Securing
a Place for a Language
in Cyberspace

By Marcel Diki-Kidiri

**United Nations
Educational, Scientific and
Cultural Organization**

Securing a Place for a Language in Cyberspace

by Marcel Diki-Kidiri
CNRS research officer,
Laboratory on "Language, Languages and
Cultures of Black Africa" (LLACAN)

The ideas, facts and opinions expressed in this publication are those of the author. They do not necessarily reflect the views of UNESCO and do not commit the Organization.

FOREWORD

L anguages are first and foremost instruments for attaining educational and cultural autonomy. They allow the transmission of knowledge from one generation to another and are a strong force in disseminating cultures and traditions between and among various ethnic groups in highly diverse geographical areas. The phenomenon of the development and disappearance of languages is part of the rise and fall of civilizations. Latin – a dead language – still exerts considerable influence on living languages and even provides the initial basis for the first standard code for computers, the ASCII code. The mother tongue is also a primary vehicle for freedom of expression.

The disappearance of languages is a phenomenon which has been present throughout History. Even in officially monolingual countries, new policies are emerging to ensure expression in endogenous languages as a human right.

According to a study undertaken by Ethnologue, Africa is the continent with the highest linguistic diversity index in the world. There is evidence suggesting that global linguistic diversity has long been in decline. Another worrying factor is that, according to some estimates, half of all languages will have disappeared by the year 2050.

Information and communication technologies (ICTs) play a key role in the linguistic transformations under way worldwide: they may provide an important vehicle for communication among the various linguistic communities. On the other hand, ICTs may be an aggravating factor in the marginalization of languages in cyberspace. There are approximately 6,000 languages in the world, but 12 languages account for 98% of Internet webpages. English, with 72% of webpages, is the dominant language,

according to a survey by O'Neill, Lavoie and Bennet in 2003.

After all, the challenge facing the international community is to overcome these tremendous obstacles in order to ensure the creation of a multilingual and culturally diverse cyberspace.

To this end, UNESCO – with the assistance of the Latin Union and the intellectual contribution of the expert Marcel Diki-Kidiri – is publishing this technical document.

It is hoped that this publication, consistent with the Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace adopted by the General Conference of UNESCO at its 32nd session, will facilitate decision-making conducive to the inclusion of new languages in cyberspace.

Abdul Waheed Khan
Assistant Director-General
for Communication and Information

TABLE OF CONTENTS

Introduction

Cyberspace is open to all languages of the world, since its infrastructure is not subject to a central authority which can decide how it should be used. Suffice it, in principle, to link a computer to an Internet access provider in order to post on the Web text, graphic or audio data in the language of one's choice. However, implementing this principle, which is a fundamental factor of democracy at the global level, requires a number of technical conditions and human and financial resources, which we shall examine in this study. In writing this article, our aim has been to give as simple as possible an answer to the following question: how to ensure that a language which is poorly endowed in linguistic and/or information technology (IT) resources, not to mention human resources, may find its proper place in cyberspace and be active there?

By extension, any language with a greater or lesser level of resources may find answers in this study if it is only weakly present in cyberspace. In what follows, we describe the lengthiest trajectory, that of a language which does not even have a written form; at the same time, other, better endowed languages will find descriptions of the stages of interest to them along the same path.

First of all, we must answer a preliminary question. What do we mean by a poorly endowed language? This refers to languages which do not have enough, or any, of the basic resources with which the world's major languages are endowed, such as stable spelling in a given writing system, reference works (grammars, dictionaries and works of literature), works for the broad public (print and audiovisual press, films, songs and music), technical and teaching materials (technical and scientific publications, textbooks and manuals), various types of everyday communications (posters, advertisements, circulars, explanatory leaflets, instructions and so on), and a sizeable number of IT applications in the language. With regard to human resources, a poorly endowed language may become an endangered language if it is only spoken by a small number of speakers. In order to save

it, the number of speakers must be increased by teaching it by all technical means possible.

Fortunately, a poorly endowed language does not necessarily have all these handicaps at once. It may be spoken as the majority language, be written, be taught at school, but still be sorely lacking in IT resources, or not even have basic linguistic resources in sufficient quantity and quality. It is therefore more exact to say that poorly endowed languages may be poorly endowed in a great variety of ways – from highly endangered languages to emerging languages which already have many such resources, but in quantities deemed insufficient or incomplete. Numerous languages on all continents correspond to this definition of poorly endowed languages. In Europe, there is Breton, Occitan and even Basque; in America, almost all Amerindian languages; in Asia, Myanmar,[1] to name just a handful among hundreds; in Oceania, practically all of the native languages of the islands of Polynesia, Micronesia and Melanesia; in Africa, where fully one third of all of the languages of the world, that is some 2,000 languages, are spoken, the best endowed languages (Afrikaans, Swahili, Hausa and so on) may be counted on the fingers of one hand and are emerging, and hence poorly endowed, languages.

We shall take as our starting point the most unfavourable hypothesis: that of a language which was once the bearer of a flourishing culture, but is spoken today by a mere handful of elderly people in a tiny village in the remote countryside, somewhere in central Africa, for example, far away from the coast. Let us call this language Ndeka. Such a language has practically no chance of gaining access one day to cyberspace because it is in danger of disappearing forever when the last of its dwindling number of speakers dies. Let us suppose that one day, a young student originally from this village believes that SPIP,[2] Wikipedia and many other tools might offer a fantastic opportunity for preserving the memory of this language, and hence of his village and the culture of his ancestors. It would even be the

---

1   New name for the language formerly known as Burmese.
2   SPIP (*Système de Publication pour l'Internet Partagé ou Participatif*): a free software content management system designed for website publishing, oriented towards online collaborative editing.

best way of teaching this language to the people of his generation, thus giving it a fresh lease on life. He comes to ask us for help with his project. With this idea in mind, our article is meant to be didactic and to accompany, step-by-step, all those who may join us at any given stage on the path to putting all poorly endowed languages, whatever their fate, into cyberspace.

1. Developing linguistic resources

1.1 From speech to writing

The objective of this first stage is to endow the language with a minimum of linguistic resources: spelling in a given writing system, a written grammar, a dictionary and as large a collection of texts as possible. It is thus the linguist's task which we hope to outline, focusing on the most indispensable intermediate objectives.

1.1.1 Collection of texts

The first thing to do is to find speakers of the language and record as many texts as possible: life stories, folktales, proverbs, songs, poems, legends, conversations, narratives and so on. In order to make these recordings, it is best to learn field research techniques and, as far as possible, to use professional-grade equipment in order to obtain the best possible sound quality because the recordings will be subjected to various operations. It would be a pity if the quality of the sound were to deteriorate when copied! Under normal working conditions, the recordings are translated with the assistance of speakers of the language and transferred to a text database.

1.1.2 Phonetic transcription

Using the International Phonetic Alphabet (IPA), linguists can transcribe any sound from any language of the world very precisely. The texts collected can thus be transcribed. It must be understood that phonetic transcription reproduces very closely each sound as pronounced at a given moment. In French, for instance, in the word *papa*, the first *p* is slightly more aspirated than the second because it comes at the beginning of the word rather than between vowels. This difference is usually ignored because it has no consequences for the meaning of the word; however, a close phonetic transcription

would reflect it as [pʻapa].[3] Phonetic transcription is indispensable for a proper analysis of the sounds of the language, but it is not at all the best way to write for everyday purposes.

1.1.3   Analysis and indication of phonemes

If the phonetic transcription is done properly, taking great care to record very faithfully the sounds actually pronounced, it should be easy to perform a phonological analysis, the purpose of which is to identify the distinctive sounds of the language known as *phonemes*, sounds which give rise to a difference in meaning when they change. For example in French, /p/[4] and /b/ are said to be different phonemes because the mere difference in pronunciation between them makes it possible to distinguish between the words *pain* "bread" and *bain* "bath".

On the other hand, some people pronounce the word *roi* "king" with a tongue-tip trilled *r*, whereas others pronounce it with an uvular ("French") *r*. These two *r*'s are phonetically distinct and transcribed respectively as [r] and [ʀ]. As this phonetic difference can never entail a difference of meaning in French, however, the two sounds represent a single phoneme, which may be noted /r/. Another example is the sequence of sounds [gz] found in the French word *exact*, pronounced [egza], and the sequence of sounds [ks] in the word *extra*, pronounced [ekstʀa], which both represent the same phoneme[5], which may be noted /x/.

The phonological analysis of a language thus makes it possible to establish a full list of the phonemes of the language and to determine how to note them. This list of phonemes can then be used to write texts. This is known as *phonological notation*, as opposed to

---

3   Square brackets [ ] are used to indicate phonetic transcription.
4   Oblique slashes // are used to indicate a phoneme or sequence of phonemes.
5   In most systems of transcription, /x/ is the *phoneme* found in the German *Bach* or the Spanish *jota*. But in some languages, it can refer to a different single sound such as [š] in Catalan *caxa* or a click in *Xhosa*. However, although phonemes are minimal discriminant units, they do not always reflect single sounds. Many of them are known (e.g. in African languages) to stand for complex articulation or sound sequences. In the case of French, the phoneme /x/ is predictably pronounced [ks] or [gz] depending on its context.

phonetic transcription. Phonological notation shows the phonemes rather than the mere sounds. Such a notation is thus much more economical and provides a sound basis for the development of a spelling system.

1.1.4   Analysis and indication of tones

A syllable may be pronounced with a relatively low, intermediate or high tone of voice (or *register*), as with the musical notes *do*, *re* and *mi*. A majority of African languages use this variation in tone to express differences of meaning with an otherwise identical sequence of phonemes. For instance, in Sängö[6], the words [ká] "over there", [kā] "wound" and [kà] "and" are three quite different words and in no sense homonyms. Such languages are known as *tone languages*.

A tone language may use as many as five different registers or tone heights, but most African tone languages use only two or three. With two tone heights, high and low tones are distinguished. With three tone heights, high, intermediate and low tones are distinguished. With more distinctions, one can add super-high and super-low tones. Tone analysis is used to determine whether or not a language is a tone language and, if it is, the number and frequency of the different tones and the best means of notation.

In tone languages, words necessarily consist of at least one vowel and a tone. Thus, a form such as *ka*, given above, only acquires meaning when associated with the specific tone with which it is to be pronounced. Logically, therefore, arrangements should be made to note the tones in the spelling system of the language. However, not all tone languages make the same use of their tones. Some use the tones not only to distinguish lexical items, but also for verb forms and to indicate person or aspect – in other words grammatical functions. In such languages, tones perform a function which is much too important to be ignored. On the other hand, in languages which use tones solely for lexical purposes and in relatively limited fashion (for

---

6   Sängö is the national language of the Central African Republic, where it shares official status with French.

example, in order to distinguish only short words, in which ambiguity may be higher), the decision has often been taken not to note the tones in everyday spelling. Decisions thus need to be made on a case-by-case basis. Even so, it should be stressed that in the case of linguistic research on tone languages, proper indication of tones is absolutely indispensable because, in such studies, account must be taken of all information pertaining to the language, and tones constitute an important and defining dimension of tone languages.

