

Revue coéditée
par l'Agence
de la francophonie
et la Communauté
française de Belgique

23
novembre 2003

Le traitement informatique des langues africaines

Cahiers du Rifal

Réseau
international
francophone
d'aménagement
linguistique

3

Avant-propos

5

Marcel Diki-Kidiri et
Edema Atibakwa Baboya :

Les langues africaines sur la toile.

33

Anneleen Van der Veken et
Gilles-Maurice de Schryver :

*Les langues africaines sur la Toile :
étude des cas haoussa, somali, lingala
et isixhosa.*

46

Chantal Enguehard et Chérif Mbodj :

*Flore : un site coopératif pour recueillir et
diffuser les noms des plantes dans les
langues africaines.*

56

Andrei Popescu-Belis :

*Constitution de banques de textes
multilingues : un mécanisme fondé sur le
standard XML.*

62

Jian Yang :

*Tri informatique pour le lingala et le
hausa dans le projet BTML.*

73

Oliva Ramavonirina :

*Les caractères spéciaux dans l'orthographe
courante : le cas du malgache.*

76

Edema Atibakwa Baboya :

*L'orthographe des langues de la République
démocratique du Congo : entre usages et
norme.*

84

Jean de Dieu Karangwa :

Le kiswabili à l'épreuve de la modernité.

101

En bref - Nouvelles du Rifal

Louis-Jean Rousseau :

*Le Rifal et la normalisation internationale
en matière de langue.*

Louis-Jean Rousseau :

*Le Rifal et l'informatisation des langues
et l'ingénierie linguistique.*

Avant-propos

LES LANGUES africaines, dont on estime le nombre à environ 2000, représentent un tiers des langues du monde. Elles constituent donc une composante importante et irremplaçable du patrimoine linguistique de l'humanité et de sa diversité écolinguistique. Par ailleurs, le développement et la mondialisation des technologies de l'information et de la communication (TIC) offrent, en principe, à toutes les langues du monde un même espace globalisé de communication et d'échange: le cyberspace. Cependant, toutes les langues du monde ne font pas usage de cet espace offert, car pour y accéder, il faut avoir fait l'objet d'un traitement informatique, traitement qui relève de l'aménagement linguistique. Dès lors, la première question que l'on se pose ici se rapporte à l'utilisation des langues africaines dans le cyberspace. Marcel Diki-Kidiri et Edema Atibakwa, dans « Les langues africaines sur la Toile », explorent plus de 3000 sites pour ne retenir que ceux qui traitent des langues africaines. Leur analyse montre que s'il existe bien une abondante documentation sur les langues africaines sur la Toile, il y a très peu de sites qui utilisent une langue africaine comme langue de communication. Bien que de nombreux facteurs puissent être pris en compte pour expliquer cet état des faits, l'un des facteurs dominants serait l'inexistence de cybercommunautés linguistiques capables d'intensifier leurs échanges dans leurs langues via la Toile. Cette conclusion sera modérée, nuancée, voire corrigée par une étude différente faite par Gilles Maurice de Schryver et Anneleen Van der Veken, « Les langues africaines sur la Toile: étude des cas haoussa, somali, lingala et isixhosa ». Ces auteurs ont exploré plutôt les forums de discussion pour y découvrir un taux d'utilisation tout à fait satisfaisant de trois langues africaines largement diffusées: le kiswahili, le hausa et le lingála. Ce constat démontre bien l'importance de l'existence d'une cybercommunauté pour le développement d'une langue dans le cyberspace.

Par définition, une cybercommunauté ne peut se construire que par le moyen d'une activité qui mobilise et motive ses membres. Si les débats des forums constituent une attraction certaine, les sites offrent un mode d'expression et de communication plus construit, où des textes plus soignés sont publiés pour être exploités en tant que documents structurés. Un site comme celui que nous présentent Chantal Enguehard et Chérif Mbodj, « Flore, un site coopératif pour recueillir et diffuser les

noms des plantes dans les langues africaines », propose aux visiteurs une banque de connaissances sur la flore des pays du Sahel, avec une foule de renseignements dont les noms des plantes dans plusieurs des langues africaines de cette région. De tels sites, et beaucoup d'autres encore, nécessitent le recours à des techniques de banques de données pour la gestion des données et leur restitution à la demande. Avec son article « Constitution de banques de textes multilingues: un mécanisme fondé sur le standard X.M.L. », Andrei Popescu-Belis décortique pour nous le mécanisme de la constitution d'une banque de données qui utilise le standard X.M.L. En effet, il est important d'adhérer à des formats standard internationaux si l'on veut faciliter les échanges entre différentes banques de données qui ont ainsi une double vocation, à savoir servir les besoins locaux et permettre les échanges avec l'extérieur. Ceci est particulièrement important pour les langues africaines majeures, dont beaucoup sont parlées dans plusieurs pays. De ce fait, la coopération entre les institutions de recherche et d'enseignement qui s'occupent de ces langues dans chaque pays revêt un avantage inestimable pour une synergie dynamique autour des langues concernées. En prenant en compte les caractéristiques propres aux langues africaines, dont la plupart s'écrivent avec des diacritiques tonals et des caractères spéciaux, il est nécessaire de résoudre un certain nombre de problèmes techniques incontournables comme l'ordre de tri alphabétique, graphématique et tonal. Or, rares – pour ne pas dire inexistantes – sont les logiciels qui incorporent des algorithmes de tri tenant compte des langues africaines. C'est pourquoi l'exposé de Jian Yang sur « Le tri informatique pour le lingála et le hausa dans le projet B.T.M.L. » retient toute notre attention. Le B.T.M.L. est un nouveau système multiplateforme de gestion de banques de données terminologiques, textuelles multimédias et multilingues. C'est peut-être la première fois que les spécificités des langues africaines sont prises en compte dès la conception d'un produit de cette envergure. Généralement, les localisations sont faites *a posteriori* au prix de maints contorsionnements qui amènent la langue locale à s'adapter tant bien que mal au produit plutôt que l'inverse.

