



Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

# Comment assurer la présence d'une langue dans le cyberspace ?

Par Marcel Diki-Kidiri



PIPT — Programme Information pour tous  
Secteur de la communication et de l'information

Comment assurer la présence d'une langue dans le cyberspace ?



Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

# Comment assurer la présence d'une langue dans le cyberspace ?

Par Marcel Diki-Kidiri

Chargé de recherche au CNRS,

Laboratoire « Langage, Langues et Cultures d'Afrique Noire »  
(LLACAN)

**PIPT – Programme Information pour tous**

Les idées, les éléments d'information et les opinions exprimés dans cette publication sont ceux des auteurs. Ils ne reflètent pas nécessairement les vues de l'UNESCO et n'engagent pas l'Organisation.

**Note bibliographique recommandée :**

UNESCO. Programme Information pour tous (PIPT).  
Comment assurer la présence d'une langue dans le cyberspace ?  
Edité par la Division de la société de l'information, Secteur de la communication  
et de l'information (Editeur : Claudio Menezes) – Paris, UNESCO, 2007 : 80 p. ;  
21 cm. (CI-2007/WS/1)

I – Titre

II – UNESCO

III – Programme Information pour tous

Publié en 2007

Par l'Organisation des Nations Unies

pour l'éducation, la science et la culture.

7, place de Fontenoy F-75352 Paris 07 SP, Paris, France

Internet : [www.unesco.org/webworld](http://www.unesco.org/webworld)

Coordinateur de la publication : Boyan Radoykov

Composé et imprimé dans les ateliers de l'UNESCO

© UNESCO 2007

*Printed in France*

Tous droits réservés

(CI-2007/WS/1 – CLD 31084)

# Avant-propos

Les langues sont avant tout des instruments permettant d'acquérir une autonomie éducative et culturelle. Elles permettent la transmission de connaissances entre générations, en étant un vecteur pour la dissémination des cultures et traditions entre et parmi les différents groupes ethniques dans des zones géographiques diverses. Le phénomène du développement et la disparition des langues fait partie de l'évolution et du déclin des civilisations. Le latin – une langue morte – joue encore aujourd'hui une influence majeure sur les langues vivantes, y compris étant la base initiale pour le premier code standard pour les ordinateurs, le code ASCII. La langue maternelle est également le premier véhicule pour la liberté d'expression.

La disparition de quelques langues est un phénomène présent dans le cours de l'Histoire. Même dans des pays officiellement déclarés comme monolingues, des nouvelles politiques sont en train d'émerger pour assurer l'expression dans les langues endogènes comme un droit humain.

D'après une étude développée par « Ethnologue », l'indice de diversité linguistique par région montre que l'Afrique est le continent avec l'indicateur le plus élevé dans le monde. Il y a des évidences indiquant que la diversité linguistique globale est à présent en déclin depuis longtemps. Un autre élément d'inquiétude : d'après quelques estimations, la moitié des langues aura disparu vers l'année 2050.

Les technologies de l'information et de la communication jouent un rôle clé dans les transformations linguistiques ayant lieu dans le monde : elles peuvent être un important véhicule de communication entre les différentes communautés linguistiques. Par contre, les TICs peuvent être un facteur aggravant de la marginalisation des langues dans

le cyberspace. Il y a approximativement 6 000 langues dans le monde, dont 12 langues correspondent à 98 % des pages web sur l'Internet. L'anglais, avec 72 % des pages web, est l'idiome dominant, d'après une expérience menée par O'Neill, Lavoie et Bennet en 2003.

Après tout, le défi devant la communauté internationale est de franchir les obstacles gigantesques pour assurer la création d'un cyberspace multilingue et culturellement diverse.

Pour ce faire, l'UNESCO – avec l'aide de l'Union latine et la contribution intellectuelle de l'expert Marcel Diki-Kidiri – publie le présent document technique, dans le cadre de la série de publications du « Programme Information pour tous ».

Cohérente avec la « Recommandation sur la promotion et l'usage du multilinguisme et l'accès universel au cyberspace », adoptée par la Conférence générale de l'UNESCO lors de sa 32<sup>e</sup> session, la présente publication contribuera à faciliter la prise de décision en faveur de l'inclusion de nouvelles langues dans l'espace numérique.

**Abdul Waheed Khan**  
Sous-Directeur général  
pour la communication et l'information

# Sommaire

<b>Introduction</b>	7
<b>1. Élaboration de ressources linguistiques</b>	11
1.1. De l'oral à l'écrit	11
1.2. Développement terminologique	17
<b>2. Élaboration des ressources informatiques</b>	19
2.1. Systèmes d'écriture et orthographes	19
2.2. Jeux et polices de caractères	21
2.3. Codage des caractères	22
2.4. Premières ressources de base	26
2.5. Localisation des logiciels	32
<b>3. Élaboration des contenus culturels</b>	35
3.1. Contenus culturels	35
3.2. Préparer les voies d'accès au cyberspace	37
<b>4. Développer la communauté des utilisateurs</b>	41
4.1. Création d'une communauté d'utilisateurs	41
4.2. Aspects politiques et socioculturels	42
<b>5. Conclusion</b>	49
<b>Références</b>	51
<b>Annexe 1: Jeux de caractères AFRFUL</b>	53
<b>Annexe 2: Jeux de caractères AFRLIN</b>	67



# Introduction

---

Le cyberspace est ouvert à toutes les langues du monde car son infrastructure n'est pas soumise à l'autorité d'un pouvoir central qui déciderait de son utilisation. Il suffit, en principe, de disposer d'un ordinateur relié à un fournisseur d'accès Internet pour afficher sur la Toile des données textuelles, iconographiques ou sonores dans la langue de son choix. La mise en oeuvre de ce principe qui est un facteur fondamental de la démocratie au niveau mondial, nécessite, cependant, que soient réunies un certain nombre de conditions techniques et de ressources humaines et financières que nous allons examiner dans la présente étude. En rédigeant celle-ci, nous avons voulu répondre aussi simplement que possible à la question suivante : comment faire en sorte qu'une langue peu dotée en ressources linguistiques et/ou informatiques, voire en ressources humaines, puisse arriver à trouver une place dans le cyberspace et y être active ?

*Par extension, toute langue ayant plus ou moins de ressources peut trouver des réponses dans cette étude si elle est faiblement représentée dans le cyberspace. Nous décrivons ici le parcours le plus long, celui d'une langue n'ayant même pas de représentation écrite, mais d'autres langues mieux loties traverseront les étapes qui les concernent sur le même chemin.*

Il nous faut d'emblée répondre à une question préalable. Qu'est-ce qu'une langue peu dotée ? Il s'agit d'une langue qui ne dispose pas suffisamment, voire pas du tout, des ressources essentielles dont sont généralement dotées les grandes langues du monde, à savoir : une orthographe stable dans un système d'écriture donné, des ouvrages de référence (grammaires, dictionnaires, ouvrages littéraires), des oeuvres de diffusion massive (presse écrite et audiovisuelle, films, chansons et musique), des ouvrages techniques et d'apprentissage (publications techniques et scientifiques, ouvrages didactiques, manuels), divers supports de communication du quotidien (affiches, publicités, courriers, notices, modes d'emploi, etc.), ainsi qu'un nombre abondant d'applications informatiques dans cette langue. Sur le plan des ressources humaines, une langue peu

dotée peut devenir une langue en danger de disparition si elle n'est plus parlée que par un petit nombre de locuteurs. Il devient alors nécessaire pour la sauver, d'augmenter le nombre de ses locuteurs en l'enseignant par tous les moyens techniques possibles.

Heureusement, une langue peu dotée n'a pas nécessairement tous ces handicaps à la fois. Elle peut être majoritaire, écrite, enseignée à l'école et pourtant manquer cruellement de ressources informatiques ou même de ressources linguistiques en quantité et en qualité suffisantes. Il est donc plus juste de dire que les langues peu dotées le sont de manière très variée – depuis les langues en grand danger de disparition, jusqu'aux langues émergentes qui possèdent déjà une bonne partie de ces ressources, mais en nombre estimé insuffisant et incomplet. De nombreuses langues sur tous les continents répondent à cette définition des langues peu dotées. On peut citer en Europe le breton, l'occitan ou encore le basque, en Amérique, la quasi-totalité des langues amérindiennes, en Asie, le myanmar<sup>1</sup>, pour n'en citer qu'une parmi des centaines, en Océanie, la quasi-totalité des langues autochtones des îles polynésiennes, micronésiennes et mélanésiennes. En Afrique, où se parlent un tiers des langues du monde, soit environ 2000 langues, les langues les mieux dotées (afrikaans, kiswahili, hausa, etc.) se comptent sur les doigts d'une main et font partie des langues émergentes, donc des langues peu dotées.

Nous partons donc de l'hypothèse la plus défavorable, à savoir, une langue qui fut jadis porteuse d'une culture florissante, mais qui n'est plus parlée aujourd'hui que par une poignée de personnes âgées dans un petit village au fin fond de la brousse africaine, disons en Afrique centrale pour être aussi loin que possible de toutes les côtes. Appelons cette langue le ndeka. Une telle langue n'a pratiquement aucune chance d'accéder un jour au cyberspace, car elle risque de disparaître à tout jamais à la mort de son dernier locuteur, et ils ne sont plus très nombreux à la parler. Voici qu'un jour un jeune étudiant issu de ce village trouve que des outils comme SPIP, Wikipedia et bien d'autres encore pourraient offrir une formidable opportunité

---

1 Nouvelle dénomination de la langue appelée birmane autrefois.



de conserver la mémoire de cette langue et de ce fait celle de son village et de la culture de ses ancêtres. Ce serait même le moyen le plus sûr de faire apprendre cette langue aux jeunes de sa génération et ainsi lui redonner une nouvelle vie. Il vient nous trouver pour nous demander de le guider dans son projet. Avec cette hypothèse en tête, notre exposé se veut didactique, afin d'accompagner, pas à pas, tous ceux qui nous rejoindront à n'importe quelle étape du chemin pour porter jusque dans le cyberspace toutes les langues peu dotées, quelle que soit leur fortune.



# 1. Élaboration de ressources linguistiques

## 1.1. De l'oral à l'écrit

L'objectif de cette première étape est de doter la langue d'un minimum de ressources linguistiques, à savoir: une orthographe dans un système d'écriture, une grammaire écrite, un dictionnaire et un recueil de textes aussi fourni que possible. C'est donc un travail de linguiste que nous allons évoquer dans les grandes lignes, tout en nous concentrant sur les objectifs intermédiaires les plus indispensables.

### 1.1.1. Recueil de textes

La première chose à faire est d'aller trouver les locuteurs de la langue et d'enregistrer autant de textes que possibles, des récits de vie, des contes, des proverbes, des chansons, des poèmes, des légendes, des conversations, des narrations, etc. Pour faire ces enregistrements, il vaut mieux apprendre des techniques de recherche sur le terrain et utiliser autant que possible des appareils de qualité professionnelle, afin d'obtenir la meilleure qualité sonore possible, car le son ainsi enregistré sera par la suite soumis à de nombreux traitements. Il serait dommage que la qualité du son se dégrade dès la deuxième copie! Dans des conditions de travail normales, les enregistrements sont immédiatement traduits avec l'assistance des locuteurs de la langue et transférés dans une base de données textuelles.

### 1.1.2. Transcription phonétique

En utilisant l'alphabet phonétique international (API) les linguistes peuvent transcrire très exactement n'importe quel son de n'importe quelle langue du monde. Les textes recueillis peuvent donc être ainsi transcrits. Mais il faut bien comprendre que la transcription phonétique reproduit très fidèlement chaque son tel qu'il est pro-

noncé en un moment donné. Or, en français, dans un mot comme *papa* le premier *P* est légèrement plus explosif que le second parce qu'il est en début de mot et l'autre entre deux voyelles. Cette différence est habituellement négligée car elle n'entraîne aucune conséquence pour le sens du mot, mais une bonne transcription phonétique la reflètera comme ceci [p<sup>˘</sup>apa]<sup>2</sup>. La transcription phonétique est indispensable pour permettre une bonne analyse des sons de la langue, mais elle n'est pas du tout la meilleure façon de l'écrire pour un usage courant.

### 1.1.3. Analyse et notation des phonèmes

Si la transcription phonétique est faite correctement, c'est-à-dire avec minutie et fidélité par rapport aux sons réellement prononcés, il sera alors aisé de procéder à une analyse phonologique, dont l'objectif est d'identifier les sons utiles de la langue qu'on appelle les *phonèmes*. Il s'agit de sons qui entraînent une différence de sens quand ils changent. Par exemple, en français, on dira que /p/<sup>3</sup> et /b/ sont deux phonèmes différents parce que leur différence de prononciation permet à elle seule de distinguer les mots « pain » et « bain ».

Par contre, certaines personnes prononcent le mot *roi* avec un *R* roulé et d'autres avec un *R* grasseyé. Ces deux *R* sont phonétiquement différents et transcrits respectivement [r] et [R]. Pourtant cette différence phonétique n'entraînant aucune différence de sens, ces deux sons ne représentent qu'un seul phonème que l'on conviendra de noter /r/. Un autre exemple, la suite de sons [gz] que l'on trouve dans le mot « exact » prononcé [egza] et la suite de sons [ks] du mot « extra » prononcé [ekstRa] représentent toutes deux le même phonème que l'on conviendra de noter /x/.

Ainsi, l'analyse phonologique d'une langue aboutit à établir la liste complète des phonèmes de cette langue et la manière de les noter. On peut alors utiliser cette liste de phonèmes pour écrire un texte. On parle alors d'une *notation phonologique* par opposition à la *trans-*

---

2 On utilise les crochets carrés pour indiquer que le mot est transcrit phonétiquement.

3 On utilise les barres obliques pour indiquer un phonème ou une suite de phonèmes.

*cription phonétique*. Dans une notation phonologique, ce sont les phonèmes, et non plus les sons, qui sont représentés. La notation est donc bien plus économique et fournit une bonne base pour l'élaboration d'une orthographe.