1.1.5 Development of a spelling system

Phonological and tonological studies give rise to a system of notation for the language which accurately represents the distinctive sounds of the language as economically as possible. This makes it possible to write up to 90% of what one says, without phonetic refinements, and this is sufficient to make it tempting to use phonological notation alone and not to conduct studies on spelling. Studies of spelling are indispensable, however, if a language is to be endowed with a conventional system capable of meeting all of its speakers' written expression needs. Unlike phonological notation, which only reflects the distinctive sounds of the language, a well-designed spelling system also involves the indication of grammatical relations and ideas. Consider the following French examples:

(a) /ɛl ʃãt/
(b) *elle chante* "she sings"
(c) *elles chantent* "they sing"

Example (a) gives the phonological notation that corresponds to both (b) and (c). If French were written only with phonological notation, it would not be possible to differentiate between the singular and the plural, as in the conventional spelling above.

Let us take another example in Sängö, a language which has high, intermediate and low tones. In phonological notation, these tones are noted respectively by an acute accent [á], a macron [ā], and a grave accent [à], but these accents are not available on ordinary machines in the Central African Republic. A spelling system is nevertheless designed for the general public and must be capable of being written

by hand, on typewriters and on computers. Conventional spelling and a set of rules for writing the language in the most economical way are thus essential.

The following table shows the various steps required to devise an optimum and stable spelling system for a given language, in this case, Sängö:

Phonetic transcription	Phonological notation	Spelling 1 (47%)	Spelling 2 (53%)	Translation
1. [sùkúlà]	/sùkúlà/	Sukûla	sukûla	wash
2. [sā^{n}gɔ̄]	/sāngō/	Sängö	sahngo	Sängö (language)
3. [^{m}básá^{m}báláá]	/mbásámbáláa/	Mbâsâmbâlâ	mbâssambala	seven
4. [māī^{n}gɔ̄]	/māīngō/	Mäïngö	mayhngo	change
5. [kʷà] ~ [kᵁà]	/kùà/	kua	kua	work
6. [kʷā] ~ [kᵁā]	/kūā/	küä	kwa	hair
7. [kʷá] ~ [kᵁá]	/kúá/	kûâ	kwâ	dead

It should be noted that the phonetic transcription reflects the pronunciation of the words as recorded. It is thus useful for linguistic research but too detailed to serve as the basis of an everyday spelling for the general public. The phonological notation in the second column uses only distinctive sounds, or phonemes, and is thus better suited to serve as the basis for an everyday spelling. The systematic marking of all tones is an obstacle to a practical system of writing and to rapid reading. Furthermore, diacritics – symbols used to mark the tones (acute accent, macron and grave accent) – are not available on all machines and can only be reproduced properly with a computer, which is not easily available to all users of a poorly endowed language. The spelling convention of the third column makes it possible to achieve a savings of 47% with regard to the proportion of tones that need to be indicated and to write them with the diaeresis and circumflex accent, two diacritics which are found on a dead key on most of the Western keyboards of typewriters and computers used in the Central African Republic. Even so, this convention leaves 53% of tones which need to be indicated with accents. In order to reduce still further the proportion of tones to be indicated, we introduced a spelling reform, the results of which may be seen in the fourth column. This new convention makes it possible

for 53% of tones in a given text not to be indicated. By this stage, Sängö (or Sahngo)[7] would have finally acquired a truly optimum and stable spelling system, should the reform be adopted.

1.1.6 Elaboration of reference works

Analysis of the texts collected make it possible to draw up a complete grammar covering phonology, syntax, types of utterance and discourse. Reference grammars are usually written to serve as a foundation for a number of other more pedagogical works, such as grammars for learners of the language, and in some cases, textbooks.

The study of vocabulary taken from the same texts and possibly supplemented by targeted questionnaires makes it possible to compile a reference dictionary which is as complete as possible, and which may be used as a source for wordlists or small learners' dictionaries. The same texts may be used as content for readers and short stories in order to encourage the creation of other works of literature based, if not on oral tradition, then at least on observations from everyday life in the speech community.

The production of reference works and functional materials in the language is not a marginal stage which may be omitted. Poorly endowed languages always need such works, even when they are emerging languages, and all the more so when they are endangered languages. Furthermore, such works are fundamental to strengthening and, indeed, creating community awareness among speakers, especially when only a few remain. They are not only indispensable for teaching the language to young people, and hence boosting their learning capacity, but also for making adults literate and thus opening the doors for them to more knowledge.

---

7   Here, the *h* inserted before *ng* means that the intermediate tone of the *a* carries over onto the *o*. This is the application of a more general rule which makes it possible not to use diacritics to indicate the tone in certain contexts.

1.2 Developing terminology

In order for a speech community to use its language as a means of communication in cyberspace, the language must have the necessary terms to express the realities of this new space. For example, terms such as *email*, *URL*, *certified copy*, *go online*, *connect* (machine), *download*, *post*, *networks*, *site*, *webpage*, *surf*, and so on will be found to be absolutely essential. The vocabulary needed for the computer hardware one uses is in and of itself a sizeable wordlist, which must be created from scratch, especially since in most cases, such hardware is not part of traditional cultural objects, and hence is not known and there are no words for it. This being the case, methods for developing terminology need to be put into use in the community in order to endow the language with culturally acceptable neologisms. Terminological development is a continuous, permanent activity which will be increasingly practised by the speech community itself as it develops its own growing cyberspace culture. This is not, therefore, merely a "stage" on the path to access to cyberspace, but rather an ongoing process of consolidation which needs to be launched at a given moment and pursued indefinitely.

2. Developing IT resources

2.1 Writing and spelling systems

Before considering that stage, it may be useful to review certain concepts that are all too often poorly understood, or confused, especially by non-specialists: writing systems, spelling systems (also called spelling conventions or simply orthographies), characters, character sets and fonts of characters.

2.1.1 Writing systems

Consider the following forms of writing: Latin, Arabic, Hebrew, Chinese, Ethiopian, hieroglyphics, cuneiform and Mayan. Each uses a set of specific symbols with its own combinational rules. They thus constitute different writing systems. Unlike the Americas or Europe, where the number of writing systems genuinely in use may be counted on the fingers of one hand, Africa has a good dozen or so, the most common of which are the Latin, Arabic, Ethiopic, N'Ko and Tifinagh. To write a language that has never been written down before, a writing system must be chosen to ensure that the speech community may use it immediately. In general, account is taken of the writing system that is most common in the region.

2.1.2 Spelling systems

A single writing system may be used to write quite different languages. Latin writing is thus used to write most of the languages of Western Europe, the Americas and Africa. However, each language uses the resources of the writing system differently, according to its own phonological, syntactic, enunciative and semiotic structures. The resulting spelling rules constitute a unique spelling system specific

to each language. By way of example, consider the use of consonant doubling in the following four languages:

- In Italian, gemination is important because it may entail a difference in meaning. Consonant doubling is used to note gemination and makes people pronounce the double consonant strongly: *tutto* "all", *oggi* "today".

- In French, the doubling of a consonant has merely etymological or aesthetic value and in no way entails a particular pronunciation of the consonant. Compare, for instance, the words *addition* "addition, bill", *apprécier* "appreciate", *atteler* "hitch up" with the words *adorer* "adore", *apercevoir* "perceive", *atelier* "workshop". Whether these words are written with or without double consonants, the pronunciation and meaning are the same.

- In English, even if it is not a general rule, a double consonant after a vowel often serves to vary the pronunciation of the vowel in question. Thus, /i/ is pronounced [aɪ] in *write* and *hide*, but [ɪ] in *written* and *hidden*. Doubling the /t/ and /d/ is not used to indicate a strong pronunciation of those consonants, as in Italian, but merely to vary the pronunciation of the preceding vowel /i/.

- In Sängö (or Sahngo), one of the reformed spelling rules is that a single consonant is doubled to indicate the beginning of a sequence of identical tones. Accordingly, instead of writing *täsëmä* "row", in which the intermediate tone is indicated by a diaeresis on each of the three syllables of the word, *tassema* is written. The doubling of the /s/ suffices to indicate that the intermediate tone of the first syllable is carried over onto all of the following syllables until the end of the word. Similarly, *mbâsâmbâlâ* "seven", which has a series of high tones, is rewritten *mbâssambala* under the same rule. The doubling of the /s/ indicates that the high tone of the first /â/ remains valid for all of the following syllables until the end of the word.

These four examples show how specific the spelling rules may be to each language, even when the languages use the same writing system.

2.2 Character sets and fonts

A *character* is any discrete and meaningful symbol that is part of a writing system. Note that in the Latin writing system, the space between words is a character, as are an acute accent, any letter, a comma or a digit.

2.2.1 Character sets

Once the spelling system of the language has been established, an exhaustive list of the characters needed to write the language in accordance with the system can be compiled. This list is known as a *character set*. A comparison with the standard characters used in computers will show whether all characters of the set are available or not. The easiest situation is when all characters of the set are available in the standard character set, in which case, nothing special needs to be done. Unfortunately, all too often the characters needed are not available in standard computer character sets. These are then known as *special characters*. A character set may contain, for example, seven vowels and 24 consonants, several diacritics (circumflex, diaeresis and tilde), 10 digits (from 0 to 9), mathematical operators (+, −, %, <, >, =), punctuation marks and, finally, symbols such as the asterisk (*), the paragraph mark (§) or currency symbols ($, £, €). Another character set might contain five vowels instead of seven, and 30 or 40 consonants instead of 24, and so on.

2.2.2 Fonts

A font is first of all a set of definitions of character shapes that have been designed in a single artistic style. Fonts nowadays are programmes which include, naturally, a range of character sets, but also rules for the depiction of characters on-screen and computer functions enabling them to be "managed" intelligently. The creation of a professional-quality font requires advanced knowledge and high technical skill. Fonts are therefore industrial products that are

protected under registered commercial names, such as Times New Roman, Garamond, Arial and so on. Free fonts are rarely satisfactory, even though there are some excellent ones distributed as freeware. If special characters not found in existing fonts must be created in order to write a language, an expert or a specialized company must be requested to design fonts with the required character set. Users will thus have in addition a guarantee that such fonts comply with international standards.

2.3 Character encoding

It is important to understand how characters are encoded by the computer, if only in order to be able to ask specialists the right questions when fonts are to be made with special characters for a language. Most African languages which have been only recently written down use the Latin alphabet and numerous characters from the range of characters defined by the International Phonetic Association (IPA) and adapted by the International African Institute (IAI). Only this situation, to the exclusion of other writing systems, will be examined here. In any case, the question of special characters does not arise for other writing systems, such as N'Ko, Ethiopic and Tifinagh, because all characters in those systems are standardized and encoded; any not so standardized are merely variants of the standard characters and not special characters with separate values.