Un ordre de tri? Cela suppose qu'il y a au minimum une liste de caractères arrêtée, si ce n'est une orthographe complètement stabilisée, normée et standardisée. Il y a bien

entendu une différence conceptuelle entre la norme et le standard. Le premier est un idéal référentiel, tandis que le second est un compromis moyen pragmatique. Et les deux ont tendance à converger vers un point d'équilibre instable qui se situe quelque part entre une norme standardisée et un standard normalisé. Ceci explique le flottement que l'on constate dans les systèmes orthographiques qui ne sont pas encore totalement stabilisés. En effet, en tant qu'activité d'aménagement linguistique, l'élaboration d'une orthographe vise souvent à établir une norme d'écriture pour une langue donnée. Et pour diverses raisons, cette norme, idéal référentiel dont on peut parfaitement justifier le bien fondé, peut être appliquée de façon incomplète par l'usage, et même être bien différente du standard, compromis pragmatique, issu des pratiques des écrivains à travers le temps. C'est exactement ce qu'illustrent, d'une part, l'article d'Oliva Ramavonirina à propos du sort des caractères spéciaux dans l'orthographe de la langue malgache, « Les caractères spéciaux dans l'orthographe courante: le cas du malgache », et d'autre part, celui d'Edema Atibakwa sur les langues de la République démocratique du Congo, « L'orthographe des langues de la République démocratique du Congo: entre usages et norme ». On voit bien que la question orthographique est une étape nécessaire dans l'activité d'aménagement linguistique qui doit précéder le traitement informatique des langues, lequel, à son tour, leur permet d'investir le cyberspace.

L'activité d'aménagement linguistique, dans son ensemble, est une tâche de longue haleine, comme nous me rappelle l'article de Jean de Dieu Karangwa « Le kiswahili à l'épreuve de la modernité », article qui décrit la longue marche du kiswahili vers la modernisation. À la lecture de cet article, on mesure tout le travail qu'il y a à faire et tout le chemin qu'il faut parcourir, toutes les embûches, les déconvenues, les pièges qu'il faut surmonter, quand on s'engage sur la voie de l'aménagement d'une langue, même au niveau gouvernemental, voire international. L'aménagement linguistique, qu'il s'applique aux langues africaines ou à d'autres langues du monde, n'est jamais une sinécure. Il demande des moyens importants et la coopération solidaire de tous ceux qui se sentent concernés par la préservation de la diversité écolinguistique, par la sauvegarde des cultures, par la gestion des langues en tant que patrimoines et moyens de communication, par la

réduction de la fracture numérique et technologique et par une certaine conception humaniste des sociétés du savoir, de l'information et de la formation.

*Pour l'unité mixte de recherche Langage, langues et cultures d'Afrique noire (LLACAN),
INALCO, C.N.R.S. et Université Paris 7,
Marcel Diki-Kidiri,
Edema Atibakwa,
Christian Chanard.*

Les langues africaines sur la Toile

La révolution informatique est considérée à juste titre comme la plus grande révolution de notre ère depuis l'invention de l'imprimerie pour la communication et la diffusion de la culture du savoir et des idées. Dans quelle mesure les langues africaines participent-elles à cette révolution? En étudiant l'usage qui est fait des langues africaines sur la Toile, on voudrait montrer la configuration de la place qu'elles y occupent.

Termes-clés:

langues africaines; sites africains; NTIC; Afrique.

Introduction

LA LUTTE contre la fracture numérique a été et reste l'un des grands chantiers de la Francophonie, constamment repris de sommet en sommet. C'est ce qui fonde les efforts consentis au sein des réseaux francophones (notamment Rint, Riofil puis Rifal) pour apporter à leurs membres du Sud la formation nécessaire à l'utilisation des NTICs pour développer du contenu en français et dans les langues partenaires. L'étude présentée ici voudrait faire le point sur la place réelle occupée par les langues africaines sur la Toile. Elle ne concerne pas l'usage des langues africaines dans les courriers électroniques, ni dans les forums de discussion, mais uniquement sur les sites auxquels on peut accéder avec un navigateur ordinaire.

On estime le nombre des langues africaines à environ 2000 dont une centaine dites « majeures » en raison soit de leur grande diffusion soit du nombre de leur locuteurs. Cette étude se limite à 64 langues sélectionnées parmi les plus parlées des langues majeures, qu'elles soient de grande diffusion ou non. Pour cela, nous avons suivi les statistiques compilées par Mathias Brenzinger, de l'*Institut für Afrikanistik* (Université de Cologne), sur les 103 langues africaines majeures du XXI^e siècle: *Major African Languages in the 21st Century*. Université du Bénin, Lomé. Symposium au sein du 3^e Congrès mondial de Linguistique africaine, 22 Août 2000.

1 Mode de recherche

La recherche sur la Toile a donc été menée en utilisant le nom de chaque langue comme mot-clé. Ce qui a nous a conduits à devoir résoudre les problèmes de dialectologie et d'écriture multiple des noms de langue.

Certaines langues se présentent avec plusieurs dialectes nettement distingués et différemment nommés. Le cas le plus simple est celui du kirundi et du kinyarwanda qui sont deux langues différentes du point de vue socio-politique alors que d'un point de vue strictement linguistique, il s'agit d'une seule et même langue. Ce cas est comparable à celui de la langue sotho parlée au Lesotho sous l'appellation de « southern sotho » ou « sesotho » et en Afrique du Sud sous l'appellation de « northern sotho » ou « pedi », et par conséquent, ces deux parlers sont considérés comme deux langues différentes.

Le peul représente un cas plus complexe où le mot « peul » n'est utilisé qu'en français pour désigner un ensemble linguistique que les Anglophones appellent « fulani ». Les locuteurs natifs, eux, utilisent dans leurs parlers, les termes de « fulfulde » pour les parlers allant du Mali au Tchad, « pulaar » pour les parlers de l'ouest (Sénégal et Mauritanie) et enfin « pular » pour le parler de Guinée. En dépit de ces différents parlers, les locuteurs natifs se considèrent bel et bien comme une seule communauté linguistique et culturelle, les *Hal-Pulaar*. Et bien entendu, on retrouve tous ces termes sur la Toile! Ce qui complexifie la recherche.