### 1.1.4. Analyse et notation des tons

On peut prononcer une syllabe sur une hauteur de voix (ou *registre*) relativement grave, moyenne ou aiguë comme les notes de musiques *do ré mi*. La majorité des langues africaines exploitent cette variation de hauteur de voix pour exprimer des différences de sens avec une même combinaison de phonèmes. Ainsi, en *sängö*<sup>4</sup>, les mots [ká] (là-bas), [kā] (plaie), et [kà] (et) sont trois mots très différents et il ne s'agit en aucun cas d'homonymes. On qualifie ces langues de *langues à tons*.

Une langue à tons peut exploiter jusqu'à cinq registres différents, mais la plupart des langues africaines n'en exploitent que deux ou trois avec deux registres, on distingue un ton haut et un ton bas. Avec trois registres, on distingue un ton bas, un ton moyen et un ton haut. Au-delà on distingue, en plus, un ton supra-haut et/ou un ton infra-bas. C'est à l'issue d'une analyse tonologique que l'on détermine si une langue est à tons ou non et, dans l'affirmative, le nombre de registres utilisés, la fréquence des tons, la meilleure façon de les noter, etc.

Dans une langue à tons, un mot est nécessairement composé d'au moins une voyelle et un ton. Ainsi, une forme comme *ka*, évoquée plus haut, ne revêt de sens que lorsqu'on précise le ton sur lequel le mot doit être prononcé. Il faudrait donc logiquement envisager de noter les tons dans le système orthographique de la langue. Toutefois, toutes les langues tonales ne font pas un usage identique des tons. Certaines se servent des tons non seulement pour distinguer des mots au niveau du lexique, mais également des formes verbales et des modalités personnelles et aspectuelles, donc des

---

4 Le *sängö* est la langue nationale de la République centrafricaine et partage le statut de langue officielle avec le français.

valeurs grammaticales. Dans de telles langues les tons occupent une place bien trop importante pour être négligés. Par contre, dans les langues qui ne recourent aux tons qu'au niveau du lexique et de façon relativement limitée (par exemple, pour ne distinguer que des mots brefs où l'ambiguïté peut être plus élevée), on a souvent pris le parti de ne pas noter les tons dans l'orthographe courante. Il est donc nécessaire de prendre une décision au cas par cas. Toutefois, il faut souligner que pour les travaux de recherche linguistique sur une langue à tons, la notation des tons est absolument indispensable, car dans ce genre de travaux, on doit prendre en compte toutes les informations sur la réalité de la langue, et les tons constituent une dimension importante et définitoire d'une langue à tons.

### 1.1.5. Élaboration d'une orthographe

Les études phonologiques et tonologiques aboutissent à un système de notation de la langue qui représente correctement les sons utiles, et ce, de façon aussi économique que possible. Il permet à 90 % d'écrire ce qu'on prononce, sans raffinement phonétique, et cela suffit pour que l'on soit tenté de se contenter de la notation phonologique et de faire l'économie d'études orthographiques. Cependant, celles-ci s'avèrent réellement indispensables pour doter une langue d'un système conventionnel capable de prendre en compte l'ensemble des besoins d'expression écrite des locuteurs. En effet, à la différence de la notation phonologique qui ne reflète que les sons utiles de la langue, une orthographe bien conçue intègre aussi la notation des relations grammaticales et des idées. Considérons, en français, les exemples suivants :

- a) /ɛl ʃãt/
- b) elle chante
- c) elles chantent

La phrase a) présente la notation phonologique qui correspond aussi bien à la phrase b) qu'à la phrase c). Si le français était écrit uniquement avec une notation phonologique, il n'y aurait aucun moyen de distinguer ici le singulier du pluriel comme le permet la convention orthographique.

Prenons un autre exemple en sängö, une langue qui distingue un ton haut, un ton moyen et un ton bas. En notation phonologique, ces tons sont notés respectivement par un accent aigu [á], un accent plat [ā] et un accent grave [à], mais ces accents ne se trouvent pas sur les machines courantes en Centrafrique. Une orthographe est pourtant destinée à un large public, et l'on doit pouvoir écrire la langue aussi bien à la main qu'à la machine ou à l'ordinateur. C'est pourquoi il faut élaborer une convention orthographique comportant un ensemble de règles qui permettent d'écrire la langue de façon plus pratique.

Le tableau suivant montre les différentes étapes nécessaires à l'élaboration d'une orthographe optimale et stable pour une langue, en l'occurrence, le sängö :

Transcription phonétique	Notation phonologique	Orthographe 1 (47 %)	Orthographe 2 (53 %)	Traduction
1. [sùkùlá]	/sùkùlà	Sukùla	sukùla	laver
2. [sã <sup>o</sup> gõ]	/sãngõ	Sängö	sahngo	sango (langue)
3. [m <sup>b</sup> ásám <sup>b</sup> álá]	/mbásám <sup>b</sup> áláa	Mbâsâmbâlâ	mbâssambala	sept
4. [mã <sup>i</sup> gõ]	/mãingõ	Mäingö	mayhngo	évolution
5. [k <sup>w</sup> à] ~ [k <sup>u</sup> à]	/kùà	kua	kua	travail
6. [k <sup>w</sup> ā] ~ [k <sup>u</sup> ā]	/kūā	küä	kwa	poil
7. [k <sup>w</sup> á] ~ [k <sup>u</sup> á]	/kúá	kûâ	kwâ	mort

On remarque que la transcription phonétique reflète la prononciation des mots telle qu'ils sont enregistrés au magnétophone. Elle est donc propre à la recherche linguistique, mais trop fine pour servir de base à une orthographe courante destinée au grand public. La notation phonologique de la deuxième colonne ne retient que les sons utiles, les phonèmes, et de ce fait, elle est plus propice à servir de base à une orthographe courante. Mais le marquage de tous les tons constitue un obstacle pour une écriture pratique et une lecture rapide. En outre, les diacritiques – symboles utilisés pour marquer les tons (accent aigu, accent plat et accent grave) – ne sont pas disponibles sur toutes les machines et ne peuvent être rendus correctement qu'avec un ordinateur, ce qui n'est pas à la

portée de tous les utilisateurs d'une langue peu dotée. La convention orthographique de la troisième colonne permet d'économiser 47 % de la fréquence des tons et de les écrire avec le tréma et l'accent circonflexe, deux symboles qui figurent en touche libre sur la plupart des claviers occidentaux des machines à écrire, et des ordinateurs utilisés en Centrafrique. Malgré cela, cette convention laisse tout de même 53 % de la fréquence des tons à noter avec des accents. Pour réduire encore davantage le taux de fréquence, nous avons effectué une réforme orthographique dont les résultats sont visibles dans la quatrième colonne. Cette nouvelle convention permet d'économiser 53 % de la fréquence des tons dans un texte. On peut dire qu'à ce stade, le sāngö (ou sahngo<sup>5</sup>) est enfin doté d'une orthographe vraiment optimale et stable.

### **1.1.6. Réalisation d'ouvrages de référence**

L'analyse des textes recueillis permet d'élaborer une grammaire complète, qui couvre la phonologie, la syntaxe, les types d'énoncés et le discours. Généralement, une grammaire de référence est écrite pour servir de base à plusieurs ouvrages plus didactiques, tels que des grammaires d'apprentissage de la langue et, le cas échéant, des manuels de classe.

L'étude du vocabulaire tiré des mêmes textes et éventuellement complété par des enquêtes ciblées, permet de rédiger un dictionnaire de référence aussi complet que possible, lequel servira de base à des lexiques ou des petits dictionnaires d'apprentissage. Les mêmes textes vont servir de contenus à des livres de lecture, des nouvelles, afin d'inciter à la création d'autres ouvrages littéraires si ce n'est à partir de la tradition orale, du moins à partir de l'observation de la vie de tous les jours dans la communauté de locuteurs.

La réalisation d'ouvrages de référence et d'ouvrages utilitaires dans la langue n'est pas une étape marginale dont on peut faire l'économie. Les langues peu dotées ont toujours besoin d'ouvrages de

---

5 Ici, le H inséré devant NG signifie que le ton moyen du A vaut aussi pour le O. C'est l'application d'une règle plus générale qui permet de ne pas recourir aux diacritiques pour informer sur les tons du mot dans certains contextes.

ce genre, même quand elles font partie des langues émergentes et à plus forte raison quand elles sont en danger de disparition. En outre ces ouvrages sont fondamentaux pour renforcer, voire créer, une conscience communautaire chez les locuteurs, surtout quand il n'en reste que peu. Ils sont non seulement indispensables pour enseigner la langue aux jeunes et ainsi augmenter leur capacité d'apprentissage, mais également pour alphabétiser les adultes et par ce moyen leur ouvrir la voie à plus de connaissance.

## 1.2. Développement terminologique

Pour qu'une communauté de locuteurs utilise sa langue comme moyen de communication dans le cyberspace, il est absolument indispensable que cette langue soit dotée des termes techniques nécessaires pour exprimer les réalités de ce nouvel espace. Par exemple, des termes tels que *courrier électronique*, *adresse électronique*, *copie conforme*, *se connecter*, *brancher* (les appareils), *télécharger*, *publier*, *réseaux*, *site*, *page Web* (page de la Toile), *naviguer*, etc. se révéleront complètement indispensables. Rien que le vocabulaire nécessaire pour nommer le matériel informatique que l'on utilise constitue déjà un lexique conséquent qu'il faut entièrement créer, surtout que dans la majorité des cas ce matériel ne fait pas partie des objets culturels traditionnels et n'est donc ni connu, ni nommé. Dès lors, des méthodes de développement terminologique devront être mises en oeuvre au sein de la communauté pour doter la langue de néologismes culturellement acceptables. Le développement terminologique est une activité continue et permanente qui sera de plus en plus générée par la communauté des locuteurs, elle-même au fur et à mesure qu'elle développera une culture accrue du cyberspace. Il ne constitue donc pas une « étape » sur le chemin de l'accès au cyberspace, mais une activité de consolidation qu'il est nécessaire de commencer à un moment donné et de poursuivre indéfiniment.





## 2. Élaboration des ressources informatiques

### 2.1. Systèmes d'écriture et orthographes

Avant d'aborder cette étape, il est utile de préciser certains concepts trop souvent mal compris ou confondus, surtout par les non-spécialistes, à savoir ce qu'on appelle un système d'écriture, un système orthographique (également appelée une convention orthographique ou tout simplement une orthographe) un caractère, un jeu de caractères et une police de caractères.

#### 2.1.1. Les systèmes d'écriture

Considérons l'écriture latine, l'écriture arabe, l'écriture hébraïque, l'écriture chinoise, l'écriture éthiopique, l'écriture hiéroglyphique, l'écriture cunéiforme, l'écriture maya. Chacune utilise un ensemble de symboles spécifiques avec des règles d'assemblage qui lui sont propres. Il s'agit donc de systèmes d'écriture différents. Contrairement au continent américain et au continent européen où les systèmes d'écriture réellement utilisés se comptent sur les doigts d'une main, l'Afrique compte une bonne douzaine de systèmes d'écriture, dont les plus communs sont l'écriture latine, l'écriture arabe, l'écriture éthiopique, l'écriture n'ko, et l'écriture tifinagh. Lorsque l'on se propose d'écrire une langue qui n'a jamais été écrite, il convient de savoir dans quel système d'écriture le faire pour que la communauté des locuteurs puisse immédiatement s'en servir. En général, on tient compte du système d'écriture le plus courant dans la région.

#### 2.1.2. Les systèmes orthographiques

Un même système d'écriture, permet d'écrire des langues très différentes. Ainsi l'écriture latine sert à écrire la majorité des langues d'Europe occidentale, d'Amérique et d'Afrique. Cependant, chaque langue exploite différemment les ressources du système d'écriture

en fonction de ses propres structures phonologiques, syntaxiques, énonciatives et sémiotiques. Les règles d'écriture qui en résultent constituent un système orthographique propre à chaque langue. À titre d'exemple, considérons le doublement de consonne simple dans les quatre langues suivantes :

- En italien, la gémination est importante car elle peut entraîner une différence de sens. Par conséquent, le doublement de la consonne sert à noter la gémination et oblige à prononcer fortement la consonne doublée : *tutto* (tout), *oggi* (aujourd'hui).
- En français, le doublement d'une consonne n'a qu'une valeur étymologique ou esthétique et n'entraîne pas du tout une prononciation particulière de la consonne. Comparons par exemple les mots *addition*, *apprécier*, *atteler* avec les mots *adorer*, *apercevoir*, *atelier*. Que l'on écrive ces mots avec ou sans consonnes doubles, la prononciation et le sens restent les mêmes.
- En anglais, même s'il ne s'agit pas d'une règle générale, le doublement de la consonne après une voyelle sert souvent à varier la prononciation de la voyelle en question. Ainsi, le /i/ se prononce [ai] dans *write* (écrire) et *hide* (cacher), mais [i] dans *written* (écrit) et *hidden* (caché). Le doublement du /t/ et du /d/ ne sert pas à signaler une prononciation forte de ces consonnes comme en italien, mais tout simplement à varier la prononciation de la voyelle /i/ qui les précède.
- En sängö (ou sahngo), l'une des règles de l'orthographe réformée utilise le doublement de la consonne simple pour marquer le début d'une suite tonale identique. Ainsi, au lieu d'écrire *täsēmā* (rangée), où le ton moyen est marqué par un tréma sur chacune des trois syllabes du mot, on écrit *tassema*. Le doublement du /s/ suffit à indiquer que le ton moyen de la première syllabe se reporte sur toutes les syllabes qui suivent jusqu'à la fin du mot. De même, *mbâsâmbâlâ*, (sept), qui comporte une suite de tons hauts, est réécrit *mbâssambala* en application de la même règle. Le doublement du /s/ suffit à indiquer que le ton haut du premier /â/ reste valable pour toutes les syllabes suivantes jusqu'à la fin du mot.