2.3.1 Octet-based encoding

Imagine an electric wire on which the only way to create an event is to turn the current on and off. The value will be null (0) when the current is turned off and positive (1) when the current is turned back on. At a given point A along the wire, only those two values are possible, so that the answer to the question "is the current on?" will be "yes" or "no". Point A may thus be seen as a minimal unit of information storage, which may only have two different values: 0 or 1. This unit is represented by a binary digit called a *bit*.[8] While a single bit can be used to express two values (0, 1), two bits can express four values

8   Portmanteau word combining the beginning of *binary* and the end of *digit*: *b- + -it > bit.*

(00, 01, 10, 11) and three bits, eight values (000, 001, 010, 011, 100, 101, 110, 111).

The progression is thus exponential. The following table shows the number of values that may be expressed according to the number of bits used:

Number of bits	1	2	3	4	5	6	7	8
Number of values	2	4	8	16	32	64	128	256

If, by convention, a different letter is attributed to each of the values obtained with 3 bits, for example, the result may be:

000 = a	001 = b	010 = c	011 = d	100 = e	110 = f	111 = g

It would not be possible to encode all of the letters of the alphabet. In reality, characters have never been encoded on only three bits; rather they are first encoded on seven bits and then on eight bits. The characters encoded on seven bits are numbered from 0 to 127 and make up a character set called ASCII. This set contains the 26 basic letters of the Latin alphabet, but no letters with accents. By adding a single bit, 128 additional characters, numbered 129 to 255, have been defined. This extension made it possible to include special characters, which are necessary for printing, for graphics and for various languages using European alphabets. The character sets thus obtained each contain 256 characters encoded from 0 to 255, but are all partially different. They correspond to the ISO-8859-*n* standard (in which *n* stands for the number of a particular set). These codes are always represented in the machine by a series of bits. A group of eight bits is called an *octet*.

It is the character fonts which establish a link between the codes (series of 0 and 1) representing the characters in the computer and the corresponding shapes (called *glyphs*) which appear on the screen. In this way, the Times New Roman font reads the code 097 corresponding to "LATIN SMALL LETTER A" and shows on-screen one of the following variants of <a>: <a, **a**, *a, **a***>, according to the choice of the user. These four <a>s have in common a single style characteristic of the Times New Roman font, which has thick and thin strokes and above all the little foot, called serif, at the base of each letter. Similarly for the Arial font, which has straight stems without serifs.

2.3.2 Limits to octet-based encoding

With the simultaneous rapid increases in computer capacity, international exchanges and the needs of publishers, the limits to encoding on octets were rapidly reached. One of the main drawbacks of using this kind of encoding lies in the need for special characters. Thus, the ISO-8859-1 character set (also known as Latin-1) associates code 198 with "LATIN CAPITAL LETTER AE", which is the ligature <Æ>. However, there may be no need for this letter in a given poorly endowed language, which, on the other hand, may need a non-existent character, such as "LATIN SMALL LETTER OPEN E" or <ɛ>. By using appropriate software tools, users may redefine an encoded character. They may thus replace <Æ> with <ɛ> in a local font. Therefore, in order to read a text written in this language with this local font, the said font must be installed on the computer. However, a well-made font is an industrial product, which is often marketed and may not be distributed free of charge. Occasional users may not wish to invest in a new font whenever they use a different language on their computer. Even if fonts were free, every document sent (to a colleague, publisher or printer) would have to be accompanied by a local font that could be "discarded" after use! Such constraints are only a small part of the drawbacks of local fonts that nevertheless serve their users well when shared locally. In any case, with a maximum of 256 possible codes, encoding with a single octet is clearly insufficient for the 2,000 basic characters of Chinese, and this holds true for Japanese and Korean.

ISO 8859-1 Latin-1

0	32		64	@	96	`	128	160		192	À	224	à
1	33	!	65	A	97	a	129	161	¡	193	Á	225	á
2	34	"	66	B	98	b	130	162	¢	194	Â	226	â
3	35	#	67	C	99	c	131	163	£	195	Ã	227	ã
4	36	$	68	D	100	d	132	164	¤	196	Ä	228	ä
5	37	%	69	E	101	e	133	165	¥	197	Å	229	å
6	38	&	70	F	102	f	134	166	¦	198	Æ	230	æ
7	39	'	71	G	103	g	135	167	§	199	Ç	231	ç
8	40	(	72	H	104	h	136	168	¨	200	È	232	è
9	41	)	73	I	105	i	137	169	©	201	É	233	é
10	42	*	74	J	106	j	138	170	ª	202	Ê	234	ê
11	43	+	75	K	107	k	139	171	«	203	Ë	235	ë
12	44	,	76	L	108	l	140	172	¬	204	Ì	236	ì
13	45	–	77	M	109	m	141	173	-	205	Í	237	í
14	46	.	78	N	110	n	142	174	®	206	Î	238	î
15	47	/	79	O	111	o	143	175	-	207	Ï	239	ï
16	48	0	80	P	112	p	144	176	°	208	Ð	240	ð
17	49	1	81	Q	113	q	145	177	±	209	Ñ	241	ñ
18	50	2	82	R	114	r	146	178	²	210	Ò	242	ò
19	51	3	83	S	115	s	147	179	³	211	Ó	243	ó
20	52	4	84	T	116	t	148	180	´	212	Ô	244	ô
21	53	5	85	U	117	u	149	181	µ	213	Õ	245	õ
22	54	6	86	V	118	v	150	182	¶	214	Ö	246	ö
23	55	7	87	W	119	w	151	183	·	215	×	247	÷
24	56	8	88	X	120	x	152	184	¸	216	Ø	248	ø
25	57	9	89	Y	121	y	153	185	¹	217	Ù	249	ù
26	58	:	90	Z	122	z	154	186	º	218	Ú	250	ú
27	59	;	91	[	123	{	155	187	»	219	Û	251	û
28	60	<	92	\	124			156	188	¼	220	Ü	252	ü
29	61	=	93	]	125	}	157	189	½	221	Ý	253	ý
30	62	>	94	^	126	~	158	190	¾	222	Þ	254	þ
31	63	?	95	_	127		159	191	¿	223	ß	255	ÿ

2.3.3 Encoding with multiple octets

In order to overcome the limitations of single-octet encoding, the Unicode Consortium and the International Standards Organization (ISO) have developed the ISO/IEC-10646 standard, which uses two octets to encode a set of 65,536 characters, known as the UCS-Unicode (Universal Character Set), with enough room to handle all the writing systems of the world! This is ideal for Chinese ideograms, but for Latin characters, which had hitherto been satisfactorily handled by single-octet encoding, two-octet encoding offers much more capacity than is really necessary. In order to ensure compatibility with existing documents, several encoding formats have been proposed for those characters encoded with two octets. The most popular is the UTF-8 encoding format (Unicode Transformation Format).

However, the conversion is usually effected automatically by the application software, rarely by the user. It is important to understand that in the Unicode standard each character may be defined only once, such that a single digital code may only refer to a single character definition and vice versa.

2.4 Primary basic resources

2.4.1 Character sets and fonts

The first IT resource that should be developed for a language which has none is the definition of a character set on the basis of a list of characters needed for the spelling system of the language. If the character set needed contains no special characters, any ordinary software may be used immediately. However, if special characters are needed, before proceeding to create a special font, it should be ascertained whether those characters are not already included in special fonts available over the Internet, such as Lucida Sans Unicode, Gentium and Doulos SIL[9] or the African fonts of Progiciels BPI[10]. If they are not available, a character set will have to be defined.

---

9   See the SIL site: http://scripts.sil.org/SILFontList.

10   See the Progiciels BPI site http://www.progiciels-bpi.ca/tcao/apercu.html.

It is highly advisable to ascertain whether that set could also be used for other languages of the region in order to secure a better return on the investment made and ensure that multilingual texts can be produced in future without having to change fonts with each language. Two good examples of the definition of single-octet (but Unicode-compatible) character sets are provided by Progiciels BPI. Each of these two character sets (AFRFUL and AFRLIN) covers several languages.[11]

Once the character set has been defined, the corresponding font must be created. Nowadays, experts strongly advise against the creation of single-octet fonts because of the limitations referred to above. It is best to commission a specialist to create a font which contains the desired characters and complies with the Unicode standard. Such recourse is all the more indispensable if the language uses a non-Latin writing system that is not yet represented in Unicode! The http://www.freelang.com/ site offers visitors a free font creation service. The font created is free of charge and shared on the site. A number of programmes are available on the Internet for the creation of local fonts, in particular on the following sites:

1. http://scripts.sil.org/TypeDesignResources (ready-to-use fonts and tools for creating them);

2. http://scripts.sil.org/SILFontList (ready-to-use fonts for downloading);

3. http://scripts.sil.org/SILEncore_Glyphs (for visualizing font glyphs).

2.4.2 Virtual keyboards

Once a font containing the selected set of characters has been obtained, it may be installed on the computer in accordance with the font installation procedure required for the operating system. It is then easy to use these fonts, since most modern word processing

---

11 See also the Burkina Faso site: http://www.abcburkina.net/sedelan/index.htm for the languages of Burkina Faso.

programmes have a function allowing the insertion of special characters. For instance, in Word, a window of glyphs (characters) opens, and suffice it to click on the character needed to insert it into the text. Unfortunately, this procedure is somewhat fastidious and considerably slows down the typing of a text in a poorly endowed language. It is therefore indispensable to design more practical key sequences which give more fluid access to the special characters of a given font.

It is interesting to note that the dominant languages of the industrialized countries each have a keyboard which has been specially created for them, and even for regional variants of the same language. For instance, the French Canada (Quebec) keyboard is not the same as the French France keyboard. Not only do they differ with regard to the position of the letters A, Z, Q and W, but they are also quite independent of each other concerning the layout of punctuation marks and, naturally, special characters (diacritics, accented letters, currency symbols and so on). In Africa, most of the major African languages (Hausa, Fulfulde and so on) do not have their own keyboards. There is little hope of a physical keyboard being designed for poorly endowed languages anytime soon. The simplest solution is thus to design a virtual keyboard.

On the basis of the physical keyboard used in the region in which the poorly endowed language is spoken, the first step is to draw up a table of easy-to-remember keystroke sequences. For example, one may wish to type the sequence "<" + "O" (ignoring the quote marks and the plus sign) to produce a SMALL LETTER OPEN O <ɔ> and the sequence "SHIFT" + ">" + "O" to get the corresponding upper-case <Ɔ>, as in the following table:

"<" + "O" = <ɔ>

"SHIFT" + ">" + "O" = <Ɔ>

These sequences are ergonomic only if the LESS-THAN and GREATER-THAN signs "<" and ">" are on the same key, the latter above the former. They may then be used as dead keys. In order to use them to produce "<" and ">", suffice it to strike the space-bar

after typing one of them. The following lines should thus be added to the table:

"<" + "SPACE" = "<"

"SHIFT" + ">" + "SPACE" = ">"

The characters "<" and ">" may then be used as general modifiers (dead keys) to type other special characters in lower- and upper-case. It is more complicated to put one or more diacritics on a character that has already been modified, for example an acute accent to indicate a high tone and a tilde to indicate nasality. It must be ascertained that the software used supports multi-step commands such as "ACUTE" + "TILDE" + "DEAD-KEY" + "BASE CHARACTER" = "SPECIAL CHARACTER + ACUTE + TILDE". One of the best software programmes for creating virtual keyboards is currently Tavultesoft KeymanTM. Virtual keyboards for African languages created with Keyman are available at: http://scripts.sil. org/SILKeyboards.