Le cas le plus complexe est cependant celui de la langue berbère, également appelée « amazighe ». Elle comprend de très nombreux dialectes et parlers régionaux dont les quatre plus importants sont le *tamazight*, le *tarifit*, le *tachelhit* au nord du Sahara, et le *touareg* qui inclut le *tamachek* ou *tamajaq* au sud du Sahara. Ces différents parlers portent de nombreux noms qui relativisent la clarté de ces distinctions. Ce qui démultiplie d'autant la recherche.

Notre objectif n'étant pas de régler les problèmes identitaires entre langues et dialectes, nous relèverons tous les sites où il est clairement question d'un parler, quel que soit son statut. Toutefois, les sites relevés à partir de variantes purement orthographiques seront regroupés dans un même comptage. Par exemple on comptera ensemble les sites relatifs au hausa, hawsa, haoussa, etc.

2 Critères de sélection

Nous avons utilisé Copernic 2001 Plus^(tm) pour rechercher les sites avec chaque nom de langue comme mot-clé. Or tous les sites récoltés par cette méthode ne se rapportent pas nécessairement à la langue. Nous devons donc visiter chaque site et éliminer tous ceux qui ne permettent pas d'identifier formellement le nom de la langue comme renvoyant bien à une langue. En effet, le même mot qui désigne une langue désigne aussi le peuple qui la parle, la culture de ce peuple, etc. Par exemple, le mot « bambara » désigne aussi bien la langue que le peuple et l'art de celui-ci, tandis que dans les langues bantu ce type d'ambiguïté est quasiment inexistant du fait que les noms de langues comportent généralement un préfixe spécifique (*ki-*, *li-* ou *se*) qui les rend univoques. Par exemple: *kiswahili*, *lingála*, *sesotbo*).

Par ailleurs, certains noms de langues fortuitement homographes de noms désignent des sociétés, des marques ou des associations n'ayant rien à voir avec la langue. Ainsi *ewe* est plus connu sur la Toile comme une marque de laine écossaise que comme langue d'Afrique de l'ouest, tandis que le mot-clé *akan* rapporte des sites aussi bien sur la langue et la culture *akan* que sur une société polonaise du même nom.

En conséquence, à l'issue d'un premier tri entre les 3 332 pages Web visitées, seules les 1 374 qui traitent explicitement des langues sont retenues, soit 41,24 % de l'ensemble des sites récoltés par le métamoteur (cf. fig. 1 et 2).

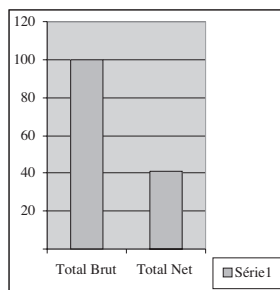


Figure 1

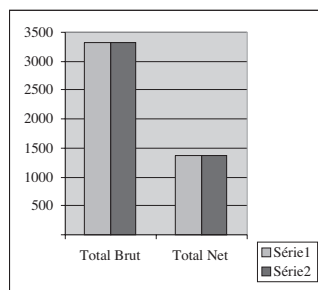


Figure 2

3 Grille d'analyse

Le contenu des sites visités a été soumis à la grille d'analyse suivante :

- mention simple ;
- description ;
- documentation ;
- échantillon ;
- cours de langue ;
- textes ;
- sites ;
- langue de communication -français (LC-F) ;
- langue de communication - anglais (LC-A) ;
- langue de communication - autres (LC-O).

Cette grille s'est constituée progressivement à partir de l'examen du contenu des sites au fur et à mesure du dépouillement. Elle est commentée de façon détaillée ci-après.

3.1 Mention simple

La mention simple d'une langue peut se manifester de deux façons différentes: soit dans un texte quelconque, soit sur une carte géographique de pays, délimitant l'espace d'extension de la langue. Dans un texte, le nom de la langue est situé dans un contexte linguistique explicite comme ceux-ci :

la langue X

Dictionnaire X -français et français X

Cet artiste chante en langue X

Vous pouvez apprendre à parler le X grâce à notre cours interactif sur CD-Rom

Parlez-vous X ?

Déclaration universelle des Droits de l'Homme. Version X.

L'enseignement en X...

Le X se parle dans l'est de...

La famille linguistique du X

etc.

Un autre contexte significatif est représenté par les sites génériques qui donnent des listes de noms de langues avec leur code ISO en deux ou trois lettres, ou encore la liste des langues parlées dans un pays. Dans tous les cas, le visiteur n'a aucun mal à identifier le mot-clé recherché comme étant

bien un nom de langue sur le site visité. Et c'est la seule information pertinente qu'il aura.

Sur les 1 374 sites qui traitent de langues, 383, soit 27,87 %, ne mentionnent que le nom des langues sans aucune information qui soit de nature à révéler quoi que ce soit de leur substance, de leur statut social ou juridique.

3.2 Documentation

Ce critère s'applique aux sites qui sont dédiés à des ressources documentaires sur les langues, comme par exemple, des bibliographies, des cédéroms, des ressources multimédia, mais aussi des librairies électroniques. Nous classons également ici les sites qui font de la publicité pour des produits linguistiques non directement utilisables en ligne, comme par exemple, des méthodes d'apprentissage de langues sur cédérom qu'il faut acheter avant de savoir comment l'utiliser. De nombreuses publicités sur les langues africaines sont de ce type et ne permettent pas d'accéder directement à la langue de façon interactive.

Certaines documentations générales qui concernent presque toutes les langues ne sont reprises ici que lorsque la langue recherchée est directement concernée. Par exemple, les codes des noms de langue (codes ISO 639-1 et 639-2, code MARC), les polices de caractères et, le cas échéant, les logiciels spécifiques à l'écriture de certaines langues.

La documentation apporte donc au visiteur du site une quantité d'informations qui vont bien au-delà du simple nom de la langue, sans pour autant l'informer davantage sur la matérialité de la langue.

Ce critère s'applique à 152 sites sur 1374, soit 14,79 % des sites qui traitent des langues africaines (Cf. figures 3 et 4, ci-dessus).