Ces quatre exemples montrent bien à quel point les règles orthographiques sont propres à chaque langue, même quand elles utilisent le même système d'écriture.

## 2.2. Jeux et polices de caractères

On appelle *caractère* tout symbole discret et significatif faisant partie d'un système d'écriture. Notons que dans le système d'écriture latine, l'espace qui sépare deux mots est un caractère au même titre que l'accent aigu, une lettre, une virgule ou un chiffre.

### 2.2.1. Les jeux de caractères

Une fois l'orthographe de la langue établie, on peut faire la liste exhaustive des caractères nécessaires pour écrire la langue dans les règles de cette orthographe. Cette liste constitue ce qu'on appelle un *jeu de caractères*. En le comparant avec les caractères standard utilisés en informatique, on peut vérifier si tous les caractères du jeu en font partie ou non. Le cas le plus facile à traiter est celui où tous les caractères du jeu sont présents dans les caractères standard, car dans ce cas, il n'y a rien de spécial à faire. Malheureusement, force est de constater que bien souvent les caractères dont on a besoin ne se trouvent pas dans les jeux de caractères standard des ordinateurs. On parle alors de *caractères spéciaux*. Un jeu de caractères peut comporter, par exemple, sept voyelles et vingt-quatre consonnes, quelques diacritiques (circonflexe, tréma, tilde), dix chiffres (de 0 à 9), des opérateurs mathématiques (+, -, %, <, >, =), des signes de ponctuation et enfin des symboles comme l'astérisque (\*), le paragraphe (§) ou encore les symboles monétaires (\$, £, €). Un autre jeu de caractères pourrait contenir cinq voyelles au lieu de sept et trente ou quarante consonnes au lieu de vingt-quatre, etc.

### 2.2.2. Les polices de caractères

Une police de caractères est d'abord un ensemble de définitions de formes de caractères dessinées dans un même style artistique. Les polices de caractères sont aujourd'hui des programmes qui comprennent, certes, un stock de jeux de caractères, mais aussi des règles de représentation de ces caractères à l'écran et des fonctions infor-

matiques pour les gérer « intelligemment ». La création d'une police de caractères de qualité professionnelle nécessite des connaissances pointues et une grande compétence technique. C'est pourquoi les polices de caractères sont des produits industriels protégés par des marques déposées, tels que *Times New Roman*, *Garamond*, *Arial*, etc. Les polices gratuites sont rarement satisfaisantes, bien qu'il en existe d'excellentes dans le domaine des logiciels libres. En cas de besoin de créer des caractères spéciaux pour écrire une langue après les avoir vainement cherchés dans les polices existantes, il est nécessaire de s'adresser à une personne compétente ou à une entreprise spécialisée pour faire fabriquer des polices avec le jeu de caractères de son choix. L'utilisateur aura, en plus, la garantie que ces polices respectent les normes internationales.

## **2.3. Codage des caractères**

Il est important de bien comprendre comment les caractères sont codés en informatique, ne serait-ce que pour pouvoir poser les bonnes questions aux spécialistes lorsqu'on souhaite faire fabriquer des polices de caractères comportant des caractères spéciaux pour sa langue. La plupart des langues africaines récemment écrites utilisent l'alphabet latin enrichi de nombreux caractères tirés du stock de caractères définis par l'Association internationale de phonétique (API) et adaptés par l'Institut africain international (IAI). À titre d'exemple, nous allons nous limiter à ce cas de figure, sans prendre en compte d'autres systèmes d'écriture. Du reste, la question de caractères spéciaux ne se pose pas pour les autres systèmes d'écriture tels que le n'ko, l'éthiopique, le tifinagh, puisque tous les caractères de ces systèmes sont normalisés et codés, ceux qui ne le seraient pas sont des variantes de caractères normalisés et non des caractères spéciaux ayant des valeurs distinctes.

### **2.3.1. Le codage sur un octet**

Imaginez un fil électrique sur lequel la seule façon de créer un événement est de couper le courant puis de le rétablir. On obtiendrait une valeur nulle (0) quand le courant est coupé et une valeur positive (1) quand le courant est rétabli. En un point A donné, le long du fil, seules ces deux valeurs sont possibles et permettent de répondre « oui

ou non » à la question « le courant passe-t-il? ». On peut donc voir le point A comme une unité minimale de stockage d'informations qui ne peut prendre que deux valeurs: 0 ou 1. On représente cette unité par un chiffre binaire (*binary digit* en anglais) appelé *bit*<sup>6</sup>. Si avec un seul bit on peut exprimer 2 valeurs (0, 1), avec 2 bits on peut exprimer 4 valeurs (00, 01, 10, 11) et avec 3 bits, 8 valeurs (000, 001, 010, 011, 100, 101, 110, 111). La progression est donc exponentielle. Le tableau suivant montre le nombre de valeurs que l'on peut exprimer selon le nombre de bits utilisés :

Nombre de bits	1	2	3	4	5	6	7	8
Nombre de valeurs	2	4	8	16	32	64	128	256

Si par convention nous décidons d'attribuer une lettre différente à chacune des valeurs obtenues avec 3 bits, nous aurions par exemple ceci :

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

Il ne nous serait pas possible de coder toutes les lettres de notre alphabet. En réalité, on n'a jamais codé les caractères sur 3 bits, mais d'abord sur 7 bits, puis sur 8 bits. Les caractères codés sur 7 bits sont numérotés de 0 à 127 et constituent le jeu de caractères appelé ASCII. Ce jeu contient les 26 lettres de l'alphabet latin de base et aucune lettre accentuée. L'ajout d'un seul bit a permis de définir 128 caractères supplémentaires numérotés de 129 à 255. Cette extension a permis d'intégrer des caractères spéciaux, nécessaires aussi bien pour l'imprimerie et le dessin de graphiques que pour diverses langues utilisant des alphabets européens. Les jeux de caractères ainsi obtenus contiennent chacun 256 caractères codés de 0 à 255, mais sont tous partiellement différents. Ils correspondent à la norme ISO-8859-*n* (où *n* indique le numéro d'un jeu particulier). Ces codes sont toujours représentés dans la machine par une suite de bits. Un paquet de 8 bits est appelé un *octet*.

Ce sont les polices de caractères qui établissent le lien entre les codes (suite de 0 et de 1) représentant les caractères dans l'ordinateur et

<sup>6</sup> Mot valise composé par le début de binary et la fin de digit : bi- + -it > bit.

les dessins correspondants (appelés *glyphes*) qui s'affichent à l'écran. Ainsi, la police *Times New Roman* va lire le code 097 correspondant au « a minuscule latin » et affichera à l'écran l'un des « a » suivants : a, **a**, *a*, **a**, selon le choix de l'utilisateur. Ces quatre « a » ont en commun le style propre à la famille de la police *Times New Roman* caractérisé par le plein et le délié et surtout le petit empattement (appelé *serif* en anglais) à la base de la lettre. Il en sera de même pour la police *Arial*, mais avec des lettres « bâton », sans empattements.

### 2.3.2. Limite du codage sur un octet

Avec l'accélération simultanée de la puissance des ordinateurs, des échanges internationaux et des besoins de l'édition, les limites du codage sur un octet ont rapidement été atteintes. L'un des inconvénients majeurs de l'utilisation de ce type de codage réside dans la nécessité d'utiliser un jeu de caractères spécifique pour une langue particulière qui a besoin de caractères spéciaux. Ainsi, le jeu ISO-8859-1 (aussi dénommé Latin-1) associe au code 198 la lettre « *Lettre majuscule latine æ* » c'est-à-dire la ligature « *Æ* ». Mais cette lettre peut être inutilisée par une langue peu dotée qui aurait par contre besoin d'un caractère inexistant tel que le « e ouvert », soit « *ξ* ». En utilisant les outils logiciels adéquats, un utilisateur lambda a la possibilité de redéfinir un caractère codé. Il pourrait donc remplacer « *Æ* » par « *ξ* » dans une police locale. En conséquence, pour lire un texte écrit dans cette langue avec cette police locale, ladite police doit être installée sur la machine. Cependant, une police bien faite est un produit industriel, souvent commercialisé et que l'on ne peut donc pas distribuer gratuitement. L'utilisateur occasionnel n'a pas non plus envie d'investir dans une police nouvelle chaque fois qu'il change de langue de consultation ! Même si l'on dispose de polices gratuites, tout envoi d'un document (à un collègue, un éditeur ou un imprimeur) devra obligatoirement être accompagné d'une police locale « jetable » après usage ! Ces contraintes ne représentent qu'une petite partie des inconvénients des polices locales, qui rendent, par ailleurs, d'appréciables services à leurs utilisateurs lorsqu'elles sont localement partagées. De toutes les façons, avec un maximum de 256 codes possibles, le codage sur un seul octet est évidemment insuffisant pour coder les 2000 caractères de base de l'écriture chinoise, et ceci vaut aussi pour le japonais et le coréen.

# 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

Source: Projiciels BPI - <http://www.projiciels-bpi.ca/tcaoc/apercu.html>

### 2.3.3. Le codage sur plusieurs octets

Pour dépasser les limitations du codage sur un octet, le consortium Unicode et l'Organisation internationale de normalisation (ISO) ont mis au point la norme ISO/IEC-10646 qui code sur deux octets un jeu de 65 536 caractères, appelé « JUC-Unicode – Jeu universel de caractères<sup>7</sup> » (en anglais: *Universal Characters Set* – UCS). De quoi loger tous les caractères de tous les systèmes d'écriture du monde! C'est parfait pour les idéogrammes chinois, mais pour les caractères latins qui se contentaient jusqu'ici d'un codage sur un octet, le codage sur deux octets est un habit un peu trop ample. Aussi, pour assurer la compatibilité avec les documents existants, plusieurs formats de codage ont été proposés pour ces caractères codés sur deux octets. Le plus populaire d'entre eux est le format de codage UTF-8.

Cependant, il est rare que l'on ait besoin d'assurer la conversion soi-même, les logiciels d'application s'en chargent généralement seuls. Ce qu'il est important de bien comprendre avec la norme Unicode, c'est que chaque caractère ne peut être défini qu'une fois, en ce sens qu'un même code numérique ne peut renvoyer qu'à une seule définition de caractère et vice versa.

## 2.4. Premières ressources de base

### 2.4.1. Jeu et police de caractères

La première ressource informatique à élaborer pour une langue qui n'en a pas est la définition d'un jeu de caractères à partir de la liste des caractères retenus pour l'orthographe de cette langue. Si le jeu de caractères dont on a besoin ne contient aucun caractère spécial, on peut utiliser immédiatement tous les logiciels courants. Par contre, si des caractères spéciaux sont nécessaires, alors, avant de se lancer dans la création d'une police particulière, il faudra vérifier que ces caractères ne sont pas déjà présents dans des polices spéciales

---

7 Voir la page du site du consortium Unicode : < <http://www.unicode.org/fr/charts/>>.



disponibles sur Internet, tels que *Lucida Sans Unicode*, *Gentium*, *Doulos SIL*<sup>8</sup> ou les polices africaines des Progiciels BPI<sup>9</sup>. À défaut, il sera alors nécessaire de définir son propre jeu de caractères. Il est fortement conseillé de vérifier que ce jeu pourra servir aussi pour d'autres langues de la région afin de mieux rentabiliser les efforts et de garantir la possibilité de produire plus tard des textes multilingues sans avoir à changer de polices d'une langue à l'autre. Deux bons exemples de définition de jeux de caractères 8 bits (mais compatibles Unicode) sont fournis par les Progiciels BPI. Chacun de ces deux jeux (*aifful* et *afrlin*) couvre plusieurs langues<sup>10</sup>.

Une fois le jeu de caractères défini, il faut créer la police correspondante. De nos jours, les experts déconseillent vivement la création de nouvelles polices 8 bits en raison des limitations évoquées plus haut. Il vaut mieux s'adresser à un spécialiste pour créer une police contenant les caractères voulus et respectant la norme Unicode. Ce recours est d'autant plus indispensable si on utilise un système d'écriture non latin et non encore représenté dans Unicode! Le site <<http://www.freelang.com>> propose un service gratuit de création de polices à la demande de ses visiteurs. La police créée est libre, gratuite et partagée sur le site. Plusieurs outils logiciels sont disponibles sur l'Internet pour créer des polices locales, notamment sur les sites suivants :

- 1) <http://scripts.sil.org/TypeDesignResources> (polices prêtes à l'emploi et outils pour en créer) ;
- 2) <http://scripts.sil.org/SILFontList> (polices prêtes à l'emploi à télécharger) ;
- 3) [http://scripts.sil.org/SILEncore\\_Glyphs](http://scripts.sil.org/SILEncore_Glyphs) (pour visualiser les glyphes des polices).

---

8 Voir le site de la SIL : < <http://scripts.sil.org/SILFontList>>.

9 Voir le site des Progiciels BPI : <<http://www.progiciels-bpi.ca/tcao/apercu.html>>.

10 Voir également le site burkinabé : <<http://www.abcburkina.net/sedelan/index.htm>> pour les langues du Burkina Faso.

## 2.4.2. Claviers virtuels

Une fois que l'on dispose d'une police de caractères correspondant au jeu de caractères que l'on s'est choisi, on peut l'installer sur son ordinateur en suivant la procédure d'installation des polices relative au système d'exploitation utilisé. Il est ensuite aisé d'utiliser ces polices, la plupart des traitements de texte modernes possédant une fonction permettant l'insertion de caractères spéciaux. Par exemple, dans le traitement de texte Word, une fenêtre de glyphes s'ouvre et il n'y a qu'à sélectionner le caractère désiré pour l'insérer dans le texte. Malheureusement cette façon de procéder est fastidieuse et ralentit considérablement la saisie d'un texte dans une langue peu dotée. Il est donc indispensable d'élaborer des combinaisons de touches plus pratiques permettant d'accéder aisément aux caractères spéciaux d'une police particulière.