Keyman can be installed on the computer in order to manage keyboards and corresponding fonts. It is possible to install several keyboards and to move from one to another by means of a simple keystroke sequence, such as "ALT" + "K". It is necessary to check that the keystroke combination chosen is not already used for something else by one of the programmes on the computer.

2.4.3 Computer-assisted text analysis programmes

When special characters are being used, basic tools used for alphabetic sorting, character encoding conversion, text alignment and other operations should be borne in mind. Such tools will prove very useful for manipulating and processing texts to be produced in the poorly endowed language for posting on the Web. The most effective tools for language processing are produced by Progiciels BPI and are listed below.

Table 1 below presents the following suite of five computer-aided corpus processing (CACP) programmes: Alibi, Concorde, Recode, Ventile, Vocable.[12]

Programme	TABLE 1 – Description of CACP programmes
Recode	*Recode* is an automatic conversion programme for character sets. It can process and re-encode some 175 different character sets and a dozen surfaces of files. Depending on *the* pair of character sets defined in the call command (from a source character set to a target character set), Recode re-encodes the input files. Since each character set may be converted to most of the 174 others, tens of thousands of different conversions are possible.
Ventile	Ventile is a programme that produces text statistics. With regard to text elements, Ventile counts the number of paragraphs, sentences, words and characters per file. With regard to text statistics, Ventile gauges absolute frequencies and provides three separate measures of key trends (mode, median and arithmetical average) and five measures of dispersion (minimum, maximum, quartile deviation, average deviation and standard deviation). The statistical results are displayed digitally in the form of a table and graphically as a histogram.
Vocable	Vocable is a programme that produces vocabulary lists. This tool is used to process texts and compile regularly updated vocabulary lists. These lists may be sorted in regular alphabetical order (from left to right), in inverse alphabetical order (from right to left) or by order of decreasing frequency. Each word may show an indication of its frequency of occurrence and a list of all text references.

12   Source: Progiciels BPI (2005) http://www.progiciels-bpi.ca/tcao/apercu.html.

Concorde	Concorde is a programme that produces context-based word concordances. This tool goes through text files and produces an alphabetical list of features with their immediate context. Features of a concordance may be either words in sentence-context or characters in word-context. Feature words may be sorted either in regular alphabetical order or in inverse alphabetical order.
Alibi	Alibi is a bi-textual (or bilingual) alignment programme. This tool goes through two files in tandem and automatically aligns the text constituents taken from the pair of files. Examples of text constituents that Alibi can align are paragraphs, sentences or words from the pair of texts.

Such software cannot be designed by amateurs. These programmes have proven to be highly effective and have the advantage of performing equally well under the most popular operating systems (Windows and Linux). In association with character sets, fonts and virtual keyboards, these programmes constitute basic resources for the processing of a poorly endowed language.

2.5 Software localization

Software localization involves adapting a programme designed in a particular language in a given country to the language and culture of another country. Each programme is made up of two parts: the computer code and the human-machine interface. There is never any question of touching the code except in exceptional cases, such as a need to change the direction of the writing or the order of alphabetic sorting when no arrangements have been made for these to be localizable. Most programmes today are designed to be localized and the human-machine interface is thus easily identifiable and accessible, either by means of a file provided for this purpose or through appropriate tools.

Elements of a programme which need to be localized are menus (scroll-down, fixed, floating or contextual), dialogue boxes, notifications, the online help function, navigation buttons and commands, graphics (logos, flags, coats-of-arms, illustrations, site appearance or "skin"

and design), units of measurement (distance, weight, volume, time, currency, density and so on, all to be converted to regional standards), tutorials and documentation (installation, user and reference manuals). The objective of localization is naturally to enable users to work in a culturally and linguistically familiar computer environment, which is thus easy to master. The users' language thus becomes a working language of computer use. Software localization in a poorly endowed language thus helps to enhance the prestige of that language in the eyes of users and, in particular, of its own speakers.

If software programmes are to be translated into a poorly endowed language, persons who are not software proprietors should localize them as freeware, such as programmes distributed under a General Public Licence (GPL).[13] The distribution of GNU/Linux (Ubuntu Linux,[14] for example, for African countries) offers the world an online collaborative platform known as Rosetta,[15] which allows anyone to choose a freeware programme and translate it into the desired language. A set of project management tools, provided under the name Launchpad,[16] constitutes a precious aid to users wishing to localize programmes. Such a facility offered free of charge to all the languages in the world is indeed an extremely rare and generous initiative – this is a genuine windfall for poorly endowed languages, which may harness it to gain easier access to the status of a working language of cyberspace. However, software localization is not an indispensable stage for access in cyberspace. A poorly endowed language may gain access to cyberspace without any such localization projects. As long as the language has the primary basic resources at its disposal (cf. section 2.4), work can begin on developing cultural and linguistic content for cyberspace.

---

13 For further information on freeware licences, see http://www.gnu.org/licenses/license-list.fr.html.
14 See http://www.ubuntu.com.
15 See https://launchpad.net/rosetta.
16 See http://launchpad.net.

3. Developing cultural content

3.1 Cultural content

Cultural content may be based on text, sound or image. All such content may be digitized and posted on the Web or shared through cyberspace.

3.1.1 Text

In addition to the corpus used to describe the language and create its first works of reference, it is a good idea to collect or create even more texts, of all kinds, from personal advertisements to legendary epics, translations of works of literature and newspaper articles. These texts will subsequently undergo various types of processing by means of the computer-aided corpus processing (CACP) programmes. It would be useful to disseminate the results of such processing on the Web because they may help to enhance the study of the language itself.

3.1.2 Sound

Audio documents may consist of recordings of lists of words or spoken phrases, songs, recitals, folktales, legends, interviews and media reports, not to mention instrumental music and sounds of everyday life and nature which are representative of the local cultural environment. Digitization of these audio documents requires professional-quality recording equipment, which may be very expensive for someone from the Third World who speaks a poorly endowed language. Works of music, in particular, may require costly equipment. The same holds for nature recordings of birds singing or the chirping of insects at night with a view to studying the local environment and ecology. If such documents on the environment contain a running commentary in the local language, the value of that language will be enhanced.

Background noise and other parasitic sounds must be filtered out of all audio files thus recorded. Lengthy studio work is thus required to segment and calibrate each portion of the sound recording in order to make the files lighter and more easily transmissible. Audio documents can now be broadcast over the Internet and in cyberspace, as are radio programmes or films, but without pictures, such audio documents, however interesting they may be, will be rather like old-fashioned black-and-white television in the heyday of colour.

3.1.3 Image

Iconographic documents bring together drawings, logos, photographs and videos. While drawing software is extremely common, still and motion picture cameras are somewhat expensive. All the same, they are indispensable for filming dances, ceremonies, everyday scenes, panoramic vistas, documentaries and so on. These images may be edited and retouched in order to improve their quality. Image files are nowadays digital as soon as they are created. Nevertheless, it is imperative to ensure that the image quality is excellent, because some processes require several copies, which may slightly impair the quality of the image to be disseminated.

Whenever possible, it is preferable to combine sound, image and text in the production of documents for dissemination on the Internet. However, if a poorly endowed language has not yet been reduced to writing, it is always possible to create sound-and-image documents, such as films and/or photographs with audio commentary, and post them on the Web.

3.2 Ensuring access to cyberspace

3.2.1 Internet sites

Internet sites are the best way to boost the presence of a language in cyberspace, because they are infinitely variable, ranging from simple personal pages to multimedia portals offering a broad range of different services. Before embarking upon the creation of a site, it would be useful to bear in mind the following information.

A site written in a poorly endowed language can only be read by speakers of that language. It may not be very easy if the speakers cannot read or write their own language. The site may encourage them to learn to do so. When a spelling system has only just been devised for a poorly endowed language, it is preferable to plan to create a bilingual site using both the poorly endowed language and a language more widely spoken in the region.[17] In such case, care must be taken to ensure that the poorly endowed language is used for the navigation buttons and commands and in all messages for visitors to the site, so that it will not become a mere curio and lose its status as a working language on the Web.

Navigation buttons, command menus, titles, logos, decorative texts, animations and site presentation styles are all elements that must be made ready beforehand, using specialized programmes such as Button StudioTM. Text files to be posted on the Web are written in HTML (HyperText Markup Language). To write a page on the Web, a simple word processing programme such as Open Office (freeware), which saves files in HTML format may be used. There are also numerous other tools which are better geared to the creation of sites, from the simplest, often free of charge, such as NetscapeTM Composer, to more sophisticated commercial ones like DreamweaverTM.

Webpage design is a matter of taste and culture.

3.2.2 Email and instant messaging

These are the most widely used communication tools on the Internet. However, for security reasons, it is very difficult to gauge the rate of use of a given language in global email exchanges. The use of a poorly endowed language in this utility may well go unnoticed. On the other hand, localization of a least one email or instant messaging programme constitutes a real contribution to the equipping of the poorly endowed language.

---

17   See for example the YSB Sahngo site: http://sango.free.fr.

3.2.3 IP telephony and mobile phones

Internet Protocol (IP) telephony enables long-distance voice telephone calls to be made in real time over the Internet. This technology may thus be freely used by speakers of a poorly endowed language. They merely need to download a suitable programme, such as SkypeTM, currently the most popular, or Google Talk, VoipBuster or Messenger, and install it on their computer. The only potential resource for a poorly endowed language is localization of the software's user interface. The same is true for mobile phones – which can help to equip poorly endowed languages only by having their user interface localized, which has been done in South Africa.

3.2.4 Forums and mailing lists

The creation of a forum or a mailing list on a site enables the members of a speaker community of a poorly endowed language to dialogue with each other in writing. Provided the software environment supports the necessary special characters, these facilities may prove to be an excellent means of ensuring that the poorly endowed language thrives in cyberspace. Without a community that engages in active exchanges in it, the language risks being relegated to the status of an inert decorative element in cyberspace. Surely this is not the ultimate purpose of efforts to ensure that a poorly endowed language may take its rightful place in cyberspace. Numerous commercial and non-commercial portals offer facilities for creating community forums free of charge. Some of the most popular forums are:

- http://www.yahoogroups.com

- http://www.google.com

- http://www.ubuntu.com

- http://www.free.fr

- http://lists.kabissa.org/mailman/options/a12n-entraide/ mdkidiri%40free.fr.

