consideration, since it is by means of comprehensible idioms that ICT can convey information for development or other purposes. Additional considerations, such as the cultural appropriateness of themes and images, and the approach to communication within the language (for example, dealing with dialects, and contemporary as opposed to formal styles), are also important aspects of localised content.

One can divide localised web content into two parts based on origin, namely that produced locally and that produced elsewhere but targeted at the local audience. Both are important, but our main concern is the former. Ballantyne (2002) discusses content in terms of this division as well as the target of the content. It may be useful to seek to develop collaboration among local and international content developers, wherever they may be, in terms of how best to address their common intended audience.

Localisation of content in major international languages is the focus or concern of a wide range of commercial and non-profit activities. This represents a significant industry, with a number of commercial and professional associations,⁹ but has not touched Africa in any significant way, apart from localisation in Arabic and ELWCS, and has not seriously addressed African languages.

At present, there appears to be little web content of any sort developed in African languages, whether in Africa or elsewhere, and relatively little content in any language coming out of Africa. The issue of localised content is discussed in more detail in chapter 10.

Localisation of user interfaces

The localisation of user interfaces includes the translation of basic computer software such as browsers and word processors into different languages, including commands, dictionaries and help files. (The capacities of software to handle diverse language needs are considered as an aspect of equipping systems.) Increasingly, it also involves interactive web content (so-called 'Web 2.0').

In addition to translation, other important issues in localisation include development of terminology (which in some ways is an extension of translation), conventions for the display of certain information and the cultural appropriateness of themes used in the software. In many cases, localisation may relate to some, but not all, of these issues. Localisation in Arabic, and to a very limited extent in some of the more widely spoken African languages, is being undertaken by commercial software companies, notably Microsoft as part of its local language programme. While such initiatives serve to promote localisation in general, they are concerned only with major languages and large markets. Localisation may in some cases entail

higher costs for users than can be supported, especially in ICT4D contexts. This has led many to emphasise the use of free/open-source software (FOSS) as a less expensive way of promoting local expression and localisation in more and more languages.

The localisation of computer software has various aspects and presents some complexities. One might suggest that in this area, we are speaking about the following four things:

- ◆ Operating systems (for example, Windows, UNIX, Linux)
- ◆ Software applications or packages completely translated (such as OpenOffice, Microsoft Office, or specialised programs such as web browsers)
- ◆ Partial package, in which only the most frequently encountered commands are translated. This is the strategy used by Microsoft in its language interface packs (LIPs), which will typically involve translating only the most-used commands and help files, or about eighty per cent of the total.
- ◆ Production tools for the target language without commands or help files translated, in other words, including only language settings, input methods (keyboard layout) and spelling dictionaries. This may be regarded as an extension of the category of equipping systems.

These aspects, with the exception of the operating system, represent different levels of localisation of commonly used software. Most of the current focus with regard to African languages is at these levels. While operating systems are translated into the major international languages, relatively little is being done in this regard for African languages.

Although Microsoft has increased the number of languages in which it offers its software and has announced projects for others, there is still considerably more localisation in FOSS. The OpenOffice suite is being localised into about 100 languages, and work on some of these is already complete.¹⁰

Other proprietary general office software, such as Corel's WordPerfect, has focused on major languages, sometimes also developing production tools in less widely used languages.

The evolution of increased software localisation has thus far been driven mainly by improved internationalisation (as already discussed) and increased demand in major markets. With regard to text, progressive improvements in the ability to render text in diverse writing systems – notably through Unicode as the single standard for all scripts – facilitates multilingual production and use of content. (Unicode and scripts are further discussed in chapter 6.) Moreover, as the linguistic background of users diversifies (as ICT becomes less expensive and more readily available in more places and forms), the demand for diverse language

interfaces intensifies and the potential pool of contributors to fill the demand for localisation expands.

Localisation has also been undertaken by various companies and projects internationally for their specific uses, although with some exceptions, localisation has not featured prominently in ICT4D projects in Africa.

Localisation as a project

Localisation, in its broader sense of a process and enterprise, takes into consideration several other matters, such as the following:¹¹

- ◆ Factors necessary for localisation, including a standardised orthography, locale data, and organisation and resources to accomplish localisation in the stricter technical sense;
- ◆ Aspects of sustainability in the long term, including follow-through and marketing of localised software and follow-up with the user community and with respect to updates;
- ◆ Attention to issues of user skills (ranging from basic literacy to computer literacy);
- ◆ The impacts of the localisation of ICT on other aspects of society, economy and culture.

A framework to facilitate taking all these important matters into consideration is proposed in the next chapter.

Overlapping regional contexts: localisation where?

Africa is a multilingual continent, but there is no software or even internet content in the vast majority of its many languages – even in most of the major and more widely spoken ones. Linguistic diversity is evident in every country on the continent, resulting from the history of population movement and the overlay of colonial languages.

Sub-Saharan Africa

Most countries south of the Sahara have no single majority language. Several of the countries are characterised by having scores or even hundreds of different languages. The E1WCs introduced during colonisation – English, French and Portuguese – serve as official languages and facilitate communication to one degree or another across wide areas, but are primarily second languages (L2s) of the more

educated and urbanised segments of society, and do not have the same connections with African cultures as indigenous African languages.¹² Most of the people with less facility in the ELWCs live in rural areas and include a higher percentage of women than men.

Software and content in ELWCs cannot, therefore, satisfy the needs of the majority of the African population, and even in the limited (mainly urban) locations where they might, many people would effectively still be linguistically excluded, and everyone would have their language options restricted by the lack of content and software in African languages. In addition, any effort to use ICT for development purposes would be hindered to the extent that the working languages are limited to ELWCs. Expressed differently, the needs for localisation in Africa correspond to the hopes for ICT to play a full and effective role in development on the continent.

The numbers of computers and the quality of internet connections on the continent are increasing, along with interest in localising software and content in African languages. The reasons are not only related to development, but include the same reasons in which there is such interest elsewhere. The amount of internet material in African languages is increasing slowly, and there are active efforts to localise software, particularly in South Africa, East Africa and Nigeria. However, there have been limited connections among the diverse efforts, and there is limited knowledge beyond that of a few specialists.

This comes at a time when there is increased interest in localisation around the world in both commercial software companies (Bergmann 2005) and the FOSS movement. The time is opportune to facilitate localisation in Africa, given the need, the budding local interest and the international resources potentially available.

Arabic-speaking world

The need for localisation in Arabic is very real but of a different nature, even though many countries in this region were also colonised, had a similar overlay of English and French, and were first introduced to ICT in those languages. Unlike the languages of sub-Saharan Africa, Arabic already has a significant amount of localised software and content, as one would expect for a major international language.¹³ Furthermore, the Arabic-speaking world, including the countries of North Africa, generally has better infrastructure and ICT indicators than Africa south of the Sahara. The challenges in this region are thus less daunting than in sub-Saharan Africa.

Nevertheless, the range of localised software is still arguably limited, and a corresponding level of localisation dealing with local themes and idioms is lacking. Building the capacity of developers to localise Arabic software and content for their diverse user communities, particularly those outside the major cities, is a goal in

such cases, and will be the focus of the work of the project with respect to Arabic. There is also a need to produce Arabic electronic dictionaries for FOSS applications such as OpenOffice; these could be localised to particular countries, much as English or French dictionaries differ among locales.

Within the Arabic-speaking world, this book focuses on the countries of North Africa, while acknowledging the important linguistic, cultural and historic connections of this region with the rest of the Middle East.

Who localises?

The question of who benefits, or potentially might benefit, from localisation has already been raised in discussing why localisation is important. It is also useful to briefly consider who does localisation and would thus most immediately benefit from increased attention to localisation in Africa in terms of information, networking and tools.

The question has as many dimensions as there are types of localisation, yet a simple answer might be that anyone that is motivated to connect African languages with the content and interactive language of ICT, that has the knowledge and means to do so, and that actually initiates or participates in some aspect of localisation is a localiser. The profile of a localiser would also include higher than average education, a working knowledge of ICT, and knowledge of at least two languages – a dominant language (or languages) used in ICT and one or more in which to localise. This is a very select group of people, especially in Africa.

In terms of the origin and location of localisers, one might identify three broad categories:

- ◆ Africans in Africa;
- ◆ Africans residing in other parts of the world;
- ◆ Non-Africans with a strong knowledge base and interest related to African localisation, including knowledge of a relevant language.

In some parts of the continent that are better off economically, educationally and in terms of technical infrastructure, such as North Africa or South Africa, the first category is the most significant, although the latter two categories can in some ways reinforce the first. In certain contexts, however, the categories of localisers from outside Africa may initiate or drive localisation efforts – for instance, African expatriates may develop a project in their home languages, and commercial interests or international development organisations may play a role.

Another valid characterisation would be to say that content localisers require language skills but less depth in technical skills, while software localisation requires

both. The fundamental concern of equipping systems – whether at the level of designing a project using ICT for development or for education, for instance, or managing a cybercafé – mainly requires an awareness of internationalisation issues and familiarity with local language needs.

People who localise also range in skills sets, so that groups of individuals with complementary skills in language and technology may form a logical team. This implies some level of organisational skills to coordinate efforts and plan actions. Since localisation implies products destined for a market (whether those products are free or not), marketing is another concern. This means that the question of who localises may also involve people who each bring some of the necessary skills to a collaborative effort.

Returning to the description of who localises, we might then emphasise the motivation to work on localisation as the first defining characteristic of people who localise. Such people probably work most effectively in groups with complementary skills sets, and the effectiveness of such individuals and groups would be enhanced by support systems in the form of education and training opportunities, policies emphasising the importance of localisation, information resources such as those developed by the project, and tools to facilitate the work of localisation for people without high levels of technical skills.

What is the current state of localisation across the African region?

One of the goals of this book is to provide a better idea of what is happening with respect to localisation in Africa in order to inform localisers, policy-makers and experts in ICT for development, and to serve as a benchmark of sorts for evaluating the effectiveness of future localisation efforts.

In general, one can say that there is considerable potential for localisation in Africa, but that despite growing interest, the current level of activity tends to be small, with some differences between regions in the degree and character of localisation initiatives and related local or multilingual ICT efforts.

Chapters 10 to 12 cover recent and current localisation efforts under three headings: the internet, software localisation, and mobile technology and specialised applications. Chapter 13 addresses the topic of achieving sustainable localisation. Chapters 3 to 9 set the context for these discussions.

This page intentionally left blank

Introducing 'localisation ecology'

ONE OF THE PREMISES of this book is that there is more to localisation at any level than might at first be apparent. This is especially true in a complex environment characterised by multiple languages, weak or non-existent standards (such as for orthographies), fluctuating policies, many cross-border languages, sociolinguistic trends, uneven access to computers and the internet, and a range of factors that define the digital divide and other so-called 'analogue divides'.¹⁴ Not only are the technical and project aspects of localisation affected by such factors, but localisation has the potential to impact upon all these factors. Indeed, localisation is ultimately a key to bridging the digital divide by improving access to ICT and increasing its relevance, thereby enhancing the potential of this technology to assist development and education.

An ecological perspective on the environment for localisation

Localisation is one of the keys to bridging the digital divide – that is, to improving access to ICT and increasing its relevance. At the same time, the localisation process is affected and conditioned by various factors, including most notably those that define the digital divide and other analogue divides that separate the more and less advantaged parts of the world. Understanding how these factors – individually or in combination – constrain or facilitate localisation initiatives is important both for profiling the actual state of localisation in Africa and for suggesting how to assure the sustainability of localisation projects. This is arguably a complex system, with characteristics that behave in ways that are more than the sum of its parts.

To help understand this system, we propose using an ecological approach, which we refer to as 'localisation ecology', as a conceptual framework for accounting for the impact and interaction of the diverse factors involved. We introduce a possible model to facilitate such applications.

An abstract framework of this nature, although it does not address the immediate and relatively straightforward tasks of translating software or content, is nevertheless also intended to have practical utility in planning for longer-term

localisation projects. The localisation ecology framework is further intended to provide a schema for ongoing evaluation of localisation in Africa. Beyond the framing question of the digital divide, this ecological model can hopefully also be utilised in considering the impact of localisation on the analogue divide.

Ecology

The concept of 'ecology', which originally referred to the study of the interaction of organisms in the natural environment, has gradually found wider application, first in various models to describe interactions between human societies and the natural environment (human, social, cultural, landscape and political ecologies),¹⁵ and later in more abstract conceptualisations to describe multifaceted processes in society and individual life (for example, family ecology, cognition and linguistic ecology).

This evolution and broader application of ecology as a conceptual tool is entirely appropriate, given its origins in holistic and systems thinking in the early 20th century.¹⁶ We continue this process by proposing 'localisation ecology', and in so doing, find that other work provides a ready foundation.

In 1970, Einar Haugen proposed using the metaphor of ecology to understand the dynamics of languages and how these relate to other factors (Fill 2001). This ushered in a period in which the study of language ecology or linguistic ecology was popular.

The notion of linguistic ecology has not been without its critics – Harlech-Jones (1993), for instance, is sceptical of the extent to which the analogy to a natural ecology is meant and focuses instead on a conflict perspective to make sense of the situation of African languages. Given that ecology has been more widely used over the last few decades, however, it seems altogether appropriate for organising thinking about the interrelationship of factors in localisation without presuming the nature of those relationships. Although there have been fewer publications in recent years carrying either of these terms ('language ecology' or 'linguistic ecology'), it is an accepted approach to analysing languages in relation to one another and to various social factors.

The idea of applying the concept of ecology to various aspects of the use of ICT and the digital divide, although obviously more recent, is also not new. For instance, Matwyshyn (2003) discusses the gender gap in the use of technology in terms of human or social ecology. Another example is the World-Information.org website, which discusses 'digital ecology' in terms of 'information ecosystems' that aim at 'understanding the production, distribution, storage, accessibility, ownership, selection and use of information in technologically determined environments'.¹⁷

Our interest in developing a model of localisation ecology is in some ways anticipated by Robert Chaudenson's (1987, 2003) discussion of 'integrated language

management¹⁸ in terms quite similar to those of ecology. His particular focus on the factors related to the choice of orthography is still relevant and can be expanded to other technical and linguistic considerations from the point of view of language planning and management (see also the discussion of the PLETES model later in this chapter).

The ecology of localisation

Haugen (2001/1972) originally defined language ecology as: ‘the study of interactions between any given language and its environment’. This definition serves as a useful starting point for this discussion.

Localisation is undertaken within specific contexts. The general rationales for localisation have already been discussed. However, in specific countries or for specific languages, the motivations and hopes of localisers may address conditions, needs, opportunities and aspirations that are particular to their area. In addition, localisation is undertaken in an environment of sociocultural, linguistic, policy and legal, educational, technological and economic factors and trends. The framework of localisation ecology ideally provides a way of accounting for such factors and trends, as well as their potential interaction.

Localisation ecology, therefore, is first of all a way of understanding:

- ◆ Factors that affect the potential for localisation and specific efforts to localise;
- ◆ How these factors affect localisation (by facilitation or limitation, for example);
- ◆ Interaction among these factors in ways important to localisation.

It is important, however, to remember that localisation is not only affected by various factors, but is also a process that can introduce new dynamics into other spheres of activity, such as the use of ICT, education, the development of languages and the evolution of the sociolinguistic situation. The model, like the real-world situation it is intended to reflect, is therefore thoroughly interactive.

When we consider a particular localisation effort, we usually think of a group of people or an organisation dealing with a range of specific tasks and requiring the input of some resources and perhaps advice and information to achieve them. In fact, the effort is dependent on other parts of what is actually a system. If this fact is not clear when the effort is launched, it is likely to become so later, as obstacles and new phases of localisation are encountered. The immediate tasks of localisation are therefore simply part of a bigger system and thus aspects of a range of concerns that must be taken into account for successful and sustainable localisation initiatives.

The encoding of the Tifinagh script into Unicode, although not strictly speaking a localisation project, is a case in point. The proposal to encode the script had not progressed beyond an initial draft for several years. Then, in 2003, a

decision by the Moroccan government to use Tifinagh in teaching the Tamazight (Berber) language in schools highlighted the need for capability to use the script freely in computing and the internet, which was hindered by reliance on legacy encodings. In effect, an education policy decision revived the effort to encode Tifinagh, and these two factors along with others may lead to localised content and software in the Tamazight language, which in turn will have other outcomes.

Other examples similarly illustrate the interaction of factors. In cases where a government or a donor announces a project to establish rural telecentres (as in Ghana in 2005¹⁹) or to supply computers to schools (as in Rwanda in 2005) (Nsengiyumva & Stork 2005), the availability or lack of localised software makes a big difference to the options for the project and what it can provide. Such programmes could in turn provide impetus and resources for localisation.

Localising software also depends on some levels of standardisation of orthographies, terminology and dictionaries, which might in turn perhaps benefit from government language policies as well as other institutional programmes on local languages (for example, at universities, literacy agencies or non-governmental organisations such as SIL International), but which might also be catalysed by localisation initiatives.

Such examples point to another important feature of such an approach to understanding the interaction of factors, namely scalability. Decisions and actions related to ICT and localisation in Africa may be taken at various levels, whether at a localised level or a broader national or regional level. The interaction of donors and national governments on questions of ICT for development or education, for example, affects the environment for interactions at local levels. Conversely, the localisation of software in specific languages – a process that may involve only a few actors at a local level – can impact the discussion of language and ICT policies at national and regional levels, possibly even involving international donors.

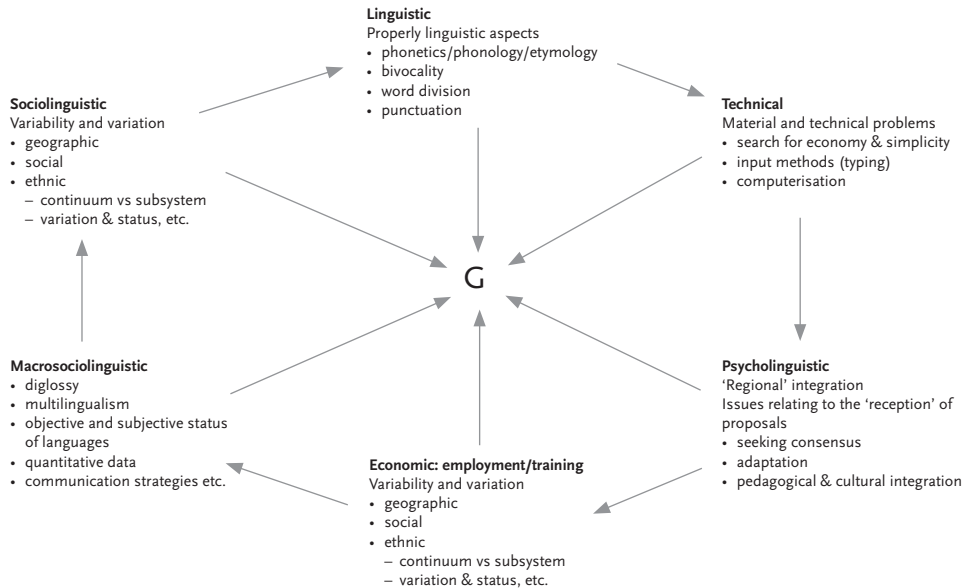
The PLETES model

A number of specific factors may be discussed in imagining a dynamic and scalable model of localisation ecology. In any ecological system, one could say that everything is directly or indirectly interrelated, but for the purposes of understanding and analysis, it is helpful to identify or specify a limited number of key factors (or categories of factors) and relationships. This was the approach of Otis Duncan (1959), for instance, who considered four key elements in his human ecology analysis, namely population, organisation, environment and technology.²⁰

Perhaps the closest model to what we propose is that developed by Chaudenson (2003) for language management. Even though it was not framed as

FIGURE 3.1 *Model of language management*

Language planning and development: a model for the determination of a planning 'operation' (here the choice of a grapheme 'G' but the principle of the model can be maintained for any language planning action).



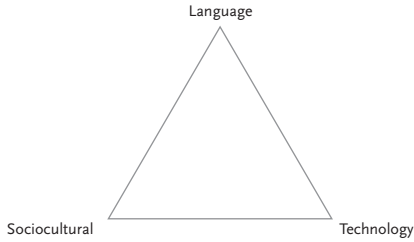
Source: Translated from Chaudenson (2003)

an 'ecology', it does illustrate the interrelationship of various factors in a decision relating to language management. The model includes the following elements: linguistic, technical, psycholinguistic (individual reactions), economy (in the sense of economy of usage) and sociolinguistic (Figure 3.1).

In any such model of interacting factors, there is always a degree of simplification and a selection of aspects to emphasise the particular type of situation to be described. Chaudenson (2003) focused on a relatively specific matter in which four aspects of linguistics (namely, aspects of the language itself, sociolinguistics, psycholinguistics and economy of use) are considered separately alongside the technical factor. Social dimensions – in this case, the ways in which people interact with an element of orthography – are implied in other factors.

In the case of localisation ecology, we can start by suggesting that the fundamental factors are language, technology, and society or sociocultural aspects. As in the Chaudenson model, the latter is sometimes considered to be implied, but it merits attention as a separate category. Indeed, although the two factors of language and technology are the immediate focus in accomplishing the translation part of a localisation project, the social and cultural dimension is at the same level

FIGURE 3.2 *Three basic factors in localisation ecology*



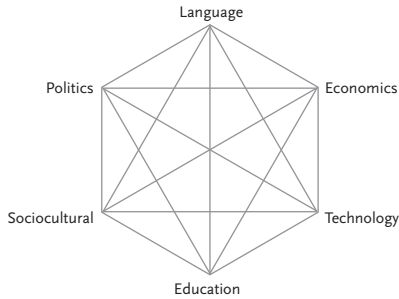
of importance when one considers the user dimensions as well as the impact of localised technology on society.

Each of these three categories is very broad and includes subdivisions, which are useful for highlighting the importance of interrelationships. It is possible, for instance, to develop ways of using a language on a computer without considering the users (other than those working on the project), or to develop new systems for users without considering the dimension of languages (other than a dominant one, as will be discussed with respect to 'digital divide' projects). Figure 3.2 is a simple representation of these three factors and their interconnections.

We also know that other factors affect the potential and results of localisation – in effect, one must 'think outside the triangle'. Among these factors, three emerge as especially important: policies and the process that produces them (politics); financing, markets and resource availability (economics); and the schooling and training of people in general skills such as literacy and the use of ICT (education). By adding these three categories, we then have six headings for factors, or groups of factors, that can be considered as key to localisation:

- ◆ Political: policies, decision-making processes and the interplay of interests leading to those, the legal and licensing environment;
- ◆ Linguistic: the linguistic situation in the country or region and aspects of each language, the number of languages spoken, their distribution and body of speakers, whether there is a standardised orthography for each language, whether the languages are characterised by diverse dialects;
- ◆ Economic: standards of living, resources available for various kinds of business, public, social and philanthropic investment, individual and family income levels;
- ◆ Technological: electricity and communications infrastructures, availability of computers (and types and kinds of operating systems), internet connectivity, the ways in which these factors differ across the territory of a country;
- ◆ Educational: systems of education (whether formal or informal), school infrastructure;

FIGURE 3.3 *The PLETES model*



- ◆ Sociocultural: demographics, social structure, ethnic groups, culture(s), popular and individual attitudes.

These six categories and the connections between and among them make the model a useful tool for understanding the environment for localisation. For convenience of reference, we may refer to it as the PLETES model from the first letters of each of the six factors (see Figure 3.3). The model is not intended to be definitive – there may be other ways of illustrating the same issues and the relationships among the key factors – but it serves at least as the basis for discussion and as a means of keeping the key factors in plain view during the discussion.

A list of combinations of pairs of factors helps to show the coverage of the PLETES model. Although some of these combinations may not directly affect the process of localisation, all of them in one way or another shape the environment for localisation, especially in relation to projects undertaking localisation and their potential for sustainable results:

- ◆ Language and society: sociolinguistics and applied linguistics; numbers of languages, distribution of speakers, attitudes about languages; the ways in which different languages are used in different situations and by different groups; languages in popular culture; sociolinguistic factors as the key to understanding the expressed and latent needs for localised interfaces and content, and the way in which people receive the products of localisation;
- ◆ Language and education: literacy rates, by language in a multilingual context; literacy in first (spoken) languages as an important consideration in gauging the immediate potential usership of localised software; language(s) of instruction in formal education systems, which will have an effect on literacy rates in those languages as well as on the development of the languages;
- ◆ Language and politics: language policy and planning, from legislation to implementation, relating to official attitudes towards localisation as well as to matters such as the standardisation of orthographies and government

support for African language documentation, periodicals and other media that produce resources that can be used in aspects of localisation;

- ◆ Language and technology: ways in which the technology supports the languages, including Unicode, keyboard layouts, potential for software localisation and advanced applications; ways in which the languages support the technology, including terminology; the translation aspect of localisation; computational linguistics;
- ◆ Language and economics: resources for language work (including documentation, corpus development and terminology); economics of localisation;
- ◆ Sociocultural and education: rates of school attendance and completion; who is educated; numbers of people with skills in particular areas, which is a factor in knowing what proportion of the population could actually take advantage of localised software;
- ◆ Sociocultural and political: who makes the policies; the nature of the interests, which may be important in understanding official attitudes to localisation as well as matters such as language policy, and in turn suggests approaches to addressing those as needed;
- ◆ Sociocultural and technology: who has physical access and rights to use the technology; attitudes to technology; the impact of technology on culture and society;
- ◆ Sociocultural and economic: fundamental and generally longstanding socioeconomic issues, including the foundations of the analogue divide that often parallels or conditions the digital divide;
- ◆ Education and politics: educational policy;
- ◆ Education and technology: education about technology; technology in education; efforts to put computers in schools or give laptops to children, which are examples of cases where this is a primary dynamic;
- ◆ Education and economics: investment in education; budgets (for example, for schools, teacher training, materials development and books for students);
- ◆ Technology and politics: ICT policy and planning, including national information and communications infrastructure plans; issues related to software licensing and intellectual property;
- ◆ Economics and politics: economic policy, including development, budget and donor priorities;
- ◆ Economics and technology: the economics of ICT, including issues such as the relative resources available for investment in ICTs, the attractiveness of outsourcing strategies and the marketing of localised software.

In the real world, of course, many factors interact. For instance, sociolinguistic and technical factors contribute to popular impressions that ICT can be used only in

ELWCs and foreign languages. The educational system, by focusing on the official languages (ELWCs) to the exclusion or marginalisation of L1s, may reinforce this notion. Policies concerning language of instruction in schools and language policy generally mandate such approaches, and economic factors (including other development priorities, budget realities, shortage of trained teachers and costs of materials) limit the resources that can be devoted to L1 instruction. The level of literacy in L1s, then, may be a factor in manifest demand for localised software.

Another example is the way in which the distribution of speakers of different languages within a country or region correlates with access to computers and connectivity (with language, technology, society, policy and economics also being relevant). This is a concern when thinking ahead to marketing localised software.

Yet another example, in the case of the localisation of open-source software, is the way in which economics and policies, as well as the quality of translations, affect actual or potential competition with proprietary software.

A localisation effort might not actively consider all these factors, nor have to address them, but they are part of the environment in which it operates. On a small scale, it is certainly possible that motivated individuals and groups with the necessary skills and minimal resources can begin, and even bring to conclusion, the translation of some software without deliberation over such factors. As the goals become more ambitious, however, and sustainable results are sought, the environment for localisation becomes an increasingly important consideration, and a systematic means of viewing that environment becomes necessary.

Dynamic complexes within localisation ecology

In order to make sense of such a long array of connections, which becomes even more complex as one considers multiple factors, it is helpful to highlight several key relationships or complexes in the system. These include the relationships between language, technology and society; intersections and non-intersections between language policy and ICT policy; and factors in sustainable localisation.

Language–technology–sociocultural interrelationships

The language–technology–sociocultural interrelationships are illustrated in the triangle already introduced in Figure 3.2 within the larger PLETES model (Figure 3.4). While other factors cannot be ignored, this triangle represents the core set of dynamics of localisation: sociolinguistics (the languages that people use and how), the connections between language and technology, and the interaction of people with the technology. On the one hand, localisation as translation (language-

FIGURE 3.4 *The three key factors of localisation in the PLETES model*

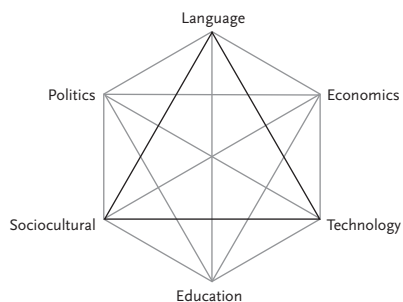
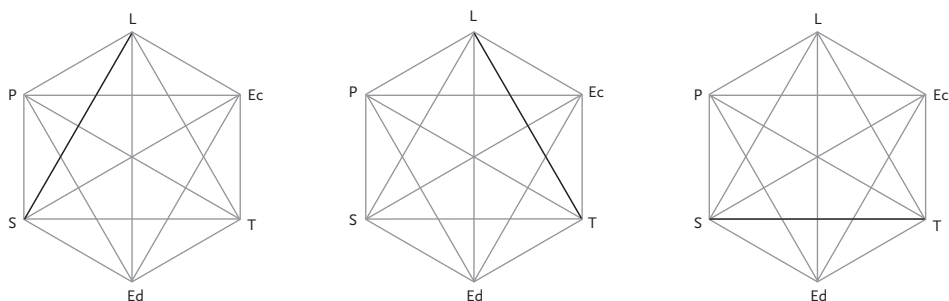


FIGURE 3.5: *Applied linguistics, translation in localisation, and social uses of ICT (illustrated from left to right)*



technology) involves attention to the cultural dimensions of communication, and on the other, information and communication technology is developed and localised for people (in the interrelationship between language, technology and sociocultural aspects). Developing keyboard layouts, for instance, entails more than simply finding a way to facilitate the input of the characters used to write a language, but also requires consideration of user expectations and existing practice.²¹

Most localisation initiatives, at least in their early stages, focus mainly on the set of concerns related to technical process, localisation for language and cultural factors, and end-users.

Applied linguistics, translation in localisation and social uses of technology

Another way of looking at the core set of dynamics illustrated in Figure 4 is to consider them as three dynamics or specialisations that often operate independently of one another: applied linguistics and sociolinguistics; the translation aspects of localisation; and the social applications of technology. Each of these is illustrated in Figure 3.5. In effect, localisation brings together aspects of all three.

Language policy and ICT policy

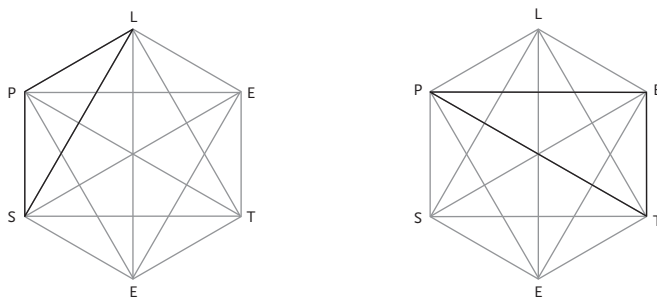
The model can also be used to illustrate two main policy dynamics that relate to localisation. Policy, particularly the ICT policies of governments and donors and the language policies of governments, may affect any or all of the main concerns of localisers. Language policy arguably involves mainly politics, sociocultural concerns and of course language. Other factors are also implicated, but these seem to be the main ones. It could be argued that technology policy is concerned with high-level economic concerns in terms of both the investments required and the anticipated returns. An example is the efforts to increase connectivity through a project such as the Leland Initiative (funded by USAID), in which the technical issue of expanding the bandwidth available to certain African countries was accompanied by consideration of the regulatory environment for the use of that bandwidth (in other words, policy and economic considerations). Here, too, other factors are also important, but arguably not to the same extent. The two are depicted in the PLETES model in Figure 3.6.