Il est intéressant de remarquer que les langues dominantes des pays industrialisés ont un clavier créé spécialement pour chacune d'entre elles, voire pour certaines variétés régionales d'une même langue. Par exemple, le clavier du français du Québec n'est pas le même que celui du français de France. Non seulement tous ces claviers se distinguent par la position des lettres A Z et Q W, mais encore, ils présentent une très grande liberté dans la disposition des caractères de ponctuation et naturellement des caractères spéciaux (diacritiques, caractères accentués, symboles monétaires, etc.). En Afrique, aucune des grandes langues africaines (*kiswahili*, *zulu*, *hausa*, *fulfulde*, etc.) ne dispose de son propre clavier. Alors, dans le cas des petites langues peu dotées, il ne faut pas espérer disposer de si tôt d'un clavier physique. La solution la plus simple est donc d'élaborer un clavier virtuel.

En se basant sur le clavier physique utilisé dans la région où se parle la langue peu dotée, on commence par créer une table de combinaisons de touches faciles à mémoriser. Par exemple, on décide qu'il faut frapper la suite de touches « < »+« O » (sauf les guillemets et le +) pour obtenir un « o ouvert minuscule », soit « ɔ » et les touches « Maj »+ « > »+« O » pour obtenir la majuscule correspondante « Ɔ ». Considérons la table de combinaisons suivante :

« < » + « O » = « ɔ »

« Maj » + « > » + « O » = « ɔ »

Ces combinaisons ne sont ergonomiques que si les chevrons « < » et « > » sont superposés sur une même touche, avec le second placé au-dessus du premier. Ils sont alors utilisés comme des touches mortes. Aussi, pour pouvoir les utiliser à nouveau en tant que chevrons brisés, il suffit de taper la touche « barre d'espacement » après les avoir tapés. Il faut donc ajouter les lignes suivantes à la table :

« < » + « espace » = « < »

« Maj » + « > » + « espace » = « > »

Les caractères « < » et « > » peuvent alors être utilisés comme modificateurs généraux pour obtenir d'autres caractères spéciaux en minuscule et en majuscule. La chose se complique lorsque l'on doit mettre un ou deux diacritiques sur un caractère déjà modifié, par exemple un accent aigu pour indiquer un ton haut et un tilde pour indiquer une nasalité. Il est alors nécessaire de s'assurer que le logiciel utilisé pour interpréter la table supporte l'enchaînement des commandes du type : « accent aigu » + « tilde » + « modificateur » + « caractère de base » = « caractère modifié accentué nasal ». Actuellement, l'un des meilleurs logiciels pour la création de claviers virtuels est Keyman<sup>(TM)</sup> de la société Travultsoft. Des claviers virtuels pour les langues africaines réalisés avec ce logiciel sont disponibles à l'adresse suivante : <http://scripts.sil.org/SILKeyboards>.

Il suffit d'installer Keyman sur son ordinateur pour gérer les claviers et polices correspondantes. Il est possible d'installer plusieurs claviers et de passer de l'un à l'autre par une simple combinaison de touches telle que « Alt + K ». Il faut simplement s'assurer que la combinaison de touches choisie n'est pas déjà utilisée pour autre chose par un des logiciels installés sur l'ordinateur.

### 2.4.3. Logiciels de traitement de corpus assisté par ordinateur

Dès lors que l'on utilise des caractères spéciaux, il faut penser aux outils de base qui permettront d'effectuer des tris alphabétiques, des conversions de codes des caractères, des alignements de textes, etc. Ces outils se révéleront très utiles pour la manipulation et le traitement des textes à produire dans la langue peu dotée en vue d'un affichage sur la Toile. Les outils les plus performants en matière de traitement des langues sont les Progiciels BPI cités ci-après.

Le tableau 1 ci-dessous présente la suite des cinq progiciels de traitement de corpus assisté par ordinateur (TCAO) suivants: **Alibi**, **Concorde**, **Recode**, **Ventile**, **Vocable**.<sup>11</sup>

Progiciels	TABLEAU 1 – Description des progiciels de TCAO
<b>Recode</b>	<i>Recode</i> est un progiciel de conversion automatique de jeux de caractères. Il permet de traiter et de recoder environ 175 jeux de caractères différents et une douzaine de <i>surfaces</i> de fichiers. Selon la paire de jeux de caractères indiquée dans la commande d'appel (d'un jeu de départ vers un jeu d'arrivée), <i>Recode</i> effectue un recodage du ou des fichiers d'entrée. Chaque jeu de caractères pouvant être converti vers la plupart des 174 autres, plusieurs milliers de conversions différentes sont alors possibles.
<b>Ventile</b>	<i>Ventile</i> est un progiciel de production de statistiques textuelles. En matière d'éléments textuels, <i>Ventile</i> permet de compter le nombre de paragraphes, de phrases, de mots et de caractères par fichier. En matière de statistiques textuelles, <i>Ventile</i> produit les mesures de fréquence absolue, trois mesures de tendance centrale (le mode, la médiane et la moyenne arithmétique) et cinq mesures de dispersion (le minimum, le maximum,

<sup>11</sup> Source: <http://www.progiciels-bpi.ca/tcao/apercu.html>.

l'écart quartile, l'écart moyen et l'écart type). Les résultats statistiques sont affichés numériquement sous la forme d'un tableau et graphiquement sous la forme d'un histogramme.

---

**Vocable**      Vocable est un progiciel de production de listes de vocabulaire. Cet outil permet de dépouiller des textes et de fabriquer la liste des vocables actualisés dans ces textes. Ces listes peuvent être triées en ordre alphabétique habituel (de gauche à droite), en ordre alphabétique inverse (de droite à gauche) ou en ordre de fréquence décroissante. Chaque vocable peut être accompagné de la fréquence de ses occurrences et de la liste de toutes ses références textuelles pour chaque occurrence.

---

**Concorde**      Concorde est un progiciel de production de concordances de mots en contexte. Cet outil permet de dépouiller des fichiers textuels et de produire une liste alphabétique des éléments-vedettes avec leur contexte immédiat. L'élément-vedette d'une concordance peut être basé sur les mots dans leurs contextes phrastiques ou sur les caractères dans leurs contextes de mots. Les mots-vedettes peuvent être triés en ordre alphabétique habituel ou en ordre alphabétique inverse.

---

**Alibi**            Alibi est un progiciel d'alignement bi-textuel (ou bilingue). Cet outil permet de dépouiller en parallèle deux fichiers et d'aligner automatiquement les constituants textuels provenant de cette paire de fichiers. Comme constituants textuels, Alibi peut aligner les paragraphes, les phrases ou les mots de ces deux textes.

---

La réalisation de tels logiciels n'est pas à la portée du premier venu. Ceux-ci se révèlent être très performants et ont l'avantage de fonctionner aussi bien sous les systèmes d'exploitation les plus populaires (Windows et Linux). Associés aux jeux de caractères, aux polices et aux claviers virtuels, ces logiciels constituent des ressources de base pour le traitement d'une langue peu dotée.

## 2.5. Localisation des logiciels

On entend par localisation de logiciels l'activité qui consiste à prendre un logiciel conçu dans une langue et un pays étranger et à l'adapter à la langue et à la culture d'un autre pays. Tout logiciel est composé de deux parties : le code informatique, et l'interface homme-machine. Il n'est jamais question de toucher au code sauf dans des cas exceptionnels comme la nécessité de modifier le sens de l'écriture ou l'ordre de tris alphabétiques lorsque ceux-ci n'étaient pas prévus pour être localisables. Aujourd'hui la plupart des logiciels sont prévus pour être localisés et par conséquent, la partie interface homme-machine est aisément identifiable et accessible, soit grâce à un fichier prévu à cet effet, soit au moyen d'outils appropriés.

Sont concernés par la localisation d'un logiciel, les menus (déroulants, fixes, flottants ou contextuels), les boîtes de dialogues, les notifications, les informations d'aide en ligne, les boutons et commandes de navigation, les graphismes (logos, drapeaux, armoiries, illustrations, apparence du site, ou *skin* en anglais et design), les unités de mesure (distance, poids, volume, temps, monnaie, densité, etc., à convertir en fonction des normes régionales), les tutoriels et la documentation (manuels d'installation, d'utilisation, et de référence). L'objectif de la localisation est bien évidemment de permettre à l'utilisateur de travailler dans un environnement informatique qui lui est culturellement et linguistiquement familier et donc facile à maîtriser. En conséquence, sa langue devient une langue de travail des technologies de l'information et de la communication. Ainsi, la localisation des logiciels dans une langue peu dotée contribue à valoriser cette langue aux yeux des utilisateurs, notamment de ses propres locuteurs.

Lorsqu'on souhaite traduire des logiciels dans une langue peu dotée, à moins d'en être le propriétaire, il vaut mieux localiser des logiciels libres de droits, comme ceux distribués sous licence GPL (*General Public Licence*)<sup>12</sup>. La distribution GNU/Linux (Ubuntu Linux<sup>13</sup>, par

---

12 Pour plus d'informations sur les licences des logiciels libres, voir <http://www.gnu.org/licenses/license-list.fr.html>.

13 Voir <http://www.ubuntu.com>.

exemple, pour les pays africains) offre au monde entier une plateforme collaborative en ligne appelée *Rosetta*<sup>14</sup> qui permet à quiconque de choisir un logiciel libre et de le traduire dans sa langue. Un ensemble d'outils de gestion de projet est fourni sous le nom de *Launchpad*<sup>15</sup> et constitue une aide précieuse pour l'utilisateur qui se lance dans la localisation d'un des logiciels. Un tel dispositif offert gratuitement à toutes les langues du monde est vraiment une initiative extrêmement rare et généreuse, une véritable aubaine pour les langues peu dotées qui peuvent en profiter pour accéder plus facilement au statut de langue de travail du cyberspace. Cependant, la localisation de logiciels n'est pas une étape indispensable pour accéder au cyberspace. Une langue peu dotée peut y accéder sans pour autant avoir mis en place un projet de localisation. Dès lors que la langue dispose des premières ressources de base (cf. section 2.4.), des contenus culturels et linguistiques pour le cyberspace peuvent déjà commencer à être élaborés.

---

14 Voir <https://launchpad.net/rosetta>.

15 Voir <http://launchpad.net>.



## 3. Élaboration des contenus culturels

### 3.1. Contenus culturels

Les contenus culturels sont de nature textuelle, sonore ou iconographique. Tous peuvent être numérisés et portés sur la Toile ou transportés à travers le cyberspace pour être partagés.

#### 3.1.1. Le texte

Outre le corpus utilisé pour décrire la langue et créer ses premiers ouvrages de référence, il convient de collecter ou de créer encore davantage de textes, de tout genre, depuis les petites annonces jusqu'aux récits d'épopées légendaires, traductions d'œuvres littéraires ou encore rédaction de journaux. Tous ces textes feront ensuite l'objet de divers traitements informatiques grâce aux logiciels de traitement de corpus assisté par ordinateur (TCAO). Il serait intéressant de diffuser les résultats de ces traitements sur la Toile car ils peuvent servir à améliorer l'étude de la langue elle-même.

#### 3.1.2. Le son

Les documents sonores peuvent être constitués d'enregistrements de listes de mots ou de phrases vocalisées, de chansons, de récitals, de récits de contes et légendes, d'interviews, de reportages, mais également de musique instrumentale, de sonnaillies et de bruits de la vie et de la nature témoignant de l'environnement culturel local. La numérisation de tous ces documents sonores nécessite du matériel d'enregistrement de qualité professionnelle qui peut être très onéreux pour qui vient du tiers-monde et parle une langue peu dotée. La création musicale en particulier peut demander des appa-



reils coûteux. Il en est de même pour les reportages naturalistes captant les chants des oiseaux et les cris des insectes nocturnes dans le souci d'étudier l'environnement local et l'écologie naturelle. Il suffit que ces documents sur l'environnement soient accompagnés de commentaires dans la langue locale pour que celle-ci s'en trouve valorisée.

Tous les fichiers audio ainsi obtenus doivent être filtrés pour être débarrassés des souffles et autres bruits parasites. Un long travail en studio s'impose ensuite pour découper et étalonner chaque section de l'enregistrement sonore afin de constituer des fichiers plus légers, plus facilement transportables. Aujourd'hui, il est possible de diffuser des documents sonores sur Internet et dans le cyberspace comme on le fait pour une station radio ou pour les films. Mais sans l'image ces documents sonores, aussi intéressants soient-ils, ressemblent à la télévision en noir et blanc à l'heure de la couleur.

### **3.1.3. L'image**

Les documents iconographiques regroupent les dessins, les logos, les photos, et la vidéo. Si les logiciels de dessins sont extrêmement courants, les appareils de photographie et les caméras de qualité sont un peu chers. Toutefois, on ne peut pas s'en passer pour filmer une danse, une cérémonie, une scène de vie, une vue panoramique, un reportage, etc. Toutes ces images peuvent être montées et retouchées pour en améliorer la qualité. Les fichiers image sont aujourd'hui numérisés dès leur création. Il faut néanmoins s'assurer de la très bonne qualité de l'image, car les traitements nécessitent plusieurs copies et cela peut ternir quelque peu la qualité de l'image diffusable.

Chaque fois que cela est possible, il est préférable d'associer son, image et texte dans la production de documents destinés à la diffusion sur l'Internet. Mais si une langue peu dotée n'était pas encore écrite, il est toujours possible de créer des documents sonores et iconographiques, à l'instar d'un film et/ou de photos commentées et de les placer sur la Toile.

## 3.2. Préparer les voies d'accès au cyberspace

### 3.2.1. Le site Internet

Le site Internet est la meilleure façon d'affirmer la présence d'une langue dans le cyberspace, car on peut le décliner à l'infini, depuis la page personnelle jusqu'au portail multimédia offrant des services divers et variés. Avant de se lancer dans la création d'un site, il convient de prendre en compte les observations suivantes :

Un site écrit dans une langue peu dotée ne peut être lu que par les locuteurs de cette langue. Cela devient difficile si ces locuteurs ne savent pas lire et écrire dans leur langue. Le site peut alors servir à les y encourager. Quand on vient de mettre au point une orthographe pour la langue peu dotée, il est préférable d'envisager de créer un site bilingue utilisant simultanément la langue peu dotée et une langue de plus grande diffusion parlée dans la région<sup>16</sup>. Dans ce cas, il faut absolument veiller à ce que la langue peu dotée soit présente dans les boutons et commandes de navigation ainsi que dans tous les messages destinés au visiteur du site afin d'éviter qu'elle ne se retrouve en position d'objet de curiosité et perde son statut de langue de travail sur la Toile.