4. Developing the user community

4.1 Creation of a user community

The existence of a user community literate in its own language is absolutely necessary to ensure that a language stays alive when it eventually gains access to cyberspace. Indeed, if efforts are made to place poorly endowed languages into cyberspace, this is not so that they should function as mere accessories or extras, but that they may live and work productively there for their speakers, who are directly concerned by the mastery and promotion of their languages in cyberspace. It is therefore vital to teach the speakers of a poorly endowed language, in particular young people, to read and write their language. Introductory courses on the use of computers and office and communication software should be accompanied by teaching of the poorly endowed language, for example as part of community-based activities, if not at school. Local associations are generally in a better position to provide this sort of training than schools, which depend heavily on the national education system. Unfortunately, local associations all too often have little money and no support, especially when it comes to poorly endowed languages. Naturally, any support given to local associations involved in the computer-related development of languages helps to promote linguistic and cultural diversity in cyberspace.

4.1.1 Regenerating the community

When a poorly endowed language has only a few speakers left, it is absolutely indispensable to increase the number of speakers by teaching the language to young people by all possible means: cultural activities, reading and cultural centres, local association activities, and, if possible, at school. A living environment must be created anew for the language. The more the language is taught and its prestige enhanced, the better it will be able to produce cultural

content that can be posted on a website. It should be remembered that that requires a tremendous amount of work, considerable investment of both financial and human resources, and an unfailing determination in the long term. Cultural associations may be very active at the local level to rehabilitate an endangered language and ensure growing interest in it, especially among administrative and political decision-makers.

4.1.2 Consolidating the community

Take the example of a poorly endowed language which nevertheless has a sizeable number of speakers. Educating people in their own language is a long-term undertaking which requires mobilizing numerous human and financial resources. In this case, ensuring that the language has access to cyberspace is one of the ways of mobilizing and dynamizing the speaker community of that language. Consolidating the speaker community must not therefore be merely a preliminary activity to ensuring the language's access to cyberspace, but rather must be an ongoing activity that will keep the language alive in cyberspace. Here, much more than in the case of endangered languages which do not have such sizeable numbers of speakers, it is necessary to involve political and administrative decision-makers in the process of boosting the prestige of the poorly endowed language. Otherwise, the efforts of local associations will not be sufficient to ensure its development on a sustainable basis, owing to a lack of substantial means over the long term.

4.2   Political and sociocultural aspects

4.2.1 Convincing arguments

As explained above, an association's efforts to develop a speaker community for a poorly endowed language is limited by the level of funding that it can mobilize. It is therefore essential to sensitize political leaders at all levels to the advantages to be derived from supporting efforts to make poorly endowed languages functional and to build the capacities of their speakers. With the support of local administrations, it is easier to envisage the teaching of the languages

at school, especially primary school, where the local language can best be used as the language in which children first acquire knowledge. Bilingual education can succeed, even locally, only if it is provided within the legal framework of the national education system. A number of experiments in bilingual education involving African languages in Burkina Faso, Senegal, Mali, Burundi and Rwanda have shown that children learn better in their mother tongue and obtain better school results. In the longer term, they become better educated citizens. Well-educated individuals are much more likely to be able to have a global vision and a deeper understanding of the complex problems of the modern world and to help to find sustainable solutions to them in their walk of life and field of activities. Such people are more apt to complete socio-economic projects, whether for themselves or for their families, villages or regions. Multilingual education involving the mother tongue is the key to economic development.

In addition to the educational advantage, policies conducive to linguistic diversity may generate jobs in translation, the production of textbooks, publishing and the management of local associations. Linguistic diversity is thus clearly linked to economic potential, and the development of poorly endowed languages is a contributing factor. When a poorly endowed language is spoken throughout a region or country, the whole national language policy is concerned by its development. Furthermore, such a policy must respect linguistic rights, which are an integral part of basic human rights.

4.2.2 Authoritative instruments

Beyond national political authorities, support should also be sought from international organizations working to promote and enhance linguistic diversity worldwide, which presupposes at the very least a favourable attitude to the efforts of local associations to develop poorly endowed languages. In the last 40 years, international organizations have held numerous meetings on the languages of local populations, often called "national languages" in opposition to the European languages inherited from colonialism, especially in Africa and the Americas. These meetings have resulted in the drafting and adoption of a number of "international instruments" which may be invoked in support of concrete activities. Among the various events held by

States and international organizations, Amadou Touré[18] notes, *inter alia*, for Africa:

1. UNESCO meeting on "The Use of Vernacular Languages in Education" (Monographs on Fundamental Education), Paris, 1953;
2. UNESCO Regional Conference on the Planning and Organization of Literacy Programmes in Africa, Abidjan, March 1964;
3. Congress of the West African Linguistics Society (WALS), Accra, 1965;
4. Meeting of a group of experts organized by UNESCO for the unification of the alphabets of national languages: Fulfulde, Hausa, Kanuri, Mandingue, Songhay-Zarma and Tamasheq, Bamako, 28 February to 5 March 1966;
5. Intergovernmental Conference on Cultural Policies in Africa, UNESCO-Organization of African Unity (OAU), Accra, 1975;
6. Meeting organized by the Agency for Cultural and Technical Cooperation (ACCT) on the promotion of national languages, Yaoundé, 1977;
7. ACCT International Conference on "Linguistic research, employment and the teaching of languages in Africa: ways of strengthening cooperation between States", Yaoundé, 1983;
8. Meeting of experts on "Definition of a strategy for the promotion of African languages", Conakry, 1984;
9. UNESCO meeting of experts on "Promotion of African languages as instruments of culture and continuing education", Yaoundé;
10. Meeting on the draft charter for the promotion and use of African languages in Education, Accra, August 1996;
11. Intergovernmental Conference on African Language Policies (UNESCO-OAU-ACCT), Harare, 1997;
12. African consultation on the establishment of the African Academy of Languages (ACALAN), Bamako, 25 to 27 May 2001;

---

18   Working document for the symposia of the African Academy of Languages 2006.

13. Conference on "Multilingualism for Cultural Diversity and Participation of All in Cyberspace", Bamako, May 2005;

14. 33rd session of UNESCO's General Conference, which adopted a draft resolution initiated by the African Academy of Languages (ACALAN) and submitted by Benin concerning the holding of five regional conferences on cross-border languages and lesser-used languages in Africa, Paris, October 2005.

A number of major gatherings have led to the drafting of standard-setting instruments and regulatory texts, such as:

1. Organization of African Unity (OAU) Charter, 1963;
2. Pan-African Cultural Manifesto of Algiers, 1969;
3. Cultural Charter for Africa, OAU, Port Louis, 1976;
4. Results of the First Conference of Ministers of Culture of OAU, Port Louis, 1986;
5. Lagos Plan of Action for the Economic Development of Africa, 1980;
6. Language Plan of Action for Africa, Addis Ababa, 21-25 July 1986;
7. OAU ten-year plan on languages and oral traditions, 1987;
8. Treaty establishing the African Economic Community (AEC), Abuja, 1991;
9. Regional plan for collecting oral traditions in southern Africa, Harare, 1993;
10. Linguistic development plan in the French-speaking area 1990-2000, ACCT, 1993;
11. Programme of Action for the Decade of Education in Africa, Harare, 1999;
12. Treaty establishing the African Union (AU), Lomé, 2000;
13. 31 C/Resolution 11 on the African Academy of Languages, adopted by the General Conference of UNESCO at its 31st session, 2001;
14. Decision CM/Dec. 613 (LXXIV) of the 37th summit of OAU Heads of State and Government establishing the African Academy of Languages, July 2001;
15. UNESCO Universal Declaration on Cultural Diversity adopted by the General Conference of UNESCO at its 31st session, 2001;

16. Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace adopted by the UNESCO General Conference at its 32nd session, October 2003;

17. Final documents of the World Summit on the Information Society: Geneva Declaration of Principles and Geneva Plan of Action, Geneva, 2003; and Tunis Commitment and Tunis Agenda for the Information Society, Tunis, 2005;

18. Decision AU/Dec. 92 (VI) on the Second Decade of Education for Africa (2006-2015) and the Framework of the Plan of Action for the Second Decade (EX/CL/224 (VIII) Rev.2) of the Sixth Ordinary Session of the Assembly of the African Union, Khartoum, 2006;

19. Decision AU/Dec. 94 (VI) on the Revised Charter for the Cultural Renaissance of Africa of the Sixth Ordinary Session of the Assembly of the African Union, Khartoum, 2006;

20. Decision AU/Dec. 95 (VI) on the Statutes of the African Academy of Languages of the Sixth Ordinary Session of the Assembly of the African Union, Khartoum, 2006;

21. Decision AU/Dec. 96 (VI) on the linkage between culture and education of the Sixth Ordinary Session of the Assembly of the African Union, Khartoum, 2006;

22. Decision AU/Dec. 98 (VI) to declare 2006 as the Year of African Languages of the Sixth Ordinary Session of the Assembly of the African Union, Khartoum, 2006.

A number of institutions have been established, including:

– African Cultural Institute (ACI), Dakar, Senegal;
– Centre for Linguistic and Historical Studies through Oral Traditions (CELHTO), Niamey, Niger;
– Centre for Research and Documentation on Oral Traditions and African Languages (CERDOTOLA), Yaoundé, Cameroon;
– Eastern African Centre for Research on Oral Traditions and African National Languages (EACROTANAL), Zanzibar, United Republic of Tanzania;
– International Centre for Bantu Civilizations (CICIBA), Libreville, Gabon;

–   African Bureau of Educational Sciences (ABES), now the Pan-African Institute of Education for Development (IPED), Kinshasa, Democratic Republic of the Congo.

Most of these institutions, it must be said, have fallen short of expectations, owing to insufficient funding with which to operate to capacity. It is therefore not so easy to win their support for local activities designed to enhance the prestige of a poorly endowed language. However, with a little perseverance and a great deal of good will, they will be won over to the cause, which is no mean achievement! Indeed, an inventory of existing standard-setting instruments and an examination of the reports of the various meetings held over this long period show that all the countries of Africa, at one time or another, and to varying degrees, have taken initiatives to promote their national languages. Every effort should therefore be made to encourage them to become even more actively involved in concrete action to promote linguistic diversity. Only through strong and enduring political will can the necessary funds be secured to ensure the provision of multilingual education and the development and use of a nation's languages in all areas of life so that cultural, economic and social goods may be generated for the well-being of all.

5. Conclusion

In order to promote and bolster linguistic and cultural diversity in cyberspace, the most underprivileged languages need help to gain access to it. If it is possible to do this with a small, oral, unwritten, endangered language, there is all the more reason why this should be possible with all poorly endowed languages which are in somewhat better circumstances.

The first stage consists in undertaking the necessary studies in order to develop the linguistic resources that are indispensable: a list of phonemes, an alphabet, a spelling system, a grammar, a dictionary and a collection of texts.

The second stage involves work on computerization of the language in order to identify or develop compatible IT resources: a character set in at least one font, a virtual keyboard and corpus processing programmes, which may also be used to fine-tool linguistic analysis of the language and enhance its linguistic resources.