3.3 Description

Le critère de description s'applique lorsqu'un site donne des informations de type sociolinguistiques et linguistiques. Les informations sociolinguistiques décrivent l'usage de la langue dans la société et sont du genre :

- nombre de locuteurs;
- pays d'origine ou aire d'extension de la langue;
- son statut juridique ou ses usages sociaux;
- sa classification linguistique;
- etc.

Les informations linguistiques vont d'une étude ponctuelle (par exemple, *la structure du verbe en ambarique*) jusqu'à une esquisse linguistique (phonologie, syntaxe, lexique) plus ou moins développée qui apporte beaucoup plus d'informations encore sur la langue. Grâce à ces informations, on commence à avoir une petite idée de la manière dont telle ou telle langue structure la pensée, du moins sur les points étudiés.

Il est important de noter que même dans ce cas, les informations sont données *sur* la langue africaine en utilisant comme langue de description une autre langue, en général une langue européenne (allemand, anglais, espagnol, français, néerlandais, polonais, portugais, italien). La description linguistique fragmentaire souvent proposée ici comporte généralement peu de données matérielles sur la langue.

Ce critère s'applique à 148 des 1374 sites qui traitent des langues, soit 14,40 %, comme le montrent les figures 3 et 4 ci-après.

3.4 Échantillons

Par définition, ce critère permet d'accéder à la matérialité de la langue, car les sites classés ici nous livrent des échantillons de la langue plus ou moins importants qui vont des simples citations de mots dans la langue à des lexiques interactifs. Nous considérons comme citation l'usage ponctuel d'un mot dans une langue africaine alors que tout l'environnement est dans une autre langue. C'est par exemple le cas des sites dont le nom est donné dans une langue africaine alors que tout le site est en français ou en anglais. C'est aussi le cas des sites où l'usage d'une langue africaine se limite à un slogan ou une enseigne, voire un nom propre, en cette langue. Bien entendu, on peut avoir des citations plus longues que cela!

Un autre type d'échantillon est celui des sites qui traduisent dans un grand nombre de langues un même mot ou une même expression. Par exemple « comment dit-on *paix* dans 2 000 langues, comment dit-on *bonjour* dans 800 langues, comment dit-on *je t'aime* dans 350 langues, comment dit *Joyeux Noël* dans deux cents langues » etc. Ces sites ne concernent bien évidemment pas que les langues africaines mais un bon nombre des langues africaines sont incluses dans ce type de listes.

Certains sites comportent des fichiers sonores permettant de saisir encore mieux la matérialité de la langue. D'autres proposent des phrases usuelles dans la langue traitée. La matière linguistique fournie est déjà plus riche et permet d'avoir une meilleure idée de la langue, même si elle demeure encore trop sommaire. Entrent dans cette catégorie les esquisses grammaticales qui s'accompagnent des exemples dans la langue.

Plus élaborés encore sont les lexiques interactifs qui permettent au visiteur d'obtenir instantanément la traduction d'un mot qu'il entre. Cette traduction peut se faire vers ou depuis la langue africaine. Les lexiques interactifs de ce genre ne sont pas généralement très fournis et leur mode de consultation ne permet pas de lire plusieurs mots à la fois. C'est pourquoi nous les classons parmi les échantillons.

L'ensemble des sites concernés par ce critère s'élèvent à 76 sur 1374, soit 7,39% des sites traitant de langues (Cf. figures 3 et 4 plus haut).

3.5 Textes

C'est parce qu'un texte fournit davantage d'informations sur une langue qu'une liste des mots ou de phrases usuelles que nous mettons à part les sites qui proposent des textes en morceaux choisis ou des textes entiers. L'un des sites les plus remarquables de cette catégorie est le *Language Museum* qui se veut une collection des textes dans 2000 langues. On y trouve très largement représentés la traduction de textes religieux tels que le *Pater Noster*, l'*Ave Maria*, les premiers versets du premier chapitre de l'Évangile selon saint Jean), les textes juridiques comme les premiers articles de la *Déclaration universelle des droits de l'homme* et des textes narratifs avec leur traduction en anglais.

Par ailleurs le site de l'Unesco donne la traduction du texte intégral de la *Déclaration universelle des droits de l'homme* dans de nombreuses langues africaines entre autres. À l'instar de l'Unesco, certains sites offrent plusieurs versions d'un même texte dans plusieurs langues, chacune ayant ses propres pages. C'est souvent le cas pour les textes officiels d'Afrique du Sud, traduits dans les onze langues officielles de ce pays. C'est aussi le cas pour de nombreux sites religieux en mal de prosélytisme.

Certains sites universitaires consacrés essentiellement à la littérature livrent des textes en langues africaines avec leur traduction dans une langue européenne, généralement en anglais. Nous n'incluons pas ici les sites qui ne livrent que la traduction en langue européenne des contes ou poèmes africains sans en donner les textes originaux dans la langue source!

Dans notre corpus, 89 sites sur 1374, soit 8,66% des sites fournissent des textes avec traduction. On notera que la majorité des textes recueillis sont des traductions vers des langues africaines à l'exception des sites littéraires qui eux s'intéressent aux langues africaines et traduisent vers les langues européennes, mais dans les deux cas, la langue de communication utilisée reste une langue européenne.

3.6 Cours de langue

Il s'agit ici des sites qui proposent des cours de langues en ligne et non pas qui font la publicité des cours de langues non disponibles en ligne. Le plus souvent on trouve des débuts de cours ou des cours complets pour débutants qui constituent des produits d'appel pour inciter le visiteur à aller plus loin dans la découverte de la langue en achetant un produit plus élaboré.

Nous n'avons trouvé que très peu de cours interactifs soignés en ligne pour les langues africaines. Les produits d'appel sont très largement orientés vers l'enseignement des langues africaines aux touristes et résidents étrangers (anglais, espagnols, allemands ou français notamment).