Les boutons de navigation, les menus de commandes, les titres, les logos, les textes décoratifs, les animations et les styles de présentation du site sont des éléments à préparer à l'avance, avec des logiciels spécialisés comme *Button Studio*<sup>TM</sup>. Les fichiers texte destinés à être affichés sur la Toile sont écrits en langage HTML (*HyperText Markup Language*). On peut écrire une page sur la Toile avec un simple logiciel de traitement de texte comme *Open Office* (gratuit et de diffusion libre) qui est capable d'enregistrer un fichier au format HTML. Il existe également de très nombreux outils mieux adaptés à la création de sites, des plus simples et souvent gratuits comme *Nets-*

---

16 Voir par exemple, le site de l'Association YSB SAHNGO : <<http://sango.free.fr>>.

*cape*<sup>TM</sup> *Composer* aux plus sophistiqués et payants comme *Dream-weaver*<sup>TM</sup>. Pour débiter, on trouvera de quoi s'informer et se former à l'adresse suivante : <http://www.unice.fr/urfist/ResInternet.html>.

La composition de la page du site est une affaire de goût et de culture.

### **3.2.2. Le courrier électronique et la messagerie instantanée**

C'est l'outil de communication le plus utilisé de l'Internet. Mais, pour des raisons de sécurité, il est très difficile de mesurer le taux d'utilisation de telle ou telle langue dans les échanges mondiaux par courrier électronique. L'utilisation d'une langue peu dotée dans ce service passera donc inaperçue. Par contre, la localisation d'au moins un logiciel de messagerie et de messagerie instantanée constitue une vraie contribution à l'équipement de la langue peu dotée.

### **3.2.3. La téléphonie IP et la téléphonie mobile**

La téléphonie IP (*Internet Protocol*) permet de téléphoner par voix à distance en temps réel via l'Internet. Cette technologie peut donc être utilisée librement par les locuteurs d'une langue peu dotée. Il leur suffit de télécharger un logiciel adéquat comme *Skype*<sup>TM</sup>, le plus populaire du moment, ou encore *Google Talk*, *Voipbuster*, *Messenger*, etc. et de l'installer sur leur ordinateur. Seule la localisation de l'interface utilisateur du logiciel peut constituer une ressource propre pour la langue peu dotée. Il en est de même pour la téléphonie mobile dont la seule façon de contribuer à l'équipement de la langue peu dotée est de localiser l'interface utilisateur des téléphones, comme cela se fait en Afrique du Sud.

### **3.2.4. Les fora et listes de diffusion**

L'installation d'un forum ou d'une liste de diffusion dans un site permet à la communauté des locuteurs de la langue peu dotée de dialoguer entre eux par écrit. Sous réserve que l'environnement logiciel supporte les caractères spéciaux requis, cette technologie peut

s'avérer un excellent moyen de faire vivre la langue peu dotée dans le cyberspace. En effet, sans une communauté qui anime et active les échanges dans la langue, celle-ci court le risque de n'être qu'un élément de décor inerte dans le cyberspace. Or ce n'est certainement pas là le but de tout l'effort que l'on peut faire pour amener une langue peu dotée dans le cyberspace. De nombreux portails commerciaux et non commerciaux offrent la possibilité de créer des fora communautaires gratuitement. En voici quelques-uns des plus populaires :

- <http://www.yahogroups.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. Développer la communauté des utilisateurs

### 4.1. Création d'une communauté d'utilisateurs

L'existence d'une communauté d'utilisateurs instruite dans sa langue est absolument nécessaire pour faire vivre cette langue dans le cyberspace quand elle y accèdera. En effet, si on oeuvre pour amener des langues peu dotées à occuper une place dans le cyberspace, ce n'est pas du tout pour en faire du mobilier ou des figurantes, mais bien pour qu'elles y vivent et y soient productives pour leurs locuteurs, lesquels sont les premiers concernés par la maîtrise et la promotion de leurs langues dans le cyberspace. Il est donc primordial d'apprendre aux locuteurs d'une langue peu dotée, notamment aux jeunes, à lire et à écrire dans cette langue. L'initiation à l'informatique et aux logiciels de bureautique et de communication devrait être accompagnée de l'enseignement de la langue peu dotée, par exemple, dans le cadre d'activités associatives, si ce n'est à l'école. Les associations locales sont généralement plus à même de mettre en place des formations de ce type que les écoles qui dépendent lourdement du système éducatif national. Malheureusement le plus souvent les associations locales ont peu d'argent et ne bénéficient pas de soutien, surtout lorsqu'il est question de langues peu dotées. Il est clair que tout soutien apporté à des associations locales engagées dans le développement des langues en relation avec les TIC contribue à la promotion de la diversité linguistique et culturelle dans le cyberspace.

#### 4.1.1. Régénération d'une communauté

Lorsqu'une langue peu dotée n'a plus que quelques locuteurs, il est absolument indispensable de consolider le nombre des locuteurs en enseignant cette langue à des jeunes par tous les moyens disponibles : activités culturelles, centres de lecture et de culture, activités associatives, et si possible à l'école. Il est en fait question de recréer

un environnement vital pour la langue. Plus la langue sera enseignée et valorisée, plus elle pourra servir à produire du contenu culturel que l'on pourra mettre sur un site. Il faut être conscient que cela demande un travail colossal, un investissement considérable, tant en ressources financières qu'humaines et une détermination sans faille à long terme. Les associations culturelles peuvent être très actives localement pour réhabiliter une langue en danger et maintenir un intérêt croissant en sa faveur, notamment chez les décideurs administratifs et politiques.

### **4.1.2. Consolidation d'une communauté**

Prenons l'exemple d'une langue peu dotée mais qui dispose d'une population de locuteurs non négligeable. L'éducation des gens dans leur langue est une entreprise de longue haleine et qui exige de mobiliser de nombreuses ressources humaines et financières. Dans ce cas, l'accès de la langue au cyberspace est l'un des moyens de mobiliser et dynamiser la communauté des utilisateurs de cette langue. La consolidation de cette communauté ne peut donc pas être une activité préalable à l'accès de la langue au cyberspace, mais une activité permanente qui concourt à faire vivre la langue dans cet espace. Ici bien plus encore que dans le cas des langues en danger de disparition qui ne jouissent pas d'un nombre aussi considérable de locuteurs, il est nécessaire d'impliquer les décideurs politiques et administratifs dans le processus de valorisation de la langue peu dotée. Sans cela, les efforts des associations ne suffiront pas à garantir durablement son essor, faute de moyens conséquents sur le long terme.

## **4.2. Aspects politiques et socioculturels**

### **4.2.1. Des arguments pour convaincre**

Comme expliqué ci-dessus, l'action d'une association pour développer une communauté d'utilisateurs d'une langue peu dotée est limitée par les moyens financiers qu'elle peut mobiliser. Il est donc essentiel de sensibiliser les responsables politiques à tous les niveaux sur tous les avantages qu'il y a à soutenir l'instrumentalisation des langues

peu dotées et le renforcement des capacités de leurs locuteurs. Avec le soutien des administrations locales, il est plus facile d'envisager l'enseignement des langues à l'école, notamment à l'école primaire où la langue locale est la mieux placée pour servir de langue des premières acquisitions des savoirs chez l'enfant. Un enseignement bilingue ne peut réussir, même localement, qu'en étant inscrit dans le cadre juridique de l'Éducation nationale. Plusieurs expériences d'enseignement bilingue impliquant des langues africaines au Burkina Faso, au Sénégal, au Mali, au Burundi et au Rwanda ont montré que les enfants apprennent bien mieux dans leurs langues maternelles et obtiennent de bien meilleurs résultats scolaires. À terme, ils deviennent des citoyens plus instruits. Or un homme instruit est bien plus capable d'avoir une vision globale et une compréhension approfondie des problèmes complexes du monde moderne et de contribuer à y apporter des solutions durables dans sa sphère de vie et d'activités. C'est une personne qui pourra mener à bien un projet socio-économique autant pour lui-même que pour sa famille, son village et sa région. L'éducation multilingue impliquant la langue maternelle est la clé de l'essor économique.

Outre l'avantage éducatif, une politique favorisant la diversité linguistique peut générer des emplois dans les domaines de la traduction, de la rédaction de manuels, de l'édition ou encore de l'animation. La diversité linguistique est donc bien un enjeu économique et le développement des langues peu dotées y contribue. Lorsque la langue peu dotée concerne une région entière d'un pays, c'est toute la politique linguistique nationale qui est concernée par son développement. Celui-ci devra alors s'inscrire dans un plan national d'aménagement politique, mettant en oeuvre une politique linguistique nationale clairement affirmée en faveur de la diversité linguistique. En outre, une telle politique doit respecter les droits linguistiques qui font partie intégrante des droits fondamentaux de l'homme.

#### **4.2.2. Des instruments de référence**

Au-delà des autorités politiques nationales, il est recommandé de rechercher aussi des soutiens auprès d'organisations internationales qui oeuvrent pour la promotion et la valorisation de la diversité linguistique dans le monde, ce qui implique au moins une attitude

bienvueillante vis-à-vis des efforts que peuvent fournir des associations locales pour développer des langues peu dotées. Il faut savoir qu'au cours des quarante dernières années, les organisations internationales ont organisé de nombreuses rencontres sur les langues des populations locales, souvent qualifiées de « langues nationales » par opposition aux langues européennes héritées de la colonisation, notamment en Afrique et sur le continent américain. Ces rencontres ont permis l'élaboration et l'adoption de plusieurs textes, appelés « instruments internationaux », sur lesquels on peut s'appuyer pour mener des actions concrètes. Parmi les manifestations organisées par les États et les Organisations internationales, Amadou Touré<sup>17</sup> note pour l'Afrique, entre autres :

1. Réunion de l'UNESCO sur l'emploi des langues vernaculaires dans l'enseignement (Monographies sur l'éducation de base) – Paris, 1953 ;
2. Conférence régionale de l'UNESCO sur la planification et l'organisation des programmes d'alphabétisation en Afrique – Abidjan, mars 1964 ;
3. Congrès de la Société africaine de linguistique (SLAO) – Accra, 1965 ;
4. Réunion d'experts organisée par l'UNESCO pour l'unification des alphabets des langues nationales : fulfulde, hawsa, kanuri, mandingue, songhay-zarma, tamasheq – Bamako, 28 février – 5 mars 1966 ;
5. Conférence intergouvernementale sur les politiques culturelles en Afrique (UNESCO – OUA) – Accra, 1975 ;
6. Réunion organisée par l'Agence de coopération culturelle et technique sur la promotion des langues nationales – Yaoundé, 1977 ;
7. Conférence internationale de l'ACCT sur « la recherche linguistique, l'emploi et l'enseignement des langues en Afrique : les moyens de renforcer la coopération entre les États » – Yaoundé, 1983 ;
8. Réunion de l'UNESCO sur les stratégies de promotion des langues africaines – Conakry, 1984 ;

---

<sup>17</sup> Document de travail pour les colloques de l'Académie africaine des langues, 2006.



9. Réunion d'experts de l'UNESCO sur la promotion des langues africaines comme instruments de culture et d'éducation permanente – Yaoundé ;
10. Réunion sur le projet de Charte d'action pour la promotion et l'utilisation des langues africaines dans l'éducation – Accra, août 1996 ;
11. Conférence intergouvernementale sur les politiques linguistiques en Afrique (UNESCO-OUA-ACCT) – Harare, 1997 ;
12. Consultation africaine sur la création de l'Académie africaine des langues – Bamako, 25–27 mai 2001 ;
13. Réunion thématique sur le multilinguisme pour la diversité culturelle et la participation de tous dans le cyberspace, organisée par l'UNESCO en préparation du Sommet mondial sur la société de l'information – Bamako, mai 2005 ;
14. 33<sup>e</sup> session de la Conférence générale de l'UNESCO qui a adopté le projet de résolution initié par l'ACALAN, présenté par le Bénin et relative à l'organisation de cinq colloques régionaux sur les langues transfrontalières et les langues de moindre diffusion en Afrique – Paris, octobre 2005.