The third stage consists in developing and adapting cultural resources so that they may be shared in cyberspace. This means recording and digitizing as many text, sound and graphic records as possible and making them ready for posting on websites. It is also necessary to design the various ingredients of a website, such as menus, navigation bars, titles and other texts for human-machine communication. In some cases, it will be necessary to localize programmes in order to develop the language as a working tool and endow it with supplementary IT resources.

Finally, it is useful to learn to develop websites in the poorly endowed language, possibly in tandem with a more widely used language. All tools necessary for such training and tools for creating forums and localizing freeware may be found on the Internet. Once it has a website, a forum, a mailing list, IP telephony, music, still photographs and video, the lesser-used language can now be well ensconced in cyberspace, but to survive there, a community capable of using

it intensively must be developed. Assistance to local associations in developing such communities will contribute to the promotion and enhancement of the diversity of languages and cultures in cyberspace.

References

Unicode Consortium: http://www.unicode.org/fr/charts/

Online creation of fonts: http://www.freelang.com

Platform for the translation and localization of freeware: https://launchpad.net/rosetta

List of fonts: http://scripts.sil.org/SILFontList

Progiciels BPI (2005): http://www.progiciels-bpi.ca/tcao/apercu.html

SEDELAN (*Service d'édition en langues nationales de Koudougou*, Burkina Faso): http://www.abcburkina.net/sedelan/index.htm

Bilingual Sahngo-French site of the YSB SAHNGO association for the promotion of Sahngo: http://sango.free.fr

University of Nice – multimedia resources, training in website creation: http://www.unice.fr/urfist/ResInternet.html

Annex 1: AFRFUL character set[19]

The African AFRFUL character set covers the alphabets of Bambara, Ewondo, French and Fulfulde. The AFRFUL-102-BPI_OCIL character set is a set of encoded characters used for on-screen display of the following African languages: Bambara, Ewondo and Fulfulde (Peuhl). Acceptable alternate names for this set of encoded characters are afrful102bpiocil, bambara, ewondo, fulfulde and bra.

The AFRFUL-103-BPI_OCIL character set is a set of transliterated characters used for keyboard encoding of the following African languages: Bambara, Ewondo and Fulfulde (Peuhl). Acceptable alternate names for this set of encoded characters are afrful103bpiocil, tbambara, tewondo, tfulfulde and tbra. For each of the 255 characters presented in the list below, the following information is given in the respective columns:

Dec	decimal value of the character
Oct	octal value of the character
Hex	hexidecimal value of the character
Mne	two-letter mnemotechnical code of the RFC *1345 standard*
UCS2	UCS value of the character
Kbd	keyboard encoding convention for the character
AFRFUL-102-BPI_OCIL	name of the character according to the ISO *10646 standard*

19   Source: Progiciels BPI http://www.progiciels-bpi.ca/tcao/apercu.html.

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
0	000	00	NU	0000	nul	null
1	001	01	SH	0001	soh	start of heading
2	002	02	SX	0002	stx	start of text
3	003	03	EX	0003	etx	end of text
4	004	04	ET	0004	eot	end of transmission
5	005	05	EQ	0005	enq	enquiry
6	006	06	AK	0006	ack	acknowledge
7	007	07	BL	0007	bel	bell
8	010	08	BS	0008	bs	backspace
9	011	09	HT	0009	ht	character tabulation
10	012	0a	LF	000A	lf	line feed
11	013	0b	VT	000B	vt	vertical tabulation
12	014	0c	FF	000C	ff	form feed
13	015	0d	CR	000D	cr	carriage return
14	016	0e	SO	000E	so	shift out
15	017	0f	SI	000F	si	shift in
16	020	10	DL	0010	dle	data link escape
17	021	11	D1	0011	dc1	device control one
18	022	12	D2	0012	dc2	device control two
19	023	13	D3	0013	dc3	device control three
20	024	14	D4	0014	dc4	device control four
21	025	15	NK	0015	nak	negative acknowledge
22	026	16	SY	0016	syn	synchronous idle
23	027	17	EB	0017	etb	end of transmission block
24	030	18	CN	0018	can	cancel
25	031	19	EM	0019	em	end of medium
26	032	1a	SB	001A	sub	substitute
27	033	1b	EC	001B	esc	escape

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
28	034	1c	FS	001C	is4	file separator
29	035	1d	GS	001D	is3	group separator
30	036	1e	RS	001E	is2	record separator
31	037	1f	US	001F	is1	unit separator
32	040	20	SP	0020	< >	space
33	041	21	!	0021	!	exclamation mark
34	042	22	"	0022	"	quotation mark
35	043	23	Nb	0023	#	number sign
36	044	24	DO	0024	$	dollar sign
37	045	25	%	0025	%	percent sign
38	046	26	&	0026	&	ampersand
39	047	27	'	0027	'	apostrophe
40	050	28	(	0028	(	left parenthesis
41	051	29	)	0029	)	right parenthesis
42	052	2a	*	002A	*	asterisk
43	053	2b	+	002B	+	plus sign
44	054	2c	,	002C	,	comma
45	055	2d	-	002D	-	hyphen-minus
46	056	2e	.	002E	.	full stop
47	057	2f	/	002F	/	solidus
48	060	30	0	0030	0	digit zero
49	061	31	1	0031	1	digit one
50	062	32	2	0032	2	digit two
51	063	33	3	0033	3	digit three
52	064	34	4	0034	4	digit four
53	065	35	5	0035	5	digit five
54	066	36	6	0036	6	digit six
55	067	37	7	0037	7	digit seven
56	070	38	8	0038	8	digit eight

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
57	071	39	9	0039	9	digit nine
58	072	3a	:	003A	:	colon
59	073	3b	;	003B	;	semicolon
60	074	3c	<	003C	<	less-than sign
61	075	3d	=	003D	=	equals sign
62	076	3e	>	003E	>	greater-than sign
63	077	3f	?	003F	?	question mark
64	100	40	At	0040	@	commercial at
65	101	41	A	0041	A	latin capital letter a
66	102	42	B	0042	B	latin capital letter b
67	103	43	C	0043	C	latin capital letter c
68	104	44	D	0044	D	latin capital letter d
69	105	45	E	0045	E	latin capital letter e
70	106	46	F	0046	F	latin capital letter f
71	107	47	G	0047	G	latin capital letter g
72	110	48	H	0048	H	latin capital letter h
73	111	49	I	0049	I	latin capital letter i
74	112	4a	J	004A	J	latin capital letter j
75	113	4b	K	004B	K	latin capital letter k
76	114	4c	L	004C	L	latin capital letter l
77	115	4d	M	004D	M	latin capital letter m
78	116	4e	N	004E	N	latin capital letter n
79	117	4f	O	004F	O	latin capital letter o
80	120	50	P	0050	P	latin capital letter p
81	121	51	Q	0051	Q	latin capital letter q
82	122	52	R	0052	R	latin capital letter r
83	123	53	S	0053	S	latin capital letter s
84	124	54	T	0054	T	latin capital letter t
85	125	55	U	0055	U	latin capital letter u

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
86	126	56	V	0056	V	latin capital letter v
87	127	57	W	0057	W	latin capital letter w
88	130	58	X	0058	X	latin capital letter x
89	131	59	Y	0059	Y	latin capital letter y
90	132	5a	Z	005A	Z	latin capital letter z
91	133	5b	<(	005B	[	left square bracket
92	134	5c	//	005C	\	reverse solidus
93	135	5d	)>	005D	]	right square bracket
94	136	5e	'>	005E	^	circumflex accent
95	137	5f	_	005F	_	low line
96	140	60	'!	0060	`	grave accent
97	141	61	a	0061	a	latin small letter a
98	142	62	b	0062	b	latin small letter b
99	143	63	c	0063	c	latin small letter c
100	144	64	d	0064	d	latin small letter d
101	145	65	e	0065	e	latin small letter e
102	146	66	f	0066	f	latin small letter f
103	147	67	g	0067	g	latin small letter g
104	150	68	h	0068	h	latin small letter h
105	151	69	i	0069	i	latin small letter i
106	152	6a	j	006A	j	latin small letter j
107	153	6b	k	006B	k	latin small letter k
108	154	6c	l	006C	l	latin small letter l
109	155	6d	m	006D	m	latin small letter m
110	156	6e	n	006E	n	latin small letter n
111	157	6f	o	006F	o	latin small letter o
112	160	70	p	0070	p	latin small letter p
113	161	71	q	0071	q	latin small letter q
114	162	72	r	0072	r	latin small letter r

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
115	163	73	s	0073	s	latin small letter s
116	164	74	t	0074	t	latin small letter t
117	165	75	u	0075	u	latin small letter u
118	166	76	v	0076	v	latin small letter v
119	167	77	w	0077	w	latin small letter w
120	170	78	x	0078	x	latin small letter x
121	171	79	y	0079	y	latin small letter y
122	172	7a	z	007A	z	latin small letter z
123	173	7b	(!	007B	{	left curly bracket
124	174	7c	!!	007C	|	vertical line
125	175	7d	!)	007D	}	right curly bracket
126	176	7e	'?	007E	~	tilde
127	177	7f	DT	007F		delete
128	200	80	PA	0080	pad	padding character
129	201	81	HO	0081	hop	high octet preset
130	202	82	BH	0082	bph	break permitted here
131	203	83	NH	0083	nbh	no break here
132	204	84	IN	0084	ind	index
133	205	85	NL	0085	nel	next line
134	206	86	SA	0086	ssa	start of selected area
135	207	87	ES	0087	esa	end of selected area
136	210	88	HS	0088	hts	horizontal tab set
137	211	89	HJ	0089	htj	horizontal tab set with justification
138	212	8a	VS	008A	vts	vertical tab set
139	213	8b	PD	008B	pld	partial line down
140	214	8c	PU	008C	plu	partial line up
141	215	8d	RI	008D	ri	reverse line feed
142	216	8e	S2	008E	ss2	single shift two

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
143	217	8f	S3	008F	ss3	single shift three
144	220	90	DC	0090	dcs	device control string
145	221	91	P1	0091	pu1	private use one
146	222	92	P2	0092	pu2	private use two
147	223	93	TS	0093	sts	set transmit state
148	224	94	CC	0094	cch	cancel character
149	225	95	MW	0095	mw	message waiting
150	226	96	SG	0096	sga	start of guarded area
151	227	97	EG	0097	ega	end of guarded area
152	230	98	SS	0098	sos	start of string
153	231	99	GC	0099	sgci	single graphic character introducer
154	232	9a	SC	009A	sci	single character introducer
155	233	9b	CI	009B	csi	control sequence introducer
156	234	9c	ST	009C	st	string terminator
157	235	9d	OC	009D	osc	operating system command
158	236	9e	PM	009E	pm	privacy message
159	237	9f	AC	009F	apc	application program command
160	240	a0	NS	00A0	/_	no-break space
161	241	a1		0190+0300	E[`	latin capital letter open e with grave
162	242	a2		0190+0301	E['	latin capital letter open e with acute
163	243	a3		0190+0302	E[^	latin capital letter open e with circumflex
164	244	a4		0190+030C	E[\v	latin capital letter open e with caron
165	245	a5		0186+0300	O[`	latin capital letter open o with grave