Sur 1374 sites de notre corpus, il n'y a que 24 qui proposent des cours de langues soit 2,33% de l'ensemble. Pourtant les cours de langue sont l'un des moyens les plus sûrs d'augmenter à la fois le nombre et la compétence des usagers d'une langue. C'est dire l'importance de ce type de matériel pour la création et le développement d'une cybercommunauté d'utilisateurs de la langue africaine comme véhicule de communication.

3.7 Sites

Nous classons ici les sites qui utilisent effectivement les langues africaines comme moyen de communication avec le visiteur.

On peut distinguer les sites qui utilisent les langues africaines sur quelques pages ou dans quelques rubriques seulement. Par exemple sur les boutons de navigation, la partie journal d'informations (*news*) ou encore la partie forum de discussion, le reste étant en langue européenne.

Ensuite viennent les sites qui utilisent majoritairement les langues africaines. Il s'agit des sites bilingues dont le côté bilingue ne concerne qu'un nombre restreint de textes

en général traduits de la langue africaine vers une langue européenne.

Enfin, quelques sites sont rédigés en langues africaines voire dans une écriture spécifique (comme l'écriture éthiopienne, l'écriture *Nko*, etc.). On trouve ici l'utilisation maximale d'une langue africaine comme véhicule de communication sur la Toile.

Ce critère s'applique à 33 sites sur 1 374 soit 3,22 % de l'ensemble de notre corpus (Cf. figures 3 et 4 ci-après).

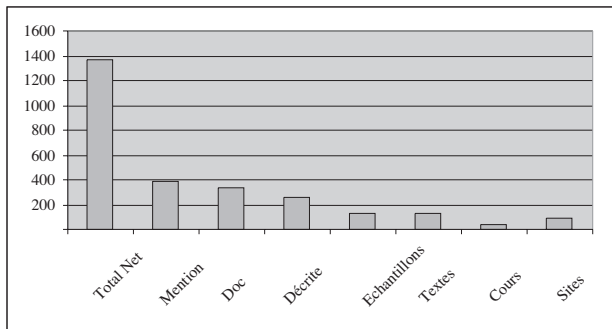


Figure 3

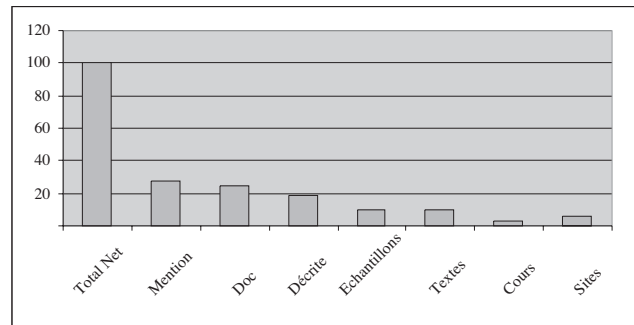


Figure 4

Total Net	Mention	Doc	Décrite	Échantillons	Textes	Cours	Sites
100	27,87	24,46	18,79	9,61	9,61	3,13	6,11
1374	383	332	254	132	132	43	84

Figure 4

3.8 Les langues de communication des sites

Nous appelons « langue de communication » la langue dans laquelle le webmestre entend communiquer avec le visiteur. C'est donc typiquement la langue utilisée par le webmestre pour écrire les annotations sur les boutons de navigation, les messages d'aide, les éventuels menus, les légendes d'images, les consignes de lecture, et bien entendu, tout texte destiné à communiquer avec le visiteur. Par exemple, dans une description linguistique, la langue

de description est la langue de communication tandis que la langue décrite n'est qu'un objet dont on parle.

Dans le cas de textes comme celui de la *Déclaration universelle des droits de l'homme*, on peut se demander si la langue du texte n'est pas la langue de communication, dans la mesure où il constitue l'essentiel du message que le webmestre veut communiquer au visiteur. En fait, nous constatons qu'en dehors du texte lui-même, tout ce qui est consigne ou annotation (si peu que cela soit) est en anglais. Par conséquent c'est bien dans cette langue que le

webmestre entend assister le visiteur dans l'identification de ce texte. La langue de communication est bel et bien celle dans laquelle le webmestre tente de dialoguer avec le visiteur et non celle du texte qu'il lui présente pour lecture.

On pourrait croire que les langues parlées dans les pays francophones comme le bambara, le sängö, le lingála, etc., sont traitées dans des sites utilisant largement le français. En fait, il n'en est rien! L'examen de la langue de communication utilisée dans 503 sites pris au hasard sur les 1374 montre que dans 376 sites, soit 74,76%, c'est l'anglais qui est utilisée comme langue de communication. Le français vient en second lieu avec 77 sites, soit 15,30%. Les langues africaines sont utilisées comme langues de communication dans les 33 sites en langues africaines, soit 6,56%. Enfin, les 17 sites qui restent, soit 3,37% se partagent entre l'allemand, l'espagnol, le portugais, le polonais, le néerlandais et l'italien. Ce sont les langues de communication « autres » (LC-O dans la figure 5).

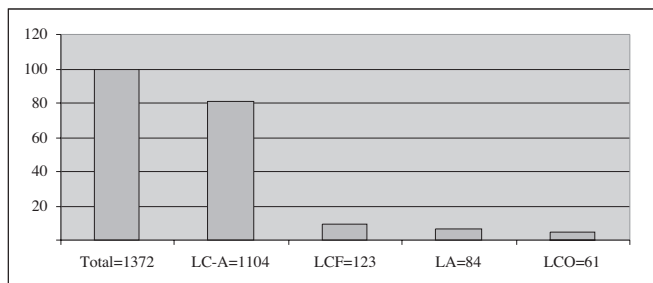


Figure 5

Total=1372	LC-A=1104	LCF=123	LA=84	Lc-O=61
100	80,47	8,07	6,12	4,45

Conclusion

Les principaux enseignements que révèle cette brève étude sont clairement les suivants :

Les langues africaines apparaissent sur la Toile beaucoup plus comme des objets d'étude (mention, documentation, description, échantillons, textes, cours) que comme des véhicules de communication (sites).

La langue de communication utilisée pour parler des langues africaines est très largement l'anglais, même pour les langues en zone francophone.