Plusieurs grandes rencontres ont permis l'élaboration d'instruments normatifs et de textes réglementaires tels que :

1. Charte de l'OUA (1963) ;
2. Manifeste culturel panafricain d'Alger (1969) ;
3. Charte culturelle de l'Afrique – OUA (1976, Port Louis) ;
4. Résultats de la 1<sup>re</sup> conférence des ministres africains de la Culture de l'OUA (Port-Louis, 1986) ;
5. Plan d'action de Lagos de l'OUA pour le développement économique de l'Afrique (1980) ;
6. Plan d'action linguistique pour l'Afrique (Addis-Abeba, 21–25 juillet 1986) ;
7. Plan décennal de l'OUA sur les langues et les traditions orales (1987) ;
8. Traité d'Abuja créant la Communauté économique africaine (1991) ;
9. Plan régional de collecte des traditions orales en Afrique australe (Harare, 1993) ;

10. Plan d'aménagement linguistique de l'espace francophone 1990-2000 (ACCT 1993);
11. Programme d'action de la décennie de l'éducation en Afrique (Harare, 1999);
12. Traité créant l'Union africaine (Lomé, 2000);
13. Résolution de la 31<sup>e</sup> session de la Conférence générale de l'UNESCO sur l'Académie africaine des langues (2001);
14. Décision CM/Déc.613 (LXXIV) du 37<sup>e</sup> Sommet des chefs d'États et de Gouvernement de l'OUA créant l'Académie africaine des langues (juillet 2001);
15. Déclaration universelle sur la diversité culturelle adoptée par la 31<sup>e</sup> session de la Conférence générale de l'UNESCO (2001);
16. Recommandations sur la promotion et l'usage du multilinguisme et l'accès universel au cyberspace adoptées par la 32<sup>e</sup> session de la Conférence générale de l'UNESCO (octobre 2003);
17. Textes fondateurs du Sommet mondial sur la Société de l'information (Genève, 2003 et Tunis, 2005);
18. Décision AU/Dec.92 (VI) de la sixième session ordinaire de la Conférence de l'Union africaine relative à une seconde décennie de l'éducation pour l'Afrique (2006–2015) et au cadre d'action de la seconde décennie (Doc. EX. CL/224–VIII Rev.2) (Khartoum, 2006);
19. Décision AU/Dec.94 (VI) de la sixième session ordinaire de la Conférence de l'Union africaine sur la Charte révisée de la renaissance culturelle africaine (Khartoum 2006);
20. Décision AU/DEC.95 (VI) de la sixième session ordinaire de la Conférence de l'Union africaine sur les statuts de l'Académie africaine des langues (Khartoum, 2006);
21. Décision AU/Dec.96 (VI) de la sixième session Ordinaire de la Conférence de l'Union africaine sur le lien entre la culture et l'éducation (Khartoum, 2006);
22. Décision AU/Dec.98 (VI) de la sixième session ordinaire de la Conférence de l'Union africaine relative à la proclamation de 2006 comme l'Année des langues africaines (Khartoum 2006).

Des institutions ont été créées, entre autres :

- L'ICA (Institut Culturel Africain) – Dakar, Sénégal;
- Le CELHTO (Centre d'études linguistiques et historiques par tradition orale) – Niamey, Niger;

- Le CERDOTOLA (Centre régional de documentation sur les traditions orales et les langues africaines) – Yaoundé, Cameroun ;
- L'EACROTANAL (*East African Center of Research on Oral Traditional and National Languages*) – Zanzibar, Tanzanie ;
- Le CICIBA (Centre international des civilisations bantou) – Libreville, Gabon ;
- Le BASE (Bureau africain des sciences de l'éducation) devenu aujourd'hui IPED (Institut panafricain de l'éducation pour le développement) – Kinshasa, République Démocratique du Congo.

La plupart de ces institutions, il faut le reconnaître, sont restées en deçà des attentes, faute de moyens financiers pour fonctionner pleinement. Il n'est donc pas évident d'obtenir leur soutien pour mener une action locale de valorisation d'une langue peu dotée. Cependant, avec un peu de persévérance et beaucoup de bonne volonté, on peut arriver à s'en faire des alliés dans sa lutte et ce n'est pas négligeable ! En effet, un inventaire des instruments normatifs existants et un examen des rapports des différentes rencontres organisées au cours de cette longue période, montrent que tous les pays d'Afrique ont, à un moment ou un autre et à des degrés divers, pris des initiatives pour la promotion de leurs langues nationales. C'est pourquoi, il convient de tout mettre en oeuvre pour les amener à s'impliquer encore davantage dans des actions concrètes en faveur de la diversité linguistique. Seule une volonté politique forte et permanente conduira à débloquer les fonds nécessaires pour organiser l'éducation multilingue et le développement et l'usage des langues dans toute la vie de la nation afin que soient générés des biens culturels, économiques et sociaux pour le bien-être de tous.



## 5. Conclusion

---

Dans le souci de promouvoir et de valoriser la diversité linguistique et culturelle dans le cyberspace, il convient d'aider les langues les plus défavorisées à accéder à cet espace. S'il est possible d'y parvenir avec une petite langue orale non écrite et menacée d'extinction, on peut, à plus forte raison, le faire avec toutes les langues peu dotées en meilleure posture.

La première étape consiste à entreprendre les études nécessaires à l'élaboration des ressources linguistiques indispensables, à savoir : une liste des phonèmes, un alphabet et une orthographe, une grammaire, un dictionnaire et un recueil de textes.

La seconde étape est consacrée aux travaux de traitement informatique de la langue afin d'identifier ou d'élaborer les ressources informatiques compatibles, à savoir : un jeu de caractères, au moins une police de caractères, un clavier virtuel, des logiciels de traitement de corpus, lesquels peuvent être utilisés aussi pour affiner l'analyse linguistique de la langue et améliorer les ressources linguistiques.

La troisième étape consiste à élaborer les ressources culturelles et à les conditionner pour les rendre partageables dans le cyberspace. Il s'agit d'enregistrer et de numériser le plus grand nombre possible de données textuelles, sonores et iconographiques et de les préparer à un affichage sur site Internet. Il est également nécessaire de créer les éléments constitutifs d'un site, tels que les menus, les barres de navigation, les titres et autres textes de communication homme-machine. Dans certains cas, la localisation de logiciels est nécessaire pour contribuer à valoriser la langue en tant que langue de travail et à la doter de ressources informatiques complémentaires.

Enfin, il convient d'apprendre à créer un site dans la langue peu dotée, éventuellement en complémentarité avec une autre langue de plus grande diffusion. On trouve sur Internet tous les outils nécessaires pour assurer un tel apprentissage, ainsi que pour créer un forum

et localiser des logiciels libres. Avec un site, un forum, une liste de diffusion, une téléphonie IP, de la musique, des images fixes et des vidéos, la petite langue est désormais installée dans le cyberspace. Mais pour qu'elle y vive, il faut développer une communauté d'utilisateurs capable de s'en servir intensément. Aider les associations locales à développer de telles communautés, c'est contribuer à promouvoir et à valoriser la diversité des langues et des cultures dans le cyberspace.



# Références

---

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

Création de polices en ligne : <http://www.freelang.com>

Plate-forme de traduction et de localisation de logiciels libres :  
<https://launchpad.net/rosetta>

Liste de polices de caractères : <http://scripts.sil.org/SILFontList>

Progiciels BPI : <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>

Site bilingue sahngo-français de l'association YSB SAHNGO pour la promotion de la langue sahngo : <http://sango.free.fr>

Université de Nice - Ressources multimédias, formation à la création de site Internet : <http://www.unice.fr/urfist/ResInternet.html>





# ANNEXE 1 : JEUX DE CARACTÈRES

## AFRFUL<sup>18</sup>

Les jeux de caractères africains *afrrul* permettent de couvrir les alphabets du bambara, de l'éwondo, du français et du fulfulde. Le jeu de caractères *AFRFUL-102-BPI\_OCIL* est un jeu de caractères en version encodée, utilisé pour l'affichage à l'écran des langues africaines suivantes : le bambara, l'éwondo, le fulfulde (ou peuhl). Les noms alias acceptables pour désigner ce jeu de caractères encodés sont *afrrul102bpiocil*, *bambara*, *ewondo*, *fulfulde* ou *bra*.

Le jeu de caractères *AFRFUL-103-BPI\_OCIL* est un jeu de caractères en version translittérée, utilisé pour l'encodage au clavier des langues africaines suivantes : le bambara, l'éwondo, le fulfulde (ou peuhl). Les noms alias acceptables pour désigner ce jeu de caractères translittérés sont *afrrul103bpiocil*, *tbambara*, *tewondo*, *tfulfulde* ou *tbra*. Pour chacun des 255 caractères présentés dans la liste plus bas, voici la signification de chacune des colonnes :

Dec	indique la valeur décimale du caractère
Oct	indique la valeur octale du caractère
Hex	indique la valeur hexadécimale du caractère
Mne	fournit le code mnémotechnique à deux lettres de la norme <i>RFC 1345</i>
UCS2	indique la valeur UCS du caractère
Kbd	fournit la convention d'encodage au clavier de ce caractère
AFRFUL-102-BPI_OCIL	donne le nom du caractère, si possible en français, conformément à la norme <i>ISO 10646</i>

<sup>18</sup> 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 nul
1	001	01	SH	0001		soh début d'en-tête
2	002	02	SX	0002		stx début de texte
3	003	03	EX	0003		etx fin de texte
4	004	04	ET	0004		eot fin de transmission
5	005	05	EQ	0005		enq demande
6	006	06	AK	0006		ack accusé de réception positif
7	007	07	BL	0007		bel sonnerie
8	010	08	BS	0008		bs espace arrière
9	011	09	HT	0009		ht tabulation horizontale
10	012	0a	LF	000A		lf interligne
11	013	0b	VT	000B		vt tabulation verticale
12	014	0c	FF	000C		ff page suivante
13	015	0d	CR	000D		cr retour de chariot
14	016	0e	SO	000E		so hors code
15	017	0f	SI	000F		si en code
16	020	10	DL	0010		dle échappement transmission
17	021	11	D1	0011		dc1 commande d'appareil un
18	022	12	D2	0012		dc2 commande d'appareil deux
19	023	13	D3	0013		dc3 commande d'appareil trois
20	024	14	D4	0014		dc4 commande d'appareil quatre
21	025	15	NK	0015		nak accusé de réception négatif
22	026	16	SY	0016		syn synchronisation



Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
23	027	17	EB	0017		etb fin de transmission de bloc
24	030	18	CN	0018		can annulation
25	031	19	EM	0019		em fin de support
26	032	1a	SB	001A		sub caractère de substitution
27	033	1b	EC	001B		esc échappement
28	034	1c	FS	001C		is4 séparateur de fichier
29	035	1d	GS	001D		is3 séparateur de groupe
30	036	1e	RS	001E		is2 séparateur d'article
31	037	1f	US	001F		is1 séparateur de sous-article
32	040	20	SP	0020	<>	espace
33	041	21	!	0021	!	point d'exclamation
34	042	22	«	0022	«	guillemet
35	043	23	Nb	0023	#	dièse-symbole numéro (croisillon)
36	044	24	DO	0024	\$	symbole dollar
37	045	25	%	0025	%	symbole pour cent
38	046	26	&	0026	&	perluète
39	047	27	'	0027	'	apostrophe
40	050	28	(	0028	(	parenthèse gauche
41	051	29	)	0029	)	parenthèse droite
42	052	2a	*	002A	*	astérisque
43	053	2b	+	002B	+	signe plus
44	054	2c	,	002C	,	virgule
45	055	2d	-	002D	-	tiret-trait d'union, signe moins
46	056	2e	.	002E	-	point
47	057	2f	/	002F	/	barre oblique

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
48	060	30	0	0030	0	chiffre zéro
49	061	31	1	0031	1	chiffre un
50	062	32	2	0032	2	chiffre deux
51	063	33	3	0033	3	chiffre trois
52	064	34	4	0034	4	chiffre quatre
53	065	35	5	0035	5	chiffre cinq
54	066	36	6	0036	6	chiffre six
55	067	37	7	0037	7	chiffre sept
56	070	38	8	0038	8	chiffre huit
57	071	39	9	0039	9	chiffre neuf
58	072	3a	:	003A	:	deux-points
59	073	3b	;	003B	;	point-virgule
60	074	3c	<	003C	<	signe inférieur à
61	075	3d	=	003D	=	signe égal à
62	076	3e	>	003E	>	signe supérieur à
63	077	3f	?	003F	?	point d'interrogation
64	100	40	At	0040	@	a commercial – arobase
65	101	41	A	0041	A	lettre majuscule latine A
66	102	42	B	0042	B	lettre majuscule latine B
67	103	43	C	0043	C	lettre majuscule latine C
68	104	44	D	0044	D	lettre majuscule latine D
69	105	45	E	0045	E	lettre majuscule latine E
70	106	46	F	0046	F	lettre majuscule latine F
71	107	47	G	0047	G	lettre majuscule latine G
72	110	48	H	0048	H	lettre majuscule latine H
73	111	49	I	0049	I	lettre majuscule latine I
74	112	4a	J	004A	J	lettre majuscule latine J
75	113	4b	K	004B	K	lettre majuscule latine K
76	114	4c	L	004C	L	lettre majuscule latine L

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
77	115	4d	M	004D	M	lettre majuscule latine M
78	116	4e	N	004E	N	lettre majuscule latine N
79	117	4f	O	004F	O	lettre majuscule latine O
80	120	50	P	0050	P	lettre majuscule latine P
81	121	51	Q	0051	Q	lettre majuscule latine Q
82	122	52	R	0052	R	lettre majuscule latine R
83	123	53	S	0053	S	lettre majuscule latine S
84	124	54	T	0054	T	lettre majuscule latine T
85	125	55	U	0055	U	lettre majuscule latine U
86	126	56	V	0056	V	lettre majuscule latine V
87	127	57	W	0057	W	lettre majuscule latine W
88	130	58	X	0058	X	lettre majuscule latine X
89	131	59	Y	0059	Y	lettre majuscule latine Y
90	132	5a	Z	005A	Z	lettre majuscule latine Z
91	133	5b	<(	005B	[	crochet gauche
92	134	5c	//	005C	\	barre oblique inversée
93	135	5d	)>	005D	]	crochet droit
94	136	5e	'>	005E	^	accent circonflexe
95	137	5f	_	005F	_	trait bas
96	140	60	'!	0060	`	accent grave
97	141	61	a	0061	a	lettre minuscule latine a
98	142	62	b	0062	b	lettre minuscule latine b
99	143	63	c	0063	c	lettre minuscule latine c
100	144	64	d	0064	d	lettre minuscule latine d
101	145	65	e	0065	e	lettre minuscule latine e
102	146	66	f	0066	f	lettre minuscule latine f
103	147	67	g	0067	g	lettre minuscule latine g
104	150	68	h	0068	h	lettre minuscule latine h
105	151	69	i	0069	i	lettre minuscule latine i