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
166	246	a6		0186+0301	O['	latin capital letter open o with acute
167	247	a7		0186+0302	O[^	latin capital letter open o with circumflex
168	250	a8		019D	N]	latin capital letter n with left hook
169	251	a9		014A	N[	latin capital letter eng
170	252	aa		004E+0302	N^	latin capital letter n with circumflex
171	253	ab	<<	00AB	<<	left-pointing double angle quotation mark
172	254	ac		004E+0308	N"	latin capital letter n with diaeresis
173	255	ad	--	00AD	\-	soft hyphen
174	256	ae		01B3	Y[	latin capital letter y with hook
175	257	af		0186+030C	O[\v	latin capital letter open o with caron
176	260	b0		00B0	_DG	degree sign
177	261	b1		025B+0300	e[	latin small letter open e with grave
178	262	b2		025B+0301	e['	latin small letter open e with acute
179	263	b3		025B+0302	e[^	latin small letter open e with circumflex
180	264	b4		025B+030C	e[\v	latin small letter open e with caron
181	265	b5		0254+0300	o[	latin small letter open o with grave
182	266	b6		0254+0301	o['	latin small letter open o with acute
183	267	b7		0254+0302	o[^	latin small letter open o with circumflex
184	270	b8		0272	n]	latin small letter n with left hook

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
185	271	b9		014B	n[	latin small letter eng
186	272	ba		006E+0302	n^	latin small letter n with circumflex
187	273	bb	>>	00BB	>>	right-pointing double angle quotation mark
188	274	bc		006E+0308	n"	latin small letter n with diaeresis
189	275	bd		0294	?^	latin letter glottal stop
190	276	be		01B4	y[	latin small letter y with hook
191	277	bf		0254+030C	o[\v	latin small letter open o with caron
192	300	c0	A!	00C0	A`	latin capital letter a with grave
193	301	c1	A'	00C1	A\'	latin capital letter a with acute
194	302	c2	A>	00C2	A^	latin capital letter a with circumflex
195	303	c3		01CD	A\v	latin capital letter a with caron
196	304	c4	A:	00C4	A"	latin capital letter a with diaeresis
197	305	c5		018E	E<	latin capital letter reversed e (schwa)
198	306	c6	AE	00C6	AE+	latin capital ligature ae
199	307	c7	C,	00C7	C\,	latin capital letter c with cedilla
200	310	c8	E!	00C8	E`	latin capital letter e with grave
201	311	c9	E'	00C9	E\'	latin capital letter e with acute
202	312	ca	E>	00CA	E^	latin capital letter e with circumflex

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
203	313	cb	E:	00CB	E"	latin capital letter e with diaeresis
204	314	cc	I!	00CC	I`	latin capital letter i with grave
205	315	cd	I'	00CD	I\'	latin capital letter i with acute
206	316	ce	I>	00CE	I^	latin capital letter i with circumflex
207	317	cf	I:	00CF	I"	latin capital letter i with diaeresis
208	320	d0		0190	E[	latin capital letter open e
209	321	d1	N?	00D1	N~	latin capital letter n with tilde
210	322	d2	O!	00D2	O`	latin capital letter o with grave
211	323	d3	O'	00D3	O\'	latin capital letter o with acute
212	324	d4	O>	00D4	O^	latin capital letter o with circumflex
213	325	d5	O?	01D1	O\v	latin capital letter o with caron
214	326	d6	O:	00D6	O"	latin capital letter o with diaeresis
215	327	d7		0152	OE+	latin capital ligature oe
216	330	d8		0186	O[	latin capital letter open o
217	331	d9	U!	00D9	U`	latin capital letter u with grave
218	332	da	U'	00DA	U\'	latin capital letter u with acute
219	333	db	U>	00DB	U^	latin capital letter u with circumflex
220	334	dc	U:	00DC	U"	latin capital letter u with diaeresis

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
221	335	dd		011A	E\v	latin capital letter e with caron
222	336	de		01CF	I\v	latin capital letter i with caron
223	337	df		01D3	U\v	latin capital letter u with caron
224	340	e0	a!	00E0	a`	latin small letter a with grave
225	341	e1	a'	00E1	a\'	latin small letter a with acute
226	342	e2	a>	00E2	a^	latin small letter a with circumflex
227	343	e3		01CE	a\v	latin small letter a with caron
228	344	e4	a:	00E4	a"	latin small letter a with diaeresis
229	345	e5		0259	e<	latin small letter turned e (schwa)
230	346	e6	ae	00E6	ae+	latin small ligature ae
231	347	e7	c,	00E7	c\,	latin small letter c with cedilla
232	350	e8	e!	00E8	e`	latin small letter e with grave
233	351	e9	e'	00E9	e\'	latin small letter e with acute
234	352	ea	e>	00EA	e^	latin small letter e with circumflex
235	353	eb	e:	00EB	e"	latin small letter e with diaeresis
236	354	ec	i!	00EC	i`	latin small letter i with grave
237	355	ed	i'	00ED	i\'	latin small letter i with acute
238	356	ee	i>	00EE	i^	latin small letter i with circumflex

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
239	357	ef	i:	00EF	i"	latin small letter i with diaeresis
240	360	f0		025B	e[	latin small letter open e
241	361	f1	n?	00F1	n~	latin small letter n with tilde
242	362	f2	o!	00F2	o`	latin small letter o with grave
243	363	f3	o'	00F3	o\'	latin small letter o with acute
244	364	f4	o>	00F4	o^	latin small letter o with circumflex
245	365	f5		01D2	o\v	latin small letter o with caron
246	366	f6	o:	00F6	o"	latin small letter o with diaeresis
247	367	f7		0153	oe+	latin small ligature oe
248	370	f8		0254	o[	latin small letter open o
249	371	f9	u!	00F9	u`	latin small letter u with grave
250	372	fa	u'	00FA	u\'	latin small letter u with acute
251	373	fb	u>	00FB	u^	latin small letter u with circumflex
252	374	fc	u:	00FC	u"	latin small letter u with diaeresis
253	375	fd		011B	e\v	latin small letter e with caron
254	376	fe		01D0	i\v	latin small letter i with caron
255	377	ff		01D4	u\v	latin small letter u with caron

Annex 2: AFRLIN character set[20]

The African AFRLIN character set covers the alphabets of French, Lingala, Sahngo and Wolof. The AFRLIN-104-BPI_OCIL character set is a set of encoded characters used for on-screen display of the following African languages: Lingala, Sahngo and Wolof. Acceptable alternate names for this set of encoded characters are afrlin104bpiocil, lingala, sango, wolof and lin.

The AFRLIN-105-BPI_OCIL character set is a set of transliterated characters used for keyboard encoding of the following African languages: Lingala, Sahngo and Wolof. Acceptable alternate names for this set of encoded characters are afrlin105bpiocil, tlingala, tsango, twolof and tlin. For each of the 255 characters listed below, the following information is given in the respective columns:

Dec	decimal value of the character
Oct	octal value of the character
Hex	hexidecimal value of the character
Mne	two-letter mnemotechnical code of the RFC *1345 standard*
UCS2	UCS value of the character
Kbd	keyboard encoding convention for the character
AFRLIN-104-BPI_OCIL	name of the character according to the ISO *10646 standard*

20 Source: Progiciels BPI http://www.progiciels-bpi.ca/tcao/apercu.html.

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
0	000	00	NU	0000	nul	null
1	001	01	SH	0001	soh	start of heading
2	002	02	SX	0002	stx	start of text
3	003	03	EX	0003	etx	end of text
4	004	04	ET	0004	eot	end of transmission
5	005	05	EQ	0005	enq	enquiry
6	006	06	AK	0006	ack	acknowledge
7	007	07	BL	0007	bel	bell
8	010	08	BS	0008	bs	backspace
9	011	09	HT	0009	ht	character tabulation
10	012	0a	LF	000A	lf	line feed
11	013	0b	VT	000B	vt	vertical tabulation
12	014	0c	FF	000C	ff	form feed
13	015	0d	CR	000D	cr	carriage return
14	016	0e	SO	000E	so	shift out
15	017	0f	SI	000F	si	shift in
16	020	10	DL	0010	dle	data link escape
17	021	11	D1	0011	dc1	device control one
18	022	12	D2	0012	dc2	device control two
19	023	13	D3	0013	dc3	device control three
20	024	14	D4	0014	dc4	device control four
21	025	15	NK	0015	nak	negative acknowledge
22	026	16	SY	0016	syn	synchronous idle
23	027	17	EB	0017	etb	end of transmission block
24	030	18	CN	0018	can	cancel
25	031	19	EM	0019	em	end of medium
26	032	1a	SB	001A	sub	substitute

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
27	033	1b	EC	001B	esc	escape
28	034	1c	FS	001C	is4	file separator
29	035	1d	GS	001D	is3	group separator
30	036	1e	RS	001E	is2	record separator
31	037	1f	US	001F	is1	unit separator
32	040	20	SP	0020	< >	space
33	041	21	!	0021	!	exclamation mark
34	042	22	"	0022	"	quotation mark
35	043	23	Nb	0023	#	number sign
36	044	24	DO	0024	$	dollar sign
37	045	25	%	0025	%	percent sign
38	046	26	&	0026	&	ampersand
39	047	27	'	0027	'	apostrophe
40	050	28	(	0028	(	left parenthesis
41	051	29	)	0029	)	right parenthesis
42	052	2a	*	002A	*	asterisk
43	053	2b	+	002B	+	plus sign
44	054	2c	,	002C	,	comma
45	055	2d	-	002D	-	hyphen-minus
46	056	2e	.	002E	.	full stop
47	057	2f	/	002F	/	solidus
48	060	30	0	0030	0	digit zero
49	061	31	1	0031	1	digit one
50	062	32	2	0032	2	digit two
51	063	33	3	0033	3	digit three
52	064	34	4	0034	4	digit four
53	065	35	5	0035	5	digit five
54	066	36	6	0036	6	digit six

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
55	067	37	7	0037	7	digit seven
56	070	38	8	0038	8	digit eight
57	071	39	9	0039	9	digit nine
58	072	3a	:	003A	:	colon
59	073	3b	;	003B	;	semicolon
60	074	3c	<	003C	<	less-than sign
61	075	3d	=	003D	=	equals sign
62	076	3e	>	003E	>	greater-than sign
63	077	3f	?	003F	?	question mark
64	100	40	At	0040	@	commercial at
65	101	41	A	0041	A	latin capital letter a
66	102	42	B	0042	B	latin capital letter b
67	103	43	C	0043	C	latin capital letter c
68	104	44	D	0044	D	latin capital letter d
69	105	45	E	0045	E	latin capital letter e
70	106	46	F	0046	F	latin capital letter f
71	107	47	G	0047	G	latin capital letter g
72	110	48	H	0048	H	latin capital letter h
73	111	49	I	0049	I	latin capital letter i
74	112	4a	J	004A	J	latin capital letter j
75	113	4b	K	004B	K	latin capital letter k
76	114	4c	L	004C	L	latin capital letter l
77	115	4d	M	004D	M	latin capital letter m
78	116	4e	N	004E	N	latin capital letter n
79	117	4f	O	004F	O	latin capital letter o
80	120	50	P	0050	P	latin capital letter p
81	121	51	Q	0051	Q	latin capital letter q
82	122	52	R	0052	R	latin capital letter r