Les cours de langues africaines sont beaucoup trop rares sur la Toile. Ce qui entrave la possibilité de développer des cybercommunautés de locuteurs utilisant les langues africaines comme véhicules de communication via l'Internet.

Les produits logiciels ou les solutions informatiques intégrant en standard des polices de caractères pour toutes les langues africaines sont rarement proposés sur les sites.

Pour corriger cette situation, il y a donc lieu de promouvoir :

- la multiplication des sites bilingues (ou multilingues) comportant le français et au moins une langue africaine comme langues de communication ;
- une plus grande diffusion de la documentation francophone sur les langues africaines, car cette documentation existe mais n'est pas systématiquement diffusée sur la Toile ;
- les cours de langues africaines de qualité à diffuser sur la Toile ;
- le développement et la diffusion de produits logiciels ou de solutions informatiques facilitant l'écriture des langues africaines et leur utilisation normale et courante dans le cyberspace.

*Marcel Diki-Kidiri et Atibakwa Baboya Edema,
Llacan (CNRS, Inalco, Université de Paris),
Villejuif, France.
kidiri@vjf.cnrs.fr
edema_atibakwa@yahoo.fr*

Bibliographie

Ngangala Balade Tongamba (J.), 2001 : «Technologie de l'information et langues africaines: Lieux des enjeux existentiels», dans *Cahiers du Rifal*, n° 22, décembre 2001, p. 3-9.

Bearth (Th.), 2001 : «Ali Akan: initiation à une langue africaine par les nouvelles technologies», dans *Cahiers du Rifal*, n° 22, décembre 2001, p. 27-32.

Chanard (Chr.), Popescu-Belis (A.), 2001 : «Encodage informatique multilingue. Application au contexte du Niger», dans *Cahiers du Rifal*, n° 22, décembre 2001, p. 33-45.

Codes ISO 639 des langues du monde (code à 3 lettres)
<http://www.w3.org/WAI/ER/IG/ert/iso639.htm>
<http://lcweb.loc.gov/standards/iso639-2/englangn.html>
<http://www.oasis-open.org/cover/nisoLang3-1994.html>
<http://www.dsv.su.se/~jpalme/ietf/language-codes.html>
<http://www.rtt.org/ISO/TC37/SC2/WG1/N-doc/N066.doc>
<http://www.rtt.org/ISO/TC37/SC2/WG1/N-doc/N065.doc>
 Codes MARC (à trois lettres)
http://www.loc.gov/marc/languages/lang_o2s.html
 Code ISO 639-1 (à 2 lettres)
<http://babel.alis.com/languages/iso639.htm>
<http://mail.gnu.org/pipermail/commit-gnue/2001-July/000938.html>
 Dictionnaires en ligne (environ 250 langues)
<http://www.yourdictionary.com/languages.html>
 Liste des traductions de la DH par langue
<http://www3.itu.int/udhr/navigate/alpha.htm>
 Base de données bibliographiques du Center of Applied Linguistics de Washington
<http://www.lmp.ucla.edu/lmd/cals.htm>
 Dire «je t'aime» en X langues
<http://www.lodace.com/mabel/jtaime.htm>
 Pater Noster dans plusieurs langues (sous forme de Base de données)
<http://www.christusrex.org/www1/pater/>
 Le projet Rosetta vise à archiver des esquisses descriptives de 1 000 langues
<http://www.rosettaproject.org:8080/live/search/browsebylangname>
 Paix dans de nombreuses langues
<http://www.cs.wm.edu/~radu/peace.html>
<http://www.mindmills.net/articles/peace.shtml>
<http://www.giftofpeace.net/PeaceLanguages.htm>
<http://users.bestweb.net/~om/yoga/peace.html>

Liste des sites étudiés

La liste qui suit présente l'ensemble des sites étudiés. Pour chaque langue, chaque site est placé sous l'élément de la grille (mention, documentation, description, etc.) qui lui est le plus applicable. Par mesure d'économie, nous n'avons pas reporté ici toutes les annotations plus fines qui ont été faites à l'intérieur de chaque élément. Il faut souligner enfin que la visite des sites ayant été faite entre juin 2002 et mars 2003, certains URL peuvent avoir été modifiés depuis lors.

Les portails collectionneurs

Dans ces portails, les langues africaines ne sont mentionnées ou exemplifiées qu'incidemment, à l'occasion du traitement de sujets concernant un grand nombre de langues à travers le monde :

Le musée des langues: échantillon de textes dans 2000 langues
<http://www.language-museum.com/>
 Références de ressources sur de nombreuses langues du monde
<http://www.vada.nl/talenen.htm>
 Salutations dans plus de 800 langues
<http://www.elite.net/~runner/jennifers/hello.htm>
 Salutations à écouter dans 55 langues du monde
<http://vraptor.jpl.nasa.gov/voyager/lang.html>
 Les quatre lois spirituelles dans 147 langues
<http://www.greatcom.org/laws/languages.html>
 Les nombres de 1 à 10 dans plus de 4500 langues
<http://www.zompist.com/numbers.shtml>
 Joyeux Noël dans plus de 350 langues
<http://www.flw.com/merry.htm>

Afrikaans (37)

Mention (6)
<http://www.gov.za/documents/2000/mdda/>
<http://www.ncrf.org.za/mpumalanga.htm>
<http://www.letsgo.com/SAF/01-teaser-01>
<http://general.rau.ac.za/sociology/index.htm>
http://members.tripod.com/lingua_consultant
http://members.tripod.com/lingua_consultant
Documentation (4)
http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm
<http://dictionaries.travlang.com/EnglishAfrikaans/>
<http://www.freedict.com/onldict/afr.html>
<http://www.yourdictionary.com/languages/germanic.html>
Description (2)
<http://www.santesson.com/eaforcurs.htm>
<http://www.science.gmu.edu/~bwallet/afrikaans.html>
Echantillons (21)
<http://dictionaries.travlang.com/AfrikaansEnglish/dict.cgi?query=month&max=50>
<http://www.freedict.com/onldict/afr.html>
Texte (1)
<http://www.gov.za/documents/2000/mdda/afrikaans.pdf>
Cours (3)
<http://www.santesson.com/eaforcurs.htm>
<http://www.geocities.com/Paris/2920/wazzit-en.html>