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
106	152	6a	j	006A	j	lettre minuscule latine j
107	153	6b	k	006B	k	lettre minuscule latine k
108	154	6c	l	006C	l	lettre minuscule latine l
109	155	6d	m	006D	m	lettre minuscule latine m
110	156	6e	n	006E	n	lettre minuscule latine n
111	157	6f	o	006F	o	lettre minuscule latine o
112	160	70	p	0070	p	lettre minuscule latine p
113	161	71	q	0071	q	lettre minuscule latine q
114	162	72	r	0072	r	lettre minuscule latine r
115	163	73	s	0073	s	lettre minuscule latine s
116	164	74	t	0074	t	lettre minuscule latine t
117	165	75	u	0075	u	lettre minuscule latine u
118	166	76	v	0076	v	lettre minuscule latine v
119	167	77	w	0077	w	lettre minuscule latine w
120	170	78	x	0078	x	lettre minuscule latine x
121	171	79	y	0079	y	lettre minuscule latine y
122	172	7a	z	007A	z	lettre minuscule latine z
123	173	7b	(!	007B	{	accolade gauche
124	174	7c	!!	007C		barre verticale
125	175	7d	!)	007D	}	accolade droite
126	176	7e	'?	007E	~	tilde
127	177	7f	DT	007F		del suppression
128	200	80	PA	0080		pad caractère de bourre
129	201	81	HO	0081		hop octet supérieur prédéfini
130	202	82	BH	0082		bph arrêt permis ici
131	203	83	NH	0083		nbh aucun arrêt ici
132	204	84	IN	0084		ind index
133	205	85	NL	0085		nel à la ligne

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
134	206	86	SA	0086		ssa début de zone sélectionnée
135	207	87	ES	0087		esa fin de zone sélectionnée
136	210	88	HS	0088		hts arrêt de tabulateur horizontal
137	211	89	HJ	0089		htj tabulateur horizontal avec justification
138	212	8a	VS	008A		vts arrêt de tabulateur vertical
139	213	8b	PD	008B		pld interligne partiel vers le bas
140	214	8c	PU	008C		plu interligne partiel vers le haut
141	215	8d	RI	008D		ri index inversé
142	216	8e	S2	008E		ss2 remplacement unique deux
143	217	8f	S3	008F		ss3 remplacement unique trois
144	220	90	DC	0090		dcs chaîne de commande d'appareil
145	221	91	P1	0091		pu1 usage privé un
146	222	92	P2	0092		pu2 usage privé deux
147	223	93	TS	0093		sts mise en mode transmission
148	224	94	CC	0094		cch annulation du caractère précédent
149	225	95	MW	0095		mw message en attente
150	226	96	SG	0096		sga début de zone protégée
151	227	97	EG	0097		ega fin de zone protégée
152	230	98	SS	0098		sos début de chaîne

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
153	231	99	GC	0099		sgci introducteur de caractère graphique unique
154	232	9a	SC	009A		sci introducteur de caractère unique
155	233	9b	CI	009B		csi introducteur de séquence de commande
156	234	9c	ST	009C		st fin de chaîne
157	235	9d	OC	009D		osc commande de système d'exploitation
158	236	9e	PM	009E		pm message privé
159	237	9f	AC	009F		apc commande de progiciel
160	240	a0	NS	00A0	/_	espace insécable
161	241	a1		0181	B[	lettre maj. latine B crose
162	242	a2		018A	D[	lettre maj. latine D crose
163	243	a3		0194	G[	lettre maj. latine Gamma
164	244	a4		0198	K[	lettre maj. latine K crose
165	245	a5		01A4	P[	lettre maj. latine P crose
166	246	a6		01AC	T[	lettre maj. latine T crose
167	247	a7		01A9	S]	lettre maj. latine ESH
168	250	a8		019D	N]	lettre maj. latine N hameçon (enj)
169	251	a9		014A	N[	lettre maj. latine ENG (Sami)
170	252	aa		0189	D]	lettre maj. latine D africain (edh)
171	253	ab	<<	00AB	<<	guillemet angulaire double vers la gauche

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
172	254	ac		0191	F]	lettre maj. latine F hameçon
173	255	ad	--	00AD	\-	tiret conditionnel (trait d'union virtuel)
174	256	ae		01B3	Y[	lettre maj. latine Y crosse
175	257	af		0193	G[	lettre maj. latine G crosse
176	260	b0	DG	00B0	_DG	symbole degré
177	261	b1		0253	b[	lettre min. latine b crosse
178	262	b2		0257	d[	lettre min. latine d crosse
179	263	b3		0263	g]	lettre min. latine gamma
180	264	b4		0199	k[	lettre min. latine k crosse
181	265	b5		01A5	p[	lettre min. latine p crosse
182	266	b6		01AD	t[	lettre min. latine t crosse
183	267	b7		0283	s]	lettre min. latine esh
184	270	b8		0272	n]	lettre min. latine n hameçon à gauche (enj)
185	271	b9		014B	n[	lettre min. latine eng (Sami)
186	272	ba		0256	d]	lettre min. latine d hameçon rétroflexe
187	273	bb	>>	00BB	>>	guillemet angulaire double vers la droite
188	274	bc		0192	f]	lettre min. latine f hameçon
189	275	bd		0294	?^	lettre latine coup de glotte

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
190	276	be		01B4	y[	lettre min. latine y crosse
191	277	bf		0260	g[	lettre min. latine g crosse
192	300	c0	A!	00C0	A`	lettre maj. latine A accent grave
193	301	c1	A'	00C1	A\'	lettre maj. latine A accent aigu
194	302	c2	A>	00C2	A^	lettre maj. latine A accent circonflexe
195	303	c3	A?	00C3	A~	lettre maj. latine A tilde
196	304	c4	A:	00C4	A»	lettre maj. latine A tréma
197	305	c5		018E	E<	lettre maj. latine E réfléchi (schwa)
198	306	c6	AE	00C6	AE+	lettre maj. latine AE (ligature)
199	307	c7	C,	00C7	C\,	lettre maj. latine C cédille
200	310	c8	E!	00C8	E`	lettre maj. latine E accent grave
201	311	c9	E'	00C9	E\'	lettre maj. latine E accent aigu
202	312	ca	E>	00CA	E^	lettre maj. latine E accent circonflexe
203	313	cb	E:	00CB	E»	lettre maj. latine E tréma
204	314	cc	!!	00CC	!`	lettre maj. latine I accent grave
205	315	cd	I'	00CD	I\'	lettre maj. latine I accent aigu
206	316	ce	I>	00CE	I^	lettre maj. latine I accent circonflexe
207	317	cf	I:	00CF	I»	lettre maj. latine I tréma
208	320	d0		0190	E[	lettre maj. latine E ouvert



Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
209	321	d1	N?	00D1	N~	lettre maj. latine N tilde
210	322	d2	O!	00D2	O`	lettre maj. latine O accent grave
211	323	d3	O'	00D3	O\'	lettre maj. latine O accent aigu
212	324	d4	O>	00D4	O^	lettre maj. latine O accent circonflexe
213	325	d5	O?	00D5	O~	lettre maj. latine O tilde
214	326	d6	O:	00D6	O»	lettre maj. latine O tréma
215	327	d7		0152	OE+	digramme soudé maj. latin OE (ligature OE)
216	330	d8		0186	O[	lettre maj. latine O ouvert
217	331	d9	U!	00D9	U`	lettre maj. latine U accent grave
218	332	da	U'	00DA	U\'	lettre maj. latine U accent aigu
219	333	db	U>	00DB	U^	lettre maj. latine U accent circonflexe
220	334	dc	U:	00DC	U»	lettre maj. latine U tréma
221	335	dd		018E +0301	E<'	lettre maj. latine E réfléchi accent aigu
222	336	de		0186 +0301	O['	lettre maj. latine O ouvert accent aigu
223	337	df		01B2	V[	lettre maj. latine V crosse
224	340	e0	a!	00E0	a`	lettre min. latine a accent grave
225	341	e1	a'	00E1	a\'	lettre min. latine a accent aigu
226	342	e2	a>	00E2	a^	lettre min. latine a accent circonflexe
227	343	e3	a?	00E3	a~	lettre min. latine a tilde

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
228	344	e4	a:	00E4	a”	lettre min. latine a tréma
229	345	e5		0259	e<	lettre min. latine schwa (e culbuté)
230	346	e6	ae	00E6	ae+	lettre min. latine ae
231	347	e7	c,	00E7	c\,	lettre min. latine c cédille
232	350	e8	e!	00E8	e`	lettre min. latine e accent grave
233	351	e9	e’	00E9	e\’	lettre min. latine e accent aigu
234	352	ea	e>	00EA	e^	lettre min. latine e accent circonflexe
235	353	eb	e:	00EB	e»	lettre min. latine e tréma
236	354	ec	i!	00EC	i`	lettre min. latine i accent grave
237	355	ed	i’	00ED	i\’	lettre min. latine i accent aigu
238	356	ee	i>	00EE	i^	lettre min. latine i accent circonflexe
239	357	ef	i:	00EF	i“	lettre min. latine i tréma
240	360	f0		025B	e[	lettre min. latine e ouvert
241	361	f1	n?	00F1	n~	lettre min. latine n tilde
242	362	f2	o!	00F2	o`	lettre min. latine o accent grave
243	363	f3	o’	00F3	o\’	lettre min. latine o accent aigu
244	364	f4	o>	00F4	o^	lettre min. latine o accent circonflexe
245	365	f5	o?	00F5	o~	lettre min. latine o tilde
246	366	f6	o:	00F6	o»	lettre min. latine o tréma

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
247	367	f7		0153	oe+	digramme soudé min. latin oe (ligature ō)
248	370	f8		0254	o[	lettre min. latine o ouvert
249	371	f9	u!	00F9	u`	lettre min. latine u accent grave
250	372	fa	u'	00FA	u\`	lettre min. latine u accent aigu
251	373	fb	u>	00FB	u^	lettre min. latine u accent circonflexe
252	374	fc	u:	00FC	u»	lettre min. latine u tréma
253	375	fd		0259 +0301	e<'	lettre min. latine schwa accent aigu
254	376	fe		0254 +0301	o[´	lettre min. latine o ouvert accent aigu
255	377	ff		028B	v[	lettre min. latine v de ronde



## Annexe 2 : Jeux de caractères AFRLIN<sup>19</sup>

Les jeux de caractères africains *afrlin* permettent de couvrir les alphabets du français, du lingala, du sahnogo et du wolof. Le jeu de caractères *AFRLIN-104-BPI\_OCIL* est un jeu de caractères en version encodée, utilisé pour l’affichage à l’écran des langues africaines suivantes: le lingala, le sahnogo, le wolof. Les noms alias acceptables pour désigner ce jeu de caractères encodés sont *afrlin104bpiocil*, *lingala*, *sango*, *wolof* ou *lin*.

Le jeu de caractères *AFRLIN-105-BPI\_OCIL* est un jeu de caractères en version translittérée, utilisé pour l’encodage au clavier des langues africaines suivantes: le lingala, le sango, le wolof. Les noms alias acceptables pour désigner ce jeu de caractères translittérés sont *afrlin105bpiocil*, *tlingala*, *tsango*, *twolof* ou *tlin*. Pour chacun des 255 caractères présentés dans la liste plus bas, voici la signification de chacune des colonnes:

Dec	indique la valeur décimale du caractère
Oct	indique la valeur octale du caractère
Hex	indique la valeur hexadécimale du caractère
Mne	fournit le code mnémotechnique à deux lettres de la norme <i>RFC 1345</i>
UCS2	indique la valeur UCS du caractère
Kbd	fournit la convention d’encodage au clavier de ce caractère
AFRFUL-102-BPI_OCIL	donne le nom du caractère, si possible en français, conformément à la norme <i>ISO 10646</i>

<sup>19</sup> 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 nul
1	001	01	SH	0001		soh début d'en-tête
2	002	02	SX	0002		stx début de texte
3	003	03	EX	0003		etx fin de texte
4	004	04	ET	0004		eot fin de transmission
5	005	05	EQ	0005		enq demande
6	006	06	AK	0006		ack accusé de réception positif
7	007	07	BL	0007		bel sonnerie
8	010	08	BS	0008		bs espace arrière
9	011	09	HT	0009		ht tabulation horizontale
10	012	0a	LF	000A		lf interligne
11	013	0b	VT	000B		vt tabulation verticale
12	014	0c	FF	000C		ff page suivante
13	015	0d	CR	000D		cr retour de chariot
14	016	0e	SO	000E		so hors code
15	017	0f	SI	000F		si en code
16	020	10	DL	0010		dle échappement transmission
17	021	11	D1	0011		dc1 commande d'appareil un
18	022	12	D2	0012		dc2 commande d'appareil deux
19	023	13	D3	0013		dc3 commande d'appareil trois
20	024	14	D4	0014		dc4 commande d'appareil quatre
21	025	15	NK	0015		nak accusé de réception négatif
22	026	16	SY	0016		syn synchronisation
23	027	17	EB	0017		etb fin de transmission de bloc

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
24	030	18	CN	0018		can annulation
25	031	19	EM	0019		em fin de support
26	032	1a	SB	001A		sub caractère de substitution
27	033	1b	EC	001B		esc échappement
28	034	1c	FS	001C		is4 séparateur de fichier
29	035	1d	GS	001D		is3 séparateur de groupe
30	036	1e	RS	001E		is2 séparateur d'article
31	037	1f	US	001F		is1 séparateur de sous-article
32	040	20	SP	0020	<>	espace
33	041	21	!	0021	!	point d'exclamation
34	042	22	«	0022	«	guillemet
35	043	23	Nb	0023	#	dièse-symbole numéro (croisillon)
36	044	24	DO	0024	\$	symbole dollar
37	045	25	%	0025	%	symbole pour cent
38	046	26	&	0026	&	perluète
39	047	27	'	0027	'	apostrophe
40	050	28	(	0028	(	parenthèse gauche
41	051	29	)	0029	)	parenthèse droite
42	052	2a	*	002A	*	astérisque
43	053	2b	+	002B	+	signe plus
44	054	2c	,	002C	,	virgule
45	055	2d	-	002D	-	tiret-trait d'union, signe moins
46	056	2e	.	002E	.	point
47	057	2f	/	002F	/	barre oblique
48	060	30	0	0030	0	chiffre zéro