If we take this depiction of the key dynamics as representative of the priorities of ICT policy-making on the one hand, and language policy-making on the other, there seems to be little intersection between the two.

Localisation work, in effect, appears to fall between these two important areas of policy concern. This may signal a need to more effectively link ICT and language policy, perhaps in 'localisation policy' frameworks, in ways that highlight the missing connections on each side.

This apparent disjuncture between the main factors considered in each of the two types of policy and the processes and considerations involved in their formulation would seem to leave localisers with a fair amount of leeway in their activities as well as little support.

FIGURE 3.6 Comparison of the main concerns of language policy (left) and ICT policy (right)



Dynamics of digital divide projects

Efforts to expand access to ICT seem to have begun with the technology–socio-cultural dimension (supplying computers and connections to communities), but it would appear that they soon encountered or expanded to other policy concerns (including regulations and relations with the other priorities of the country), economics (in terms of costs) and education (in the form of training, often in relation to schools) (see Figure 3.7). The efforts thus overlap significantly with the primary concerns of ICT policy. However, to the extent that language is not actively incorporated into their models – and many projects are not actively engaged in this issue – there is no overlap with some of the primary concerns of localisation or of language policy.

Dynamics of sustainable localisation

The concept of localisation ecology is also useful in considering the important issue of sustainability in localisation, which operates in two approximately sequential phases. The first part of sustainability may be termed ‘follow-through’ (namely, to complete a set of localisation projects such as software translation), and the second as ‘follow-up’ (namely, marketing the localised software, establishing it in the field, and responding to the reactions and suggestions of users). Follow-through requires attention not only to technical and language dimensions but also to other factors, the most important of which may be financial (in that ensuring the resource flow necessary to sustain the initial effort becomes critical when early enthusiasm encounters various limitations). Follow-up, in principle, involves issues such as communication and marketing, user skills and even policies. Figure 3.8 shows the interrelationship of factors for sustainable localisation including those

FIGURE 3.7 *Digital divide projects: from basic to more complex dynamics, without language*

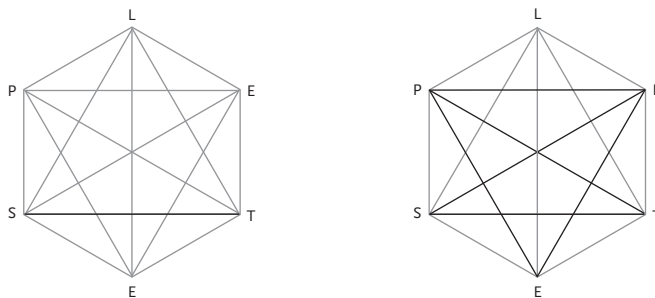
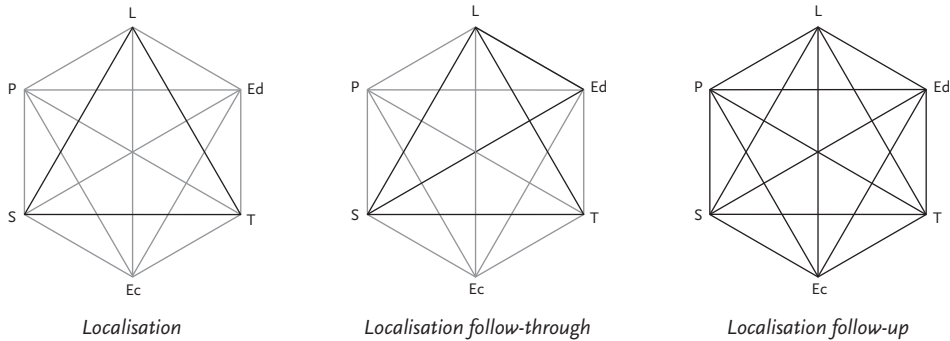


FIGURE 3.8 Localisation, localisation follow-through and localisation follow-up



basic to localisation; follow-through in a localisation project (adding the economic dimension) and follow-up, which may involve all the dimensions.

By looking at it this way, it would seem that localisation initiatives must be prepared to acquire various skills common to non-governmental or civil society organisations in order to achieve sustainable production and results.

Relevance to questions of ICT and localisation

The framework of localisation ecology relates such issues as those discussed into a complex system that affects the potential and path of localisation efforts. The concept as illustrated in the PLETES model is an abstraction, but it can be used in planning long-term, sustainable localisation efforts. In other words, the model serves as a mnemonic device to assist in considering all the various aspects of the process, anticipating issues that may arise, and identifying factors that may be important and can be used to advantage.

Localisation ecology and the processes of localisation

Once one understands more fully that various factors affect localisation, both as a technical concern and as a project, the question becomes how to use that knowledge. Perhaps this framework and the other content of this volume could be used in drafting materials and developing training workshops to enable localisation groups to proactively navigate the range of factors, influences and forces that affect the long-term success of their efforts.

Localisation ecology may also serve as a way of imagining the impact of localisation itself and considering how the localisation of a software application in a language affects the other relationships involved.

Localisation ecology and the digital divide

Localisation, and the broader agenda of bridging the digital divide of which we believe it is an indispensable part, involves several common factors. Discussing the divide in terms of various contributing factors is not new, but applying an explicitly ecological framework, including language, would seem to be a useful approach. Part of the problem in any discussion of the digital divide is that language as a factor related to access is often not fully accounted for, even though it forms the basis for communication and the coding of knowledge.

It would be interesting to consider the issue of localisation more thoroughly in discussing the digital divide and efforts to bridge it. The localisation ecology concept and its illustration in the PLETES model might be useful in doing so.

The following chapters will highlight some of these factors and interactions in more detail. In the next three chapters, the language and sociolinguistic context will first be discussed, followed by consideration of the technical contexts for localisation, namely physical access and internationalisation.

Linguistic context

THE LINGUISTIC PROFILE OF Africa is very complex in terms of the distribution of the speakers of the many African languages, current sociolinguistic trends, attitudes of social groups towards the use of various languages, the impact of language and education policies on the use of African languages and the changes within them, and the adaptation of the languages to new and introduced factors. In other words, the complexity includes sociolinguistics, policies, the geography of language and contemporary culture. As far as individual languages are concerned, other key considerations include issues of dialectal variation, the degree of interintelligibility between related languages, writing system or systems used and orthography, and the development of terminology.

Such issues are dealt with in more depth than will be attempted here in various scholarly studies of African languages from the linguistic perspective, for example Childs (2003), Heine and Nurse (2000) and Webb and Kembo-Sure (1998). This chapter attempts to broadly characterise the main elements of the linguistic environment in which African language localisation efforts are taking place, with attention to how those affect localisation.

Languages, dialects and linguistic geography

Most descriptions of African languages begin by listing the four indigenous language families that occupy the continent: Afroasiatic (formerly called Hamito-Semitic); Niger-Kordofanian (including Niger-Congo, which is by far the largest group in this family); Nilo-Saharan; and Khoisan (or Click).²² However, the linguistic taxonomy of African languages (and the genetic relationships among them), may be of practical interest to translation and localisation only much further down the taxonomic tree, where interintelligibility becomes an issue (or where similar syntactic structures open certain approaches to machine translation).

Over the vast territory of the African continent, linguistic subfamilies and languages are interspersed; the territories of speakers of different languages overlap, and the gradations in dialectal differences within many languages are more or less

gradual. In addition, indigenised languages from the Malayo-Polynesian (Malagasy) and Indo-European (Afrikaans) families are also found in Africa.

In many areas, the combination of these factors creates a complex kaleidoscope of language distributions, which is more intricate in certain parts of the continent, such as the forested regions of coastal West Africa and Central Africa (for example, according to *Ethnologue* [Gordon 2005], Nigeria accounts for over 400 languages, Cameroon over 280, and the Democratic Republic of the Congo over 200), and less so in savanna regions such as the Sahel. Over the vast distances where languages such as Arabic, the Berber languages (including Tamazight and Tamasheq), Fula and Swahili are spoken, dialectal and vernacular differentiation has occurred. In the precolonial past, and still to varying degrees today, communication among diverse groups was facilitated either by the interintelligibility of similar dialects in certain languages or by using local languages of wider communication (LWCs) as vehicular languages.

European colonisation, which imposed arbitrary borders that split many language communities and brought together linguistically dissimilar groups, also overlaid several LWCs as administrative languages (notably English, French and Portuguese).²³ These have added to Africa's linguistic complexity in several ways. Firstly, LWCs have assumed the role of lingua franca for official use as well as for contact and exchange across the continent and with the rest of the world. They have also tended to divide elite groups from the mass of society, or at least to be used as markers of elite status (see Mazrui & Mazrui 1998; Myers-Scotton 1990, 1993). To some extent, these languages have over the years also become vehicular in certain regions and urban areas, and in some cases the elite groups have adopted LWCs as their home language. In addition, Creole languages based on LWCs have become established either as vehicular languages (such as Krio in Sierra Leone) or as the L1 (such as the Creoles of Cape Verde and several Indian Ocean island states).

Languages, dialects and macrolanguages

The complex linguistic situation in Africa also raises questions about the 'boundaries' of languages as well as the very definitions of language and dialect. This is a contested terrain in the study of linguistics, with those that tend to emphasise distinctions among languages sometimes characterised as 'splitters', while those that emphasise linguistic commonalities are called 'joiners'.

The researchers of SIL International that have compiled the well-known *Ethnologue* listing of languages (Gordon 2005) may legitimately be considered as splitters. In the language profiles section of the wiki, very often several *Ethnologue* listings for variants of a given language are given for the listings we use for the

discussion of localisation. According to *Ethnologue*, Africa has 2 092 languages (Gordon 2005).

At the other end of the spectrum, the Centre for Advanced Study of African Languages (CASAS) may be considered to be a joiner. The researchers at the centre conclude from their research on groups of languages that functionally Africa has far fewer separate languages than is often claimed. The director of CASAS, Kwesi Kwaa Prah (2002, 2003), suggests that between 75 and 85 per cent of Africans speak 12 to 15 'core languages' as L1 or L2 and that these are in fact clusters of languages that are related to a greater or lesser extent.

Even splitters acknowledge that there may be degrees of interintelligibility among different languages and that the speakers of related languages may sometimes even have a concept of their fundamental unity. SIL International uses the category of 'macrolanguages' to describe such situations.²⁴

The terminology, both new and old – which includes, for example, 'language cluster' (to describe closely related but not highly interintelligible languages), 'dialect continuum' (language variation over territory, with communication becoming progressively difficult over distance) and 'dialect levelling' (reduction of differences between dialects or closely-related languages due to factors such as increased contact and intermarriage between speakers and use of the language on broadcast media) – reflects the complexity of the situation.

For localisers, this means that in certain (and possibly numerous) cases, they will have to negotiate different sets of categories in their decisions on localisation. For instance, in a language such as Fula (Fulfulde/Pular/Pulaar)²⁵ – which is spoken across much of West Africa, although by a minority in each country where it is present – there are clear differences among variants of the language, but also enough similarity to permit communication among speakers of most of the variants. Localisers will have to decide, for example, whether software should be localised for a single language, or for the nine languages into which *Ethnologue* divides it, or for some set of groupings of closely related variants. They will also have to decide whether to use the same approach to localised content in Fula as for software, or whether to apply a different set of criteria. Another consideration is whether, in developing terminology, to use forms that are likely to be understood across varieties of the language.²⁶

The case of Arabic also merits mention, as *Ethnologue* has 16 separate listings for North and East Africa alone; the total number of listings is 40. The difference in this case is that there is an established common standard form – Modern Standard Arabic – unlike the case with many other African languages.

In other cases, languages that are linguistically closely related are spoken by groups that may actually emphasise their mutual cultural differences (for example,

in the case of speakers of Teso and Turkana, whose languages are largely inter-intelligible).

This topic is further discussed in the context of a system of codes used in ICT to designate specific languages or groups of languages (see chapter 9).

Linguistic geography and localisation

The complex patterns of location of the speakers of certain languages are further complicated by the overlapping territories of different languages and the borders inherited from the colonial partition of Africa, which have further divided linguistic communities. The divisions imposed by colonial powers have in many cases led to additional changes and some divergence within languages due to the influence of different ELWCS, for instance, Wolof in Senegal and the Gambia, Hausa in Nigeria and Niger, and Tsonga in Mozambique and South Africa.

There has been discussion over the years of developing a linguistic atlas of Africa that would document the distribution of languages spoken on the continent in more detail and with more currency than earlier mapping of African languages. The African Academy of Languages (ACALAN) recently launched such a project (Alexander 2008; Afrik.com 2008) and there have also been some country-level studies such as for South Africa (Van der Merwe & Van der Merwe 2003). Such language atlas projects could be incorporated into larger localisation efforts, firstly to understand the potential ranges of use of localised software and keyboard standards, and secondly to compare the language distributions with computer and internet access.

Sociolinguistics and language change

African societies are generally multilingual, and individuals often learn several languages to varying degrees and use them in different contexts, or together in what linguists call code-switching. These societies are witnessing significant changes in the ways in which people use language, including alterations in languages used in urban areas, dialect levelling,²⁷ impoverishment or contraction of less widely spoken languages,²⁸ endangerment and extinction. It is not an exaggeration to say that the African linguistic terrain is experiencing considerable changes, sometimes in contrary directions.²⁹

The role of attitudes and perceptions is very important among both Africans and foreigners working for development and education on the continent. It is not uncommon to encounter negative attitudes towards African languages, from views of their utility merely as 'local languages' vis-à-vis ELWCS, to doubts about their

capacity for the expression of complex thought and scientific concepts. However, many Africans are committed to using and developing their mother tongues (see related discussion in the section on terminology in chapter 6).

Such divergence of opinion is to be expected in any complex society on any number of topics, and to expect otherwise in the case of language would be unrealistic.³⁰ There is certainly significant (and perhaps latent) interest in localised software and content among significant numbers of people, although it has never been surveyed to our knowledge.

Negative attitudes may discourage those that seek to use African languages in various ways, including localisation projects.³¹ This does not imply that there is a need to convince those that have no interest in localised software and content, but such attitudes may retard progress that could benefit potential users that are interested in the outcome of localisation.

The notion that the use of a particular language precludes the use of another negatively affects localisation potential, as does the view that if software or web content is provided in a widely spoken language, there is little or no need for similar offerings in other languages.³²

Such attitudes highlight the importance of education in localisation efforts.

Another area of concern is the status and future of many of Africa's less widely spoken languages, a number of which are considered endangered.³³

Many other less widely spoken African languages are also experiencing declines or degrees of contraction, however, without being close to extinction.³⁴ In all these cases, localisation may in addition to its other benefits also be a tool for language revitalisation.

Oral and literate traditions

It is often said that Africa is a region of oral tradition rather than literate and written language traditions, with the exception of the Arabic and Berber languages in North Africa and the languages from the Horn of Africa that use the script originating with Ge'ez. While there is some truth in this generalisation, it fails to take into account the widespread use of symbols for various purposes (Mafundikwa 2004), as well as the centuries-old adaptation of the Arabic script to writing many languages where Islam is important (notably in the Sahel and for Swahili). Over the last century, the writing of African languages, mainly in alphabets based on the Latin script outside North Africa and the Horn, has become more common, even as oral traditions continue.

Oral traditions

An oral tradition does not simply imply the lack of a literate tradition. While oral tradition has some obvious limitations in preserving knowledge and transmitting information over time, it involves ways of remembering and expressing that are quite impressive in their own right.

Although ICT focuses on the use of text, it is worth considering ways in which the audio capacities of the technology can be more fully exploited to accommodate the development of oral traditions. National agencies and regional institutions in Africa that exist to record and preserve oral histories (see the later discussion of language policies) could become partners in efforts to more effectively link oral traditions with ICT.

Writing African languages

Writing systems for African languages may be grouped into the following three broad categories:

- ◆ Ancient writing systems that evolved with a language or family of languages. Living examples include Arabic, Ethiopic/Ge'ez and Tifinagh.
- ◆ Writing systems invented in historically recent times within the last two centuries. Several of these are still in use (such as N'Ko, Mende KiKaKui, Vai and Mandombe), while others are not.³⁵
- ◆ The adaptation of another script for writing languages. The most widespread pre-colonial example was the use of Arabic script by learned Muslims for writing non-Arabic African languages. However, the Latin script introduced by Europeans during colonisation is certainly the most important introduced script today.

Adapting writing systems to African languages

The process of adapting a writing system from one language or family of languages to another, whether or not the latter had previously been written, frequently requires modification of the writing system to the often very different phonetics of the new language (or languages). In the case of Arabic, for instance, people in sub-Saharan Africa added modified characters to meet the needs of their languages, much as non-Arabs did in Asia. This has been done extensively with the Latin script for some African languages. Modified characters have even been added to Ethiopic/Ge'ez to accommodate certain minority languages.

Given the large number of African languages, the way in which the continent was divided during colonisation and the diverse groups (often missionary groups) involved in rendering African languages in written form, issues of standardisation

arose early on.³⁶ One such example was the efforts of the International Institute for African Languages and Cultures (IIALC 1930) to standardise the use of the Latin script at a continental level. Following independence, a series of expert meetings sponsored by UNESCO were convened to address the standardisation or harmonisation of the orthographies of African languages.³⁷ There have been similar projects at regional levels in various parts of the continent. Altogether, these ongoing efforts have had some success in guiding country-level agencies that have set the orthographies for languages within their borders.

The standardisation of the use of the Arabic script has not received as much attention, although the Islamic States Educational, Scientific and Cultural Organisation (ISESCO) sponsored an effort to develop a standardised set of modified characters for several sub-Saharan African languages (see Chtatou 1992). This initiative was based to some extent on usage in non-Arabic languages of the Middle East and has apparently not been widely accepted among speakers of the African languages it was intended for.³⁸

Contemporary situation

Several factors are important to note in considering the topic of orthographies for African languages today:

- ◆ While orthographies are relatively set for some major languages, they seem still to be in flux for others.
- ◆ One of the reasons for recent changes in the orthographies of certain languages has been the mismatch between the fonts offered on early computer systems (and even on current ones) and the characters or diacritics adopted for use in print and writing.³⁹
- ◆ There are in some cases separate systems of writing for the same language, sometimes characterised by competition or even conflict among their respective advocates. This is the case for certain languages with old written traditions as well as some that have been written only more recently. Governments have tended to ignore the usage of older non-Latin scripts where a Latin-based orthography has been adapted.⁴⁰
- ◆ There are still many African languages, mostly less widely spoken ones, that do not have formal writing systems.
- ◆ Conventions and policies concerning orthographies are generally set at a country level by governments or researchers. Coordination with other states where the same languages are spoken remains limited.
- ◆ Several countries have attempted to establish a general orthography with the intention that it be adapted to the needs of all languages spoken in its territory.⁴¹

Language and language in education policies

Policies within African countries, and indeed between and among them, affect the possibilities for localisation and are thus a critical factor in localisation ecology. ICT policies are discussed in chapter 5, but policies related to language are considered here. The following two overlapping policy areas are of particular concern:

- ◆ Language policy and related concerns of language planning and management:
 - ◆ Of particular practical interest for localisation efforts are decisions concerning orthographies, as already discussed;
 - ◆ Of interest for linking localisation with ICT4D is policy related to language use in development.⁴²
- ◆ Education policy, with specific attention to issues of African languages as media of instruction and as subject matter.

Language policy in Africa

Language policy is broadly taken to mean the set of legal and administrative mandates and guidelines concerning language use in public life, including such matters as assigning the status of 'official' or 'national' to particular languages; the use of certain languages in government, legal systems, development and education (which will be discussed in further detail in this chapter); and standards, such as official orthographies (which will also be discussed separately).⁴³

Since independence from European colonial powers, one might characterise African language-policy concerns as being marked by two features: firstly, reliance on the former colonial languages – ELWCs – for government administration and education, whether or not such reliance was codified in law or constitution;⁴⁴ and secondly, considerable discussion on the role of indigenous languages and how to use them.

The concern of many African governments to have a single common language for nation building favoured the use of ELWCs. Bamgbose (1996) characterises this as involving: 'two complementary myths: the first being that having several languages in a country (multilingualism) always divides; the other being that having only one language (monolingualism) always unites' (see also Bamgbose 1991: 14).

At the same time, discussions at national and regional levels on the use of African languages has tended to result in proposals but little follow-through.⁴⁵ This perhaps reflects a tension with the prevalent focus on a single language, as already mentioned, as well as ambivalent attitudes about the value of African languages vis-à-vis ELWCs.

In any event, language policies and planning in Africa have received little attention, to the extent that Okombo (2001), for example, refers to this area of policy as: 'a forgotten dimension in governance and development' (see also Gadelli 1999). Bamgbose (1991: 6) believes that this is due to: 'a general feeling that language problems are not urgent and hence solutions to them can wait'. He goes on to characterise the situation in these terms:

Language policies in African countries are characterised by one or more of the following problems: avoidance, vagueness, arbitrariness, fluctuations and declaration without implementation. (Bamgbose 1991: 111)

In recent years, language policy has received more attention, for example, as it affects education (as discussed in a later section of this chapter). At a regional level, ACALAN has been established, and in post-apartheid South Africa, some of these issues are being vigorously explored.

It may be instructive for governments to begin with the fundamental question of whether they want the languages spoken in their respective countries to continue to be viable for future generations.⁴⁶ This may seem a very stark question at a time when language extinction in Africa, and indeed worldwide, is a hot issue, but it serves to highlight what is actually at stake. Answering in the affirmative leads to other questions, as well as imperatives, some of which are directly related to the localisation of ICT and the need for coherent strategies to advance it.

The absence of language policies that actively support African language computing means that localisation will probably depend on initiatives by individuals, organisations and companies.

Language institutions and agencies in Africa

Agencies and organisations for research and applied linguistics exist in one form or another in each African country, often as a part of government or a university. There are also some continental and regional institutions. At the continental level, ACALAN, based in Bamako, Mali, operates with a mandate from the African Union to facilitate work with African languages. The most notable regional institutions are those established to deal mainly with oral histories, namely:

- ◆ CELHTO: Centre d'Études Linguistiques et Historiques par Tradition Orale (Centre for Linguistic and Historical Study of Oral Tradition) in Niamey, Niger;
- ◆ CERDOTOLA: Centre Régional de Recherche et de Documentation sur les Traditions Orales et pour le Développement des Langues Africaines (Regional Centre for Research and Documentation of Oral Traditions and for the Development of African Languages) in Yaoundé, Cameroon;

- ◆ CIDLO: Centre d'Investigation et de Documentation sur l'Oralité (Centre for the Investigation and Documentation of Orality) in Tananarive, Madagascar;
- ◆ EACROTANAL: Eastern African Centre for Research on Oral Traditions and African National Languages (which is currently closed) in Zanzibar, Tanzania.

Non-governmental organisations concerned with language and culture exist in many countries. On a continental scale, SIL International is the most prominent, with offices in many countries.

Educational policy and languages of instruction

The issues of using L1s in education and bilingual pedagogy in Africa are receiving increasing attention at both country and inter-African levels.⁴⁷ An extensive discussion of the rationale for L1 and bilingual (or multilingual) education is beyond the scope of this paper, but it is generally agreed that this is beneficial for school-children. However, implementation is not simple. This set of issues gives rise to considerable debate within countries, but implementation involves problems with issues such as teacher training and the availability of materials.

The ways in which educational language policy might affect localisation is a separate issue. An interesting case that has already been referred to is the way in which the decision by Morocco in 2003 to use Tifinagh in Tamazight education spurred efforts to complete the encoding of the script in the Unicode standard. L1 education and localisation efforts may be mutually supportive, especially to the extent that ICTs are introduced for educational programmes, for example, providing computers for schools, the One Laptop per Child (OLPC) project, or the involvement of ICT4D centres in adult literacy initiatives.

Basic literacy, pluriliteracy and user skills

One basic factor that contributes to the intractability of the digital divide is literacy (other factors are discussed in chapter 5). In multilingual contexts, which prevail over most of the African continent, the subject is perhaps more appropriately called 'pluriliteracy' (which means being literate in more than one language),⁴⁸ although it is seldom officially discussed in these terms. User literacy skills are grouped into the following possible profiles:

- ◆ Fully pluriliterate, that is, able to read and write in all the languages they speak (which have a writing system);

- ◆ Literate in an ELWC but not their mother tongue, this being the usual outcome of schooling conducted entirely in the ELWC;
- ◆ Varying but incomplete levels of literacy among school-leavers in the language of schooling;
- ◆ Ability to read their mother tongue or a local lingua franca to some level of proficiency learnt from literacy classes given by national programmes, development projects or traditional education (such as Koranic schools);
- ◆ Illiterate or functionally illiterate.

There are thus quite uneven levels of ability to take advantage of opportunities among potential communities of users of localised content and software in Africa. This suggests the fundamental link between education (including both literacy training and policy as it pertains to languages of instruction) and localisation. Localisation efforts may do well to associate themselves with literacy training programmes (both traditional as well as those using ICT where public telecentres have been set up for local development) and projects that aim to provide computers for schools and allow students the opportunity to encounter software and content in their LIS. In terms of localisation ecology, the factors of language, technology and education each have relevance in developing user skills.

There is also a relationship between orthography and the development of reading and writing skills. This subject is starting to receive more attention⁴⁹ and has a bearing on computer use in the numerous languages for which writing systems are not well established.

Literacy and pluriliteracy rates will take time to increase in many African countries, which would seem to indicate the desirability of making more effective use of audio and image in content and user interfaces.

Terminology and accommodation of ICT concepts

Terminology is an aspect of language change and planning that is particularly relevant to localisation. The broader area of terminology development concerns many fields, one of which is ICT. There is some attention to computer and internet terminology in this field, although technical specialists rather than linguists are often left to find or develop the terms necessary for localisation.

There are several considerations related to the ways in which languages develop or borrow terminology for new and foreign concepts, a process that Coulmas (1992) refers to as 'language adaptation'. In some cases, terms arise from the community of speakers, but where most people are unfamiliar with the technology or its details, terms are either borrowed from another language or invented,

often by individuals or groups either self-appointed or designated by an authority. Most of the theory related to this process is beyond the concerns of people who are working on translating software and need to be able to refer to various concepts, but it is useful to keep in mind when planning terminology development.

The process of developing terminology usually relies on experts in the language who have some familiarity with the technical areas for which the terminology is required. Indeed, Microsoft has used panels of experts to develop terminology and dictionaries for its localisation efforts in major African languages.

While the focus of localisation initiatives with respect to terminology is somewhat narrow and limited to meeting specific needs, such efforts should be informed by larger terminology development and in turn participate in such movements. At the same time, it should be noted that there is some debate among linguistic experts about the value of efforts to develop terminology in all scientific domains for less widely spoken languages.

The development of terminology for localisation is a subset of the concern to produce dictionaries for use with software, and both might be usefully linked to broader lexicographic projects.

Technical context I: physical access

THE REVIEW OF THE technical context for localisation in Africa complements the linguistic context described in the previous chapter and provides the background for a more detailed consideration of localisation in Africa. The technical context will be considered in the following two parts:

- ♦ Physical access: a nexus of issues relating to connectivity, infrastructure, the availability of computing devices and permissions to use all of these. Physical access complements soft access, which is facilitated by localisation.
- ♦ Internationalisation of the technology: a parallel process of technical advances and standard-setting to support diverse languages and scripts on computer applications and internet content. Internationalisation is, in a way, a prerequisite for localisation and soft access.

Although these connect in various ways – notably their fundamental connections in the broader concern with full ‘access’ to ICT as well as their actual or potential relationships with ICT for development (ICT4D) and education (ICT4E) programmes – they are dealt with in two separate chapters. In effect, all efforts to promote ICT in Africa, whether local, international or both, have the ultimate (or at least ostensible) aim of increasing access to the technology, which is a topic with very broad implications.

ICT4D/E programmes have focused in one way or another on access, and are therefore often implicated in bringing hardware to Africa, increasing connectivity and, less consistently, in FOSS and localisation. In promoting access, these programmes and projects are also important contributors in shaping the technical environment for localisation.

Within this context, the present chapter will begin by briefly discussing access and related factors. It will then discuss infrastructure issues, hardware and operating systems, and connectivity, followed by international trends in FOSS and localisation and their bearing on Africa. ICT4D/E initiatives will be referred to, as appropriate, throughout.

Physical and soft access

Access to ICT is a fundamental defining factor of the digital divide,⁵⁰ and is therefore the focus of a range of activities related to technology in Africa, including localisation. Most discussions and publications on the subject, however, have focused primarily, if not exclusively, on how people have, or may gain, proximity to and permission to use computers or other devices that are connected to the internet or another network.⁵¹ The focus has been on topics such as whether there are functional computers, whether they are located in places that are convenient for people to reach, whether the computers are connected to the internet and whether potential users can pay whatever fees may be required.

It is certainly understood that access in a more complete sense involves more; various sources have attempted to expand upon that point. For instance, TeleCommons (2000), in an early evaluation of the potential to use ICT for rural development, distinguish between 'physical access' to ICT infrastructure and applications and 'soft access', which they define as software and applications designed to enable rural African users to utilise ICTs for their own needs and uses once physical access has been established. We have adopted this simple distinction for use in this book, but there are other perspectives with respect to the complexity of access.

Bridges.org, for example, went further and defined 12 dimensions of what they called 'real access'⁵² to ICT, in which 'relevant content' specifically mentions language. In effect, several points emerge in discussing access and localisation:

- ◆ Foundations for access with respect to the hard realities, namely:
 - ◆ The starting point in discussing access involves issues such as the availability of functional computers, power to allow them to operate and connections to link them.
 - ◆ The costs of establishing and maintaining access are generally beyond local means and therefore often involve outside support or initiatives.
 - ◆ Permission to use the devices, whether by fees or other means, often represents a significant cost relative to potential users' resources and is thus a potential barrier to physical access.
- ◆ Access and localisation with respect to the way in which providing access meets the needs of the user, namely:
 - ◆ The choice of language is particularly important to two aspects of access – software interface of computing devices and interactive software on the internet.
 - ◆ The term 'soft access' is used to refer to how well the means of access anticipates user needs.
 - ◆ Localisation is the major aspect of ensuring soft access.

- ◆ User skills with respect to the way in which the user gains physical and soft access, namely:
 - ◆ User profiles, including language and literacy;
 - ◆ Implied attention to developing user skills, including basic literacy (see the discussion of language in education policies in chapter 4).

Figuratively speaking, access builds towards the user from one end, from hardware and connections to interfaces that are understandable to potential users (with other factors also involved), while user skills, in effect, build from the other end, involving education, training, experience and so on. There is thus arguably a trade-off or complementarity between soft access and user skills. More skilled or experienced users are less needy in terms of soft access, while less skilled or experienced users require more attention, including localised interfaces.⁵³

Localisation for enhanced soft access, and the enhancement of users' skills, therefore emerge as two complementary and essential elements in extending physical access to ICT into effective or 'real' access (which in turn implies connections to other concerns, such as education in L1, in terms of localisation ecology). It should be remembered, however, that skilled and multiliterate users may prefer localised interfaces rather than versions in the dominant ELWCS.

However, in the case of skilled multilingual users, soft access may also anticipate diverse user preferences. Just because educated, multilingual African users are able to use software or understand content in English or French does not mean that they will have no interest in interacting with applications in their first or other languages at some point, or that they will derive no benefit from doing so.

However, the discussion hinges on physical access, which is the main focus of this chapter.