<http://www.afrikaans-english.com/>
Site (117)

<http://afrikaans.be/>

<http://www.afrikaans.com/>

<http://padlangs.virafrikaans.com/>

<http://iafrica.com/afrikaans/>

<http://www.fliekvlooi.virafrikaans.com/>

<http://www.itweb.co.za/sections/afrikaansenuus/>

Sites multilingues dont l'afrikaans (11)

<http://www.4afrikaans.com/>

<http://iafrica.com/afrikaans/>

<http://af.wikipedia.com/>

<http://padlangs.virafrikaans.com/>

<http://de.wikipedia.org/>

<http://www.dieknoop.co.za/>

<http://www.bybel.co.za/>

<http://www.rau.ac.za/>

<http://www.fotoverhaal.co.za/>

<http://www.members.tripod.com/~vds/afroppinet/>

<http://www.museums.org.za/aftraal/afrikaans/m.htm>

LC-A (20)

Akan (6)

Mention(1)

<http://www.unizh.ch/spw/afrling/akandic/adetails.htm>

Documentation (1)

<http://www.unizh.ch/spw/afrling/akandic/adetails.htm>

Description (2)

http://www.ethnologue.com/show_family.asp?subid=50

<http://www.unizh.ch/spw/afrling/aliakan/>

Echantillons (1)

<http://www.unizh.ch/spw/afrling/aliakan/>

Cours (1)

<http://www.unizh.ch/spw/afrling/aliakan/toc.html>

LC-A (6)

Tous les sites trouvés sur cette langue sont en anglais

Amazighe / Tachelhit (29)

Mention(12)

<http://www.frebend.com/wwwboard/messages/322.html>

<http://fr.allafrica.com/stories/200201230740.html>

<http://www.rosettaproject.org:8080/live/search/addressreform?ethnocode=SHI&langname=Tachelhit>

<http://www.mailgate.org/soc/soc.culture.berber/msg04935.html>

<http://pub137.ezboard.com/fmondeberberfrm1.showMessage?topicID=94.topic>

http://www.bethany.com/profiles/p_code/1359.html

<http://perso.club-internet.fr/tilit/lexiquetamazirtfrancais.html>

http://www.bethany.com/profiles/p_code/1744.html

<http://www.globaldiversity.org.uk/condev/beldia.html>

<http://www.morocco.com/forums/printthread.php3?threadid=4263>

http://www.waac.org/amazigh/language/alaoua_tamazight-dialects.html

<http://www.waac.org/amazigh/culture/yennayer2951.html>

Documentation (6)

<http://www.azawan.com/tachelhit/schuyler/img/>

<http://www.amazigh-voice.com/Amazigh-Films.htm>

http://www.mondeberbere.com/culture/biblio/biblio_tachelhit.htm

http://www.bibliomonde.com/pages/fiche-livre.php3?id_ouvrage=128

http://www.swarthmore.edu/SocSci/jaldere1/biblio2_june01.htm

<http://www.orient-lib.com/rech.collections.phtml>

?Nbre=0&idcollections=329

Description (7)

http://www.ethnologue.com/show_lang_family.asp?code=SHI

<http://locke.ccil.org/~cowan/langtree.txt>

http://www.ethnologue.com/show_language.asp?code=SHI

http://www.ethnologue.com/show_country.asp?name=Morocco

<http://www.al-bab.com/maroc/soc/berber.htm>

http://www.ethnologue.com/show_country.asp?name=Algeria

Echantillons (1)

<http://www.morocco.com/forums/showthread.php3?threadid=4263>

Cours (3)

<http://www.mondeberbere.com/langue/Mountassir/>

<http://www.mondeberbere.com/langue/Mountassir/lecon1.htm>

<http://www.mondeberbere.com/langue/Mountassir/preface.htm>

Site (2)

<http://www.azawan.com/tachelhit/>

Site trilingue tachelhit / français / anglais

<http://www.mondeberbere.com/>

LC-F (9)

<http://www.frebend.com/wwwboard/messages/322.html>

<http://www.mondeberbere.com/langue/Mountassir/>

<http://fr.allafrica.com/stories/200201230740.html>

<http://www.mondeberbere.com/langue/Mountassir/lecon1.htm>

<http://www.mondeberbere.com/langue/Mountassir/preface.htm>

<http://pub137.ezboard.com/fmondeberberfrm1.showMessage>

?topicID=94.topic

http://www.bibliomonde.com/pages/fiche-livre.php3?id_ouvrage=128

<http://perso.club-internet.fr/tilit/lexiquetamazirtfrancais.html>

Site trilingue tachelhit / français / anglais

<http://www.mondeberbere.com/>

LC-A (21)

Site trilingue tachelhit / français / anglais

<http://www.mondeberbere.com/>

Amazighe / tamazight (26)

Mention (9)

<http://www.amazigh.nl/>

http://www.mimoun.net/_disc1/00000031.htm

<http://www.uoc.edu/euromosaic/web/document/berber/an/i1/i1.html>

<http://www.calebproject.org/ptw3/day24.htm>

<http://www.eurosur.org/ai/19/afr1936.htm>

<http://www.emi.ac.ma/depts/ginf/recherche.html>

http://www.waac.org/amazigh/language/announcement_classes.html

<http://www.ahram.org.eg/weekly/1998/384/re5.htm>

<http://www.ronhaleber.nl/marokko.html>

Documentation (5)

<http://membres.lycos.fr/bizari/>

<http://babel.uoregon.edu/yamada/lists/tamazight.html>

http://www.swarthmore.edu/SocSci/jaldere1/biblio2_june01.htm

<http://perso.club-internet.fr/tilit/lexiquetamazirtfrancais.html>

<http://www.maps2anywhere.com/Languages/>

languages_-_tamazight.htm

Description (3)

<http://www.flw.com/languages/tamazight.htm>
http://www.waac.org/amazigh/language/achab_tamazight_language_profile.html
http://www.waac.org/amazigh/language/boukous_tamazight_lin_tunisia.html