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
83	123	53	S	0053	S	latin capital letter s
84	124	54	T	0054	T	latin capital letter t
85	125	55	U	0055	U	latin capital letter u
86	126	56	V	0056	V	latin capital letter v
87	127	57	W	0057	W	latin capital letter w
88	130	58	X	0058	X	latin capital letter x
89	131	59	Y	0059	Y	latin capital letter y
90	132	5a	Z	005A	Z	latin capital letter z
91	133	5b	<(	005B	[	left square bracket
92	134	5c	//	005C	\	reverse solidus
93	135	5d	)>	005D	]	right square bracket
94	136	5e	'>	005E	^	circumflex accent
95	137	5f	_	005F	_	low line
96	140	60	'!	0060	`	grave accent
97	141	61	a	0061	a	latin small letter a
98	142	62	b	0062	b	latin small letter b
99	143	63	c	0063	c	latin small letter c
100	144	64	d	0064	d	latin small letter d
101	145	65	e	0065	e	latin small letter e
102	146	66	f	0066	f	latin small letter f
103	147	67	g	0067	g	latin small letter g
104	150	68	h	0068	h	latin small letter h
105	151	69	i	0069	i	latin small letter i
106	152	6a	j	006A	j	latin small letter j
107	153	6b	k	006B	k	latin small letter k
108	154	6c	l	006C	l	latin small letter l
109	155	6d	m	006D	m	latin small letter m
110	156	6e	n	006E	n	latin small letter n

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
111	157	6f	o	006F	o	latin small letter o
112	160	70	p	0070	p	latin small letter p
113	161	71	q	0071	q	latin small letter q
114	162	72	r	0072	r	latin small letter r
115	163	73	s	0073	s	latin small letter s
116	164	74	t	0074	t	latin small letter t
117	165	75	u	0075	u	latin small letter u
118	166	76	v	0076	v	latin small letter v
119	167	77	w	0077	w	latin small letter w
120	170	78	x	0078	x	latin small letter x
121	171	79	y	0079	y	latin small letter y
122	172	7a	z	007A	z	latin small letter z
123	173	7b	(!	007B	{	left curly bracket
124	174	7c	!!	007C	|	vertical line
125	175	7d	!)	007D	}	right curly bracket
126	176	7e	'?	007E	~	tilde
127	177	7f	DT	007F		delete
128	200	80	PA	0080	pad	padding character
129	201	81	HO	0081	hop	high octet preset
130	202	82	BH	0082	bph	break permitted here
131	203	83	NH	0083	nbh	no break here
132	204	84	IN	0084	ind	index
133	205	85	NL	0085	nel	next line
134	206	86	SA	0086	ssa	start of selected area
135	207	87	ES	0087	esa	end of selected area
136	210	88	HS	0088	hts	horizontal tab set
137	211	89	HJ	0089	htj	horizontal tab set with justification
138	212	8a	VS	008A	vts	vertical tab set

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
139	213	8b	PD	008B	pld	partial line down
140	214	8c	PU	008C	plu	partial line up
141	215	8d	RI	008D	ri	reverse line feed
142	216	8e	S2	008E	ss2	single shift two
143	217	8f	S3	008F	ss3	single shift three
144	220	90	DC	0090	dcs	device control string
145	221	91	P1	0091	pu1	private use one
146	222	92	P2	0092	pu2	private use two
147	223	93	TS	0093	sts	set transmit state
148	224	94	CC	0094	cch	cancel character
149	225	95	MW	0095	mw	message waiting
150	226	96	SG	0096	sga	start of guarded area
151	227	97	EG	0097	ega	end of guarded area
152	230	98	SS	0098	sos	start of string
153	231	99	GC	0099	sgci	single graphic character introducer
154	232	9a	SC	009A	sci	single character introducer
155	233	9b	CI	009B	csi	control sequence introducer
156	234	9c	ST	009C	st	string terminator
157	235	9d	OC	009D	osc	operating system command
158	236	9e	PM	009E	pm	privacy message
159	237	9f	AC	009F	apc	application program command
160	240	a0	NS	00A0	/_	no-break space
161	241	a1			E[`	latin capital letter open e with grave
162	242	a2			E['	latin capital letter open e with acute

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
163	243	a3			E[^	latin capital letter open e with circumflex
164	244	a4			E[\v	latin capital letter open e with caron
165	245	a5			O[	latin capital letter open o with grave
166	246	a6			O['	latin capital letter open o with acute
167	247	a7			O[^	latin capital letter open o with circumflex
168	250	a8		019D	N]	latin capital letter n with left hook
169	251	a9		014A	N[	latin capital letter eng
170	252	aa			N^	latin capital letter n with circumflex
171	253	ab	<<	00AB	<<	left-pointing double angle quotation mark
172	254	ac			N"	latin capital letter n with diaeresis
173	255	ad	--	00AD	\-	soft hyphen
174	256	ae		01B3	Y[	latin capital letter y with hook
175	257	af			O[\v	latin capital letter open o with caron
176	260	b0		00B0	_DG	degree sign
177	261	b1			e[	latin small letter open e with grave
178	262	b2			e['	latin small letter open e with acute
179	263	b3			e[^	latin small letter open e with circumflex
180	264	b4			e[\v	latin small letter open e with caron
181	265	b5			o[	latin small letter open o with grave

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
182	266	b6			o['	latin small letter open o with acute
183	267	b7			o[^	latin small letter open o with circumflex
184	270	b8		0272	n]	latin small letter n with left hook
185	271	b9		014B	n[	latin small letter eng
186	272	ba			n^	latin small letter n with circumflex
187	273	bb	>>	00BB	>>	right-pointing double angle quotation mark
188	274	bc			n"	latin small letter n with diaeresis
189	275	bd		0294	?^	latin letter glottal stop
190	276	be		01B4	y[	latin small letter y with hook
191	277	bf			o[\v	latin small letter open o with caron
192	300	c0	A!	00C0	A`	latin capital letter a with grave
193	301	c1	A'	00C1	A\'	latin capital letter a with acute
194	302	c2	A>	00C2	A^	latin capital letter a with circumflex
195	303	c3		01CD	A\v	latin capital letter a with caron
196	304	c4	A:	00C4	A"	latin capital letter a with diaeresis
197	305	c5		018E	E<	latin capital letter reversed e (schwa)
198	306	c6	AE	00C6	AE+	latin capital ligature ae
199	307	c7	C,	00C7	C\,	latin capital letter c with cedilla
200	310	c8	E!	00C8	E`	latin capital letter e with grave

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
201	311	c9	E'	00C9	E\'	latin capital letter e with acute
202	312	ca	E>	00CA	E^	latin capital letter e with circumflex
203	313	cb	E:	00CB	E"	latin capital letter e with diaeresis
204	314	cc	I!	00CC	I`	latin capital letter i with grave
205	315	cd	I'	00CD	I\'	latin capital letter i with acute
206	316	ce	I>	00CE	I^	latin capital letter i with circumflex
207	317	cf	I:	00CF	I"	latin capital letter i with diaeresis
208	320	d0		0190	E[	latin capital letter open e
209	321	d1	N?	00D1	N~	latin capital letter n with tilde
210	322	d2	O!	00D2	O`	latin capital letter o with grave
211	323	d3	O'	00D3	O\'	latin capital letter o with acute
212	324	d4	O>	00D4	O^	latin capital letter o with circumflex
213	325	d5	O?	01D1	O\v	latin capital letter o with caron
214	326	d6	O:	00D6	O"	latin capital letter o with diaeresis
215	327	d7		0152	OE+	latin capital ligature oe
216	330	d8		0186	O[	latin capital letter open o
217	331	d9	U!	00D9	U`	latin capital letter u with grave
218	332	da	U'	00DA	U\'	latin capital letter u with acute

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
219	333	db	U>	00DB	U^	latin capital letter u with circumflex
220	334	dc	U:	00DC	U"	latin capital letter u with diaeresis
221	335	dd		011A	E\v	latin capital letter e with caron
222	336	de		01CF	I\v	latin capital letter i with caron
223	337	df		01D3	U\v	latin capital letter u with caron
224	340	e0	a!	00E0	a`	latin small letter a with grave
225	341	e1	a'	00E1	a\'	latin small letter a with acute
226	342	e2	a>	00E2	a^	latin small letter a with circumflex
227	343	e3		01CE	a\v	latin small letter a with caron
228	344	e4	a:	00E4	a"	latin small letter a with diaeresis
229	345	e5		0259	e<	latin small letter turned e (schwa)
230	346	e6	ae	00E6	ae+	latin small ligature ae
231	347	e7	c,	00E7	c\,	latin small letter c with cedilla
232	350	e8	e!	00E8	e`	latin small letter e with grave
233	351	e9	e'	00E9	e\'	latin small letter e with acute
234	352	ea	e>	00EA	e^	latin small letter e with circumflex
235	353	eb	e:	00EB	e"	latin small letter e with diaeresis
236	354	ec	i!	00EC	ì	latin small letter i with grave

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRLIN-104-BPI_OCIL
237	355	ed	i'	00ED	i\'	latin small letter i with acute
238	356	ee	i>	00EE	i^	latin small letter i with circumflex
239	357	ef	i:	00EF	i"	latin small letter i with diaeresis
240	360	f0		025B	e[	latin small letter open e
241	361	f1	n?	00F1	n~	latin small letter n with tilde
242	362	f2	o!	00F2	o`	latin small letter o with grave
243	363	f3	o'	00F3	o\'	latin small letter o with acute
244	364	f4	o>	00F4	o^	latin small letter o with circumflex
245	365	f5		01D2	o\v	latin small letter o with caron
246	366	f6	o:	00F6	o"	latin small letter o with diaeresis
247	367	f7		0153	oe+	latin small ligature oe
248	370	f8		0254	o[	latin small letter open o
249	371	f9	u!	00F9	u`	latin small letter u with grave
250	372	fa	u'	00FA	u\'	latin small letter u with acute
251	373	fb	u>	00FB	u^	latin small letter u with circumflex
252	374	fc	u:	00FC	u"	latin small letter u with diaeresis
253	375	fd		011B	e\v	latin small letter e with caron
254	376	fe		01D0	i\v	latin small letter i with caron
255	377	ff		01D4	u\v	latin small letter u with caron

www.unesco.org/webworld