Basic infrastructure

In discussing efforts to expand ICT in Africa – whether localised or not – reference is often made to other basic technical and infrastructural indicators such as the number of telephone lines and the level of electrification.⁵⁴ In effect, the realities of poor communications infrastructure and the lack of reliable power sources limit even well-funded efforts to establish access to computers and the internet. These basic factors are changing slowly, while other solutions such as alternative power sources (notably solar) and the dramatic growth in access to cellphone technology are altering the equation in certain ways.

For the purposes of this study, however, these important infrastructural factors will be considered as given in order to focus attention on other technical variables that affect actual and potential localisation in Africa.

Computer hardware and operating systems

The most basic measure of ICT penetration and of the digital divide is the availability of devices in working order, or the lack thereof – primarily computers, but increasingly also portable and handheld devices – that can process, store and transfer information. In their absence, of course, any discussions of connectivity, access and localisation are moot. This is reflected in efforts such as the development of the Simputer in India (a handheld computer that has been on the market since 2002)⁵⁵ and the OLPC project to develop and supply inexpensive laptop computers to schoolchildren in poor countries.⁵⁶

By any reckoning, the numbers of computers in Africa are low in comparison with other world regions, and are often older makes, using older operating systems. There are obvious economic reasons for this. The efforts of outside agencies to address the hardware aspects of the digital divide in Africa, as they perceive it, by supplying new or used computers have limited impact. (However well designed or funded such projects might be, they address only limited objectives within far larger contexts.)

The existence in Africa of many older computers and operating systems has implications for the potential use of various kinds of software and multilingual web content. In many cases, older operating systems cannot run newer software, use Unicode fonts or take full advantage of internet connections. In time, these systems will be retired, but given the persistence of the root causes for resorting to used computers, it seems likely that Africa will continue to have a high percentage of computers in use that cannot handle the latest operating systems and software. In other words, the use of computers and systems that cannot take advantage of the most recent advances in internationalisation and localisation will continue – always a step or two behind the latest technology.

This discussion has dealt mainly with computers, although the rapid spread of mobile telephony in Africa and the increasing capacity of handheld devices should be understood together as a harbinger of significant changes in planning for ICT use on the continent, among which will be localisation. The potential for multilingual SMS is already being explored. The longer-term potential for more powerful handheld devices is discussed in chapter 12, but it is worth noting here that mobile technology is not subject to the same infrastructural limitations as computers (as discussed in a previous section of this chapter) and is naturally more affordable.

Connectivity and ICT policy

Another measure of the digital divide in Africa is the level of connectivity, indicated by the presence and quality of internet connections. While this does not directly affect localisation, it does affect the environment for localisation.

Basic connectivity at the country level was a major focus of ICT in Africa in the 1990s. Since the success of projects such as the Leland Initiative (funded by USAID) and the Internet Initiative for Africa (IIA) (funded by UNDP),⁵⁷ as well as the introduction of other infrastructures to support cellphones and VSATs, for example, attention has turned to extending connectivity from the capital cities to the more remote parts of African countries.

This evolution has been accompanied, and to varying degrees guided, by discussion and the development of policies governing the use of expanded national bandwidth. The Leland project involved policy prescriptions as much as it did infrastructure assistance, with the aim of fostering a sustainable organisational configuration to manage optimal use of the bandwidth. This was guided by a market philosophy that required the national beneficiaries of increased bandwidth to resell it to private ISPs.

More broadly, the United Nations Economic Commission for Africa (UNECA), through its African Information Society Initiative (AISI), encouraged all countries of the region to develop national information communications infrastructure (NICI) plans to help them determine how to expand connectivity.⁵⁸ This programme began in 1999, following the African Development Forum on the Challenge to Africa of Globalisation and the Information Age. NICI had four major frameworks: ICT policy development, applications (or areas) for priority attention, infrastructure building and technology selection (Ya'u 2005).

Altogether, internet connectivity and policies to guide its use have been closely interrelated.

In technical terms, bandwidth in Africa is continuing to increase (IDRC 2005), and the number of internet connections is growing rapidly (USINFO 2006). This has been accomplished largely by the use of satellite uplinks in each country, but the deployment of undersea cables around the coast of the continent is also becoming a factor, namely the SAT-3/WASC cable in West Africa and the FAST cable in southern and East Africa.⁵⁹

Despite efforts to increase connectivity to all African countries, the actual levels of connectivity between countries and within each country tend to vary significantly.

Some of the structural issues that have been mentioned affect the potential to expand access. A particular challenge is to connect rural centres. Certain countries have implemented phone tariff systems that allow dial-up access at the

same rates for all parts of the country. While this removes one of the disadvantages, connection quality and actual accessibility may still not be good.

Some of the clearest links between policy and technology are evident in relation to connectivity (in terms of the localisation ecology and the PLETES model).

This is an important consideration, since the patterns of language and connectivity or access may indicate priority areas for localisation, and the issue of access is related to the motivation for undertaking localisation.

Technical context II: internationalisation

WHILE THE CONSIDERATION OF access to ICT in Africa inevitably leads from a focus on physical access to a discussion of soft access, a separate process has sought, at a global scale over a number of years, to better enable ICT to meet such needs anywhere in the world. Internationalisation refers to the process of improving computers, systems and internet protocols to accommodate the diverse language needs of the world. As such, it enables localisation and computing in many languages. Using the term ‘internationalisation’ in a broad sense, we consider it to include work on standards such as Unicode.

The facilitating technical environment

Internationalisation – including both technical improvements and standards – may be regarded as defining the technical environment for localisation and multilingual ICT. Within the context of localisation ecology, internationalisation may also be understood to include the technical and policy-related approaches to organising language use in ICT. These factors are in continuous change and evolution. Understanding them is essential to a full consideration of localisation issues.

Technical standards enable interoperability and thus constitute an aspect of internationalisation, for instance, the adoption and progressive additions to the Universal Character Set established by Unicode (as discussed in the next section). In the sense of contributing to a predictable environment for users, technical standards facilitate soft access.

In addition to Unicode, other standards that have been adopted for various reasons may become more extensively used through localisation. One such example is the set of codes for languages codified under successive ISO 639 standardisations (see chapter 9). Others include country codes, most notably two-letter country codes (ISO 3166),⁶⁰ and four-letter and three-number codes for writing systems (ISO 15924).⁶¹ There are also guidelines for using these codes in computer applications and internet content, notably RFC 4646.⁶² Among other things, these codes are used in defining locale data (see chapter 9).

There is also a set of standards for computer keyboards, iso 9995, which includes guidelines for keyboard layouts (see chapter 8).

In technical terms, internationalisation is a way of facilitating localisation of a given product rather than building versions localised in different languages from scratch. It may therefore include specifications and allow space during software design for including commands in diverse languages and scripts. Rathke (2005) offers a succinct summary of internationalisation factors important for open-source localisation.

In broader technical terms, internationalisation includes the capacity to edit in diverse languages and scripts, for instance, on web applications or in software that is not localised.

Internationalisation also helps provide a predictable environment for users, a process which, taken a step further, involves aspects of localisation.

Handling complex scripts: from ASCII to Unicode

Representing extended Latin and non-Latin scripts on computers and the internet presents challenges in terms of space for large character sets, rendering complex characters and accommodating multiple scripts. These issues have essentially been resolved through Unicode, which is intended to cover all writing systems, as well as through increasing computer processing capacity and disk storage. However, work on certain standards is still ongoing, and the implementation of technical solutions may fall behind current standards.

Background

Character encoding for rendering text on computers initially used a seven-bit system (with 128 codepoints or spaces for characters or other information), the most commonly known version of which is the English-based ASCII. ASCII actually had its origins in the USA in the early 1960s as a telegraphic code and was subsequently adopted for use in computing. In 1972, the international standard iso 646 incorporated ASCII and defined some additional uses for other languages. Later, 8-bit encodings (with 256 codepoints) – sometimes called ‘extended ASCII’ – provided more space for diverse characters and alphabets.

The earliest means of accommodating diverse script needs involved creating fonts in which some of the characters in another character set (ASCII or 8-bit) were ‘changed’. This meant assigning certain characters to codepoints originally occupied by other characters.

In response to the use of more languages in ICT, a series of standards for 8-bit fonts was developed under ISO 8859.⁶³ Microsoft developed similar character sets, such as Windows 1252 for Latin as used in the major western European languages, also called ANSI (American National Standards Institute),⁶⁴ and Windows 1256 for Arabic. These 8-bit (or 1 byte) fonts have 256 codepoints; the lower 128 (0–127) are identical to those of ASCII, and the upper 128 (128–255) vary according to the standard. No commercial or industry standard (for example, in ISO 8859) has been developed specifically for any sub-Saharan African language or group of languages.

A separate standard was devised in 1983 for African languages transcribed in the extended Latin alphabet, namely ISO 6438 African coded character set for bibliographic information interchange, but this was apparently hardly used, if at all, even for the primary purpose indicated in its title (Clews 2003).⁶⁵ This standard was developed by a different standards technical committee from the one that had developed ISO 8859 (Clews 2003). Curiously, although ISO 6438 was developed at about the same time as the African Reference Alphabet of Niamey in 1978, the two were apparently developed separately from similar information sources.⁶⁶ Although there is much overlap between the two, the forms in ISO 6438 were retained in Unicode.

Arabic, as an international language of considerable religious and political importance, received much attention early in the process. The script presents challenges from the point of view of the direction of the writing (right-to-left, with numbers reading left-to-right, or 'bidirectional') and the changing forms of many characters when preceded or following other characters. Nevertheless, coding followed an evolution from an early 7-bit version to 8-bit encodings, ultimately including ISO 8859-6 and Windows 1256.

The UNIX-based Apple Macintosh computers followed a separate evolution from ASCII to Unicode over nearly two decades (from Macintosh Character Set to MacRoman and WorldScript). In the 1980s, some users found that Macintosh systems and software facilitated work on extended Latin and non-Latin scripts for African languages, but these do not seem to have had much impact in Africa itself, where Apple products remain relatively rare.⁶⁷

Unicode

Unicode (ISO 10646), the single encoding standard for all the world's scripts, also known as the Universal Character Set, incorporates all the characters of previous standard encodings and is designed to facilitate the use and exchange of text in any writing form across all platforms and the internet. As such, it is a central standard in the internationalisation of ICT.

Unicode can in principle define up to a million characters, although its latest version (Unicode 5.2 at the time of writing) covers all major writing systems as well as many others (but still not all) using only about 100 000. Unicode is commonly implemented in UTF-8 (Unicode Transformation Format),⁶⁸ which permits Unicode to be used in many cases with as few bits as pre-Unicode encodings.

Unicode and Africa

Since many languages in Africa use either extended Latin alphabets or non-Latin scripts (and sometimes both), Unicode in principle resolves the problems encountered with multiple non-standard coding solutions. There have nevertheless been some obstacles to realising this potential.

Firstly, although industry has moved to using Unicode, and systems have long been designed with the standard in mind, Unicode does not seem to be well understood in Africa, even among computer experts. Many technical experts are occupied with work involving only major international languages (including Arabic), and African-language experts, to the extent that they work with computers, often still resort to the panoply of legacy 8-bit encodings already mentioned.

This situation is gradually beginning to change as newer computer systems come into use, awareness of multilingual computing increases, and efforts to facilitate the use of Unicode result in more people being trained. (Among these, the efforts as from about 2000 of the French-funded RIFAL project [International Francophone Network for Language Management]⁶⁹ to help national language agencies in West Africa migrate their text banks to Unicode deserves note.)

A second problem, however, has been the lack of fonts with the necessary characters and, even when these are available, their absence from computer systems in use.

Thirdly, there has been some concern in Africa about how well the Unicode standard meets African needs, mainly with regard to the use of diacritics in Latin-based orthographies. (This is discussed in the next subsection.)

Another issue raised about Unicode was whether the disk size requirements of text in Unicode relative to 8-bit fonts are a disincentive for its use (see Paolillo 2005: 47, 72–73). This is actually not very problematic, given technical advances in handling Unicode (such as UTF-8) as well as the vast increases in disk space and computer memory to meet much larger file requirements (including images and audio).⁷⁰ However, this is an issue in relation to SMS text on cellphones, where messages with characters outside the ASCII set have a far shorter permitted length.

Unicode and diacritics in Latin transcription

While Unicode in principle meets the transcription needs of all languages written in the Latin alphabet and its variants, several issues are still under discussion, some

of which concern individual characters. For those, there is an established system for adding characters or modifying certain information.⁷¹ However, the decision by Unicode in the late 1990s to rely on 'dynamic composition' to render diacritic characters by combining base characters with one or more 'combining diacritics', rather than adding more 'pre-composed' characters for all combinations used, has drawn negative reaction by some experts. It would seem that the policy has been set and that the technology for implementing it is improving, but a brief review of the history is useful.

The issue of how to deal with diacritics in some African orthographies has received varying degrees of attention since the late 1990s. In the late 1990s, for instance, perceiving that progress in supporting dynamic composition in Windows systems was slow and that there was even less interest among developers of Macintosh and Linux systems, the Linguistic Data Consortium of the University of Pennsylvania (USA) launched a project to compile a list of character needs for African languages, with a view to determining the potential for developing alternative 8-bit standards. This effort, organised under the auspices of a body called of the African Language Resource Council (ALRC), was abandoned after a few years, largely due to advances in the field.

A similar concern, coupled with the concern that reliance on dynamic composition might disfavour African languages, motivated another project by Prologiciels BPI in Canada to develop a set of 8-bit fonts for African languages, an effort that was recognised at the Internet: Bridges to Development conference in Bamako, Mali in February 2000 (see Bourbeau & Pinard 2000).

At a planning meeting held in Bamako in 2002 for the first World Summit on the Information Society (held in Geneva in 2003), this situation was again raised and it was suggested that a series of 8-bit fonts might lead to the adoption of new standards for Africa in the ISO 8859 series.⁷² Concern was expressed that Africa had been disadvantaged in the Unicode process when Unicode had decided, before the needs of African languages had been fully addressed, not to add any more pre-composed Latin characters.

More recently, in a paper delivered at the Unicode conference in 2005, the concern was re-framed as one of handling data in African languages that use characters in combined forms (Chanard 2005). The issue raised was partly the long-term implications of using composed characters.

There are three sets of observations to be made in response to this persistent line of concern. Firstly, there does not seem to have been a thorough assessment of the actual use of diacritic and extended characters in African orthographies, and a census of that nature may perhaps not even be possible. The closest may have been a set of characters compiled as part of the ALRC effort already discussed, as well as research by John Hudson on behalf of Microsoft.⁷³ In any event, most of

this research is based on what linguistic articles, dictionaries and the like indicate as characters and combinations that are used in the orthography of whichever language is under scrutiny. The information may sometimes be contradictory, and in other cases, there have been changes in official orthographies. In yet other cases, diacritics used to indicate tones in tonal languages are either not standardised or are used only where clarity is essential or in learning materials where guidance on pronunciation is important. In summary, the extent to which pre-composed characters are either used or potentially needed is neither clear nor easily established.

Secondly, the technology for handling dynamic composition has evolved significantly. This means that the ability to position diacritics correctly over base characters and the possibility of using a base character with a combining diacritic to render a pre-composed glyph go a long way towards obviating concerns about the lack of pre-composed characters. However, the continued use of older computer systems in Africa means that, in practice, the rendering of combining diacritics remains an issue for many.

Thirdly, from a different perspective, objections to using combining diacritics are based on inadequate understanding of the technology.

Unicode and non-Latin scripts

Among the scripts of Africa, Arabic and Ethiopic/Ge'ez (used for Amharic and Tigrinya, among others in the Horn of Africa) were encoded in Unicode early in its development. Like the Latin script, both these scripts include extended ranges with characters for languages other than the primary ones used for the main languages written in them.

Two other African scripts have been added in the last few years and are now part of Unicode, namely Tifinagh in 2005 and N'Ko in 2006.

Other writing systems are being developed, notably Vai (used for the Vai language, one of the Mande group in south-western West Africa), and the process of attending to such minority scripts is being guided by the Script Encoding Initiative at the University of California at Berkeley. These scripts have value for several reasons but are not used by large populations.

The 'last mile' of internationalisation

Internationalisation, in the broad sense used here, thus includes a range of efforts and actions designed to facilitate multilingual uses of ICT, including localisation.

In chapter 2, the equipping of systems was discussed along with the notion that this represented the 'last mile' of internationalisation. In effect, this means that a whole system of standards and technical improvements has taken us to the point

where we can in theory and practice display and input text in almost any language, but for these benefits to reach end-users, planning and action are necessary.

Internationalisation and localisation

Having thus outlined the linguistic and technical contexts of localisation in Africa, including internationalisation, the following chapters discuss several aspects of internationalisation and its implementation that are important to software and internet localisation (and their use) in Africa, including Unicode and text handling, keyboard and input systems, and language codes and locale data.

This page intentionally left blank

African-language text, encoding and fonts

THIS CHAPTER FURTHER DISCUSSES the orthographies of African languages (chapter 4) and the standards for rendering text on computers (chapter 6).

Given the central importance of text as the medium for storing and transmitting information on computers and the internet, perhaps the most fundamental aspect of facilitating localisation in African languages is the capacity to reliably and recognisably display, transmit, store and retrieve the characters used in their orthographies. This concern links the longer-term processes of developing written African languages and providing education in such languages with the more recent efforts to provide internationalisation support for diverse scripts and extended characters.

Within this context, this chapter discusses non-Latin and Latin scripts (with the focus of the latter being on extended or complex Latin), the evolution of African-language text use in ICT, encoding issues and fonts.

Non-Latin scripts and ICT

Several non-Latin scripts are actively used for writing African languages, and several others are important mainly for historical and cultural reasons. The most widely used scripts that are encoded in Unicode include:

- ◆ Arabic: For the Arabic language, of course, and for transcription of some other African languages, often with adaptations to represent different phonetic systems. The latter usage is sometimes called Ajami;
- ◆ Ethiopic/Ge'ez: Used for languages of the Horn of Africa (mainly Ethiopia and Eritrea), notably Amharic, Trigrinya and Tigre;
- ◆ Tifinagh: The traditional script of the Berber languages of North Africa and the Sahel, including Tamazight and Tamasheq;
- ◆ N'Ko: A writing system created in the late 1940s that is used primarily for the Manding languages of West Africa.

A number of other scripts have not yet been encoded. Although they are generally used by small numbers of people, they may be of historical interest.

Typology of Latin-based African orthographies

As already mentioned, most African languages in their written form use the Latin script, often with additional extended characters. Because of the number of languages involved and the diversity of their orthographies in terms of characters used and support issues (such as the availability of fonts with the appropriate characters and support for combining diacritics where these are used), it is helpful to have a schema to describe them.

For this purpose, we have adapted a typology of African language orthographies similar to that proposed by Taylor (2000) before Unicode became the *de facto* international standard for handling text in ICT. In this schema, which has previously been presented by Osborn et al. (2008),⁷⁴ each higher level presents some additional complications for handling text (including display, transmission and input) over the previous category, as follows:

- ◆ Category 1 orthographies: ASCII – all characters and combinations are covered by the ASCII character set. African examples include Swahili and Zulu.
- ◆ Category 2 orthographies: Latin-1, meaning that all characters and combinations are covered by ISO 8859-1/Windows 1252 (or ANSI). On current operating systems and in most applications, there are no practical differences in terms of display and transmission of characters between Latin-1 and the ASCII set in category 1. Moreover, the keyboard layouts for typing these characters are well established for the major European languages. However, one still encounters situations where, for one reason or another, characters are not rendered properly (for example, accented characters in French). For other languages, however, input in the absence of standardised keyboards and rendering in the case of older systems or applications may pose problems. African examples include Sango.
- ◆ Category 3 orthographies: Extended Latin without combining diacritics (although perhaps with some pre-composed diacritic characters in extended ranges beyond Latin-1), which means that the orthographies are covered by the Latin ranges of Unicode without the need to support combining diacritics for tones. African examples include Fula and Gikuyu.
- ◆ Category 4 orthographies: Latin as complex script, which means that the orthographies are covered by extended Latin with the use of combining diacritics. African examples include Yoruba and Igbo.
- ◆ Category 5 orthographies: Orthographies not yet fully supported by Unicode in that one or more characters in the orthography are not yet part of the Unicode standard. There are probably few Latin-based orthographies in Africa that fall into this category.

An orthography in a given category may include some characters from one or more of the other categories, and all Latin-based scripts include some ASCII characters (category 1). Wolof, for instance, is a category 3 orthography owing to the letter *ŋ* (*eng*) used for the velar nasal *n*, but it also includes several accented characters from Latin-1 (category 2) as well as ASCII (category 1).

Varying usage within a given language may mean using different categories in practice. In the orthography of the Manding language of Bambara, for instance, tone marks over vowels are generally not used, making it a category 3 orthography. However, tones may be marked, requiring combining diacritics over the *ɛ* (open *e*) and *ɔ* (open *o*) vowels. This would make it a category 4 orthography. The difference is significant in relation to full support in different contexts, but not critical for basic support enabling the ‘last mile’ of internationalisation.

Evolution of African-language text use in ICT

Apart from Arabic, the development of the use of African languages in computing and the internet has been relatively slow for a number of linguistic, educational, policy and technical reasons, some of which have been very basic, as already mentioned in chapter 4.

It is important to bear in mind, however, 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 other ELWC. This is a reflection both of the international dominance in software and internet content of these languages inherited from colonisation, and of the widespread knowledge of these languages by those in Africa (generally elites in urban areas) that are most likely to have access to the technology.

Ironically, it would seem that Latin-based orthographies with some extended characters tend to receive less attention than non-Latin ones, even though the technical task of handling most extended Latin orthographies is generally easier (at least up to the level of category 3 orthographies, with extended characters but no combining diacritics).

African-language text in ICT

Some specific aspects of the evolving use of African languages in ICT are dealt with in this section, but a particular problem for a number of languages written with modified letters or diacritic characters – or entire alphabets – beyond the basic Latin alphabet (the 26 letters used in English) or the ASCII character set (that alphabet

plus basic symbols) has been the way in which computer systems and software handle these (see chapter 6).

Although the earliest personal computer interfaces used the English language and the ASCII character set, the potential to use a rendition of other languages was inevitable. Such use is difficult to quantify, but with advances in the capacity of systems to handle larger character sets and the elaboration of the internet, multilingual computing became possible in Africa, as elsewhere. The greater, but still limited, potential of 8-bit fonts (described in various terms, such as ANSI, as already mentioned) permitted the development of fonts for more languages.

Over the years, a number of workarounds have been observed in common use for dealing with African-language text in situations where the available fonts, font compatibility or the lack of easy input methods (related mainly to keyboard layouts) hampered the use of the official orthography, notably in e-mail and on the web. A summary of the approaches to using African text in this environment is shown in Table 7.1.

These workarounds are still being used to some extent, even though Unicode and UTF-8, as well as their accommodation in newer applications, in principle permit the use of larger (and complete) character sets. The substitute approaches are particularly evident in e-mail lists and discussion forums.⁷⁵ Despite the advantages offered by Unicode, it would seem that its potential has not yet been realised for a wide range of users and internet applications. This is sometimes due to the lack of fonts (see the discussion in a later section of this chapter), but even when fonts are available, the lack of convenient input systems becomes an issue (see chapter 8). Lack of awareness of Unicode is another issue.

In principle, web content presents the same issues, but static presentations are able to make use of Unicode, even if the HTML coding requires the input of hex or decimal codes for non-ANSI characters. The amount of web content in African languages is discussed below.

It is also worth remembering that in considering African language use in ICT, certain themes are repeated from earlier discussions, before the advent of computers, about the kinds of orthographies to use and the harmonisation of transcription. These were the subject of study and expert meetings already mentioned in chapter 4. Although this context tends to be forgotten, it often laid the foundation for current efforts.

Moreover, the use of African languages with typewriters and in typesetting preceded discussions and propositions concerning their use in computing. Most of these issues have now been forgotten, but they encountered similar concerns to those currently raised with respect to input.

TABLE 7.1 Approaches to using Latin-based orthographies with extended characters and/or diacritics (category 3 and 4 orthographies) in ICT

Standard orthography
<ul style="list-style-type: none"> ◆ Current: the 'correct' orthography according to established practice, legislation or decree ◆ Current with errors: the characters look correct or almost correct but are not; there are several cases in which similar-looking characters may be erroneously selected from the Latin or non-Latin character range ◆ Outdated: a standard that has since been changed (for instance, some usage of Pular in Guinea follows the orthography used until the mid-1980s)
Substitution solutions
<ul style="list-style-type: none"> ◆ Plain ASCII: any extended characters or diacritics are substituted with the closest approximation in ASCII; for example, web pages of the BBC in Hausa are currently in ASCII rather than the standard Boko orthography, resulting in an 'ASCII-fied' text ◆ Plain ASCII with capitals: capital letters are used to indicate extended characters ◆ Other diacritics or combinations: other diacritics or combinations are used to indicate what is not found in the available fonts, an example of the latter being the substitution of dieresis in place of the subdot in the orthographies of certain Nigerian languages
Use of image files
<ul style="list-style-type: none"> ◆ Small images for individual extended or diacritic characters: these are usually GIF or JPG images that are inserted in the text at the appropriate position ◆ Large images of entire texts: these are usually PDF files, but may sometimes also be GIF or JPG files
Hybrid approaches
<ul style="list-style-type: none"> ◆ Combinations of the above ◆ Introduction of elements of the orthography of another language; for example, in spelling based on French or English usage that is not part of the orthography

Source: Adapted from Osborn (2001)

Fonts

Fonts, as an aspect of localisation in Africa, have been a particular issue in countries and for languages that use extended Latin orthographies and/or non-Latin scripts. The following sections briefly consider these two areas.

Extended Latin orthographies

Adapting 8-bit fonts to the transcription needs of many African languages – a practice involving various individuals, organisations and projects that has been characterised as 'anarchic' (Cissé et al. 2004) – has apparently been fairly common (see chapter 6). The result has been a number of mutually non-intercompatible

special fonts, which are generally referred to as 'legacy fonts', that are still in use to varying degrees.⁷⁶

There is to our knowledge no comprehensive listing of such fonts, but several examples are listed in Table 7.2.

The Unicode standard makes provision for extended Latin characters, but fonts including these characters have only gradually become available. Other issues arise in relation to combining diacritics (such as tone marks in some cases) that concern aspects of software and the fonts themselves and affect the positioning of diacritics with respect to the base characters. A possible means of circumventing this problem is to use a single glyph to combine a base character with a combining diacritic, but this approach requires the identification of the necessary combinations and is therefore more successful when applied to specific languages than for diverse needs.

There are ongoing efforts both to increase the repertoire of characters available in fonts and to improve the ways in which they treat combining diacritics. Microsoft has enhanced a number of its fonts in its new Windows operating system, and efforts to develop free fonts have continued. A subproject of the African Network for Localisation (ANLOC) project, for instance, focuses specifically on enhancing extended Latin character support for African languages in selected free fonts.

Non-Latin scripts

There are a significant number of Arabic fonts available, either in the 8-bit ISO 8859-6 or Windows 1256 standards or in Unicode. Unicode covers these, of course, as well as some additional characters, mostly for non-Arabic languages of the

TABLE 7.2 *Some legacy 8-bit fonts for extended Latin scripts in Africa*

Location, site or organisation	8-bit legacy fonts	Origin
Mali	Bambara Arial, Bambara Times	Created in connection with an ACCT ^a workshop (late 1990s)
Matchfont.com	Font for Gikuyu	Created by Gatua wa Mbugwa (1999)
Niger	indrap98, ^b La Nigerienne	Created for Nigerian languages (late 1990s)
SIL International	Many fonts for general and country-specific use	Created in the 1990s and since converted to Unicode versions

Note:

a. ACCT: Agency of Cultural and Technical Cooperation (Agence de Cooperation Culturelle et Technique), now the OIF (International Organisation of the Francophonie)

b. INDRAP: National Institute of Documentation, Research and Pedagogical Animation (Institut National de Documentation et de Recherche Pédagogique)

Middle East, but also for certain African languages (for many of which the modified script is locally known as ‘Ajami’). It is not clear how well existing fonts accommodate African usage, in part because standards are informal.

Various non-Unicode font solutions exist for Ethiopic/Ge’ez, and Unicode fonts are naturally more satisfactory for several reasons. There are some fonts for Tifinagh and N’Ko, but relatively few so far in Unicode.

Languages without writing systems

For the numerous less widely spoken African languages that do not yet have writing systems, there are several steps to take before being able to work with text in ICT. Diki-Kidiri (2008) offers a guide to this process.

This page intentionally left blank

Keyboard and input systems

IN ORDER TO USE text on computers – for purposes such as creating or editing documents and completing forms – an efficient means of input is required. This generally, but not necessarily, takes the form of a computer keyboard. Computer keyboards are mass produced for the major languages, which leaves languages written with characters other than those provided for on production keyboards at a disadvantage. This chapter deals with the challenge of providing for the input of extended Latin and non-Latin characters.

Keyboards

Computer keyboards largely followed the design of typewriter keyboards, which were originally designed for the languages supported by ASCII and ANSI. For script requirements beyond these, methods for facilitating input have had to be devised.

However, a computer keyboard operates at a more abstract level than a mechanical typewriter, although to the user, its functioning may appear as tied to the letters indicated on the keys as in the case of a typewriter. One of the reasons for this is the configuration of keyboard layouts as part of software design, but the keyboard-driver software can be written or adapted to yield any particular character for any key or combination of keys (by using deadkeys or the Alt- or AltGr- key plus a letter key). This can be done in several ways: by the user in changing the commands or shortcuts for individual keys on the computer they are using; by anyone with a keyboard layout program such as Tavultesoft's Keyman or Microsoft's Keyboard Layout Creator (MSKLC), which is designed to be used with other software; or by a software programmer or localiser in setting the parameters for the keyboard (including possibly providing options for the eventual user) in the software itself. None of these means requires any particular attention to what is printed on the keys of the keyboard, although some key assignments might be more intuitive for users. Furthermore, the potential for standardising such solutions is important. For instance, commercial software companies and vendors of computer hardware are

naturally keen to coordinate with some kind of standard for the languages of major markets.

Another approach entails developing production keyboards (the physical keyboard hardware) and a keyboard driver (perhaps including fonts) for one or more languages, for example, the Konyin keyboard for Nigerian languages.

The following sections deal with all the options mentioned except for the first, namely the modification of individual key commands.

Keyboard layout creation

For languages with extended Latin or non-Latin scripts, but without a pre-computing input model of any kind, it is relatively easy to set keyboard shortcuts or design keyboard layouts. In fact, programs such as Keyman and `MSKLC` make it easy for anyone so disposed to design and share a particular layout.

The design of a keyboard layout, when starting with a pre-existing model (such as the English `QWERTY` or French `AZERTY` keyboards), actually begins with the choice of which letters to retain and which to add. It is possible to retain all displayed characters and to assign additional characters to another level activated by a modifier key (`Alt`-, `Ctrl`-, `AltGr`-). Keyboard design programs also allow the option of setting deadkeys, which when followed by another keystroke, produce a character or combination different from what is shown on either key. The issue of keyboards is somewhat complicated by the fact there are at least the following three levels of consideration in their design, and within those, several alternative solutions that can be followed.⁷⁷

The general approach to providing for keyboard input of extended characters and diacritics involves the following possibilities:

- ◆ Substitution, meaning that a key is reassigned.⁷⁸ This basically means that one has to change keyboards for each language used. There are two kinds of substitution:
 - ◆ Keys for letters not used in a particular language are reassigned to characters or diacritics that are used in the target language, but not in the language for which the keyboard was designed. In the case of non-Latin alphabets, this could be all the alphabetic keys.
 - ◆ Non-alphanumeric keys on the original keyboard are reassigned to letters in the target language.
- ◆ Key combinations (also called modifier keys), meaning use of two keys, usually `Alt`-, `Ctrl`-, both `Alt`- and `Ctrl`-, or `AltGr`- keys plus another, usually a letter key, that together yield something other than what is assigned to the letter key alone. In some cases, for example, the Konyin keyboard, there is a special key that functions as `AltGr`.