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
49	061	31	1	0031	1	chiffre un
50	062	32	2	0032	2	chiffre deux
51	063	33	3	0033	3	chiffre trois
52	064	34	4	0034	4	chiffre quatre
53	065	35	5	0035	5	chiffre cinq
54	066	36	6	0036	6	chiffre six
55	067	37	7	0037	7	chiffre sept
56	070	38	8	0038	8	chiffre huit
57	071	39	9	0039	9	chiffre neuf
58	072	3a	:	003A	:	deux-points
59	073	3b	;	003B	;	point-virgule
60	074	3c	<	003C	<	signe inférieur à
61	075	3d	=	003D	=	signe égal à
62	076	3e	>	003E	>	signe supérieur à
63	077	3f	?	003F	?	point d'interrogation
64	100	40	At	0040	@	a commercial-arobase
65	101	41	A	0041	A	lettre majuscule latine A
66	102	42	B	0042	B	lettre majuscule latine B
67	103	43	C	0043	C	lettre majuscule latine C
68	104	44	D	0044	D	lettre majuscule latine D
69	105	45	E	0045	E	lettre majuscule latine E
70	106	46	F	0046	F	lettre majuscule latine F
71	107	47	G	0047	G	lettre majuscule latine G
72	110	48	H	0048	H	lettre majuscule latine H
73	111	49	I	0049	I	lettre majuscule latine I
74	112	4a	J	004A	J	lettre majuscule latine J
75	113	4b	K	004B	K	lettre majuscule latine K
76	114	4c	L	004C	L	lettre majuscule latine L
77	115	4d	M	004D	M	lettre majuscule latine M

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
78	116	4e	N	004E	N	lettre majuscule latine N
79	117	4f	O	004F	O	lettre majuscule latine O
80	120	50	P	0050	P	lettre majuscule latine P
81	121	51	Q	0051	Q	lettre majuscule latine Q
82	122	52	R	0052	R	lettre majuscule latine R
83	123	53	S	0053	S	lettre majuscule latine S
84	124	54	T	0054	T	lettre majuscule latine T
85	125	55	U	0055	U	lettre majuscule latine U
86	126	56	V	0056	V	lettre majuscule latine V
87	127	57	W	0057	W	lettre majuscule latine W
88	130	58	X	0058	X	lettre majuscule latine X
89	131	59	Y	0059	Y	lettre majuscule latine Y
90	132	5a	Z	005A	Z	lettre majuscule latine Z
91	133	5b	<(	005B	[	crochet gauche
92	134	5c	//	005C	\	barre oblique inversée
93	135	5d	)>	005D	]	crochet droit
94	136	5e	'>	005E	^	accent circonflexe
95	137	5f	_	005F	_	trait bas
96	140	60	'!	0060	`	accent grave
97	141	61	a	0061	a	lettre minuscule latine a
98	142	62	b	0062	b	lettre minuscule latine b
99	143	63	c	0063	c	lettre minuscule latine c
100	144	64	d	0064	d	lettre minuscule latine d
101	145	65	e	0065	e	lettre minuscule latine e
102	146	66	f	0066	f	lettre minuscule latine f
103	147	67	g	0067	g	lettre minuscule latine g
104	150	68	h	0068	h	lettre minuscule latine h
105	151	69	i	0069	i	lettre minuscule latine i
106	152	6a	j	006A	j	lettre minuscule latine j



Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
107	153	6b	k	006B	k	lettre minuscule latine k
108	154	6c	l	006C	l	lettre minuscule latine l
109	155	6d	m	006D	m	lettre minuscule latine m
110	156	6e	n	006E	n	lettre minuscule latine n
111	157	6f	o	006F	o	lettre minuscule latine o
112	160	70	p	0070	p	lettre minuscule latine p
113	161	71	q	0071	q	lettre minuscule latine q
114	162	72	r	0072	r	lettre minuscule latine r
115	163	73	s	0073	s	lettre minuscule latine s
116	164	74	t	0074	t	lettre minuscule latine t
117	165	75	u	0075	u	lettre minuscule latine u
118	166	76	v	0076	v	lettre minuscule latine v
119	167	77	w	0077	w	lettre minuscule latine w
120	170	78	x	0078	x	lettre minuscule latine x
121	171	79	y	0079	y	lettre minuscule latine y
122	172	7a	z	007A	z	lettre minuscule latine z
123	173	7b	(!	007B	{	accolade gauche
124	174	7c	!!	007C		barre verticale
125	175	7d	!)	007D	}	accolade droite
126	176	7e	'?	007E	~	tilde
127	177	7f	DT	007F		del suppression
128	200	80	PA	0080		pad caractère de bourre
129	201	81	HO	0081		hop octet supérieur prédéfini
130	202	82	BH	0082		bph arrêt permis ici
131	203	83	NH	0083		nbn aucun arrêt ici
132	204	84	IN	0084		ind index
133	205	85	NL	0085		nel à la ligne

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
134	206	86	SA	0086		ssa début de zone sélectionnée
135	207	87	ES	0087		esa fin de zone sélectionnée
136	210	88	HS	0088		hts arrêt de tabulateur horizontal
138	212	8a	VS	008A		vts arrêt de tabulateur vertical
139	213	8b	PD	008B		pld interligne partiel vers le bas
140	214	8c	PU	008C		plu interligne partiel vers le haut
141	215	8d	RI	008D		ri index inversé
142	216	8e	S2	008E		ss2 remplacement unique deux
143	217	8f	S3	008F		ss3 remplacement unique trois
144	220	90	DC	0090		dcs chaîne de commande d'appareil
145	221	91	P1	0091		pu1 usage privé un
146	222	92	P2	0092		pu2 usage privé deux
147	223	93	TS	0093		sts mise en mode transmission
148	224	94	CC	0094		cch annulation du caractère précédent
149	225	95	MW	0095		mw message en attente
150	226	96	SG	0096		sga début de zone protégée
151	227	97	EG	0097		ega fin de zone protégée
152	230	98	SS	0098		sos début de chaîne
153	231	99	GC	0099		sgci introducteur de caractère graphique unique

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
154	232	9a	SC	009A		sci introducteur de caractère unique
155	233	9b	CI	009B		csi introducteur de séquence de commande
156	234	9c	ST	009C		st fin de chaîne
157	235	9d	OC	009D		osc commande de système d'exploitation
158	236	9e	PM	009E		pm message privé
159	237	9f	AC	009F		apc commande de progiciel
160	240	a0	NS	00A0	/_	espace insécable
161	241	a1		0190 +0300	E[	lettre maj. latine E ouvert accent grave
162	242	a2		0190 +0301	E['	lettre maj. latine E ouvert accent aigu
163	243	a3		0190 +0302	E[^	lettre maj. latine E ouvert accent circonflexe
164	244	a4		0190 +030C	E[ v	lettre maj. latine E ouvert caron
165	245	a5		0186 +0300	O[	lettre maj. latine O ouvert accent grave
166	246	a6		0186 +0301	O['	lettre maj. latine O ouvert accent aigu
167	247	a7		0186 +0302	O[^	lettre maj. latine O ouvert accent circonflexe
168	250	a8		019D	N	lettre maj. latine N hameçon (ENJ)
169	251	a9		014A	N[	lettre maj. latine ENG (Sami)
170	252	aa		004E +0302	N[^	lettre maj. latine N accent circonflexe
171	253	ab	<<	00AB	<<	guillemet gauche

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
172	254	ac		004E +0308	N»	lettre maj. latine N tréma
173	255	ad	--	00AD	\-	trait d'union virtuel (tired conditionnel)
174	256	ae		01B3	Y[	lettre maj. latine Y crosse
175	257	af		0186 +030C	O[v	lettre maj. latine O ouvert caron
176	260	b0		00B0	_DG	symbole degré
177	261	b1		025B +0300	e[	lettre min. latine e ouvert accent grave
178	262	b2		025B +0301	e[´	lettre min. latine e ouvert accent aigu
179	263	b3		025B +0302	e[^	lettre min. latine e ouvert accent circonflexe
180	264	b4		025B +030C	e[v	lettre min. latine e ouvert caron
181	265	b5		0254 +0300	o[	lettre min. latine o ouvert accent grave
182	266	b6		0254 +0301	o[´	lettre min. latine o ouvert accent aigu
183	267	b7		0254 +0302	o[^	lettre min. latine o ouvert circonflexe
184	270	b8		0272	n]	lettre min. latine n hameçon à gauche (en)
185	271	b9		014B	n[	lettre min. latine eng (Sami)
186	272	ba		006E +0302	n^	lettre min. latine n circonflexe
187	273	bb	>>	00BB	>>	guillemet droit
188	274	bc		006E +0308	n»	lettre min. latine n tréma
189	275	bd		0294	?^	lettre latine coup de glotte

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
190	276	be		01B4	y[	lettre minuscule latine y crosse
191	277	bf		0254 +030C	o[w	lettre min. latine o ouvert caron
192	300	c0	A!	00C0	A`	lettre maj. latine A accent grave
193	301	c1	A'	00C1	A'	lettre maj. latine A accent aigu
194	302	c2	A>	00C2	A^	lettre maj. latine A accent circonflexe
195	303	c3		01CD	A\v	lettre maj. latine A caron
196	304	c4	A:	00C4	A»	lettre maj. latine A tréma
197	305	c5		018E	E<	lettre maj. latine E réfléchi
198	306	c6	AE	00C6	AE+	lettre maj. latine AE (ligature)
199	307	c7	C,	00C7	C\,	lettre maj. latine C cédille
200	310	c8	E!	00C8	E`	lettre maj. latine E accent grave
201	311	c9	E'	00C9	E'	lettre maj. latine E accent aigu
202	312	ca	E>	00CA	E^	lettre maj. latine E accent circonflexe
203	313	cb	E:	00CB	E»	lettre maj. latine E tréma
204	314	cc	!!	00CC	I`	lettre maj. latine I accent grave
205	315	cd	I'	00CD	I'	lettre maj. latine I accent aigu
206	316	ce	I>	00CE	I^	lettre maj. latine I accent circonflexe
207	317	cf	I:	00CF	I»	lettre maj. latine I tréma

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
208	320	d0		0190		lettre maj. latine E ouvert
209	321	d1	N?	00D1	N~	lettre maj. latine N tilde
210	322	d2	O!	00D2	O`	lettre maj. latine O accent grave
211	323	d3	O'	00D3	O\`	lettre maj. latine O accent aigu
212	324	d4	O>	00D4	O^	lettre maj. latine O accent circonflexe
213	325	d5	O?	01D1	O\v	lettre maj. latine O caron
214	326	d6	O:	00D6	O"	lettre maj. latine O tréma
215	327	d7		0152	OE+	digramme soudé maj. latin OE (ligature OE)
216	330	d8		0186	O[	lettre maj. latine O ouvert
217	331	d9	U!	00D9	U`	lettre maj. latine U accent grave
218	332	da	U'	00DA	U\`	lettre maj. latine U accent aigu
219	333	db	U>	00DB	U^	lettre maj. latine U accent circonflexe
220	334	dc	U:	00DC	U»	lettre maj. latine U tréma
221	335	dd		011A	E\v	lettre maj. latine E caron
222	336	de		01CF	I\v	lettre maj. latine I caron
223	337	df		01D3	U\v	lettre maj. latine U caron
224	340	e0	a!	00E0	a`	lettre min. latine a accent grave
225	341	e1	a'	00E1	a\`	lettre min. latine a accent aigu
226	342	e2	a>	00E2	a^	lettre min. latine a accent circonflexe

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
227	343	e3		01CE	a\	lettre min. latine a caron
228	344	e4	a:	00E4	a»	lettre min. latine a tréma
229	345	e5		0259	e<	lettre minuscule latine schwa (e culbuté)
230	346	e6	ae	00E6	ae+	lettre min. latine ae (ligature ae)
231	347	e7	c,	00E7	c\,	lettre min. latine c cédille
232	350	e8	e!	00E8	e`	lettre min. latine e accent grave
233	351	e9	e'	00E9	e'	lettre min. latine e accent aigu
234	352	ea	e>	00EA	e^	lettre min. latine e accent circonflexe
235	353	eb	e:	00EB	e»	lettre min. latine e tréma
236	354	ec	i!	00EC	i`	lettre min. latine i accent grave
237	355	ed	i'	00ED	i'	lettre min. latine i accent aigu
238	356	ee	i>	00EE	i^	lettre min. latine i accent circonflexe
239	357	ef	i:	00EF	i"	lettre min. latine i tréma
240	360	f0		025B	e[	lettre min. latine e ouvert
241	361	f1	n?	00F1	n~	lettre min. latine n tilde
242	362	f2	o!	00F2	o`	lettre min. latine o accent grave
243	363	f3	o'	00F3	o'	lettre min. latine o accent aigu
244	364	f4	o>	00F4	o^	lettre min. latine o accent circonflexe
245	365	f5		01D2	o\	lettre min. latine o caron

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
246	366	f6	o:	00F6	o»	lettre min. latine o tréma
247	367	f7		0153	oe+	digramme soudé min. latin oe (ligature œ)
248	370	f8		0254	o[	lettre min. latine o ouvert
249	371	f9	u!	00F9	u`	lettre min. latine u accent grave
250	372	fa	u'	00FA	u\'	lettre min. latine u accent aigu
251	373	fb	u>	00FB	u^	lettre min. latine u accent circonflexe
252	374	fc	u:	00FC	u»	lettre min. latine u tréma
253	375	fd		011B	e\v	lettre min. latine e caron
254	376	fe		01D0	i\v	lettre min. latine i caron
255	377	ff		01D4	u\v	lettre min. latine u caron



UNESCO  
Secteur de la communication et de l'information  
Division de la Société de l'information  
1, rue Miollis  
75732 Paris cedex 15  
France

Tel. + 33.1.45.68.45.00  
Fax + 33.1.45.68.55.83

**[www.unesco.org/webworld](http://www.unesco.org/webworld)**

**Paris : UNESCO, 2007**