- ◆ Key sequences such as the following:
 - ◆ Deadkeys, meaning a key that when struck yields no character, but when another key is tapped yields a character or diacritic that does not appear on the keyboard. This feature is used in the Windows United States International keyboard option for accents; the apostrophe, double quote and circumflex are deadkeys, yielding accented vowels when followed by a vowel, for instance. This approach works only where the pair of keys will yield a pre-composed character and not where two characters (such as a combining diacritic on a base character) are involved.
 - ◆ Operator keys, meaning that accent keys are added after the base character. These are in effect the opposite of deadkeys.
 - ◆ Combinations of the above.

Other considerations are related to the audience and include the placement or assignment of keys for individual languages, as well as making provision for multiple languages in single layouts to satisfy the needs of multilingual countries or regions.

In general, it seems that keyboard designers take whichever approach to arranging that they consider best; opinions and preferences vary. As regards the language or languages supported, there are also alternative approaches, and the focus is often on layouts for individual languages. A case could be made for greater consensus with respect to the arrangement of keys; in multilingual societies, as in Africa, the strategy should be to support multiple languages if possible. However, efforts to devise keyboard layouts to accommodate too many languages can result in overly complicated layouts.

Keyboard design and standards

Designing keyboards to meet diverse language needs, as part of localising software or creating products, can be as simple as the preceding discussion suggests, but is also related to the larger concern with standards. Standards benefit both localisers and ultimately users by defining and meeting expectations – in other words, creating a predictable environment for programming, localising and computer use.

An international set of guidelines (iso 9995) exists for the proposal and implementation of standard keyboards for a given situation (such as a language or group of languages).⁷⁹ Among other things, it indicates that a keyboard has three groups of key assignments, namely:⁸⁰

- ◆ Group 1 is the basic layer with a base and shift (for lower and upper case).
- ◆ Group 2 is the national layer with a base and shift. There is a locking shift to access this.

- ♦ Group 3 allows for supplemental characters to be entered. This is a single plane and uses a non-locking shift.

Any longer-term strategies for keyboard development would have to consider these guidelines as well as the needs of the languages and expectations of intended users.

Keyboards for Africa

Using computer keyboards designed for Europe and North America (in particular the English QWERTY and French AZERTY keyboards) is the rule in sub-Saharan Africa. The only language covered in this survey that has well-established keyboards of its own is Arabic.⁸¹ There are input systems for Ethiopic/Ge'ez in the Horn of Africa, although no standard yet. Most of the discussion in this chapter therefore deals with efforts to provide for Latin-based transcriptions, mostly by using extended characters and diacritics (see chapters 4 and 6 for background).

Where languages essentially use the same characters that are indicated on the ELWC keyboards and in the software (in the case of category 1 and 2 orthographies), the question of new input methods generally does not arise. However, the numerous African languages that use extended characters and diacritics pose varying challenges. In the case of the many world languages for which specialised typewriters were in use at the time that desktop computers were introduced, the typewriter keyboard was adapted to the computer keyboard. There were few such languages in Africa, and to our knowledge, they have had no apparent impact on computer keyboard design.⁸² Alternative workarounds were therefore necessary.

The keyboards used in Africa were mostly designed for English or French, and a fair number of layouts have been designed for the input of specific African languages or groups of languages.

As already indicated in chapter 6, interfaces for inputting special characters and diacritics in Latin script can be created in several ways by using programs such as Tavultesoft's Keyman program or Microsoft's MSKLC utility, or simply by assigning keys within a word-processing program. These methods are not difficult to implement and are increasingly available for various languages, countries or regions.⁸³ Some examples of efforts to design keyboards for African language needs are listed in the website companion to this book.

Keyboard layouts for Latin-based scripts have received the attention of both individuals and firms. In some cases, as in Nigeria, this issue has also received official attention from government,⁸⁴ but keyboards for African languages have generally not been considered either broadly or systematically. A number of layouts have thus been designed over several years for one or other situation in Africa, and

in some cases, a keyboard driver and 8-bit font have been developed as part of a package.⁸⁵

There are, however, negative aspects to the facility for creating and disseminating keyboard layouts. Chantal Enguehard of the University of Nantes and the RIFAL project has expressed concern that the proliferation of layouts might become confusing (Enguehard 2006). She and Harouna Naroua of the University of Niamey have been researching various keyboard layouts for evaluation and comparison (Enguehard & Naroua 2008).

The discussion of keyboard layouts in Africa inevitably leads to the topic of using alternatives to the QWERTY or AZERTY keyboards or developing keyboards specifically for African languages. In one case, for instance, Chinedu Uchechukwu, a Nigerian linguist based in Germany, suggested working with the German QWERTZ keyboard, which has one more key than the QWERTY keyboard. Even a single extra key could facilitate arranging the keyboard to include the additional diacritical characters necessary to compose in Igbo. This idea and others that led to the creation of various keyboard layouts⁸⁶ were the outcome of discussions on several e-mail forums (see Table 13.1).

The only computer keyboard of which we are aware that is produced specifically for African languages is the Konyin keyboard for Nigerian languages that has already been mentioned.⁸⁷ It follows that new layouts should probably not depart much from the keyboards to which current users are already accustomed, which are generally English or French keyboards.

The entire African continent, particularly sub-Saharan Africa, currently uses keyboards originally designed for a Western European or North American environment.⁸⁸ Keyboards for English, French and Portuguese are to some extent useful, since these are often the official languages in African countries; such keyboards could form the basis for more Africanised keyboards. The current proliferation of new keyboard layouts for African languages may have some drawbacks, however, and from that process may come new concepts for producing keyboards based on QWERTY and AZERTY that are more suitable for Africa than the traditional European ones, as the Konyin keyboard attempts to do.

Nigeria, in particular, has witnessed several attempts to design keyboards to accommodate the special character needs for transcribing the many Nigerian languages. These efforts and others in francophone African countries using the AZERTY keyboard might result in at least two keyboards for Africa, each of which could accommodate more than one keyboard layout (see also chapter 9).

ANLOC currently has a subproject with the aim of developing keyboard layouts for African languages.

Alternative input methods

Alternatives to the traditional keyboards are also being developed or are already in use internationally, including graphics tablets used as keyboards or with handwriting recognition; virtual onscreen keyboards; LED keyboards that display the active characters in the keys themselves; and speech-to-text applications.

Graphics tablets

A graphics tablet is an external utility (although sometimes it is physically incorporated into a keyboard) that permits graphic signals to be sent to a software application. This is generally done with a stylus, which may be used to input text input in two ways:

- ◆ With a keyboard template superimposed and corresponding software, such that touching the locations indicated on the template for each character produces the intended character;
- ◆ Using a handwriting recognition program, such that a character written with the stylus on the tablet produces the intended character, or a written word produces the intended character string.

Some graphics tablet keyboards were developed (for production or concept) by Lee Pearce of Large-Format Computing in 2003.⁸⁹ This solution would seem especially useful for a syllabary such as Ethiopic/Ge'ez, for which a graphics tablet keyboard has indeed been developed, but it has apparently not proved popular in other contexts. Part of the reason may be that the input method using a stylus is either slow or simply unfamiliar. This method might prove more useful for a script such as Vai if potential users consider it to be more efficient for their needs than developing a keyboard-based input method editor (IME). An advantage of a graphics tablet, however, is that as a USB device, it can be used alongside any other traditional keyboard to facilitate multilingual or multi-script input.

Virtual keyboards

Virtual onscreen keyboards – either full keyboards or a limited set of keys for extended characters – are another option. Virtual keys for extended characters in interactive web applications such as forms or e-mail are fairly commonplace, but their use for African languages does not seem to have been very widespread, although these were used on the African-language e-mail sites discussed in chapter 10.

Full virtual keyboards are possible in various applications, but have the limitation of taking up screen space.

A general limitation of virtual keys is that they require the user either to switch between a standard keyboard and the virtual keys onscreen or to use only the virtual keyboard, both of which methods are time consuming.

LED keyboards

A more promising long-term keyboard solution for multilingual computing, whether in Africa or elsewhere in the world, makes use of backlit keys that indicate the assignments of keys in a particular keyboard selection and can in principle accommodate and display any keyboard arrangement. This emerging technology is being pioneered by Art Lebedev Studios in Russia under the name of Optimus.⁹⁰

Speech-to-text applications

Speech-recognition technology and its use in speech-to-text (STT) applications has interesting potential for inputting text. STT accuracy has greatly improved and become quite good. A noted commercial STT software program for English, the Dragon NaturallySpeaking program of Nuance, demonstrates its potential.

There has been some work on speech recognition for the Yoruba and Somali languages, although no products are available.

STT is further discussed in chapter 12.

This page intentionally left blank

Defining languages in ICT: tags and locales

FAIRLY EXTENSIVE LOCALISATION SOLUTIONS for diverse languages are possible with appropriate fonts and keyboards. However, a system of identification becomes necessary for localising software, adapting an interface to user preferences, and identifying web content by language and country. This is achieved through language codes and compiling locale data.

Languages and the iso 639 standards

Languages can be identified in documents on the web using certain codes, and software can be designed to insert these codes when saving in HTML. The most important of these codes are defined in iso 639. There are also supplementary language tags defined by the Language Tag Registry Unit (LTRU), a working group of the Internet Engineering Task Force. In addition, locale information, using iso 639 language tags and other information, facilitates localisation. This section discusses these, their relevance for Africa and the issues they raise about how to define 'language' in various ICT applications.

General

As of 2009, four international standards for identifying languages had been approved by the International Organisation for Standardization—iso 639-1 two-letter codes; iso 639-2 three-letter codes; iso 639-3, also three-letter codes, for all language categories identified by *Ethnologue*; and iso 639-5 for language clusters and groups. Two other standards are planned and are being formulated (see Table 9.1).

This set of standards serves several purposes, including the identification of the languages of web content and the selection of appropriate locale information, where it exists. There is a certain redundancy in parts 1 and 2 of iso 639: iso 639-1 uses two-letter codes, which mathematically provided too few identifiers to accommodate the languages of the world; iso 639-2, which uses three letters, overcame this shortcoming.

TABLE 9.1 *iso 639 categories for identifying language (current and planned)*

ISO 639-	Description	Status	Reference site
1	Two-letter codes for languages	Existed for several years before being formally adopted in 2002	http://www.loc.gov/standards/iso639-2/php/English_list.php
2	Three-letter codes for languages and collections	Adopted in 1998	http://www.loc.gov/standards/iso639-2/php/English_list.php
3	Three -letter codes for individual languages (exhaustive)	Adopted in 2007	http://www.sil.org/iso639-3/codes.asp
4	Guidelines and principles for language encoding	In draft stage in 2009	http://en.wikipedia.org/wiki/ISO_639-4
5	Three-letter codes for language groups	Adopted in 2008	http://en.wikipedia.org/wiki/ISO_639-5
6	Four-letter codes for language variations	In final draft in 2009	http://en.wikipedia.org/wiki/ISO_639-6

Like iso 639-1, however, iso 639-2 was originally created for terminology and bibliographic purposes, and therefore did not attempt to cover all languages. Moreover, there was apparently no methodology to assure systematic coverage, and it was thus possible to have code elements that were redundant and overlapping in their descriptions. For a while, iso 639-2 included both bibliographic and terminology codes for certain languages.⁹¹ A case of overlapping codes is that of *ak* for Akan, *tw* for Twi and *fan* for Fante – Twi and Fante are usually considered to be Akan languages.

The latest instalment – iso 639-3 – uses the criteria of SIL International and *Ethnologue* in attempting to identify all languages. Another problem is that of the articulation between iso 639-1 and 639-2 on the one hand and iso 639-3 on the other. In some cases, iso 639-3 uses separate codes for a single entity coded in iso 639-1 and 639-2. The category of ‘macrolanguage’ has therefore been adopted for many of the iso 639-1 and 639-2 language categories that correspond with several languages as defined in iso 639-3 (some others have been reclassified as ‘clusters’ of closely related languages in iso 639-5).

African languages in iso 639

Several African languages (or clusters of languages), including Arabic, have iso 639-1 two-letter codes. A larger number have iso 639-2 three-letter codes. There does not appear to have been any strict methodology applied in choosing the

language categories, as the first two parts include individual languages and categories that group closely related languages. Moreover, with the advent of ISO 639-3, which adopts the methodology and categories of *Ethnologue*'s list of languages, a different set of criteria has been introduced.⁹²

At some point, it would be desirable to consider a more systematic approach to selecting codes for African languages and clusters, perhaps in the process of discussing parts 4 and 6 of ISO 639. This might optimally involve linguists specialised in African languages as well as perhaps ACALAN. Indeed, since we are talking about international standards affecting Africa, and a number of African countries are affiliated to ISO (ISO 2006), it would be ideal if at least some of those countries were to participate in the process.

Locale data

Locale data are essential for certain languages used in computing and on the internet. Texin (2006) describes a locale as: 'a mechanism used in the Web, Java, and many other technologies to establish user interface language, presentation formats, and application behavior'. This is, in effect, another example of the way in which the internationalisation of ICT facilitates localisation.

General

A locale consists of basic information on certain needs and preferences, such as character ranges in Unicode, that are necessary to display text in the language, including sort order, currency units, day and date format, and decimal markers. Commonly, local data are indicated for a language and a country, using ISO 639 (language codes), ISO 3166 (country codes) and, if a language is written in more than one script, also ISO 15924 (writing system codes). Relatively few languages in Africa currently have locale data.

Completing a locale and filing it with the appropriate registry is a necessary step in localising ICT for a language. The principal registry is the Common Locale Data Repository (CLDR), managed by the Unicode Consortium,⁹³ but there is also one for OpenOffice.org.

Locale data for African languages

Relatively few African languages have locale data. There have been several efforts to redress this situation.

An early initiative was a project called Yeha, which was established by Daniel Yacob in 2001 with a focus on locale data for languages in East Africa.⁹⁴

Early in 2006, Alberto Escudero-Pascual and Louise Berthilsson of the non-governmental organisation IT46 launched an online locale generator tool – LocaleGEN⁹⁵ – to assist people in compiling locale data for OpenOffice and CLDR. This led to the filing of several more locales. In 2008, a subproject of ANLoc launched an effort to compile locale data for 100 African languages. This work is expected to continue until at least 2010

Table 9.2 lists the African languages for which locale data had been filed with CLDR by May 2009. Locale data are filed for a language and a country, with alternative writing systems also being indicated in some cases.

TABLE 9.2 *African languages filed in CLDR 1.6.1*

Language	ISO 639 code used			Country/countries filed for
	-1	-2	-3	
Afar	aa			Djibouti, Eritrea, Ethiopia
Afrikaans	af			South Africa, Namibia
Akan	ak			Ghana
Amharic	am			Ethiopia
Arabic	ar			Algeria, Egypt, Libya, Morocco, Sudan, Tunisia and several countries in south-western Asia
Atsam			cch	Nigeria
Blin			byn	Eritrea
Chewa/Nyanja	ny			Malawi
Coptic		cop		Egypt
Ewe	ee			Ghana, Togo
Ga		gaa		Ghana
Ge'ez		gez		Ethiopia, Eritrea
Hausa	ha			Nigeria, Niger, Ghana, Sudan; Latin and Arabic scripts
Igbo	ig			Nigeria
Jju			kaj	Nigeria
Kamba		kam		Kenya
Kinyarwanda	rw			Rwanda
Koro?			kfo	Nigeria? [there is an error in this locale]
Kpelle		kpe		Guinea, Liberia
Lingala	ln			Congo, Democratic Republic of Congo
Ndebele, South	nr			South Africa
Oromo	om			Ethiopia, Kenya
Sidamo		sid		Ethiopia

Language	ISO 639 code used			Country/countries filed for
	-1	-2	-3	
Somali	so			Somalia, Ethiopia, Kenya, Djibouti
Sotho, Northern		nso		South Africa
Sotho, Southern	st			South Africa, Lesotho
Swahili	sw			Kenya, Tanzania
Swazi	ss			South Africa, Swaziland
Tigre		tig		Eritrea
Tigrinya	ti			Eritrea, Ethiopia
Tsonga	ts			South Africa
Tswana	tn			South Africa
Tyap			kcg	Nigeria
Venda	ve			South Africa
Wolaytta, Walamo	wal			Ethiopia
Wolof	wo			Senegal
Xhosa	xh			South Africa
Yoruba	yo			Nigeria
Zulu	zu			South Africa

Source: CLDR 1.6.1 (as at May 2009) (<http://cldr.unicode.org/>, Accessed July 2009)

This page intentionally left blank

Internet

THE USE OF COMPUTERS to exchange and access information via internet connections has, like other computer uses, inevitably involved attempts to use diverse languages. In Africa, for various reasons already discussed (including sociolinguistic, educational and technical issues related to rendering non-ASCII characters until recently), such uses have lagged behind.

E-mail

E-mail was an obvious first step in the use of the internet in Africa, and for many it remained for long the principal use. The early development of e-mail in Africa was linked to efforts to establish connectivity in the various countries of the region (see Levey & Young 2002; Renaud 1994). Later, when World Wide Web usage was establishing itself in other parts of the world, a webpage by e-mail service was offered by Kabissa for the many users in Africa that did not have the bandwidth to download the pages via browsers.⁹⁶

Although it is difficult to ascertain the level of use of diverse languages in private e-mail, information is available to estimate the use of African languages in this medium. For instance, there were in the early 2000s two web-based e-mail services that made provision for composition in several African languages, using virtual keys for extended Latin characters, namely Africast.com and Mailafrica.net.⁹⁷

There are also indications of African language use in e-mail forums and internet chatrooms. Da Costa (2002) reported that: 'a lot of informal chat and e-mail is conducted in written indigenous languages such as Swahili'. In recent years, a number of e-mail forums have been set up in which much, if not most, of the traffic is in an African language. For instance, there are several Hausa and Swahili e-mail lists in which these languages (which are probably the most widely spoken indigenous languages on the continent) are the primary languages of communication, and Van der Veken and De Schryver (2003) found forums in Hausa, Somali and Lingala.

Esterhuysen (2002) reported a service that in 1992 handled the translation of African languages via e-mail.

Internationalisation and the web

The efforts to facilitate the use of languages in ICT include initiatives that are specific to, or have special relevance for, the internet. The UTF-8 implementation of Unicode (see chapter 6), for example, is increasingly used for multilingual web content and e-mail.

The World Wide Web Consortium (w3c)⁹⁸ sets standards for the mark-up of webpages to facilitate, among other things, diverse language content.

Some of the discussions about the way in which the web is used have implications for internationalisation and localisation. There have, for instance, been discussions over several years of how the web may evolve organically into something called the 'Semantic Web', with characteristics facilitating browsing, linking and the manipulation of information. More recently, discussions of Web 2.0 have organised the thinking of certain experts as well as commercial interests about new ways in which the web can, and indeed does, function and serve needs in increasingly interactive ways.

Web content in and about African languages

African languages are represented on the web, although not prominently, as media of communication, but the actual level of use is emerging as a topic of discussion. It is easy to form the impression that African-language content (in the languages rather than about them) is still rare and only gradually increasing. The results of several surveys give a more complete picture of the current status and evolution of African-language web content.

The surveys can be grouped into the four categories of informal surveys, censuses, statistical estimates and crawlers. Simple informal surveys, using search engines, of several years of web content by language have unsurprisingly found insufficient content in any African language to rank them as high as certain minority European languages.⁹⁹

More focused surveys yield more interesting results. For instance, an informal survey in Tanzania in 2001, as part of a larger report for the Swedish International Development Agency, estimated that ten per cent of websites with a Tanzanian focus had at least some Swahili content (Miller Esselaar & Associates 2001), but most of these sites did not have majority content in the language.

An extensive study by Diki-Kidiri and Edema (2003) involved searching, listing and counting websites. It found a significant number of sites that treat African languages in one way or another, but also showed that these generally have minimal content in the languages themselves. In effect, a large proportion of the

sites they surveyed consisted of presentations about African languages, including online dictionaries and instructional pages. Van der Veken and De Schryver (2003) used a different search methodology and statistical extrapolation. By counting hits of particular words and estimating the larger number of words that these hits might represent, based on the frequency of the searched terms in a typical text, they concluded that there might be significantly more African-language web content than was generally thought, but it is difficult to determine from such estimates what type of content they would imply.¹⁰⁰ Moreover, if Diki-Kidiri and Edema (2003) are correct that most sites with African-language content are about the languages, it is unlikely that we are discussing typical text. Dictionaries, for instance, will generally include less common words on the same basis as frequently used words and terms; in other words, this may be a context in which this statistical approach does not work.

An ongoing study undertaken by the Language Observatory has sought to more accurately evaluate diverse language content on the web. Using a technique first used for Asian languages involving a web-crawler and statistical analysis (Suzuki et al. 2002), an African-language survey was initiated in 2006.¹⁰¹ Its focus is content on websites in top-level African country domains, and a general finding of the study is that, for most African languages, there is not much content.

Analysing the character of the content in particular languages is, of course, more complex than estimating the presence of the languages on the web. The study by Diki-Kidiri and Edema (2003) seems to be the most revealing in this regard.

Another approach to investigating web content is to categorise it in terms of its origins and audience. Ballantyne (2002) used this approach in studying web content from developing countries. Although his study did not address the issue of language (or languages), his schema could be very useful in understanding the nature of the content by considering where it originated and who the intended or anticipated audience would be. For instance, the large percentage of sites with descriptive African-language content reflects the dominance on the web of non-Africans that may be interested in learning or knowing more about African languages. Looking at content in this way fosters an understanding of who is localising content for whom and how their work can best be facilitated. Furthermore, by looking at what is *not* done in terms of localisation, one could use this schema to analyse the reasons for that and what might be needed to achieve better results.

Web content about African languages

Web content about African languages deserves special comment, given its prominence in African-language web content. This is a broad category that includes a range of presentations of varying quality, from the very informal and sometimes

incomplete, to well-considered and sometimes ambitious projects. In general, this category reflects the fact that the potential audience for African-language topics has predominantly comprised people with either little or no knowledge of the language, or with knowledge of the language but little familiarity with its written form; the latter includes people who are not literate in their L1, as already mentioned in chapter 4.

Among the notable online materials about African languages are the Kamusi Online Living Swahili Dictionary,¹⁰² online descriptions of academic use such as the Hausa site at UCLA, and some efforts to use the web as an instructional tool. The latter can be categorised by audience, including L2 language learners, generally outside of Africa, with little knowledge of the language (for instance, at a university); children of expatriate Africans (this is sometimes called 'heritage language' education); and L1 literacy for Africans within Africa, including L1 literacy for the otherwise literate or the illiterate (such as the ALI project in Cameroon in the early 2000s).¹⁰³

In addition to meeting certain needs and raising the profile of African languages in general, such content approaches in principle also enhance the environment for other kinds of localisation.

New dimensions in web content

Two more recent features on the internet offer new potential for expanding African-language content, namely weblogs and Wikipedia. Weblogs, or blogs, are becoming increasingly widespread around the world, including in Africa. There are already several blogs in African languages. Blogging is a relatively easy way for individuals to produce text content in any language, given that there are free sites offering space to anyone that wishes to start a blog. As long as blogs and bloggers remain a significant feature in cyberspace, we should expect to see more content in diverse African languages.

Wikipedia is an online encyclopaedia that is expressly multilingual in its approach. There are almost 40 Wikipedia editions in African languages – with Arabic, Afrikaans, Swahili and Yoruba best represented – but many African languages have very little content. Following discussions at the Wikimania conference in 2006 on how to facilitate growth of these and other African-language editions of Wikipedia, an effort was launched to coordinate the work.¹⁰⁴

One of the topics of concern has been the lack of growth of some of the African-language editions of Wikipedia (and other Wikimedia projects), and the way in which the foundation that manages these projects should deal with cases where there are evidently no active users contributing to the editions.

Several interesting observations have been made. Firstly, some African-language editions of Wikipedia have grown quite quickly when a contributor, or a few such people, start working on them, for example, the Wolof and Yoruba editions.¹⁰⁵

Secondly, a study in Tanzania showed that even skilled internet users did not know how to create a Wikipedia page. This finding has implications for efforts to promote Wikipedia in Africa.

Internationalised domain names

While Unicode and the development of means to render complex script requirements in principle permit content in any language on the internet, another consideration is the names of domains in diverse languages and writing systems. For some years, there has been interest in multilingual or internationalised domain names for Africa, as evidenced for instance by the formation of an African chapter of the Multilingual Internet Names Consortium (MINC) called AFRIMINC.¹⁰⁶

In 2005, a project backed by ACALAN, UNDP and the Agence Intergouvernementale de la Francophonie took up the issue at a time when international discussion on the matter had become more serious.¹⁰⁷

In any event, the Internet Corporation for Assigned Names and Numbers (ICANN) has been testing alternative ways of handling internationalised multilingual domain names in non-Latin scripts such as Arabic.¹⁰⁸

This page intentionally left blank

Software localisation

SOFTWARE APPLICATIONS IN AFRICAN languages can be regarded both as a fundamental means of facilitating greater use of, and soft access to, the technology, and as a facilitator for those who wish to develop web content.

Applications and operating systems

As already discussed (chapter 2), the localisation of software includes both operating systems and software applications. Operating systems in effect run the computer hardware and host and manage the applications with which users interact. Applications include a range of programs, from common office-type software (including word processors and spreadsheets), to games, browsers and technically specialised programs such as geographic information systems (GIS).

Most work on localisation in African languages has related to software applications (such as word processors or browsers, or Office suites by Microsoft or OpenOffice), but Microsoft has projects for localising Windows, and there are several initiatives for localising Linux systems (notably Ubuntu).

Trends in proprietary software

There is a wide range of proprietary software, which is defined here as closed-source software, governed by a traditional patent and generally available commercially. Much of this is localised, if at all, only for languages for which there are significant markets. The proprietary software that has been localised for some African languages tends to include software such as word processors or internet browsers.

Microsoft Corporation (2004), whose products dominate proprietary, and indeed all, office-software markets, has devoted attention to a few major African languages. Microsoft's approach in its Language Interface Pack (LIP) project has included the translation of about eighty per cent of the commands, which are overlaid on a base system such as English.

However, one of the earliest and initially most promising of these localisations, for Swahili, has apparently not lived up to expectations.¹⁰⁹ There are various opinions about the reasons for this, ranging from lack of follow-through (see chapters 2 and 3), to the methodologies used in developing the Swahili package, and the price of the software itself. There are certainly lessons in this experience for all localisation initiatives; these need to be fully explored for the benefit of ongoing and future localisation projects.

Trends in free and open-source software

Free and open-source software (FOSS) deserves particular attention in the African context because of its demonstrated potential for localisation in diverse languages. This is due to cost advantages and the accessibility of its code. However, FOSS is as much a movement¹¹⁰ as an approach to developing and marketing software.

The FOSS movement has spread in Africa and the Arabic-speaking world much as it has elsewhere, although it is used less extensively because African user communities are generally smaller than elsewhere. Nevertheless, FOSS user groups exist in many African countries.

With a large, growing and diversifying base of enthusiasts internationally, FOSS offers an unusual range of possible support as well as some challenges for tapping it. (The PAL project hopes to facilitate communication and coordination of efforts among FOSS localisers operating in, or interested in, Africa.)

Perhaps ironically, FOSS has some support from major corporate entities such as Sun Microsystems and IBM. The initiative to develop the popular Debian Linux-based system, Ubuntu, was funded by the South African entrepreneur, Mark Shuttleworth.

Despite the potential for localisation, FOSS in Africa (as in much of the global South) is most often conceived of as a means of reducing costs and dependence on proprietary software, notably that of Microsoft. The links between FOSS and localisation in Africa, however, seem to have been somewhat slow to develop (with a few notable exceptions, such as Translate.org.za in South Africa). It seems that, at least initially, various country-level FOSS associations, as already mentioned, are focusing on promoting the use of software such as OpenOffice in the official languages. Regional groupings show the same focus.

While FOSS and localisation seem increasingly to be meeting at an international level, the question remains how to foster greater involvement of African FOSS communities in localisation processes in their respective countries and regions.

Tools to facilitate the localisation of FOSS

At a technical level, the potential for localisation of any software, especially for smaller language communities, is limited by the number of people with the requisite knowledge of programming and of the languages in question. One tactic for facilitating localisation is thus to design tools to make it easier for potential localisers that may have the motivation and the language skills, but not the technical background to translate software.

An example is the Pootle interface developed by Dwayne Bailey of Translate.org.za and Javier Sola of Khmeros to facilitate the localisation of OpenOffice by people without high levels of technical expertise.¹³¹ The object is in effect to 'lower the bar' of entry into localisation so that people with language expertise but little technical or programming background can undertake localisation projects for various FOSS applications.

This approach has also been undertaken in other contexts such as compiling locale data in an online tool. It will be interesting to monitor the effect of opening aspects of localisation to people with language skills and FOSS enthusiasm, but without much computer specialisation.

Software localisation in Africa

Early efforts

There were several efforts to localise software for African languages as early as the 1990s, for example:

DOS-based:

- ◆ Koma Kuda: word processor for Manding in the N'Ko script, a right-to-left writing system;
- ◆ Afri-Alpha: software for production in the diverse languages of Benin and West Africa (the interface was apparently in French and was not translated into any African language).

Windows:

- ◆ Somitek Hikaadiye: Somali language word processor and spellchecker (Windows 3.1);
- ◆ Oromosoft: Oromo language in the Latin-based Qubee orthography;
- ◆ Amharic WordPerfect: a locally developed version, not on a commercial level.

Such early localisation in or for African languages is an interesting subject, in part because it occurred during the period before the Windows environment changed the way in which computers operated (a change that paralleled the domination of the market by Microsoft and its Office software), and before the explosion in the use of the internet. It was a period in which there were a number of initiatives that were not widely known but nevertheless quite practical. Although such efforts have apparently not generated sustained activity in the changed environment, they still point to early recognition of the potential and may offer some useful experience (for example, in the terminologies developed and the keyboard layouts used).

In recent years, there has been more recognition of the need for localised software and efforts to localise. Initiatives such as Translate.org.za have led the way among foss localisers in Africa, and there has been some interest by the major proprietary software firm, Microsoft.

Current efforts in foss

Table 11.1 lists African languages for which there are currently active or completed projects to localise OpenOffice software.

Some other software has also been localised. For example, the non-governmental organisation, Open Knowledge Network, developed its own localised software for project purposes. Another example is a children's computer drawing program, TuxPaint, which has been localised into Swahili and more recently into Xhosa and Venda. Yet another is the Mozilla browser in Luganda.

As regards operating systems, there are some projects for localising Ubuntu Linux.

Web interfaces

Another interesting area to consider is the localisation of user interfaces for online tools such as search engines. This area is related both to software localisation, in that terminologies and user profiles need to be considered, and to web content, in that it involves the use of African languages on websites, which happen also to be user interfaces. An example is the Google program, Google in your Language, which includes several African-language versions, with several more in development.¹¹² The social networking site Facebook also has a localisation program, with several African-language versions in development

As a general rule, companies running search engines or networking sites use volunteer translators to localise interfaces. This means that the localisation depends on individuals or groups offering their time, the languages involved depend on

TABLE 11.1 *OpenOffice localisation projects*

Language	Code	Website	Responsible person
Afrikaans	af	http://af.openoffice.org/	Andreas Pauley
Amharic	am	.	Daniel Yacob
Arabic	ar	http://ar.openoffice.org/	Ossama Khayat
Kinyarwanda	rw	http://wyoming.e-tools.com/Kinyarwanda/OO/stats2.html	murf@e-tools.com
Ndebele, South	nr	.	.
Northern Sotho/Sepedi	ns	http://www.translate.org.za	Dwayne Bailey
Sotho (Southern)	st	http://www.translate.org.za	Dwayne Bailey
Swahili	sw	http://http://www.kilinux.org	Alberto Escudero-Pascual
Swati	ss	.	.
Tigrinya	ti	http://ti.openoffice.org	Mahfuz Ibrahim
Tsonga	ts	.	.
Tswana	tn	http://www.translate.org.za	Dwayne Bailey
Venda	ve	.	.
Xhosa	xh	http://www.translate.org.za	Dwayne Bailey
Zulu	zu	http://www.translate.org.za	Dwayne Bailey

Source: <http://l10n.openoffice.org/languages.html>; <http://wiki.services.openoffice.org/wiki/Languages> (Accessed 7 January 2009)

volunteers, and the quality of the translations may vary. In response to the last issue, Google attempts to verify translations in house with employees who speak the languages in question, or with other volunteers.¹¹³

This page intentionally left blank

Mobile technology and other specialised applications

WHILE MUCH OF THE focus of localisation is on computers and commonly used software, mobile technology and a range of specialised applications are also important. Mobile technology is, in effect, the latest development in the miniaturisation of technology and the blurring of the boundaries between once-separate and discrete digital technologies. Not only are telephones handheld, wireless and portable, but computing devices are as well, and these are combined with the functionality of cameras, personal data management tools and the ability to link to the internet, for example.

A number of other ICTs are worth mentioning in the broader context of internationalisation and localisation. These include audio-related technologies, geographic information systems and translation tools. Some of these are also called 'human language technologies', notably technologies for transforming language between written and spoken forms, as well as for translating among languages.

Mobile technology

Miniaturisation of ICT

Mobile technology, including cellular phones and handheld computers, is a rapidly expanding set of devices reflecting ongoing advances in technology that permit smaller devices to do cheaply what larger devices used to be required to do. At the less expensive end of the range of mobile devices, simple cellphones have become far more accessible to people with lower incomes throughout the global South. This market has attracted investment and increased interest in localisation.

At the higher end, the promise of the Simputer model with respect to relatively inexpensive handheld multilingual computing for the poor has not been realised, but with ongoing miniaturisation of the technology and reduction in costs, there may yet be future possibilities for doing so.

Localisation of mobile phones: trends in Africa

Mobile technology in the form of cellular phones has already emerged as a significant ICT in Africa. Cellular phones are increasingly widespread, to a greater extent currently than fixed-line phones, even in the rural areas of some countries. Along with this trend and the evolution of the technology to handle text messaging, there has been increasing interest in localising the user interfaces in African languages. This may be the new growth area for localisation, and certainly its importance is increased to the extent that mobile devices and computers can be used interchangeably to share and process information. Shanglee (2004) describes some of the considerations in localising cellphone technology for South African languages.

Among cellphone companies, Nokia appears to be particularly active in the area of localisation,¹⁴ and Sony-Eriksson and Samsung also market local language interfaces in South Africa. An American company, Tegic Communications (now part of Nuance), has adapted its predictive text software – which facilitates the inputting of words on a telephone keypad and is used in many models of mobile telephone – to several African languages, including Afrikaans, Arabic and Swahili, with Xhosa and Zulu in development (Senne 2006).

Commands in non-Latin African scripts, notably Arabic, are proven, and research on Amharic is well advanced; text messaging in Arabic is incorporated on many phones used in arabophone regions.

Audio dimensions: voice, text-to-speech and speech recognition

The transmission, manipulation and transformation of human speech via computing devices is theoretically possible in any language. Realising this potential could be a priority in the case of oral traditions and where many people are not literate.

Some of the audio technologies are not particularly popular in technically advanced countries. Audio e-mail or voice e-mail (sometimes called v-mail), for instance, do not seem to have become established in such countries and have found only limited use elsewhere. At this point, with advances in technology and voice over the internet possible with VoIP and mobile devices, other applications could be explored, perhaps focusing on voice commands.

Combinations of audio, image and text could be very useful for learning as well as anticipating the needs of users with lower literacy skills.

Text-to-speech (TTS) is of obvious interest in settings where people with access to the technology cannot read for one or other reason. STT is also of interest, as mentioned in the discussion of alternative input methods in chapter 8.

Africa

There has been some research interest in TTS, particularly since many Africans are not literate. In recent years, several programs have been developed for African languages. The Local Language Speech Technology Initiative (LLSTI) has coordinated the development of TTS in Swahili, Zulu and Ibibio, in conjunction with local and international partners in each case. An interesting example of the application of TTS is the Swahili version, which is used for text messaging on mobile G9phones in a Kenyan project originally pioneered by the Open Knowledge Network and the University of Nairobi.

We are not aware of any STT for African languages, although the Nigerian organisation, African Language Technology Initiative (ALT-I), has conducted some research on speech recognition of Yoruba, and there has also been some work on Somali.

Geographic information systems

GIS is a specialised technology for combining spatial information (usually maps) with other data. As the technology and its use have evolved, the potential for its localisation is becoming more attractive. GIS software has become far more accessible than in earlier versions, which required a significant amount of training to master. One of the results of this development is that it has become possible to use GIS technology outside of computer laboratories and in participatory analysis and planning for local development.

Spatial imaging is an ideal tool for land and natural-resource planning at a local level, as it is readily understood even by illiterate people, while at the same time permitting very sophisticated layering of information and analysis of data. In fact, there have been serious efforts in various parts of the global South, including Africa, to combine the use of this technology with established participatory research methodologies, such as participatory mapping, in what is known as public participatory (or simply participatory) GIS (PPGIS or PGIS).¹⁵

However, attention to the localisation of the software into the LIS of rural people is lacking. Perhaps such localisation could facilitate direct access to GIS, or at least the products of GIS, by people who could then benefit even more from it.

Many consider the commercial GIS software marketed by ESRI to be the industry standard. A number of FOSS GIS applications¹⁶ also exist, one of which, the Geographic Resources Analysis Support System (GRASS), is particularly noted.

GIS standards are governed by the ISO 19100 series (mainly concerned with standards for geographic data exchange), which have no direct bearing on the potential for localisation or the process for doing so.

Africa

The use of GIS in Africa is increasing, but to our knowledge, no GIS software has yet been localised in any African language. There is apparently an initiative to translate the GRASS software into Amharic.

As regards the potential for localisation, the designers of the GRASS GIS application are hoping that it will be translated into many languages.¹¹⁷ The software is now Unicode aware.

Computer-assisted translation

The ability to transform thought in writing or speech from one language to another with the assistance of a computer is one of the most interesting uses of ICT in multilingual contexts, but one that has thus far received relatively little attention in Africa.¹¹⁸ Technology in this area is evolving rapidly and has connections with and implications for localisation work.

Computer-assisted translation (CAT) includes machine translation (MT), or automatic translation between languages by a computer program, which aims to translate speech or text from one language into another, in general or specific settings; and translation memory (TM), which is mostly used as a tool to facilitate new translations based on previous translations of the same or similar text content. There have been significant advances in these technologies in recent years, to the point where it is relevant to ask whether we have crossed the threshold to the point where their use will be standard and useful, rather than exceptional and error-prone.

At its most basic, TM is a database of previous translations, done at some point by humans. Strings of text are matched between pairs of languages, so that retrieving terms or phrases in one language will yield the corresponding terms translated into the other language. The object of TM is predictable results from a fairly limited range of input. As such, TM is a standard tool in the localisation industry, serving corporations in industrialised countries.¹¹⁹

Unlike TM, MT has to deal with the unpredictability of human language and somehow yield useful and accurate translations. Experts have been working on its development for a number of years, employing various approaches and technologies. MT is generally considered to have two main broad approaches, which may also be combined: rules-based, in which the structure of the language is analysed

in translating words in a sentence; and statistically based, which involves data as well as examples, in some respects resembling *tm*. The former requires a thorough understanding of the grammar, and the latter, a significant amount of text data for the language pairs involved. Languages with few text resources present a challenge for the development of statistical *mt* tools.

Rules-based *mt* may find particularly useful applications in the case of closely related language pairs. This was the focus, for instance, of the Apertium shallow-transfer *mt* program.

Africa

The subcategory of *mt* referred to as 'shallow-transfer' is of particular interest. This is a simpler approach adapted for translation between closely related language pairs, an example of which is the Apertium open-source translation software.¹²⁰ Like the simpler computer-assisted dialect adjustment or adaptation (*CADA*) programs of the 1990s,¹²¹ this may find significant use among related languages within Africa.

At present, there is not much *mt* for use between African and non-African languages, apart from Arabic (especially Arabic translated into and from English), for which considerable research has been conducted, and commercial *mt* software and even online translation are already available. For the other languages of the continent, there are several projects under way, but actual working *mt* is available only for Xhosa and Pulaar/Fulfulde in pairs with English. The latter have been built and presented online¹²² through the efforts of Martha O'Kennon at Albion College (USA), using Prologue. They are rules-based and can translate only short sentences. Prof. O'Kennon is also collaborating with other individuals on languages such as Akan and Yoruba.

A number of larger-scale projects exist, notably the longstanding project for Swahili known as Salama, under the direction of Prof. Arvi Hurskainen of the University of Helsinki (Finland).¹²³ The African Language Research Project at the University of Maryland Eastern Shore (USA) has an initiative to research *mt* for African languages. There are other corpus-building efforts that envision applying their work eventually in *mt*, such as the *sax* project for Amharic (and some non-African languages) at New Mexico State University (USA).

There are apparently some *mt* specialists in South Africa, but we are not aware of any active *mt* efforts for African languages that are based in Africa.

Translation memory (*tm*) has had some application for African languages. The South African translation company, Web-Lingo, for instance, uses a *tm* program called Trados¹²⁴ in some of its work. The open-source *tm* program, Omega-*t* has an Afrikaans version.¹²⁵

This page intentionally left blank

Achieving sustainable localisation

AS LOCALISATION OF ICT in African languages becomes more important, building on the foundations discussed, it is also essential to consider what is required to facilitate its ongoing development and the accomplishment of its ends. In effect, localisation – which is introduced early in this book as a product that meets and anticipates current and future needs in Africa for linguistically diverse and contextualised use of ICT – is also a process. That process in turn has its own requirements in order to meet its purpose and achieve sustainable results.

This volume has sought, in part, to examine that process and its results. This chapter reviews the needs of localisation and localisers.

Two dimensions of the question emerge: the types of localisation and who is doing the localisation. These are considered in the following sections, which include an assessment of needs, as well as a strategic perspective on these needs related to sustainable localisation in Africa. The needs were identified by participants at a PAL workshop held in Casablanca, Morocco, from 13 to 15 June 2005.

Needs by kinds of localisation and localiser

As already discussed in chapter 2, localisation can involve several concerns, the most important of which have been identified as the equipping or enabling of systems, internet content and user interfaces (software). Localisers have been identified as falling into three main groups, namely Africans in Africa, Africans abroad and non-Africans.

Needs by localisation type

The technical prerequisites for equipping systems and for other kinds of localisation in extended Latin and non-Latin scripts are largely being met through the internationalisation of ICT (see chapter 6). The main need here, whether or not it is recognised by computer technicians, linguists and localisers in Africa, is for

greater awareness of the available resources and, in some cases, training in the use of Unicode.

The creation and translation of localised content is more demanding, mostly in respect of language skills and awareness of tools appropriate to the work, such as appropriate fonts and the means to code in extended Latin and non-Latin scripts. In any discussion of text, however, issues related to the standardisation of the written form of the language itself arise, for example, orthographies, spelling and terminologies. These are more appropriately the concern of applied linguistics, but the dynamics of localisation ecology – for example, unresolved issues related to the written form – may pose problems for certain kinds of localisation, from its achievement to its reception.

Software localisation becomes even more complicated, especially in relation to the technical skills required. One of the issues with regard to language is specialised terminologies for ICT concepts in languages that are new to the technology. Efforts to design tools to facilitate the translation of software by people with language skills but without technical expertise is a recent area that is receiving increasing attention (for example, the Pootle utility, which was mentioned in chapter 11).

This touches upon the varying needs of people involved in localisation in order to most optimally participate. Before considering that aspect in more detail, it is important to consider the dimension of sustainability, which implies a wider range of concerns as well as strategic considerations.

Needs for sustainability

Sustainability, which is a common term in international development, refers to the chances for longevity of a system or achievement. This in turn depends on the design and the way in which it takes into account the factors that affect it.

At the most basic level, the tasks of localisation involve the interface between technology and language, and that these have meaning because of sociocultural factors such as user profiles, the ways in which language is used and cultural appropriateness. As already discussed in chapter 3, this is the core set of dynamics in localisation, but these dynamics are part of the larger environment. Localisers therefore also need to understand this context and to find wider support in the environment.

Needs of localisers

One could say that localisation involves people to the extent that the process depends upon their motivation, skills and organisation. Naturally, people that work on locali-

sation come from diverse backgrounds and have diverse strengths and interests. In general, their needs are defined in the context of the localisation effort of which they are part; in other words, specific issues depend on the work being done.

Nevertheless, one might say that in an effort that, by definition, requires some mix of language, technical and organisational skills, localisers would potentially need information, support and training in one or more of these areas.

The three categories of people working on localisation (Africans in Africa, Africans abroad and foreigners) have different and complementary potential, namely:

- ◆ Localisers in Africa obviously represent the future of localisation in Africa. This is certain to be the fastest-growing group, and any attempt to evaluate localisers' needs in more detail as we move forward will naturally have to focus on them.¹²⁶ In general, one can say that localisers in this context will tend to have more strength with respect to language than technical aspects, which implies a need for some mix of training focused on the technical aspects of localisation and the tools to facilitate it.
- ◆ Africans abroad are a key group in the short term, with an actual and potential impact on localisation far beyond their numbers. This is largely because of their average educational levels and access to ICT. The issue is how to connect, organise and work with such a disparate and dispersed group.
- ◆ Non-Africans that are interested in the progress of African languages in ICT can, and do, play useful supporting roles in linking with broader localisation communities and resources, training, and collaboration on localisation initiatives.

Understanding the needs of localisers

Understanding what localisers in, and for, Africa need in order to start and follow through with a localisation initiative can be approached either by means of an overview and analysis from outside, or from the perspective of the localisers themselves. Both approaches have merit, but the approach of considering the experience of localisers benefits from their familiarity with the material and information needs of the localisation process.

One of the goals of the PAL workshop held in Casablanca in 2005 was to get feedback from localisers about localisation needs. The preparation for the workshop involved outlining anticipated topics and designing sessions to explore these.

Technical and linguistic, strategic and organisational needs

The initial breakdown suggested for analysing the needs for localisation in the PAL workshop involved the two broad categories of technical and linguistic, and strategic and organisational.

It was thought that localisers themselves could expand upon this outline. This was the approach taken before and during the workshop in Casablanca, using the PAL wiki beforehand and then workshop sessions when the participants were together.

Technical and linguistic

The main question under this heading is to identify technical and linguistic needs for localisation efforts to succeed in Africa. Technical and linguistic factors¹²⁷ together are fundamental to localisation tasks, as well as for other treatment of writing and text in ICT. It was therefore believed to be useful to consider these together.

‘Linguistic’ is understood in this context in the broader sense as meaning ‘related to language’, rather than in the narrower sense related to the structure and uses of language, although the latter meaning also has a bearing on localisation work. The pertinent linguistic issues include the standardisation of orthographies; variations in the languages through dialects, resulting in varying degrees of interintelligibility; and terminology.

On the technical side, the issues include those encountered in translating software or content, which might involve certain aspects of internationalisation, such as enabling systems, as well as filing locales.

Strategic and organisational

The main question under this heading is to identify the strategic and organisational needs for localisation efforts to succeed in Africa. It was felt that this easily overlooked category is essential to successful localisation. Too often, parallel efforts needlessly duplicate effort, initiatives begun with high hopes lack the vision to follow through, and the lack of resources becomes an obstacle. Moreover, it is important that these efforts find ways to gain support in building a favourable environment for localisation. It is therefore necessary to discuss and plan how localisation projects can more effectively coordinate their efforts.

Some of the issues in this category were believed to be common to FOSS work in general.

Needs identified at the PAL workshop

At the Casablanca workshop, an effort was made to compile a record of participants' experiences and opinions with respect to the challenges they had faced, the successes they had enjoyed, and what they believed they needed in order to be able to do more and perform better. The method used was brainstorming in small groups and then reporting to the larger group in such a way that each unique idea was recorded on a separate card. The cards were mounted on a blank wall, and the group physically rearranged them into what seemed to be natural categories. The advantage of this approach was to allow categories to emerge from the data rather than imposing a framework. The following five categories ultimately emerged:

- ◆ Government, including relations and the importance of support;
- ◆ Peer networking, including the need for collaboration in FOSS localisation and the rewards thereof;
- ◆ Standardisation, which is important in several categories, such as orthographies;
- ◆ Technical issues, including aspects of the localisation work itself;
- ◆ Sociolinguistic issues, including language-specific issues encountered in localisation.

Much of the discussion was related to OpenOffice and basic user interfaces, although various other technologies were also mentioned.

Comparison and summary

Several remarks are relevant in comparing the two lists, namely the technical and linguistic, and the strategic and organisational aspects. It is interesting that one of the strengths and interests related to localisation that was discussed in the Casablanca workshop was peer networking.

It came as something of a surprise that the localisers attending the workshop in Casablanca did not address any strategic or organisational issues. Possible explanations are that this may be a function of relative ease for localisation projects at this stage of their existence, in that such issues have not yet arisen, or it may be a topic that is of more interest from an outside perspective. However, the peer-networking process would address various questions without these having been articulated in terms we were expecting to hear.

The technical and sociolinguistic issues and standards that were mentioned more or less matched what we were expecting. The prominence that workshop participants assigned to the role of governments should have been anticipated and serves to remind us of this important concern.

Analysis of needs from a pan-African perspective

While localisers' perspectives are centrally important, it is also necessary to consider a range of short- to long-term needs from a strategic viewpoint. In effect, localisers are experts in their own contexts, but are not always attuned to larger and longer-term connections in the localisation ecology. There is thus also room for an analysis of needs in the form of an overview that takes in to account connections and commonalities among local efforts that are not readily apparent to localisers themselves, as well as longer-term trends that even a workshop of localisers might not identify.

Some of these include higher-level support from intergovernmental and donor organisations; better means of communication so that initiatives build on one another rather than duplicating efforts in ignorance of one another; and large-scale tools such as databases of information and contacts.

Cross-border languages represent a pertinent strategic issue. In the past, UNESCO sponsored a number of conferences on this topic. The need is currently driven by the potential for localising ICT, and various inter-African and international entities may sponsor activities in this field. Localisers and their initiatives can also initiate contact, and in some cases the need may be as simple as a catalyst or introduction in order to promote cross-border work.

Another issue is multi-dialect languages (many of which are also cross-border), which present questions regarding locales and alternative approaches to localising software for the wider community or variant forms.

It would be helpful from a planning perspective to cross-index the potential for localisation in particular languages and the evolving local and regional situations with regard to issues such as connectivity and physical access to ICT – in effect mapping the localisation ecology in geographic space (perhaps using GIS software). This would begin to allow an evaluation of key areas of need and the highest potential impact of localisation.

A strategic approach to localisation of this nature is lacking and, given the nature of local efforts, such an approach is unlikely without some outside help. This motivated the launching of the PAL project and the development of its internet presence through its website and lists. The project ultimately achieved some of those ends, notably in the form of a wiki with information on many African languages and their localisation.

Among these are issues presented by multi-dialect languages that do not have a single standard version as well as by cross-border languages. The questions include whether it would make sense to seek to develop common software localisation for very similar languages, and the appropriate ways of collaborating across borders when languages straddle more than one country.

Some critiques point out the lack of long-term strategies for product development and marketing. Localisation is not a process that can ever be considered complete, especially in the context of constantly evolving software. The approaches required for localisation anywhere include planning, building relationships, and seeking and using feedback. The question is how this kind of perspective can be provided for localisation efforts, policy-makers and donor initiatives.

Facilitating communication about localisation

Since September 2001, a number of e-mail lists and forums have been set up specifically to advance the discussion of topics related to computing and the internet in African languages. Before the establishment of dedicated e-mail lists and forums for African languages and ICT, these topics might have been discussed on other lists or not at all. These dedicated lists are worth mentioning, as a number of dynamics have arguably been set in motion by their existence and functioning. Table 13.1 summarises information from these lists.

TABLE 13.1 *E-mail forums on African languages and ICT*

	Date created (closed)	Format and participation	Working language	Number of subscribers	Number of postings and notes on topics
Bisharat e-mail forum					
Hausa charsets ^a and keyboards, http://www.quicktopic.com/8/H/JxKHyg9ccPUVB	2001-9-2	Message board; no subscription necessary	English	11	121, fonts, orthography, keyboards
Unicode-Afrique, http://fr.groups.yahoo.com/group/Unicode-Afrique/	2002-1-20	E-mail subscription list; open archive	French	178	1 345, orthographies, fonts, Unicode, encoding, keyboards, other projects
A12n collaboration	2002-3-21	E-mail subscription list; open archive	English	143	1 210, character sets, fonts, keyboards, encoding, technical issues
Ghanaian languages and ICT, http://www.quicktopic.com/16/H/9xffAXi7whnv	2002-7-9	Message board; no subscription necessary	English	6 ^b	78, fonts, orthography, keyboards

	Date created (closed)	Format and participation	Working language	Number of subscribers	Number of posts and notes on topics
Yoruba language and ICT (fonts, keyboards and applications), http://www.quicktopic.com/15/H/KKgbRqjUAR8	2002-7-27	Message board; no subscription necessary	English	20 ^b	351, fonts, orthography, keyboards
Igbo language and ICT (fonts, keyboards and applications)	2002-10-17	Message board; no subscription necessary	English	11 ^b	265, fonts, orthography, keyboards
A12n forum	2003-6-1 2007-12-14	E-mail subscription list; open archive	English	42	645, news, localisation, web content
A12n entraide	2003-6-1 2007-11-21	E-mail subscription list; open archive	French	18	173, news, localisation
Langues Togolaises et les NTIC, http://www.quicktopic.com/25/H/k2zuDzmgxGkc	2004-1-23	Message board; no subscription necessary	French	6 ^b	84, fonts, language instruction
Langues Sénégalaises et les NTIC, http://www.quicktopic.com/25/H/6KmBx6F8jES	2004-3-23	Message board; no subscription necessary	French	3 ^b	34, keyboards, orthographies, news
Langues Béninoises et les NTIC, http://www.quicktopic.com/27/H/UbEFBK7X46Ra	2004-7-19	Message board; no subscription necessary	French	2 ^b	52, orthography, fonts, text online
Langues Burkinabè et les NTIC, http://www.quicktopic.com/31/H/rhTwJR-2T8ar	2005-5-28	Message board; no subscription necessary	French	1 ^b	10, sample text
PanAfrLoc	2005-6-15 2006-9-15	E-mail subscription list; open archive	English & French	46	234, localisation

	Date created (closed)	Format and participation	Working language	Number of subscribers	Number of postings and notes on topics
Ibibio, Efik, Anaang and ICT (fonts, keyboards, applications), http://www.quicktopic.com/37/H/q8r5V-VqGF5Q	2006-7-28	Message board; no subscription necessary	English	4 ^b	27, fonts, keyboards, ITs
Swahili Advanced Computer Applications, http://www.quicktopic.com/37/H/mCpXBt-3d7qQN	2006-9-18	Message board; no subscription necessary	English, Swahili	1 ^b	17, advanced applications
A12n policy, http://lists.bisharat.net/mailman/listinfo/a12n-policy	2007-12-14	E-mail subscription list	English	11	16, language and ICT policy

Other selected e-mail forums on African language localisation

Africa@unicode.org	2003	E-mail subscription list	English	(not known)	(not known); encoding of African alphabets
Informatique et langues des deux Congo, http://groups.google.com/group/info-langues-congo	2005	E-mail subscription list; open archive	French	18	(not known); issues relating to languages of DRC and RC
Linux2Igbo	2003	E-mail subscription list	English		(not known); localisation of Linux operating system in Igbo

Notes: All URLs provided were correct at the time of writing, but where list hosts change, the lists can usually be located via search engines.

a. 'Charset' is an abbreviation for character set, which is a group of letters and symbols used for display and printing of a language. In this usage, it refers to technical support for Hausa orthographies.

b. Subscription is optional; anyone can read or post.

The usefulness and importance of this medium for communicating and fostering collaboration on various aspects of using African languages and ICT has become apparent from experience with Bisharat's¹²⁸ mailing lists. The importance of such a facility has been highlighted by the finding that there have sometimes been different groups working on localisation in the same country without knowing

about one another. E-mail lists are not the only way to foster communication, but they are inexpensive and effective, and when coupled with traditional conference-based approaches, as well as individual networking, they can do much to change the environment for consideration of a particular question.

For this reason, the PAL project launched a trilingual forum to attempt to encourage communication across the continent as well as across the post-colonial linguistic boundaries. There were actually three e-mail lists – one for each of the working languages of English, French and Portuguese – as well as a machine translation mechanism to facilitate following all the discussions in each of the languages.¹²⁹ A posting on any of the lists was automatically translated into the other two languages; moderator intervention was necessary to assure acceptable translation quality. The experience with this arrangement was generally positive, but managing the MT translations required more than one moderator.

Summary, recommendations and conclusion

LOCALISATION IN AFRICA IS an issue of emerging concern as well as a process that is, or should ideally be, meeting the increasing deployment around the continent of a rapidly evolving set of ICTs. In the preceding chapters, we have considered the needs and the environment for both localisation and localisers, linguistic and technical backgrounds and recent and current localisation activities. This chapter summarises the main themes and discusses some recommendations and suggestions.

The recommendations and suggestions are organised under the following headings:

- ◆ Strategic perspectives;
- ◆ Conferences and workshops;
- ◆ Training and public education on localisation;
- ◆ Information resources and networking;
- ◆ Languages, policy and planning;
- ◆ Basic localisation and ICT policy;
- ◆ Africa and ICT standards for localisation;
- ◆ Advanced applications and research.

Major themes

This book has dealt with a number of topics related to localisation in African languages, including Arabic. Before making any recommendations and suggestions on those topics, it is worth pointing out some of the themes that have emerged in the overall discussion, of which there are at least the following six:

- ◆ The importance of African languages and localisation: As the premise of this project and book, this is obvious, but it is worth repeating. The importance of languages can be argued on many levels. The focus on supporting the localisation of ICT presupposes a general disposition on the part of governments and populations to preserve and develop African languages. Although it is recognised, as already discussed in chapter 4, that there may not be

unanimity on such a question in any language community, it is assumed that there is inevitably some interest that merits a response in the area of localisation. (The important roles of government will be discussed later, especially in the next subsection.)

- ◆ Systems and connections: Localisation is carried out in an environment conditioned by other factors and processes, with connections and influences among them, including sociolinguistics, various kinds of policy, technology internationalisation and standards, and the evolution of ICT itself. The importance of understanding these systems – and this environment – was the reason for introducing the concept of localisation ecology and the PLETES model (chapter 3) and for discussing at length the linguistic (chapter 4), technical (chapter 5) and internationalisation (chapter 6) factors. These and their interrelationships are of practical importance for the success and sustainability of localisation.
- ◆ Specificity of information: In this book (and its associated website), there has been an effort to move from generalities, such as ‘African languages’, to specifics, such as stipulating which languages, in which places, and to begin to take into account the particular realities found in each country. In taking a broad pan-African perspective, one naturally looks at the trends and the larger issues, but it is equally important to keep very specific circumstances and needs in sight. This concern is also reflected in discussions of various factors in the localisation ecology and of past efforts. By being specific, one gains greater clarity with respect to the issues, needs and potential. Overall, this approach informs current activities as well as the planning and prioritisation of future localisation work.
- ◆ Communication: Sharing information about activities and resources, including peer networking, is essential to the success of localised efforts that may lack the means and knowledge to accomplish their objectives. The electronic forums discussed in chapter 13 are just one of several possible tools. Conferences and workshops are also important. These will all be dealt with further.
- ◆ Complementary importance of local prerogative and wider vision: Localisation depends on local initiative and knowledge, but these are seldom sufficient to achieve optimal ends. In some cases, broader perspectives can facilitate local work, and indeed contribute to planning localisation in ways that benefit the largest number of people and have the best chance of sustainable results. In other cases, local initiatives may be informed by a broader view of realities of which they may not have been aware. Both the local and the broader perspectives are necessary, and indeed depend on each other, which is in some ways similar to the interrelationship between localisation

and internationalisation – the first informs the latter, and the latter facilitates the former.

- ♦ Vital role of African governments and intergovernmental organisations: Official institutions in Africa, whether at the country, regional or continental level, occupy central positions in areas that affect localisation, such as language, education and technology policies. They also have the potential to support localisation directly in practical ways and by elaborating their vision of how multilingual computing could develop in their states and across the continent. Without appropriate governmental and intergovernmental attention, localisation in Africa is hampered, since certain policy roles can only be executed by governments or with government involvement.

Strategic perspectives

A useful general approach might be to develop a phased, long-term strategy emphasising various aspects of ICT and considering which should be localised, in which languages, how, in what order, when, with whom, and with what support.

To some extent, the impetus towards localisation is driven by local groups or projects, but it can equally be encouraged or catalysed from outside. External parties can also serve to cultivate interest, offer guidance or even make connections. For instance, isolated groups or focused projects working on related languages, or on the same language in different countries, might be encouraged to combine their efforts and to share their ideas, but may lack the vision or the connections to do so. The goals of the PAL project are pertinent to all these issues.

A creative strategy for the localisation of ICT in African languages will also have to take into account unique aspects of the nature and distribution of the languages, their current use and the ways in which they are evolving in the societies in which they are spoken. This is particularly true of sub-Saharan Africa, as the predominately arabophone North Africa has been able to benefit from work in many countries to develop computing and internet access in Arabic.

The interspersed of speakers of diverse languages, the use of different languages in varying social and economic contexts, and the generally multilingual nature of most African societies present a linguistic and sociocultural profile that differs from the more technically advanced countries. The situation is in some ways more comparable with multilingual South Asia, except that the older written traditions and larger bodies of speakers of most of the languages of that region have facilitated easier ICT connections. Africa will therefore have to create its own terms of reference with regard to language and technology, as well as all aspects of localisation ecology, and will ultimately have to develop its own approaches through

technologies that respond to African sociolinguistic realities and socioeconomic aspirations.

It is not at all clear that the authorities charged with developing language and technical policies are prepared for this role. There is in some cases genuine interest; ACALAN, for instance, has promoted discussion of linguistic diversity on the internet. Some basic work has been done, for example, the discussion of keyboard layouts for Ethiopic/Ge'ez in Ethiopia, and applied linguistic research has been conducted by the Centre for Advanced Study of African Languages (CASAS) and the National Association for Cameroonian Language Committees (NACALCO), among others. However, there does not appear to have been the capacity, and possibly not even the will, to articulate the potential of multilingual ICT in Africa or to propose practical steps to accomplish it. The record of follow-through with respect to language-policy proposals over the last few years is not encouraging.

Most major donors are similarly not attuned to these realities or needs. Development agencies, by and large, have not paid much attention to African languages, with the exception of support for adult literacy and, more recently, some bilingual education reforms. The situation is similar with respect to localisation.

Two immediate questions arise from this consideration: how to promote interest in strategies and prospects for localisation, and how to nurture the capacity to develop and pursue them. A third question that follows from the first two is where the vision and expertise to respond to these issues will come from.

It is possible that an African organisation will take the lead in this, perhaps in collaboration with a research organisation such as IDRC, or an international body such as UNESCO. Cooperation between external agencies with the technical vision, on the one hand, and African agencies with the policy mandate, on the other, might be an ideal combination, but the question is who will take the initiative to formulate a comprehensive vision, as opposed to general declarations.

One would hope to see African institutions of higher learning and research in the vanguard, envisioning the future, proposing policies, devising strategies and building skills, but such initiatives are not yet apparent. African academics in universities of the North, along with other Africanist scholars, may possibly take the lead. These academics and the institutions at which they are employed, in partnership with African institutions, might perhaps influence the evolution of information technology for all languages in Africa.

Another possibility is that local efforts to translate software might build momentum and encourage a broader response, in addition to achieving their particular objectives, thereby introducing dynamic new elements into the discourse on ICT and localised software applications in Africa. This would require excellent communication and organisation.

Another aspect that should not be overlooked is the proprietary and commercial sector. Microsoft, in particular, has invested in internationalisation and some localisation of its software in Africa in collaboration with government agencies in certain countries. There may be lessons to be learned from this approach as an alternative strategic model. Another category of business organisations in the field comprises cellphone companies that support African-language text messaging. A smaller but nevertheless significant group is made up of African companies that are involved in some way in localisation; this is a potential growth area. Ideally, one would hope that government agencies, donors, non-governmental organisations and businesses involved in ICT in Africa would be in general agreement about the importance of localisation and would wish to work together to facilitate this as harmoniously as possible.

Conferences and workshops

Localisation in Africa involves several factors and various agencies, organisations and individuals. Given the rapid advances in ICT, there is an ever-present need to coordinate efforts, plan developments and train people.

Apart from involving all stakeholders, the inevitable workshops and conferences that will be convened for such purposes should build progressively on previous work and ensure that important concerns are not lost from view. Conferences and workshops are too often organised without much reference to previous ones and generally fail to build on past efforts.

In that context, one might propose strategic planning for localisation meetings. The strategic plan could take the form of a reference framework for diverse organisations, and could perhaps be designed and implemented at a high level, involving the participation of key organisations and agencies. This would require broad agreement on the goals of localisation, although the meetings would facilitate a cumulative process of working towards the broad goals, the outcome of which might not yet be clear.

The following four broad concerns related to localisation (listed from the local to the more general level) could benefit from attention during the coordinated series of meetings:

- ◆ Localisation workshops, or technical workshops for software and content localisers: Regional, country-specific and even language-specific meetings would be useful in addition to workshops such as the PanAfrican Localization workshops in Casablanca, Morocco and Tshwane, South Africa, the Africa Source events and the workshop organised in Marrakech, Morocco by the non-governmental organisation, Tactical Tech. The scope of

the workshops would depend on particular needs and interests in various parts of the continent. The purpose would be to train specialists in aspects or phases of the localisation process, from the basics such as Unicode and content creation, to the creation of locales, the translation of diverse software applications, follow-up issues in localisation and advanced applications (see the relevant subsection in this chapter).

- ◆ Application and adaptation of localisation for development and business environments, including content development: One of the main purposes of localisation is to help bridge the digital divide, but exactly what that would entail in particular instances might have to be explored in meetings of localisers and development agencies. The Open Knowledge Network was an interesting example in this regard; it attempted to use diverse languages and allow the various partner agencies to share information. Broader inter-agency consideration of localisation in ICT4D contexts will be both helpful and necessary as more localised software becomes available and the potential for localising specialised applications becomes more of a reality.
- ◆ Language issues and ICT: This category includes expert meetings for language planning in order to consider how to treat multidialectal languages, cross-border languages and language clusters in localisation and more generally in ICT. Such meetings might be loosely modelled on expert meetings about African languages and orthographies that were facilitated by UNESCO in the 1960s and 1970s, or a smaller one in Okahandja, Namibia in 1996,³⁰ with ICT as an added context. There is evidently a need for more deliberation on aspects of language use in computing and on the internet. The issues to be addressed include basic matters such as orthographies, the relationships of dialects and closely related languages, and the possibility of selecting a standardised version of a language for various kinds of work. Other issues include the representation of African languages in international standards such as ISO 639 and in locales (as registered in CLDR). Terminology development and harmonisation might also be considered (as discussed in subsequent subsections).
- ◆ Strategic planning for localisation in Africa: This category is closely related to the discussion of broad visioning and planning over the long term, at continental, regional and country levels, as already discussed earlier in this chapter. The sessions on ICT at the World Summit on the Information Society African planning meetings in Bamako, Mali in 2002 and Accra, Ghana in 2005 were along these lines. This category could include planning meetings for conferences and other gatherings in the categories mentioned in the preceding three bullets. It might also include less formal meetings among high-level decision-makers.

Such meetings could sometimes perhaps be organised as part of larger events or co-located with other meetings in order to optimise participation and minimise travel costs for delegates. The topics covered and the purposes of the meetings would vary according to local, country, regional and continental needs.

Geographic scope and strategy for meetings

In addition to Africa-wide meetings, a pan-African strategy for conferences and workshops on localisation would have to pay particular attention to facilitating and sponsoring workshops at national and regional levels. In fact, it might be best at the present time to prioritise regional meetings, with secondary attention to periodic continent-wide meetings. The regional meetings should include some dedicated to working languages other than English – notably French, Arabic³¹ and Portuguese.

To give localisers and others interested in African-language computing and internet accessibility the chance to address issues particular to their circumstances, and the languages in which they localise, would help support local efforts and enrich periodic continent-wide meetings. E-mail lists and websites, such as those set up by the PAL project, could be used to maintain pan-African communication.

A conference on aspects of localisation in Nigeria, for instance, has been suggested, and this might be the place to initiate such efforts. Several organisations are working on various aspects of localisation, including software translation, keyboard development, text-to-speech software and speech recognition. These initiatives include efforts focusing on the three main languages in Nigeria, namely Hausa, Yoruba and Igbo, with some notable efforts with respect to other languages, including Ibibio.

An initial regional meeting in one part of Africa could provide experience on which to model similar meetings elsewhere on the continent.

Beyond Africa, IDRC has discussed the possibility of a global conference on localisation, including Asia (involving the PAN Localization project, a regional initiative to develop local language computing in Asia), Africa and perhaps indigenous communities of the Americas and Oceania. A meeting on such a scale would be productive in terms of permitting wider exchanges and networking so as to expand local, regional and continental networks.

Training and public education on localisation

Education for localisation – that is, training of localisers and public education on localisation – is a set of concerns beyond workshops and the declarations often issued by such meetings.

Training

In any field, it is common to hear proposals for more training to build skills and enable people and teams to achieve certain ends. Localisation is no exception. Training topics might vary between training events, but as with planning workshops, attention should be given to how these fit into the larger strategy of developing skills and producing localisation.

In terms of intended beneficiaries, training for localisation should focus on people already involved in localisation initiatives, motivated people that would like to be involved, and people in neither group whose work logically connects with localisation (such as people planning ICT4D/E projects or setting up telecentres).

Investing in training for the future

Most discussions and meetings related to localisation focus on immediate projects and measurable results in the short to medium term. This is true of training for localisation (or related areas such as ICT4D) as much as it is for other goals.

When considering issues such as the sustainability of localisation and the potential use of advanced ICTs with African languages (a later section of this chapter), a longer-term need for highly professional skills becomes apparent. It is worth calling for investment in educating a generation of African experts in localisation and in language and computer science, although this goes far beyond the aims of the PAL project or its successor, ANLOC. These interdisciplinary areas are receiving increased attention elsewhere in the world, but less so in Africa. Furthermore, few Africans are involved in discussing the internationalisation of ICT, for instance, and even fewer are involved in research on machine translation or speech recognition. Unless this is addressed now by institutions and donors with the means to do so – by providing scholarships and investing in research programmes – Africa will remain handicapped over the long term with respect to multilingual ICT, an area from which it could arguably benefit and to which it could make a significant contribution.

Public education

The ‘public’ in public education on localisation includes several groups: computer users in Africa in general, especially those that work in a technical capacity but are not formally involved in localisation, as well as others that occupy key decision-making positions within the localisation ecology (such as policy-makers, educators and business people). A public education approach could help raise general

awareness about localisation, increase knowledge about specific aspects of ICT and even motivate action.

Public education can be accomplished through conventional public relations and development communications approaches. The internet (including websites and mailing lists) can of course be used to advantage (see also the next section), but it may also be useful to develop a public education campaign to increase awareness and to attract the attention of the media in Africa.

One example would be to organise thematic Years of Localisation in Africa campaigns to focus attention on particular aspects of localisation and multilingual ICT in Africa. Possible topics might include Unicode, locales, digitising texts for dissemination via the web, and advanced applications. Different approaches might be adopted for different topics. Unicode might be addressed by raising awareness, while a topic such as locales might be associated with goals for measurable output or even conferences. The process of declaring a year dedicated to a particular theme is relatively easy, but would require active support from major continental organisations in order to be successful. Experience with the Year of African Languages in 2006 would offer some lessons in this regard.

Information resources and networking

The focus of this section is on providing structures for information to promote and assist localisation, the means for discussing and learning about localisation throughout the relevant environment (namely, localisation ecology), and ways to facilitate and enhance peer networking among localisers themselves. In a sense, these communication strategies might be called meta-strategies, in that they are intended to help facilitate other strategies discussed in this chapter through development, access, modifications and implementation.

Various resources – about FOSS and even about African languages – are available to localisers online, but with few exceptions, adequate resources on localisation in African languages have been lacking, whether of a general nature or focused on specific languages. The exceptions include some resources for Arabic and Ethiopic/Ge'ez, as well as a small number of other languages, including the limited resources available from Translate.org.za. This indicates a need for web resources for localisers and localisation in Africa.

As regards e-mail lists, there are several dedicated to various aspects of localisation and multilingual ICT in Africa (as discussed in chapter 13), but these tend to be rather specialised and are divided into lists that use English or French as the working language.

Apart from such resources, which facilitate finding and exchanging information, there are people and their networks. The cultivation and enhancement of localisation networks is a wider and perhaps more dynamic need. Networks are also served in many other ways, including meetings, as already discussed. The internet provides a ready means of maintaining networks and enhancing communication, and therefore deserves special attention.

Contribution of the PAL project

In response to this situation, one of the objects of the PAL project was to develop an online resource for African localisers. After considering various approaches, an approach involving a website based on a wiki as well as associated e-mail lists (as discussed in chapter 13) was adopted. The wiki is being merged on the ANLOC website, and will serve as a repository for updated information for this volume. As a resource, it is intended to be the first port of call for information on localisation in a particular language or country, as well as to provide links to relevant websites and tools. The wiki is intended not only to serve localisers but also others that are able to support localisation or promote multilingual ICT, for instance, as part of projects on the digital divide. The wiki includes the following sections:

- ◆ Major languages;
- ◆ Writing systems (or scripts);
- ◆ Countries;
- ◆ Organisations;
- ◆ Localisation resources.

These sections are intended to provide information in the context of different perspectives on localisation. Languages (which are the core issue of localisation) cross borders and countries (where policies are made) and include diverse languages. Scripts (the form of text used in ICT) are used for many languages and modified in various ways. Various organisations deal with languages and localisation across the continent. There is a range of general localisation resources that are useful for African localisation.

Languages, policy and planning

This section deals with language policy and planning as they relate to localisation. Any such discussions have to take into consideration both countries (since this is the level at which policies are made and planning is done) and the languages themselves.

Localisation and African linguistic diversity

A strategy to support localisation in Africa must begin with a sense of the scope of the project and an awareness of the possible extent of such support, beyond focusing on existing initiatives, so as to engage in proactive outreach to potential localisers where localisation initiatives have not yet begun. With these considerations in mind, the PAL project sought to organise information and contacts by both language and country. This approach was informed by the logic that what has been identified and listed is not easily overlooked, and that these tangible points of departure could readily be built upon.

Given the limited means available for localisation, it will be necessary to prioritise efforts and resources, whether at the language or country level or according to broader considerations. This approach might disfavour less widely spoken languages, including endangered languages, at least in the short term.

Perhaps the most pragmatic strategy for languages not prioritised for software development, at least initially, would be to incorporate spellcheckers for such languages into software localised for more widely spoken languages. Even an interim solution of that nature would be a significant challenge for languages with relatively few speakers and resources, but arguably a realistic goal. Since Africa's linguistic diversity exists in multilingual contexts, such combinations of more and less widely spoken languages might actually be a logical approach to localisation.

Standardisation of languages and orthographies

A related issue, both within individual African countries and among neighbouring countries that share cross-border languages, is that of standardisation. In cases where there are dialectal variations, it might be helpful for localisation (and eventually for computer users) to identify and develop standard versions, or at least orthographies and terminologies, where this has not yet been done. This function is generally undertaken by governments (for example, for Runyakitara in Uganda), but in at least one case (N'Ko for Manding languages), the efforts to develop a standard are local.

Standardised orthographies have not yet been developed for all African languages, even for certain languages with more-established writing systems (as discussed in chapter 4). This has been a recurring issue in Africa since independence, along with the harmonisation of transcriptions for related languages, languages that cross borders, and diverse languages within each country. These issues must receive continued attention and be resolved so that the languages can be consistently used in all text-based computer applications and content.

Educational policy

Recommendations concerning educational policy with respect to languages of instruction and the teaching of African languages in schools are beyond the scope of this book, but it is clear that there is mutual benefit of cooperation between localisation initiatives and the use of African languages in education (education being a key part of localisation ecology). Installing localised software on computers in schools would offer learners different ways of studying and interfacing with the technology. Pluriliterate students can more effectively interact with multilingual African web content and software and contribute to building a rich array of such material.

Externally funded ICT4E projects, such as the OLPC project, could have a positive impact on localisation and programming for language in education.

Beyond education in schools, the potential to use ICT for basic literacy and L1 literacy for those that have been educated only in ELWCs can be explored.

It is also worth exploring possible connections with the development of online dictionaries and second-language instructional modules. Collaboration between African-language programmes at Northern universities and applied linguistics and educational programmes in Africa could possibly result in the creation of online language resources for Africans.

Basic localisation and ICT policies and programmes

Another substantial source of support for all aspects of localisation could be through other ICT activities, notably country-level ICT policies, FOSS user groups and even development projects.

Country ICT policies

Country-level policies for the development of ICT, such as NICIS, could make more specific mention of and commitment to localisation. This could include moral or material support for initiatives such as the development of web content in indigenous languages; the translation of software; participation in setting standards affecting the use of the languages; and training related to language and computing.

Governments have enormous potential influence through the examples they set, for instance, the languages in which their content is posted. South Africa is an interesting example in this regard, with some official sites in several languages, although some other governments also have indigenous language content. Governments could insist that foreign-funded ICT4D projects take into account the

multilingual nature of populations in their choice of software, fonts and keyboards for the computers they deploy.

Role of international development organisations

A number of donor agencies and other organisations that are involved in international development have taken an interest in various ICT4D and ICT4E projects. Although some of these organisations, including IDRC, OLPC and Geekcorps, have attended to aspects of localisation in web content and computer interfaces, they are well placed to do far more, for example:

- ◆ Equipping computer systems in telecentres (with basic fonts and keyboard options, and localised software if available);
- ◆ Providing diverse language content;
- ◆ Designing localised web content to be printable for community reading in locations distant from telecentres;
- ◆ Seeking collaboration concerning localised software with localisation groups where they exist in beneficiary countries;
- ◆ Supporting new localisation efforts.

Bringing FOSS communities more fully into localisation

The often-mentioned distance between linguists and computer technicians risks being repeated between some FOSS user groups and localisers in Africa. This is a good time to build bridges, as both FOSS and localisation are receiving more attention. To that end, it is worth proposing that FOSS groups in various countries be identified and contacted about their interest in developing localisation agendas.

Linking various localisation currents: FOSS, ICT4D and commercial interests

There seem to be two or three different levels of operation related to localisation that should be linked. The first is localisation focused mainly on the language communities within specific countries or groups of countries. This is the level at which the PAL project and other localisation initiatives have tended to operate. In general, these are local non-governmental initiatives focusing on content and FOSS applications.

Another is localisation in specific development contexts, as the Open Knowledge Network project attempted in some parts of Africa and South Asia. This and the previous level overlap in Africa itself, but localisation for specific language communities focuses on general FOSS applications and may also involve Africans abroad, while localisation in development contexts involves development agencies designing and implementing ICT4D projects. Both levels should in principle involve

linguists and computer technicians, which could be linked so as to promote the cross-fertilisation of ideas and fruitful collaboration.

A third level is that of business and commercial interests. International proprietary software firms (notably Microsoft) and African software companies show an interest in localisation in certain markets, while firms in other industries may consider the localisation of content or interfaces.

Linking commercial interests with localisation initiatives for specific language communities or for development might present challenges in reconciling the approaches, but could also offer unforeseen benefits. Seeking coordination or cooperation among diverse groups with common interests might be very positive for African languages if localisation is to have broader impact and longer-term sustainability.

FOSS, proprietary software and limited localisation resources

Africa tends to have limited resources for ICT and language work, as in other domains. Another means of linking the various localisation streams might be to promote collaboration among proprietary and open-source efforts with respect to basic resources for localisation in minority languages, which would constitute a historic compromise of sorts for these languages and their speakers. This might take the form of favouring the development of resources such as dictionaries in ways that are most likely to benefit the largest number of people, as early as possible. Such efforts would target the least widely spoken and most poorly resourced languages.

Mobile technology

The rapid spread of cellphones in Africa and the miniaturisation of computer technology opens up a new dimension for localisation, which therefore deserves to be considered in localisation policies and programmes. This technology area is discussed further in a later section of this chapter.

Africa and ICT standards for localisation

As discussed in previous chapters, the requirements for success in localisation include standards, which facilitate the use of diverse languages in ICT, the translation of software and the production and use of linguistically diverse web content. Recommendations concerning language-related standards have already been mentioned. This section focuses on technology-related standards affecting

localisation and multilingual computing in Africa, ranging from keyboards to coding. Greater involvement by governments and inter-African agencies is recommended at all levels in establishing standards that affect localisation in African languages, including iso and its relevant technical committees.

International standards for language and African participation

The role of iso and African participation – or lack thereof – in establishing standards for languages was discussed in chapter 9. There is a need to find ways to facilitate and encourage African governments and standards bodies, as well as language and applied linguistics agencies, to take a more active interest in these issues. This key aspect of localisation ecology is being shaped almost entirely outside Africa.

Language coding in terms of iso 639 is a particularly important issue (see chapter 9), since new codes are sometimes needed. Another important issue relates to the new elements of the standards that are being developed (iso 639-4 and iso 639-6). Without input from Africa or from experts familiar with the realities of African languages and localisation, these needs may fail to be identified. The means must be found to analyse the system of codes as it applies to African languages and language groups in order to ensure appropriate coverage for locales and diverse localisation needs.

Unicode is another area of concern related to international standards. Although the character ranges already encoded for major scripts used in Africa are quite extensive, additional characters may be needed, while certain minority scripts have not yet been encoded. However, there is hardly any African representation in the standards process for this area, and the deliberations and decisions are left to other countries.

Locales

Locales are categorised by internationally standardised codes (in terms of language, country and sometimes script) and are fundamental to localisation, yet there are still relatively few languages with locales. This area not only needs more attention but also coordination in the case of some language groups or macrolanguages where there are various code options. A campaign to increase the number of African-language locales filed should be planned and coordinated, taking these concerns into account.

The ANLOC subproject to create 100 African locales, as discussed in chapter 9, represents an important step in this direction, but it is just a small beginning.

Keyboard standards for Africa

As discussed in chapter 8, the issue of keyboard layouts for African languages whose orthographies include extended character needs is receiving increased local attention. The initiatives include layouts included in software localisation, keyboard layouts designed with programs such as Keyman and *MSKLC*, and a production keyboard for Nigerian languages. The question is how to promote a degree of standardisation in keyboard layouts so as to benefit localisers and computer users. Although many keyboard layouts are independent of other software, they are nevertheless a consideration in software localisation and keyboard manufacture.

Current localisation efforts related to keyboards should begin to consider the larger and longer-term keyboard evolution issues, at least to the extent of familiarising themselves with other keyboard layouts already in existence in the same region or country, or for the same or similar languages. The general evolution of African keyboards might ideally accommodate multiple language usage in a particular country, group of countries or region.

Other areas of interest include keyboard layouts associated with localising production software such as OpenOffice and Microsoft Office, as well as facilitating input in African languages into non-localised software. The question of coordinating these efforts then arises. Such discussions might provide the basis for standards development.

At a higher policy level, the issue of who would have the authority to consider this kind of question requires attention, as does the issue of how this authority would operate and whether it would involve only government agencies, or language specialists, or commercial interests, or a combination of these. It is again important to consider what is already being done and used in different circumstances.

In developing keyboards for wider and longer-term use, it is essential to bear in mind some basic given factors as well as the iso 9995 guidelines. It is firstly necessary to understand the habits of computer users with regard to keyboard use. It is reasonable to assume that multilingual and pluriliterate people in Africa will use more than one language in their computer use, possibly even during a single session. It is also important to remember that many computers will be used in public access places such as cybercafés and telecentres, and that provision would therefore have to be made for multiple language preference at each computer station. While these considerations may not seem particularly pressing at present, when software and websites are generally available only in *ELWCs*, such needs should be anticipated. The potential for diverse use of the technology will be progressively realised as localised software and interactive content become more widely available in African languages.

In this context, we face a dual imperative, to provide on the one hand the most effective and useful keyboard layout for the end-user in particular situations, and on the other, appropriate ‘massification’ for the complex multilingual market. A single keyboard should therefore satisfy the needs of as wide a usership as possible, without creating a layout that is cumbersome to use in any particular language. Keyboard strategies for African languages will have to balance such criteria in one of two ways: either a single keyboard that satisfies the needs of as wide a usership as possible, without presenting a layout that is cumbersome to use in any particular language; or a combination of specialised language-specific keyboard layouts alongside a general keyboard that facilitates input of characters used in many languages. In either case, each keyboard layout does not exist in isolation, but as part of a system requiring attention to the harmonisation of layouts for the benefit of people who may use more than one keyboard for input in different languages.

A comprehensive and forward-looking strategy should also consider alternative means of input, ranging from the emerging LED keyboards, to graphics tablets and speech-to-text facilities (the latter is considered in the next section).

Advanced applications, tools and research

Localisation, as applied to software, most often refers to computer software for general use. It is important to go beyond that definition and consider the localisation of specialised and advanced applications, the development of software tools that allow one to do more with language, as well as research to develop these aspects and advance the use of technology for African languages. In other words, localisation efforts might focus on specialised and advanced uses of ICT, which might sometimes facilitate other localisation efforts.

This section discusses mobile and advanced technologies and suggests how to promote research in these areas.

Mobile technology

Mobile technology, facilitated by ongoing advances in miniaturisation, was discussed in chapter 12 and mentioned in a previous section of this chapter as an important new area of localisation that is rapidly gaining importance in Africa, at least as far as cellphones are concerned. The potential for SMS, e-mail and other text in African languages from mobile devices, as well as the possible interfaces with voice (such as the Swahili TTS, as already mentioned), may require attention to localisation standards.

Complex scripts raise the issue not only of compatibility with computers but also of standards. What sorts of links can there be, and should there be, between companies involved in localising cellphones and handhelds and the localisation initiatives related to computer software? In the case of Microsoft, there may be some synergisms between their various efforts related to computer and mobile software, but where does FOSS localisation fit in?

With increasing miniaturisation and innovation related to handhelds, a key question is whether the evolution of mobile devices for African contexts should be discussed, given that the mobile sector is dominated by commercial interests. The commercial nature of the industry is not an issue, but the apparent lack of connection between the mobile industry and other localisation efforts is worth examining. Moreover, the potential to develop localised software for handheld devices should be researched.

Geographic information systems

GIS technology permits digital manipulation of data and map images, as discussed in chapter 12. It merits further mention as a possible priority technology that can be localised for local development planning, analysis and education. GIS is being used for participatory development projects in Africa and other parts of the world, and the inevitable increase in the use of GIS in Africa suggests the potential for localisation.

Given that there is at least one good FOSS GIS program available free of charge (namely, GRASS), it is worth recommending that computers in all community and school telecentres and government offices in Africa be equipped with GIS software and that a programme to localise such software into major African languages be launched. The biggest issue with a supply-side approach such as this would be to address the need for training in order to take advantage of the full potential of the application.

This is not to exclude proprietary GIS products, but cost is a factor in Africa. In any event, any project to encourage the widespread use of GIS in any form will certainly increase the demand for GIS in all forms.

A first and immediate step in localising GIS should be to explore, together with African localisers working on other projects, the possibility of localising the GRASS software. The localisation in Amharic that has already begun could form the starting point. Any such effort should involve language specialists, GRASS technicians and possible funders.

GIS might also be a useful tool in language and localisation planning given the geographic distribution of the speakers of various languages.

Cutting-edge language technologies

African-language contexts are characterised by multiple languages, traditions that are often described as more oral than written (such that written forms have often not yet been standardised), and low levels of literacy. In these circumstances, conventional uses of ICT that focus almost exclusively on text will fail to take advantage of the many talents and ideas that naturally come forth in many languages that do not have a well-established written tradition. Moreover, minority languages will be at a continued disadvantage, because the spoken language will rarely be written, and the few recordings that exist may never be transcribed in ways that are accessible to native speakers.

In principle, new technologies offer potential new uses for any language, and the most disadvantaged languages stand to gain the most. Language and ICT-technology policies must take this aspect into consideration. This in turn requires clarity of vision and communication among policy-makers, researchers, activists and native-speaking communities.

Some of the most advanced technologies dealing with transformations of speech and text may paradoxically be the most appropriate for African languages. These include the following, as discussed in chapter 12:

- ◆ Speech synthesis and text-to-speech (TTS);
- ◆ Speech recognition and speech-to-text (STT);
- ◆ Advanced uses of audio, including digital audio;
- ◆ Machine translation (MT) and translation memory (TM).

These areas could benefit not only from basic research but also from innovative thinking on long-term strategies for using such technologies and adapting them to African realities.

Planning and research

One idea would be to promote basic and applied research on the areas mentioned in a long-term applied research programme by a consortium of institutions in and outside Africa. The philosophy would not be to 'catch up', but rather to 'go ahead' in the sense that ICT might find new and innovative uses in African contexts.

The idea would be to target, over the next decade or so, a range of cutting-edge technologies with specific aims. Some suggestions follow, and it should be noted that for some of them, a standardised orthography would be a prerequisite:

- ◆ Enhance and diversify the currently limited number of TTS applications. This might involve the Local Language Speech Technology Initiative (LLSTI), among others.

- ◆ Develop STT programs for a select number of languages and investigate ways of streamlining production of such software for a variety of languages, making it reliable across dialectal differences for languages without standard dialects.
- ◆ Research the potential for African-language content and applications of various types that do not rely primarily on text, namely audio only and audio with images. The object would not only be to enhance soft access by those with poor literacy skills and the disabled but also to explore new ways of interacting with and using the technology for all users in Africa and beyond. The possibility of involving Africa's oral history centres in such research should be considered.
- ◆ Develop a strategy for compiling corpora for more African languages. This could exploit extant sources, such as they are, and research new ways of developing texts in African languages. Parallel corpora with translation into an ELWC could be especially useful for translation software.
- ◆ Explore the uses of translation software (including both MT and TM) for African languages, including those less widely spoken. Examples of issues that might be investigated include: the possibility of developing TM banks online for use in more than one location; whether the evolving range of approaches to machine translation could be adapted to different language situations on the continent; and the ways in which this set of technologies could address basic hurdles to translating material into African languages, for instance, for government services, education and development.
- ◆ Involve and train African linguists and computer scientists, with the aim of empowering a new generation of African research in natural-language processing and localisation.
- ◆ Link efforts at multilingual ICT innovation to basic localisation and ICT4D/E, with the ultimate aim of facilitating a range of work to serve African needs. This might involve developing new partnerships for both research and practice, for example, localised GIS for rural development, or localised Foss distributed to new telecentres, which can then be involved in feedback on performance and improvements.

Facilitating other kinds of localisation

Software to facilitate the translation of software has been mentioned in chapters 6 and 12. Some of the above-mentioned applications could similarly be used to facilitate localisation and content-creation in African languages. They may also find uses in localising software.

Where the production of text in diverse languages is involved, the potential use of translation software and STT to reduce the cost and time required to develop material for web content or even print (such as school materials) should be investigated.

Finally, the use of well-established scanning technologies, with optical character recognition (OCR) adapted as necessary for extended Latin and non-Latin scripts, could allow the digitisation of much published material on and in African languages for dissemination (with permission) to native-speaking communities via the web.

Conclusion

This volume has presented a survey of localisation in Africa and its further potential. A model of localisation ecology was introduced as a key conceptual element. Background information on African languages and on aspects of ICT has been presented to establish the context for the current situation and the resulting recommendations.

While the general message of this volume is that the localisation of ICT in African languages is important, and that there is considerable potential to use ICT for various purposes in African languages, it is also recognised that the path is not straightforward. Basic technological and educational situations are not favourable, resources for localisation are limited, and policies are not actively supportive. While it is possible to make recommendations, we are still left with some major questions, such as:

- ◆ How will localisation in Africa develop in its next phases?
- ◆ Which aspects of localisation should receive priority?
- ◆ Who will lead the localisation efforts at the language, country, regional, working-language (ELWC) and continental levels, and how can these efforts be coordinated?
- ◆ How can initiatives for localisation best be encouraged, coordinated and supported?
- ◆ How can policy and institutional support be gained and maintained?
- ◆ How can African expatriate and foreign volunteer support for localisation best be used to develop skills on the continent?
- ◆ Where will the resources for localisation come from, particularly for less widely spoken and resource-poor languages?
- ◆ How will the information in this book and its companion website be used to assist these efforts, and how might it be further developed?

The information in this volume is ultimately intended to serve as a resource to assist in considering strategic questions such as these, as well as to assist individual localisation initiatives. Specific information is intended to benefit locally directed efforts, while the aggregation of specifics shows patterns and connections in the larger whole. This contribution is thus part of a process, and as with any process, there are cycles of evaluation and vision, of review and revision.

Africa is the second-largest continent, with some of the greatest linguistic diversity. However, it is not yet well placed to take full advantage of new ICTs, let alone to shape them, in order to respond to the realities and aspirations of its rapidly growing population. The expanding multilingual potential of ICT is also encountering a language-policy and sociolinguistic environment that is not well positioned to take advantage of these advances.

However successfully localisation is achieved, is not an end in itself. At the beginning of the volume, the digital divide was discussed, as well as the ways in which localisation, by increasing access to ICT and making it more relevant, can significantly contribute to bridging that divide. The vision for localisation and the consideration of how to sustain it must therefore address the ways in which research on localisation and the products of that process can link with ICT4D and ICT4E projects.

Localisation is not merely dependent on other forces, in terms of localisation ecology, for its success and contributions. It represents a new dynamic in social, economic, technological, educational, linguistic and political development in Africa no less than in other parts of the world. The extent to which that dynamic can effectively benefit larger processes is dependent on attention, planning, action and indeed unity.

The latter point deserves special mention in closing. The theme of unity has concerned individual African states, Africa-wide gatherings (some under the banner of pan-Africanism) and continental bodies such as the African Union. Promoting the use of individual African languages at national levels has often encountered debates about the effect of such policy on national unity. Paradoxically, at the continental level, the discussion of the promotion of African languages has frequently resulted in favourable declarations and even action plans, but these have ultimately not resulted in much action by individual countries, regional organisations, or for that matter, major donors.

Taking a pan-African approach to the issue of localisation – for practical reasons related to the need for collaboration on rapidly advancing multilingual ICT – also resonates in principle with the object of Africa-wide conferences and statements on the promotion of African languages over the last four decades. However, localisation initiatives may risk confronting similar ideological roadblocks, at the level of individual governments and even ICT authorities, as those faced by

previous pan-African initiatives related to Africa's indigenous languages. This may be especially true when the issue of prioritising some of the more widely spoken languages is raised.

It is therefore important to emphasise that the diversity represented by localised content and software is not a disunifying force, but rather a common enterprise; starting with some languages does not mean that others will be excluded, but rather that resources and capacities will be developed to handle them all. The very nature of ICTs is additive, in that advances for one language enhance rather than hinder the opportunities for work in other languages.

In the same way that Unicode as a single character-coding system facilitates greater use of diverse scripts, so a pan-African approach to localising ICT in many parts of Africa may paradoxically yield the best results for each language individually and for all of them collectively. Small, disjunct projects and initiatives that are unaware of the importance of the localisation ecology in Africa and uninvolved in broader world discussions of local language computing are unlikely to achieve sustainable results. If they are linked together, however, sustainability as well as other goals become possible.

In the long run, the hope and potential of ICT, through localisation and adaptation to the languages and modes of communication of people on the African continent, is to advance development in the broadest and most encompassing sense of revealing potentialities. It is hoped that this volume will in some small but significant way further that aim in and for Africa.

- * Idea developed by Mohamed Mrayati, seminar on the role of translation in socioeconomic development, Union des traducteurs arabes, Tripoli, Lebanon, July 2009.
- † Tunis Agenda for the Information Society, Document WSIS-05/TUNIS/DOC/6 (rev. 1), paragraph 53: www.itu.int/wsis/documents/doc_multi.asp?lang=en&id=226710
- 1 ICT in its broadest sense also includes broadcast technologies (radio, which in Africa is often local community radio, and television) and telephony. Since these technologies are based on transmission of the spoken word, they can be used with any language or dialect. However, the newer ICTs rely largely on digital text (computers and the internet, as well as human language technologies), which are the focus of this book.
- 2 This observation is frequently made. Herbert (1992: 1) is among the more recent sources.
- 3 The term 'European language of wider communication' (ELWC) was introduced by Bokamba (1995). The term 'europhone', which has been coined more recently, is sometimes used to refer to European languages and speakers of these languages in Africa. 'Language of wider communication' (LWC) is an established term that refers to any vehicular language, generally in contexts where it is a second or additional language. Many African languages, including Arabic, serve the role of a LWC or local lingua franca in certain contexts. ELWCs are dominant in web content and software worldwide.
- 4 According to one estimate, up to ninety per cent of people in some countries do not speak the official languages (Mackey 1989: 5, quoted in Robinson 1996: 5).
- 5 One of the strategies for education and literacy training in Africa has been to use instruction in LIS primarily as a 'bridge' to learning in the official language; this is sometimes referred to as a 'subtractive bilingual' approach. The localisation of ICT, as discussed in this volume, is not conceived in terms of such limited ends, although it is certainly true that people who learn to use computers in a language that is very familiar to them would more readily be able to acquire computer skills in an additional language.
- 6 The International Organization for Standardization (ISO) and the Unicode Consortium, which began as an industry association, coordinated their efforts in the mid-1990s and established a single coding system. This is sometimes called the Universal Character Set (UCS), but is commonly referred to simply as Unicode, a practice that is followed in this volume.
- 7 This is a subject that cannot be treated in depth here but merits brief discussion. The devaluation of all aspects of indigenous cultures in Africa was a fundamental feature of European and North American interaction with Africa for centuries, during which the slave trade and colonisation were rationalised. While such attitudes are no longer acceptable today, and indeed there is increased appreciation of African cultures in the world today, little value has been attributed to African languages outside the limited circles of linguistic specialists. As late as the 1970s, a major introductory text on Africa gave little attention to African languages other than to suggest that their future was in doubt (Bohannon & Curtin 1971; this statement was modified in later editions – see Bohannon & Curtin 1995). Chaudenson (2004) notes that the subject of language has been almost

entirely absent from the discourse on development in Africa, and Brock-Utne (2005) calls attention to the negative attitudes towards multilingualism in Africa among foreign donors, who tend to regard it as a hindrance to development.

- 8 The focus here is mainly on the written languages, but it is important to acknowledge the importance of audio and non-text images – whether alone or in combination with text – in localisation and multilingual computing.
- 9 For instance, the Globalization and Localization Association (GALA), the Localization Industry Standards Association (LISA) and the Institute of Localisation Professionals (TILP). There is a trade magazine, *Multilingual* (formerly *Multilingual Computing*), as well as an academic journal, *Localisation Focus*.
- 10 Accessed July 2009, <http://lion.openoffice.org/languages.html>. In general, open-source software and operating systems have been localised to a greater extent than proprietary software (*The Economist*, Open source's local heroes, 4 December 2003).
- 11 Even more broadly, at a meta level, one might also include the development of tools to facilitate the process of localisation. This is different from the internationalisation of the technology.
- 12 This is not to suggest that EIWCs in Africa have no connection, but that it is different and, for obvious reasons, less deep.
- 13 Apart from being the mother tongue of a large population in North Africa (where Egypt, for example, is the most populous arabophone country), Arabic is also a major world language with significant numbers of speakers outside the continent. Some African localisation issues therefore involve potentially large markets and can draw on significant and diverse resources.
- 14 'Analogue divide' is a term that emerged after discussion of the 'digital divide' to refer to the longstanding socioeconomic gap between 'haves' and 'have-nots', which is not related at all to ICT or access to digital technology.
- 15 Otis Duncan's (1959) description of the 'ecological complex' linking natural ecology and human social systems into a single model was a key step.
- 16 One might note that the South African, Jan Smuts, articulated the concept of 'holism' in 1926. Smuts was a scholar and politician who served as his country's prime minister twice in the early 20th century.
- 17 Accessed July 2009, <http://world-information.org/wio/readme/992006691>
- 18 In French, '*aménagement linguistique intégré*'.
- 19 Ghana News Agency (2005) About 230 rural communities to get ICT centres. Accessed June 2009, <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=94999>
- 20 Other examples include models by Campbell and Olson (1991), expanding upon the work of Duncan (1959), as well as Rambo (1983).
- 21 The marketing of the Konyin keyboard, for instance, includes the phrase, 'Does not change how you type! No cryptic codes to remember! No training required!' – an explicit recognition of the importance of this sociocultural factor.
- 22 The classical work on the subject of African-language families, although it has its critics, is that of Joseph Greenberg (1970).

- 23 German and Italian have effectively disappeared from use where they were once introduced; Spanish has been retained in Equatorial Guinea, but it is no longer in official use in Western Sahara.
- 24 Macrolanguages, which 'joiners' might in some cases simply call 'languages', but which in other cases may approximate language clusters, is a category that arose in the process of reconciling different parts of the ISO 639 standard for codes representing languages.
- 25 Pular is the variety of Fula spoken in Guinea and vicinity; Pulaar is the variety spoken in Senegambia and vicinity.
- 26 Fula's rich lexicon based on verbal roots includes many synonyms, not all of which are used in the same way or to the same degree by all Fulaphone groups. Moreover, particular forms are sometimes used in particular localities alongside more widely used forms (the counting system in Mali being one example, and pronouns in Pular of Guinea being another). A strategy for terminology development might have to consider such situations.
- 27 This process involves, in effect, a blurring of dialectal differences as a result of factors such as marriage, the movement of people and broadcast media.
- 28 The phenomenon of speakers not mastering the language fully and, in the extreme, no speakers or group of speakers mastering the full range of the language at all.
- 29 Among the recent sources that survey language change in contemporary Africa is one by Batibo (2005).
- 30 Bernard (1996) mentions such diversity of opinion in a discussion of whether linguists should work to preserve indigenous languages.
- 31 The author encountered the opinion that there is no huge demand in Ghana for Ghanaian language interfaces or software from at least two sources. The expectation that there should be large-scale demand manifest before providing interfaces or beginning localisation work for various languages fails to understand the issue of latent demand.
- 32 The author has encountered this attitude among some development professionals.
- 33 The UNESCO *Red book of endangered languages* lists over 180 languages in Africa that it considers endangered (Accessed July 2009, http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/Africa/AF_index.cgi). Batibo (2005) explores the status of endangered and contracting African languages.
- 34 There have been references, for instance, to Igbo – a Nigerian language spoken by at least 18 million people – being endangered, based on perceptions of how the language is and is not being used and passed on – see *Daily Champion*, Igbo: Endangered language (Editorial), 20 December 2004 (Accessed December 2004, <http://allafrica.com/stories/200412201271.html>); Lotanna AA (2005) Revitalising the Igbo language, *Daily Champion*, 28 November 2005 (Accessed December 2005, <http://allafrica.com/stories/200511280234.html>). This obviously stretches the definition of 'endangered' too far, but it also reflects popular interest and concern among many Igbo speakers.
- 35 There are still people who develop and propose new alphabets for one or more African languages (see, for instance: Secka 2005; Agence de Presse Sénégalaise 2005).
- 36 This history of adapting the Latin or Arabic scripts to African languages has been explored in some cases, but not extensively nor for all African languages. The work on the history of Hausa orthographies by Philips (2000) is an exception. Sadembouo (1991) mentions the impact of diverse missionary groups on the orthographies of Beti languages.

- 37 A list of many of these conferences, as well as some others dealing with other aspects of language policy in Africa, is available online at <http://www.bisharat.net/Documents/>. (Accessed July 2009).
- 38 A number of experts are interested in new research on this topic (Fallou Ngom, personal communication, 2006).
- 39 Roger Blench (personal communication, 2006) notes, for instance, that much of what Kay Williamson (1984) compiled on orthographies for several Nigerian languages may not be in current use.
- 40 One exception was the Naira currency notes in Nigeria, which up until 2007 included the amount of the note in Hausa, written in Ajami. This was the only indigenous language represented on the currency. Nigeria has since decided to move to Latin transcriptions of Hausa, Yoruba and Igbo on its currency.
- 41 Particular note should be taken of the process in Cameroon, where an effort to develop an alphabet has apparently met with some success (Tadadjeu 1993; Tadadjeu & Sadembouo 1984).
- 42 This has been dealt with to a very limited extent in the literature, for instance, by Ongarora (2002), Prah (2000), Robinson (1996) and Simala (2002), as well as in this study; however, language in development will be considered as part of the larger issue of language policy.
- 43 Halaoui (2001) distinguishes between language policy and language management, and in fact, a thorough consideration of this subject would merit a nuanced analysis along those lines. However, for the purposes of understanding the localisation ecology, it is sufficient to treat the subject as a single problem, while acknowledging the underlying complexity.
- 44 Many African countries do not have a legislated official language (Gadelli 1999). This fact was borne out by research of national language policies for the PAL project, which referenced Jacques Leclerc's (2005) *L'aménagement linguistique dans le monde* website (Accessed July 2009, <http://www.tlfq.ulaval.ca/axl/afrique/afracc.htm>). This is not particular to Africa, as numerous countries elsewhere (such as the USA) have not found it necessary to designate an official language.
- 45 The ACALAN website lists the many declarations and plans of action issued by conferences and meetings in Africa that have not been acted on (Accessed July 2009, <http://www.acalan.org>).
- 46 Helen Ladd, professor of public policy and economics at Duke University, proposed a similar question regarding the South African government: '... part of the broader language policy they need to grapple with is should all eleven [official] languages remain as viable languages?' (N Aziz interview, 'South Africa: rainbow nation pursues "elusive equity"', AllAfrica.com. 15 December 2004. Accessed June 2009, <http://allafrica.com/stories/200412150903.html>). In other words, she was not talking even about languages in danger of extinction, for which such questions cannot be escaped, but rather of the official and widely used languages of the country. This volume proposes similar questions by other African countries.
- 47 It is worth noting that there have been numerous conferences and meetings over the years to discuss aspects of the use of African languages in education. Two of the earliest such meetings took place in Abidjan, Côte d'Ivoire and Ibadan, Nigeria in 1964 (Sow 1977). Two of the most recent include one on bilingual education in Windhoek, Namibia in August

2005 and one on languages and education in Africa held in Oslo, Norway in June 2006. A partial list of such meetings is available at <http://www.bisharat.net/Documents> (Accessed July 2009).

- 48 There are actually two terms used for this: 'multiliteracy' is also, and perhaps more frequently, used to describe literacy in multiple media; 'pluriliteracy' has been used in some European literature in the stricter sense of the ability to read more than one language. The latter term is used here.
- 49 For example, see Joshi and Aaron (2005). *Ethnologue* has a list of resources on this topic at [http://www.ethnologue.com/LL_docs/index/Orthography\(Literacy\).asp](http://www.ethnologue.com/LL_docs/index/Orthography(Literacy).asp) (Accessed June 2009).
- 50 Access is also an important issue where disabilities are involved, but this report will not address that dimension of access in Africa.
- 51 The roots of this go back to the beginning of discussions of access and the digital divide. In their preface to Afnan-Manns & Dorr's (2003) study of the digital divide, Dorr and Besser write: 'Since its conceptual beginnings in the 1990s, "digital divide" discourse has centred on access to technology, predicating solutions based on additional hardware and wiring.'
- 52 This presentation is no longer on their website, but can be viewed at <http://web.archive.org/web/20041119054155/http://www.bridges.org/digitaldivide/realaccess.html> (Accessed June 2009).
- 53 The author has had the experience of using software in German and Chinese, with little or no mastery of these written languages. Extensive previous experience with the software in English and French enabled him to 'access' this software for basic uses.
- 54 Other non-technical factors that impinge on levels of ICT usage in Africa include literacy and income level (disposable income for use in accessing ICT at for-pay telecentres, or for purchasing computers and internet connections).
- 55 The Simputer project began in India several years ago as a means of addressing the digital divide (Accessed July 2009, <http://www.simputer.org/>).
- 56 Spearheaded by Nicholas Negroponte and the Massachusetts Institute of Technology Media Laboratory, this project has a web presence at <http://laptop.media.mit.edu/> and <http://laptop.org/> (Accessed July 2009).
- 57 The Leland initiative, known as the Africa Global Information Infrastructure project, was formally established in 1995, with the target of extending full internet connectivity to 20 or more African countries (<http://www.usaid.gov/leland/>). IIA was founded in 1996. The two coordinated their efforts to extend connectivity to the maximum number of counties possible (Okpaku 2003).
- 58 Accessed July 2009, http://www.uneca.org/aisi/nici/country_profiles/
- 59 IDRC's (2005) *Acacia atlas* discusses these cables, as do the *Balancing Act* reports on the internet in Africa (2004, 2005, 2006).
- 60 There are also three-letter and three-number codes (Accessed July 2009, http://userpage.chemie.fu-berlin.de/diverse/doc/ISO_3166.html). In addition, there is a draft for country territories (Accessed July 2009, http://en.wikipedia.org/wiki/ISO_3166).
- 61 Accessed June 2009, <http://www.unicode.org/iso15924/iso15924-codes.html>
- 62 Accessed July 2009, <http://www.ietf.org/rfc/rfc4646.txt>. Earlier versions were RFC 1766 and RFC 3066.

- 63 There are 15 in all (Accessed July 2009, http://en.wikipedia.org/wiki/ISO/IEC_8859).
- 64 ANSI is something of a misnomer, as the institute has never formally adopted the drafts of this standard. Nevertheless, they were used as 'Windows ANSI', and the term is commonly heard.
- 65 ISO 6438 is copyrighted and is not available for viewing online. Uncopyrighted versions can be viewed at <http://www.itscj.ipsj.or.jp/ISO-IR/039.PDF> (Accessed July 2009), before ISO 6438 was adopted in 1983, and <http://anubis.dkuug.dk/jtc1/sc2/open/02n3129.PDF> (Accessed July 2009) after 1983. To our knowledge, the only formal use of ISO 6438 is its availability through the UNIMARC bibliographic database (Riley personal communication, 18 March 2009).
- 66 Although largely the same, there were a number of differences in character form between the two; see, for instance, http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=AfrGlyphVars (Accessed July 2009).
- 67 A notable exception is the Senegalese non-governmental organisation, Associates in Research & Education for Development (AREDE), which was headed for many years by an American linguist, Sonja Fagerberg-Diallo. An early example of the kind of use of extended Latin that was possible with Macintosh computers was a learning manual for the Pular of Guinea, produced by Dr Fagerberg-Diallo in 1986.
- 68 There are also other UTFs, including UTF-16 and UTF-32; the number indicates the number of bits. Some background is given at <http://en.wikipedia.org/wiki/UTF-8> (Accessed July 2009).
- 69 Accessed July 2009, <http://www.rifal.org>
- 70 The author is indebted to Mark Davis, Doug Ewell and Steve Summit for clarifying this matter on the Unicode list (September 2006). They responded to a thread entitled 'Unicode & space in programming & L10N', which can be read at <http://unicode.org/mail-arch/unicode-ml/y2006-m09/> (Accessed June 2009).
- 71 A recent example was the sample glyph used for the Y (upper case Y with hook, which is used for the ejective *y* sound in Fula and Hausa), in which the side on which the hook is shown was changed to reflect local usage in West Africa. A discussion of this aspect of this character can be read at <http://scripts.sil.org/HooktopYVariants> (Accessed July 2009). This was apparently an inheritance from the divergence years before between the current practice in Africa (as reflected in the African Reference Alphabet) and the glyph form retained in ISO documents (per ISO 6438) (Accessed July 2009, <http://en.wikipedia.org/wiki/%C6%B3>).
- 72 Accessed July 2009, <http://www.bisharat.net/Documents/Bamako2002-workshop.htm>
- 73 SIL would certainly be capable of making such a global summary from its work in various offices based around the African continent, but to our knowledge, has never done so. The idea of censusing characters used in African orthographies also attracted a project on internationalised domain names (IDN) in 2005, but this was never implemented. The PAL project also considered developing a database on extended characters used in African orthographies, but the uneven quality of data available and the fact that orthographies are sometimes in flux were disincentives.
- 74 This was also discussed on the A12N-collaboration list at <http://lists.kabissa.org/lists/archives/public/a12n-collaboration/msg01059.html> (Accessed July 2009). A12N is an

abbreviation for Africanisation that follows the common technical abbreviations for internationalisation (i18n) and localisation (l10n) – the first and last letter with the number of intervening letters inserted between.

- 75 Examples include several Yahoo! groups with significant Hausa content, as well as a Senegalese forum in which there is Pulaar and Wolof content.
- 76 New ones are still occasionally created, even though Unicode makes them unnecessary. For example, in 2006 a new 8-bit font was announced for the Ewe language in Togo (Togocity.com 2006).
- 77 This outline benefited from information from Cunningham (personal communication, 2006) and Hosken (2003).
- 78 An example of different assignment of keys is the set of differences between the QWERTY and AZERTY keyboards. The placement of the A, Z, Q and W keys, among others, differs between the two layouts. Similarly, it is possible to reassign keys in a keyboard driver without changing what is printed on them in a customised keyboard layout.
- 79 Accessed July 2009, http://en.wikipedia.org/wiki/ISO/IEC_9995
- 80 This description is from a webpage on the IBM website entitled 'Globalize your on demand business' (Accessed July 2009, <http://www-306.ibm.com/software/globalization/topics/keyboards/iso.jsp>).
- 81 These are commonly combined with a QWERTY or AZERTY Latin layout, such that one sees both Arabic and Latin letters on the key caps.
- 82 Williamson (1984: 66) mentions some typewriter keyboards for Nigerian languages along with strategies for typing with English keyboards. In the 1980s, the IBM company developed some typeballs with what we now call extended Latin characters for its Selectric typewriter. Mann and Dalby (1987) proposed a lower-case-only keyboard for typewriters and computers based on the Niamey African Reference Alphabet, but this did not become popular (Accessed July 2009, http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=IntlNiameyKybd). There is actually a keyboard layout based on the Mann-Dalby Niamey keyboard, but it includes upper-case characters as well.
- 83 See, for instance, the Tavultsoft site (Accessed July 2009, <http://www.tavultsoft.com>) or the keyboard projects links (Accessed July 2009, <http://www.bisharat.net/A12N/Projects>).
- 84 The Nigerian Information Technology Development Agency (NITDA) developed a keyboard layout.
- 85 This was the case in Mali, for instance, where the 8-bit fonts Bambara Arial and Bambara Times were developed by a project facilitated by the French agency ACCT during the late 1990s.
- 86 These include several keyboard layouts by Andrew Cunningham of the OpenRoad project (Accessed July 2009, <http://www.openroad.net.au/languages/files/>). See also the discussion on the Igbo language and ICT forum (Accessed July 2009, <http://www.quicktopic.com/17/H/tCcDxVXHgQxN>).
- 87 Accessed July 2009, <http://www.konyin.com/>. It is designed for use with Microsoft Windows software.
- 88 For instance, in a large cybercafé in Bamako, Mali in 2000, the author encountered French, English and German language keyboards.
- 89 Accessed July 2009, <http://www.bisharat.net/A12N/Projects/#tabl>

- 90 According to the Art Lebedev website (Accessed July 2009, <http://www.artlebedev.com/portfolio/optimus/>): 'Every key of the Optimus keyboard is a stand-alone display showing exactly what it is controlling at this very moment.'
- 91 A description of how this situation arose is given at www.loc.gov/standards/iso639-2/develop.html (Accessed July 2009).
- 92 Each of the profiles of the major languages (on the PAL wiki) includes information on ISO 639 codes for that language.
- 93 Accessed July 2009, <http://www.unicode.org/cldr/>
- 94 Accessed July 2009, <http://yeha.sourceforge.net/>
- 95 Accessed July 2009, <http://www.it46.se/localegen/>
- 96 This service ended in 2006.
- 97 Both of these have since ceased to function.
- 98 Accessed July 2009, <http://www.w3.org/>
- 99 A simple survey of websites by language conducted in 2000 by Vilaweb, the website of a Barcelona newspaper (Pastore 2000), listed no African languages among the top 31 languages. A follow-up to the Vilaweb survey, which ranked the top 48 languages on the web, found Afrikaans to be 42nd, after languages such as Basque and Slovenian, and Swahili last, following Frisian and Faeroese among others (Mas 2003).
- 100 A more recent survey by UNESCO (2005) on linguistic diversity on the internet repeats the information summarised in this section for Africa.
- 101 Accessed July 2009, <http://gii2.nagaokaut.ac.jp/gii/blog/lopdiary.php/lopdiary.php?catid=154&blogid=8>
- 102 Accessed July 2009, <http://kamusiproject.org/>
- 103 ALI stands for 'Apprentissage des Langues Africaines par l'Internet' (Learning African languages on the internet) (Tonye, Soundjock & Mbede [2002]; Accessed July 2009, <http://www.kabissa.org/archives/a12n-forum/msg00187.html>). This project has been suspended since 2004 due to lack of funding (Tonye, personal communication, 2007). The ALI project should not be confused with the ALI Akan online programme, which is for L2 learners of Akan and is based in Switzerland (ALI in that case stands for African Languages on the Internet).
- 104 This effort uses a Yahoo! group at <http://tech.groups.yahoo.com/group/afrophenewikis/> (Accessed July 2009).
- 105 See <http://wo.wikipedia.org/> for Wolof and <http://yo.wikipedia.org/> for Yoruba (Accessed July 2009).
- 106 The director, Dr Nii Quaynor of Ghana, also served as At-Large Director of ICANN from 2000 until 2003.
- 107 An organisational meeting, convened by Mouhamet Diop of the Senegalese company Next SA, was held in Dakar on 7 September 2005 to launch this effort.
- 108 This involves testing two main alternative ways of handling non-ASCII characters and scripts (Crawford 2006).
- 109 Gachenge B, Why Microsoft Swahili version failed, *Business Daily* (Nairobi), 12 March 2008. Accessed June 2009, <http://allafrica.com/stories/200803111159.html>
- 110 For background, see Rasch (2000).
- 111 Accessed June 2009, <http://translate.sourceforge.net/wiki/pootle/index?DokuWiki=04bo203d28c36961cbc8160b690994ba>

- 112 A brief list is given at <http://lists.kabissa.org/lists/archives/public/a12n-forum/msg00478.html> (Accessed July 2009).
- 113 Denis Gikunda (personal communication, 2009).
- 114 Nokia has localised 'menu text and predictive input' for at least one phone model in Afrikaans, Arabic and Swahili and 'menu text only' in Zulu, Xhosa, Sesotho, Yoruba, Hausa and Igbo (Accessed July 2009, <http://www.europe.nokia.com/A4160009>).
- 115 Accessed July 2009, <http://www.iapad.org/>
- 116 A list of such resources is available at <http://opensourcegis.org/> (Accessed July 2009).
- 117 Markus Neteler (personal communication, 2005). For further information on GRASS, see <http://grass.itc.it/> (Accessed July 2009).
- 118 The International Association for Machine Translation (IAMT), for instance, is composed of three regional associations, one each for the Americas, Europe and the Asia-Pacific region, but none in Africa, a continent that alone accounts for about a third of the world's languages.
- 119 It is important for the end-users of products – people who may need to refer to manuals and instructions in a language other than the original – to have material that is clear and consistent.
- 120 Accessed July 2009, <http://xixona.dlsi.ua.es/apertium-www/?id=whatisapertium>
- 121 Jeff Allen (personal communication, 2006).
- 122 Accessed July 2009, <http://mokennon2.albion.edu/language.htm>
- 123 Accessed July 2009, <http://www.njas.helsinki.fi/salama/>
- 124 Accessed July 2009, <http://www.trados.com/>; <http://www.sdl.com/en/products/translation-memory/sdl-trados/>
- 125 Accessed July 2009, <http://www.omegat.org/omegat/omegat.html>
- 126 As indeed the PAL and ANLOC projects have done.
- 127 In the framework of the PLETES model, this would refer to two points in the localisation dynamic.
- 128 Bisharat is an initiative begun in 2000 and incorporated in 2004, with the purpose of research, advocacy and networking on the use of African languages in ICT for development.
- 129 Accessed July 2009, <http://lists.panafri1ion.org/mailman/listinfo/pal-en>; <http://lists.panafri1ion.org/mailman/listinfo/pal-fr>; <http://lists.panafri1ion.org/mailman/listinfo/pal-pt>. The translation was provided by Tradauto, a service of FUNREDES.
- 130 Accessed July 2009, <http://www.bisharat.net/Documents/>
- 131 It is our understanding from previous experience that many people working on localisation of Arabic often use English or French as working languages. Nevertheless, the possibility of working in Arabic is considered.

References

- Afnan-Manns S & Dorr A (2003) *Re-evaluating the bridge: An expanded framework for crossing the digital divide through connectivity, capability, and content*. 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. Accessed June 2009, http://www.newliteracies.gseis.ucla.edu/publications/re-eval_bridge.pdf
- Afrik.com (2008) Linguistic atlas to provide Africa with a geo-linguistic map. 11 September 2008. Accessed 20 June 2009, <http://en.afrik.com/article14471.html>
- Agence de Presse Sénégalaise (APS). (2005) Un professeur d'Anglais à la retraite invente une écriture dénommée Typafrica. APS 06-05-2005. Accessed July 2005, http://www.aps.sn/artifice.php?page=&id_article=8389
- Alexander N (2008) The African Academy of Languages (ACALAN): A linguistic renaissance for the continent? *Openspace* 2(3). Accessed 20 June 2009, http://www.osisa.org/files/openspace/2_3_language_p055-059_neville_alexander.pdf
- Balancing Act Africa* (2004) African internet country profiles, Part 1, West Africa. London: Balancing Act. Accessed June 2009, <http://www.balancingact-africa.com/profile1.html>
- Balancing Act Africa* (2005) African internet country profiles, Part 2, East Africa. London: Balancing Act. Accessed June 2009, <http://www.balancingact-africa.com/profile2.html>
- Balancing Act Africa* (2006) African internet country profiles, Part 3, Southern and Central Africa. London: Balancing Act. Accessed June 2009, <http://www.balancingact-africa.com/profile3.html>
- Ballantyne P (2002) *Collecting and propagating local development content: Synthesis and conclusions*. IICD in association with the Tanzania Commission for Science and Technology funded by DFID, Research Report No. 7. Accessed June 2009, <http://www.ftpicd.org/files/research/reports/report7.pdf>
- Bamgbose A (1991) *Language and the nation: The language question in sub-Saharan Africa*. Edinburgh: Edinburgh University Press
- Bamgbose A (1996) Pride and prejudice in multilingualism and development. In R Fardon & G Furniss (Eds) *African languages, development, and the state*. London: Routledge
- Batibo H (2005) *Language decline and death in Africa: Causes, consequences, and challenges*. Clevedon, UK: Multilingual Matters
- Bergmann F (2005) Open-source software and localization. *MultiLingual Computing* 16(2): 70. Accessed June 2009, <http://www.project-open.com/whitepapers/oss-110n/>
- Bernard HR (1996) Language preservation and publishing. In NH Hornberger (Ed.) *Indigenous literacies in the Americas: Language planning from the bottom up*. New York: Mouton de Gruyter
- Bohannon P & Curtin P (1971) *Africa and Africans* (revised edition). Garden City, NY: Natural History Press
- Bohannon P & Curtin P (1995) *Africa and Africans* (4th edition). Prospect Heights, IL: Waveland Press

- Bokamba EG (1995) The politics of language planning in Africa: Critical choices for the 21st century. In M Putz (Ed.) *Discrimination through language in Africa: Perspectives on the Namibian experience*. Berlin: Mouton de Gruyter
- Bourbeau L & Pinard F (2000) Observations, réflexions et perspectives de l'informatisation des langues Africaines. Paper presented at the Internet: Bridges to Development Conference, Bamako, Mali (21–26 February), organised by the ANAIS Network. Accessed June 2009, <http://web.archive.org/web/20060103201531/http://www.progiciels-bpi.ca/man/bam2000/>
- Brock-Utne B (2005) Language-in-education policies and practices in Africa with a special focus on Tanzania and South Africa: Insights from research in progress. In AMY Lin & PW Martin (Eds) *Decolonization, globalization: Language-in-education policy and practice*. Clevedon: Multilingual Matters
- Campbell DJ & Olson JM (1991) *Framework for environment and development: The kite*. Center for Advanced Study of International Development Occasional Paper No. 10, Michigan State University
- Chanard C (2005) Pour une transcription pérenne des langues Africaines. Paper presented at the 27th Conference on Internationalization and Unicode, Berlin (6 April)
- Chaudenson R (1987) Pour un aménagement linguistique intégré: Le cas de la graphie des créoles français. *Etudes Créoles* 10(2): 143–58
- Chaudenson R (2003) Langues et numérisation: Français, créoles, langues Africaines. In I Ndaywel è Nziem (Ed.) *Les langues Africaines et créoles face à leur avenir*. Paris: L'Harmattan
- Chaudenson R (2004) De Ouagadougou (1988) à Ouagadougou (2004) en passant par Libreville (2003). In *Penser la Francophonie: Concepts, actions et outils linguistiques*. Actes des premières journées scientifiques communes des réseaux de chercheurs concernant la langue Ouagadougou, Burkina Faso (31 May–1 June). Paris: Agence Universitaire de la Francophonie.
- Childs GT (2003) *An introduction to African languages*. Amsterdam: John Benjamin
- Chtatou M (1992) *Using Arabic script in writing the languages of the peoples of Muslim Africa*. Rabat: Institute of African Studies
- Cissé T, Mbodj C, Van Campenhoudt M & Wane M (2004) Expérimentation de normes de balisage en langues partenaires. In *Penser la Francophonie: Concepts, actions et outils linguistiques*. Actes des premières journées scientifiques communes des réseaux de chercheurs concernant la langue, Ouagadougou, Burkina Faso (31 May–1 June). Paris: Agence Universitaire de la Francophonie. Accessed June 2009, <http://www.termisti.refer.org/ouagadougou.PDF>
- Clews J (2003) iso 6438, the Niamey keyboard, and iso/TC46/SC4/WG1. Message posted on Unicode@unicode.org list, 13 May. Accessed June 2009, <http://unicode.org/mail-arch/unicode-ml/y2003-m05/0256.html>
- Coulmas F (1992) *Language and economy*. Oxford: Blackwell
- Crawford S (2006) Testing IDNs. CircleID. Accessed June 2009, http://www.circleid.com/posts/testing_internationalized_domain_names_idns/
- Da Costa P (2002) African content on the web. In L Levey & S Young (Eds) *Rowing upstream: Snapshots of pioneers of the information age in Africa*. Johannesburg: Sharp Sharp Media

- Diki-Kidiri M (2008) Securing a place for a language in cyberspace. Information Society Division, Communication and Information Sector, UNESCO Accessed June 2009, <http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?catno=149786>
- Diki-Kidiri M & Edema AB (2003) Les langues Africaines sur la toile. *Cahiers du Rifal* 23: 5–32
- Duncan OD (1959) Human ecology and population studies. In PM Hauser & OD Duncan (Eds) *The study of population: An inventory and appraisal*. Chicago: University of Chicago Press
- Enguehard C (2006) Message posted on Unicode-Afrique, 18 April. Accessed 21 June 2006, <http://fr.groups.yahoo.com/group/Unicode-Afrique/message/1050>
- Enguehard C & Naroua H (2008) Evaluation of virtual keyboards for West African languages. Proceedings of the Sixth International Language Resources and Evaluation. Accessed 21 June 2009, http://www.sciences.univ-nantes.fr/info/perso/permanents/enguehard/recherche/LREC2008_Enguehard_Naroua.PDF
- Esselink B (1998) *A practical guide to software localization: For translators, engineers and project managers*. Amsterdam: John Benjamins
- Esterhuysen A (2002) Networking for a purpose: African NGOs using ICT. In L Levey & S. Young (Eds) *Rowing upstream: Snapshots of pioneers of the information age in Africa*. Johannesburg: Sharp Sharp Media
- Fill A (2001) Ecolinguistics: State of the art 1998. In A Fill & P Mühlhäusler (Eds) *The ecolinguistics reader*. London: Continuum
- Gadelli KE (1999) Language planning: Theory and practice. Evolution of language planning cases worldwide. Paper prepared for the Languages Division, Education Sector, UNESCO, Paris
- Gordon RG Jr (Ed.) (2005) *Ethnologue: Languages of the world* (15th edition). Dallas: SIL International. Accessed June 2009, <http://www.ethnologue.com/>
- Greenberg JH (1970) *The languages of Africa* (3rd edition). Bloomington, IN: Indiana University Press
- Halaoui N (2001) Aménagement et politique linguistiques: la politique des langues au Bénin. *Language Problems and Language Planning* 25(2): 145–166
- Harlech-Jones B (1993) What ecology? *Logos* 13: 11–24
- Haugen E (2001/1972). The ecology of language. In A Fill & P Mühlhäusler (Eds) *The ecolinguistics reader*. London: Continuum. Originally published in AS Dil (Ed.) 1972. *The ecology of language: Essays by Einar Haugen*. Stanford: Stanford University Press
- Heine B & Nurse D (Eds) (2000) *African languages: An introduction*. Cambridge: Cambridge University Press
- Herbert RK (1992) Language in a divided society. In RK Herbert (Ed.) *Language and society in Africa: The theory and practice of sociolinguistics*. Johannesburg: Witwatersrand University Press
- Hosken, M (2003) *An introduction to keyboard design theory: What goes where?* SIL International. Accessed June 2009, <http://scripts.sil.org/KeybrdDesign>
- IDRC (International Development Research Centre) (2005) *The Acacia atlas: Mapping African ICT growth*. Ottawa: IDRC
- IILAC (International Institute for African Languages and Cultures) (1930) *Practical orthography of African languages* (revised edition) London: IILAC

- ISO (International Organization for Standardization) (1988) *ISO 3166:1988. Codes for the representation of names of countries* (3rd edition). August 1988
- ISO (International Organization for Standardization) (1998) *ISO 639-2:1998. Codes for the representation of names of languages* (Part 2: Alpha-3 code) (1st edition)
- ISO (International Organization for Standardization) (2002) *ISO 639-1:2002. Codes for the representation of names of languages* (Part 1: Alpha-2 code)
- ISO (International Organization for Standardization) (2004) *ISO 15924:2004. Information and documentation: Codes for the representation of names of scripts*. January 2004
- ISO (International Organization for Standardization) (2006) *ISO and Africa*. Accessed June 2009, http://www.iso.org/iso/en/comms-markets/developingcountries/PDF/iso_and_africa.PDF
- Joshi RM & Aaron PG (Eds) (2005) *Handbook of orthography and literacy*. Hahwah, NJ: Lawrence Erlbaum Associates
- Levey LA & Young S (Ed.) (2002) *Rowing upstream: Snapshots of pioneers of the information age in Africa*. Johannesburg: Sharp Sharp Media
- Mackey WF (1989) Status of languages in multinational societies. In U Ammon (Ed.) *Status and function of language and language varieties*. Berlin: Walter de Gruyter
- Mafundikwa S (2004) *Afrikan alphabets: The story of writing in Africa*. West New York, NJ: Mark Batty
- Mann M & Dalby D (1987) *A thesaurus of African languages: A classified and annotated inventory of the spoken languages of Africa with an appendix on their written representation*. London: Hans Zell Publishers
- Mas J (2003) La salut del català a internet. *Softcatalà*. Accessed June 2009, <http://www.softcatala.org/articles/article26.htm>
- Matwyshyn AM. (2003) Silicon ceilings: Information technology equity, the digital divide and the gender gap among information technology professionals. *Northwestern Journal of Technology and Intellectual Property* 2(1). Accessed June 2009, <http://www.law.northwestern.edu/journals/njtip/v2/n1/2>
- Mazrui AA. & Mazrui AM (1998) *The power of Babel: Language and governance in the African experience*. Chicago: University of Chicago Press
- Microsoft Corporation. (2004) Microsoft enables millions more to experience personal computing through local language program. 16 March. Accessed June 2009, <http://www.microsoft.com/presspass/press/2004/mar04/03-16LLPPR.asp>
- Miller Esselaar & Associates. (2001) *A country ICT Survey for Tanzania*. Final report prepared for the Swedish International Development Cooperation Agency, Accessed November 2004, <http://www.milless.co.za/downloads/Sida%20report%20-%20Tanzania.PDF>
- Myers-Scotton C (1990) Elite closure as boundary maintenance: The evidence from Africa. In B Weinstein (Ed.) *Language policy and political development*. Norwood, NJ: Ablex
- Myers-Scotton C (1993) Elite closure as a powerful language strategy: The African case. *International Journal of the Sociology of Language* 103: 149–63
- Nsengiyumva A & Stork C (2005) Rwanda. In A Gillwald (Ed.) *Towards an African e-index: Household and individual ICT access and usage across 10 African countries*. Research ICT Africa. Accessed June 2009, <http://link.wits.ac.za/papers/e-index-rwanda.PDF>
- Okombo DO (2001) Language policy: The forgotten parameter in African development and governance strategies. Inaugural lecture, University of Nairobi

- Okpaku JO Sr (2003) Towards a road map for ICT development in Africa. In JO Okpaku Sr (Ed.) *Information and communications technologies for African development: An assessment of progress and the challenges ahead*. Accessed June 2009, http://www.ictdevlibrary.org/downloads/18_un_ict_task_force_african_development_roadmap.pdf
- Ongarora DO. (2002) African languages in development: Prospects and encumbrances. In FR Owino (Ed.) *Speaking African: African languages for education and development*. Cape Town: Centre for Advanced Studies of African Society
- Osborn D (2001) The knotty problem of using African languages for e-mail and internet. *Balancing Act's News Update* 69. Accessed 21 June 2009, http://www.balancingact-africa.com/news/back/balancing-act_69.html
- Osborn D, Anderson D & Kodama (2008) Support for modern African languages and scripts in Unicode/ISO 10646: Where are we today? Paper presented at the 32nd Internationalization and Unicode Conference, San Jose, California (10 September)
- Paolillo J (2005) Language diversity on the internet. In J Paolillo, D Pimienta, D Prado, et al. (Ed.). *Measuring linguistic diversity on the internet: A collection of papers*. No. CI.2005/WS/06. Montreal: UNESCO. Accessed June 2009, <http://unesdoc.unesco.org/images/0014/001421/142186e.pdf>
- Pastore M (2000) Web pages by language. *ClickZ Stats*, 5 July. Accessed June 2009, http://web.archive.org/web/20060314155743/www.clickz.com/stats/sectors/demographics/article.php/5901_408521
- Philips JE (2000) *Spurious Arabic: Hausa and colonial Nigeria*. African Studies Program, University of Wisconsin
- Prah KK (2000) *Mother tongue for scientific and technological development in Africa* (3rd edition). Cape Town: Centre for Advanced Studies of African Society
- Prah KK (2002) Language, neo-colonialism, and the African development challenge. *TRicontinental* No. 150 (Havanna, Cuba)
- Prah KK (2003) Going native: Language of instruction for education, development and African emancipation. In B Brock-Utne, Z Desai & M Qorro (Eds) *Language of instruction in Tanzania and South Africa*. Dar-es-Salaam: E&D Publishers
- Rambo AT (1983) *Conceptual approaches to human ecology*. East-West Environment and Policy Institute Research Report No. 14. Honolulu: East-West Environment and Policy Institute
- Rasch C (2000) *A brief history of free/open source software movement*. Open Knowledge. Accessed June 2009, <http://www.glennmcc.org/foss/brief-open-source-history.html>
- Rathke E (2005) Internationalization for localization (118n for 110n). OpenOffice.org Conference 2005, Koper/Capodistria, Slovenia. Accessed July 2009, http://ooocon.kiberpipa.org/media/index-talk-2005.html#Localization_Eike_Rathke
- Renaud P. (1994) *Le projet RIO, historique, organisation, partenaires*. Accessed July 2009, <http://www.unitar.org/isd/publications/rio%5Cprog-rio94.html>
- Robinson CDW (1996) *Language use in rural development: An African perspective*. Berlin: Mouton de Gruyter
- Sadembou E (1991) Préalables à la standardisation des langues africaines. In N Cyffer, K Schubert & H-I Weier (Eds) *Language standardization in Africa*. Hamburg: Helmut Buske Verlag
- Secka, PM (2005) Local alphabet to be launched. *The Independent* (Banjul), 25 February. Accessed March 2005, <http://www.qanet.gm/Independent/independent.html>

- Senge P (2006) *The fifth discipline: The art and practice of the learning organization* (revised edition). New York: Doubleday/Currency
- Senne D (2006) Zulu, Xhosa sms made possible. *ITWeb*. Accessed June 2009, <http://www.itweb.co.za/sections/telecoms/2006/0606231042.asp?S=Cellular&A=CEL&O=FRGN>
- Shanglee R (2004) Localization in African languages: Translators face linguistic challenges as they localize modern technology. *Multilingual Computing and Technology* 15(1): 61
- Simala IK (2002) Empowering indigenous African languages for sustainable development. In FR Owino (Ed.) *Speaking African: African languages for education and development*. Cape Town: Centre for Advanced Study of African Languages
- Smuts J (1926, reprinted 1973) *Holism and evolution*. Westport, CN: Greenwood
- Sow AI (Ed.) (1977) *Langues et politiques de langues en Afrique noire: l'Expérience de l'UNESCO*. Paris: Nubia
- Suzuki I, Mikami Y, Ohsato A & Chubachi Y (2002) A language and character set determination method based on n-gram statistics. *ACM transactions on Asian language information processing* 1(3): 269–278. Accessed June 2009, <http://portal.acm.org/affiliated/citation.cfm?id=772759&dl=guide&coll=ACM&CFID=15151515&CFTOKEN=6184618>
- Tadadjeu M (1993) Cameroon. In RH Hartell (Ed.) *The alphabets of Africa*. Dakar: United Nations Educational, Scientific and Cultural Organisation and SIL
- Tadadjeu M & Sadembouo R (Eds) (1984) General alphabet of Cameroon languages, adopted by the National Committee for the Unification and Harmonization of the Alphabets of Cameroon Languages, Yaoundé, Cameroon (7–9 March 1979). Faculty of Letters and Social Sciences, Department of African Languages and Linguistics, University of Yaoundé
- Taylor C (2000) Typesetting African languages: Report of an investigation. Accessed November 2009, <http://www.archive.org/details/TypesettingAfricanLanguages>
- TeleCommons Group. (2000) Rural access to information and communication technologies: The challenge of Africa. Paper prepared for the African Connection Secretariat, with support from the Information for Development Program (infoDev). Accessed June 2009, <http://www.telecommons.com/reports.cfm?itemid=122>
- Texin T (2006) *What is wrong with locales?* Accessed June 2009, <http://www.i18nguy.com/locales/Locales.pdf>
- Togocity.com. (2006) L'écriture informatisée de la langue Ewe est maintenant possible. 8 March. Accessed June 2009, http://www.togocity.com/article.php?id_article=914
- Tonye E, Soundjock ES & Mbede J (2002) Apprentissage des langues Africaines par l'internet (ALI): Les cas des langues camerounaises. Paper presented at the Global Conference on Community Networking in the Digital Age, Montreal, Canada (7–12 October)
- UNESCO (United Nations Educational, Scientific and Cultural Organisation) (2005) *Measuring linguistic diversity on the internet*. Accessed June 2009, http://portal.unesco.org/ci/en/ev.php-URL_ID=20804&URL_DO=DO_TOPIC&URL_SECTION=201.html
- USINFO (American Information Web) (2006) Internet connections growing fastest in Africa. US Department of State, Bureau of International Information Programs. Accessed June 2009, <http://newsblaze.com/story/20060427183152sop.nb/topstory.html>
- Van der Merwe IJ & Van der Merwe JH (2003) *Linguistic atlas of South Africa: Language in space and time*. Sun Press. <http://usinfo.state.gov/dhr/Archive/2006/Apr/28-229950.html>
- Van der Veken A & De Schryver G-M (2003) Les langues Africaines sur la toile: Etude des cas Haoussa, Somali, Lingala et Xhosa. *Cahiers du Rifal* 23: 33–45

- Webb V & Kembo-Sure (Eds) (1998) *African voices: An introduction to the languages and linguistics of Africa*. Cape Town: Oxford University Press
- Williamson K (1984) *Practical orthography in Nigeria*. Ibadan: Heinemann Educational Books
- Yacob D (2004) Localize or be localized: An assessment of localization frameworks. Paper presented at the International Symposium on ICT: Education and Application in Developing Countries, Addis Ababa, Ethiopia (19–21 October)
- Ya'u YZ (2005) Confronting the digital divide: An interrogation of African initiatives to bridge the gap. In JO Adesina, Y Graham & AO Olukoshi (Eds) *Africa and development challenges in the new millennium: The NEPAD debate*. London: Zed Books

Index

A

analogue divide 17–18, 131m14
see also digital divide
 African Academy of Languages (ACALAN) 34, 39, 75, 83, 110
 African Development Forum 47
 African Language Research Project 95
 African Language Resource Council 53
 African-language text and ICT
 encoding and fonts/typography 53,
 57–62&tab, 63, 69, 119, 121
 diacritics in 53–54, 58–59, 61tab, 62, 66–68,
 135n71
 orthographies xii, 9, 12, 17, 19–20, 22, 37–38,
 41, 57–61&tab, 68, 98, 100, 112, 117, 125,
 132n36, 135n73
 writing systems 36, 54, 57, 75
see also character encoding; language
 technologies; literate/written systems;
 software localisation
 African languages xi–xiii, 5–6, 9, 11, 13, 17, 24,
 32–33, 127
 Afrikaans 32, 89tab, 92, 95
 Akan 74, 95
 Amharic 87, 89, 93, 95
 Arabic xi, 9–10, 13, 33, 35–36, 51, 54, 57, 62,
 68, 74, 76tab, 89, 92, 95, 109, 115, 131m13
 Berber languages (Tifinagh) 19–20, 32, 35,
 40, 54, 57, 63
 endangerment/extinction/preservation of
 34–35, 39, 107, 117, 132n33&34
 Ethiopic 54, 57, 63, 68, 115
 Fula 32–33, 132n26
 Hausa 34, 61tab, 76tab, 79, 82, 113
 interintelligibility of 6, 31–34, 100
 Ibibio 93, 113
 Manding (N'ko) 57, 59, 63, 87, 117
 negative attitudes to/promotion of 34–35, 38,
 128, 130n7
 policy 20, 27&fig, 28, 30, 32, 37–39, 109–110,
 116, 125, 128, 133n44&46
 Somali 71, 77tab, 87, 93

Swahili 32, 35, 58, 77tab, 79, 81–82, 86, 88,
 89tab, 92–93, 95, 123
 taxonomy of 31
 Wolof 34, 59, 77tab, 82
 Yoruba 71, 77tab, 82, 93, 95, 113
 Xhosa 88, 89tab, 92, 95
 Zulu 58, 77tab, 89tab, 92–93
see also language/linguistic geography;
 localisation: language/linguistic context

African Union 128

Apple MacIntosh 51

B

Bamgbose, A 38–39
 benefits of localisation *see under* localisation
 bilingualism 40

C

Cameroon 32, 39
 Centre for Advanced Studies of African Society
 (CASAS) 33, 110
 Centre d'Études Linguistiques et Historiques par
 Tradition Orale (CELHTO) 39
 Centre Régional de Recherche et de
 Documentation sur les Traditions Orales
 et pour le Développement des Langues
 Africaines (CERDOTOLA) 39
 Centre d'Investigation et de Documentation sur
 l'Oralité (CIDLO) 40
 character encoding
 ANSI 51, 58, 60, 65
 ASCII 50–52, 58–61tab, 65
 free fonts 62
 ISO 49–51, 53, 58, 62, 69, 73–77tab, 94, 121
 Latin-1 58–59
 optical character recognition 127
 Unicode 6, 9, 11, 19, 23, 46, 49–55, 58, 60,
 62–65, 75, 80, 83, 94, 98, 112, 115, 121,
 129–130n6
 UTF-8 60, 80
see also African-language text; language
 technologies

Chaudenson, Robert 18–20
 closed source/proprietary software 85, 88, 111, 120, 124
 colonial influence on language 32, 34–35, 38
see also European languages of wider communication
 computer-assisted translation 1, 91, 94, 98
 machine translation xi, 1, 31, 42, 79, 89, 94–95, 106, 114, 125–126, 138
 software 10, 21, 24–26, 28, 85, 95, 98, 110, 113, 118, 120, 126
 translation memory 94–95, 125–126
 Web-Lingo 95

D

Democratic Republic of the Congo 32
 development 1–2, 10, 13, 15, 17, 34, 38, 41, 118–119, 124, 126, 129
see also economic/resource context of localisation; localisation projects; sustainable localisation
 digital divide 1, 17, 22, 25, 28, 30, 44, 46–47, 112, 116, 128, 134
 and gender gap 18
 language as a factor in 30, 44–45
 and physical access 43–46, 48–49, 99, 102, 122, 128
 and poor communications infrastructure 45–46
 and soft access 43–45, 49, 126
see also localisation ecology

E

East Africa 13, 33, 47
 Eastern African Centre for Research on Oral Traditions and African National Languages (EACROTANAL) 40
 ecology
 digital/information ecosystems 18
 human/social 18
 linguistic 18–19
see also ICT localisation: ecology
 economic/resource context of localisation 13, 22, 24–25, 27–28, 46, 107, 110, 114–116, 120, 127

and donor organisations/development agencies 20, 102–103, 110–112, 114–115, 119, 121
 and enabling/support systems 3, 15, 111, 118–119
 education policy 31, 34, 38, 59, 109
 and languages of instruction 40–41, 45, 118, 133
 educational/training context of localisation 6, 15, 17, 20, 22–25, 28–29, 40–41, 43, 52, 107, 111–114, 117, 127, 130
 and public education 114–115
Ethnologue 6, 32–33, 73–75
 European languages of wider communication (ELWCS) 5–6, 10, 12–13, 25, 32, 34, 38, 41, 59, 68, 118, 122, 126
 and nation building 38

F

first language/mother tongue communication (11) xiii, 5–6, 25, 33, 35, 40–41, 45, 82, 93, 118, 130
see also educational policy: languages of instruction

G

Gambia 34
 Geekcorps 119
 geographic information systems 91, 93–94, 102, 124, 126
 Geographic Resources Analysis Support System (GRASS) 93–94, 124, 138
 ISO 19100 series 94
 Ghana 20, 112

H

Haugen, Einar 18–19
 Horn of Africa 35, 68

I

IBM 86
 information and communication technology (ICT) xi–xii, 1, 130
 digital xi, 7, 91, 124–125

- networks/networking 7–8, 14, 44, 52, 88,
106–107, 113, 115–116
policy 25, 27&fig, 28, 47, 109
see also internet; localisation; localisation
ecology
- integrated language management 18–21&fig
PLETES model 19, 23&fig, 26&fig, 29–30, 48,
108
- International Development Research Centre
(IDRC) xi, 47, 110, 113, 119
- International Francophone Network for
Language Management 52
- International Institute of African Languages and
Cultures 37
- international standards for localisation 6, 9, 12,
20, 22–23, 36–37, 49–51, 53, 58, 62, 65, 67,
73–74&tab, 75–76&tab, 77tab, 94, 98, 100,
101, 107–108, 112, 117–118, 120–123, 125
International Organisation for
Standardization (iso) 73, 121, 130n6
see also iso codes
- internet xi, 1, 5, 7, 9, 13, 15, 41, 44, 46–47, 55,
59–60, 79, 109, 112, 116
connectivity 47–48, 79, 102
Facebook 88
Google 88–89
Internet Engineering Task Force 73
see also localisation, technical context: web
content; World Wide Web
- Islamic States Educational, Scientific and
Cultural Organisation 37
- iso codes 49–50, 51, 53, 58, 62, 67, 94, 112
iso 639 73–76, 121
- K
- Kenya 93
- keyboards 9, 24, 26, 50, 55, 58, 60, 65–69, 73,
88, 110, 113, 119, 121–123, 131n21, 136n78&82
LED 71, 123
virtual 70, 79
see also software localisation
- L
- language boundaries 32
and language/linguistic geography 31–32
and Creole languages 32
and cross-border languages 102, 112, 117
and dialects/sub-families 31–34, 100, 112, 117,
126
and language atlas projects 34
and language clusters 33, 73–75, 112
and localisation 34
see also geographic information systems
- language technologies
African Language Technology Initiative 93
and codes 34, 49, 55, 73–74&tab, 75
and online dictionaries 10–11, 14, 20, 42,
81–82, 118
and terminology development 10, 20, 24, 31,
33, 35, 41–42, 88, 98, 100, 112
see also character encoding; computer-
assisted translation; locales
- language technologies, audio-related 91, 126
audio e-mail 92
digital audio 125
Local Language Speech Technology Initiative
93, 125
speech-to-text xi, 9, 70–71, 92–93, 125–126,
114, 123, 125–126
text-to-speech xi, 92–93, 123, 125
- linguistic diversity *see* multilingualism
- literacy xii, 12, 20, 22–23, 25, 40–41, 45, 82,
92–93, 118, 126
pluriliteracy 40–41, 118, 122
see also multilingualism
- literate/written systems 35–37, 109, 116–117, 125
and Arabic script 36–37, 54
and extended Latin script 36–37, 52–53,
57–61tab, 62&tab, 66, 68, 79, 97, 127
and non-Latin script 57, 59, 62, 66, 83, 97,
127
see also African-language text
- locales 55, 73, 75&tab, 100, 115, 121
Common Locale Data Repository 75–76, 112
Language Tag Registry Unit (LTRU) 73
LocaleGEN 76
written system codes 75–76&tab,
Yeha 76
- localisation in Africa xi–xiii, 1–3, 5–7, 14, 35, 128

- advanced applications and research 107, 110, 112, 114–115, 123
- current state of 15
- language/linguistic context of 3, 5–6, 12, 31, 39, 107, 109
- see also under* localisation ecology
- networks 116
- policy 27, 29–30, 38, 59, 103, 107–110, 116, 118,
- political context of 22–24, 27, 128
- sociocultural context 21–23 *&fig*, 27, 109
- see also under* localisation ecology
- strategic planning perspective on 29, 47, 100, 102–103, 107, 109, 111–112, 124–125
- see also* international standards for localisation
- localisation ecology xii, 3, 17, 38, 98, 102, 108, 110, 115, 118, 121, 127–129
- and language-technology-sociocultural interrelationships 18–21 *&fig*, 22–23 *&fig*, 24–26 *&fig*, 27, 28–30, 41, 48–49, 55, 108
- see also* integrated language management
- localisation projects 12, 18–19, 28–29 *&fig*
- African Network for Localisation Project (ANLOC) 62, 69, 76, 114, 116, 121
- and evaluation/feedback of 18, 44, 103
- and ICT4d 2, 11–12, 38, 40, 43, 112, 114, 118–119, 126, 128
- Internet Initiative for Africa 47
- Leland Initiative 27, 47, 134 *n*57
- One Laptop per Child (OLPC) 40, 46, 118–119
- Pan African Localization Project (PAL) xi–xii, 1, 8, 97, 99–101, 106, 109, 111, 113–114, 116–117, 119
- and role of governments 101–102, 109, 111, 118, 121
- and telecentres 20, 41, 114, 119, 122, 124, 126
- and user skills 12, 28, 40–41, 45
- see also* sustainable localisation
- localisation, technical context 3, 7, 43, 101, 127
- bandwidths 27, 47, 79
- equipping systems 8, 10–11, 54–55, 119
- hardware and operating systems 7–8, 43, 46, 85, 88
- and internationalisation 6, 9, 11, 30, 43, 49–51, 54–55, 59, 75, 80, 83, 91, 97, 100, 108–109, 111, 114
- and web content 8–10, 46, 60, 80–82, 85, 115, 118–120, 127, 137 *n*99
- see also* internet; language technologies; localisation ecology: interrelationships; mobile/cellphone technology; software localisation
- localisers
- categories of 14, 97, 99
- skills of 15, 98–99, 114
- M
- Madagascar 40
- Mali 39, 53, 62 *tab*, 112
- macrolanguages 32–33, 74, 121, 132 *n*24
- see also* African languages: interintelligibility of
- marketing 15, 25, 28
- Microsoft Corporation 10, 42, 51, 53, 62, 65, 68, 85, 111, 120, 124
- and language interface packs 11, 85
- and OpenOffice 11, 14, 75–76, 85–89 *tab*, 101, 122
- Middle East 14, 37, 63
- mobile/cellphone technology 15, 46–47, 91, 111, 120, 123–124
- and e-mail 7, 9, 60, 79–80, 123
- Nokia 92, 138 *n*113
- Samsung 92
- Simputer 46, 91–92
- Sony Eriksson 92
- monolingualism 38, 59
- Morocco 40, 97, 99, 111
- Mozambique 34
- multilingualism and ICT 2, 11–12, 17, 34, 38, 40, 45, 49, 54, 67, 80, 91, 109–110, 114–119, 122–123, 126, 128
- code-switching in 34
- and multidialect languages 102, 112
- Multilingual Internet Names Consortium 83
- and sms/texting 46, 111
- N
- National Association for Cameroonian Language Committees (NACALCO) 110
- national information communications infrastructure (NICI) 47, 118

- Niger 34, 39, 62*tab*
 Nigeria 13, 32, 34, 61*tab*, 68, 113
 North Africa 13–14, 33, 35, 109
- O
 oral tradition and ICT 35–36, 39–40, 92, 125–126
- P
 peer networking 101, 106, 108, 115
 and e-mail forums 103–106, 113
 PLETES model *see under* integrated language
 management
 Progiciels BPI 53
- R
 Rwanda 20
- S
 second language communication (L2) 13, 33, 82,
 118
 Senegal 34
 scalability *see* ICT localisation in Africa: inter-
 relationships of factors in
 Shuttleworth, Mark 86
 SIL International 20, 32, 40, 62*tab*, 74
 sociolinguistics 6, 17, 24–25, 34, 101, 108, 127
 software localisation in Africa 12–13, 15, 20, 25,
 28–29, 33, 35, 44–45, 55, 73, 85, 95, 102, 110,
 112, 118, 122–123, 126
 early efforts 87–88
 free open software (FOSS) movement 11,
 13–14, 25, 43, 86–87, 93, 95, 100–101, 115,
 118–120, 124, 126
 graphics *tablets* 70, 123
 Open Knowledge Network 88, 93, 112, 119
 Pootle interface 87, 98
 TuxPaint 88
 Ubuntu 85–86, 88
 user interfaces 8, 45, 68, 73, 75, 88, 92, 101,
 118, 123
 see also computer-assisted translation;
 keyboards; Microsoft Corporation:
 language interface packs;
 Translate.org.za
- South Africa 13–14, 34, 39
 southern Africa 47
 speech recognition *see* language technologies,
 audio-related
 sub-Saharan Africa 12–13, 37, 51, 68–69, 109
 Sun Microsystems 86
 sustainable localisation 12, 15, 17, 19, 23, 25,
 28–29, 47, 97, 108, 120, 128–129
 and needs assessment 97–102
 Swedish International Development Agency 80
- T
 Tactical Tech 111
 Tanzania 40, 80, 82
 Tavultesoft 65, 68
 Tegic Communications 92
 TeleCommons 44
 Texin 75
 Translate.org.za 86, 88, 115
 translation *see under* language technologies
- U
 UNDP 47, 83
 UNESCO 37, 102, 110, 112
 United Nations Economic Commission for
 Africa (UNECA) African Information Society
 Initiative 47
 USAID 27
- W
 West Africa 33, 47, 52, 87
 World-Information.org 18
 World Summit on the Information Society 53, 112
 World Wide Web 7–10, 60, 79
 Semantic Web 80
 Weblogs 82
 Wikipedia 82–83
 World Wide Web Constortium 80
see also internet; localisation, technical context:
 web content
- Y
 Year of African Languages 115

AFRICAN LANGUAGES IN A DIGITAL AGE

Challenges and opportunities for
indigenous language computing

With increasing numbers of computers and diffusion of the internet around the world, localisation of the technology, and the content it carries, into the many languages people speak is becoming an ever more important area for discussion and action. Localisation, simply put, includes translation and cultural adaptation of user interfaces and software applications, as well as the creation and translation of internet content in diverse languages. It is essential in making information and communication technology more accessible to the populations of the poorer countries, increasing its relevance to their lives, needs, and aspirations, and ultimately in bridging the 'digital divide'.

Localisation is a new and growing field of inquiry. This book identifies issues, concerns, priorities and lines of research, and is intended as a baseline study in defining localisation in Africa and the ways in which it is important for long term development.

Techies, geeks, P2P experts, etc. as well as researchers and development organisations, this book is for you.

DON OSBORN is the founder of the Bisharat Language, Technology & Development Initiative and lived in worked in West Africa for over a decade while working as Associate Director for Agriculture in the Peace Corps in Niger.

ISBN 978 0 7969 2249 6



9 780796 922496

Canada



www.idrc.ca



www.hsrbpress.ac.za