Ecbantillons (4)

http://libyamazigh.org/tamazight_language.htm
<http://perso.club-internet.fr/tilit/lexiquetamazirtfrancais.html>
<http://membres.lycos.fr/bizari/chiffres.htm>
http://perso.club-internet.fr/tilit/Anglais_Tamazight.html

Texte (2)

PN: <http://www.christusrex.org/www1/pater/JPN-tamazight.html>
 DH: <http://www.unhchr.ch/udhr/lang/tzm.htm>

Cours (3)

<http://elgvanche.net/tamazgha/cursotamazight.htm>
<http://elgvanche.net/tamazgha/indicecursotamazight.htm>
<http://www.tamurt-imazighen.com/tamazgha/agherbaz.html>

Site (3)

Site tamazight - français - anglais
<http://www.mondeberbere.com/>
 Site anglais - Tamazight
<http://www.geocities.com/Athens/Parthenon/9860/tamaz.html>
<http://www.tamurt-imazighen.com/tamazgha/agherbaz.html>
LC-F (5)

<http://perso.club-internet.fr/tilit/lexiquetamazirtfrancais.html>
http://perso.club-internet.fr/tilit/Anglais_Tamazight.html
http://www.mimoun.net/_disc1/00000031.htm
<http://www.emi.ac.ma/depts/ginf/recherche.html>

Site tamazight - français - anglais

<http://www.mondeberbere.com/>
LC-A (15)

Site tamazight - français - anglais

<http://www.mondeberbere.com/>
 Site anglais - Tamazight
<http://www.geocities.com/Athens/Parthenon/9860/tamaz.html>
<http://www.tamurt-imazighen.com/tamazgha/agherbaz.html>
LC-O (3)

Espagnol

<http://elgvanche.net/tamazgha/cursotamazight.htm>
<http://elgvanche.net/tamazgha/indicecursotamazight.htm>
 Néerlandais
<http://www.amazigh.nl/>

Amazighe / Tarifit (29)

Mention (10)

<http://freepages.cultures.rootsweb.com/~tarifit/guest.html>
<http://freepages.cultures.rootsweb.com/~tarifit/news.html>
<http://freepages.cultures.rootsweb.com/~tarifit/links.html>
<http://freepages.cultures.rootsweb.com/~tarifit/stories.html>
<http://www.christusrex.org/www3/ethno/Moro.html>
<http://www.al-bab.com/maroc/soc/berber.htm>
<http://www.geocities.com/Athens/Parthenon/9860/tamaz.html>
<http://www.ee.umd.edu/~sellami/DEC95/research.html>
<http://www.eurosur.org/ai/19/afr1904.htm>
http://www.bethany.com/profiles/p_code3/1087.html

Documentation (3)

<http://www.geocities.com/lameens/darja/links.html>
http://let.kub.nl/people/aissati/publications_in_Tarifit.htm
http://let.kub.nl/people/aissati/teaching_methods_and_materials.htm
Description (9)

http://www.ethnologue.com/show_lang_family.asp?code=RIF
<http://www.flw.com/languages/tarifit.htm>
<http://freepages.cultures.rootsweb.com/~tarifit/>
<http://www.rz.uni-frankfurt.de/~vajkonny/berber.html>
http://www.ethnologue.com/show_language.asp?code=RIF
<http://www.ling.arts.kuleuven.ac.be/sle2001/abstracts/weberramdani.htm>
<http://cwis.kub.nl/~fdl/general/people/ye-rramd/Research.htm>
http://www.ethnologue.com/show_country.asp?name=Morocco
http://www.ethnologue.com/show_language.asp?code=RIF

Ecbantillons (4)

<http://www.flw.com/merry.htm>
<http://freepages.cultures.rootsweb.com/~tarifit/vocab.html>
<http://freepages.cultures.rootsweb.com/~tarifit/flora.html>
<http://freepages.cultures.rootsweb.com/~tarifit/body.html>

Cours (2)

<http://freepages.cultures.rootsweb.com/~tarifit/proguide.html>
<http://www.mondeberbere.com/langue/cours.htm>

Site (1)

<http://www.amazigh.nl/links/links.php?katid=5>
LC-F (4)

Site anglais - espagnol - français

<http://freepages.cultures.rootsweb.com/~tarifit/news.html>
<http://freepages.cultures.rootsweb.com/~tarifit/links.html>
<http://freepages.cultures.rootsweb.com/~tarifit/stories.html>
<http://www.mondeberbere.com/langue/cours.htm>

LC-A

Site anglais - espagnol

<http://freepages.cultures.rootsweb.com/~tarifit/proguide.html>
<http://freepages.cultures.rootsweb.com/~tarifit/body.html>
 Site anglais - espagnol - français
<http://freepages.cultures.rootsweb.com/~tarifit/news.html>
<http://freepages.cultures.rootsweb.com/~tarifit/links.html>
<http://freepages.cultures.rootsweb.com/~tarifit/stories.html>

LC-O (8)

Espagnol

<http://www.eurosur.org/ai/19/afr1904.htm>

Allemande

http://www.koeppe.de/html/e_fab.htm

Site anglais - espagnol

<http://freepages.cultures.rootsweb.com/~tarifit/proguide.html>
<http://freepages.cultures.rootsweb.com/~tarifit/body.html>

Site anglais - espagnol - français

<http://freepages.cultures.rootsweb.com/~tarifit/news.html>
<http://freepages.cultures.rootsweb.com/~tarifit/links.html>
<http://freepages.cultures.rootsweb.com/~tarifit/stories.html>
 Allemand

<http://www.rz.uni-frankfurt.de/~vajkonny/berber.html>

Amharique / amharic (60)

Mention (9)

http://www.extense.com/bin/x2cgi_view.cgi?userID=64393794&view=on&query=amharique&url=http%3A

%2F%2Fwww.quid.fr%2Fweb.php%3Fweb%3D%2FWEB%2FETATS%2FET%2FQ029450.HTM#marker
<http://www.univ-orleans.fr/EXT/ASTEX/astex/doc/fr/web/html/astex003.htm>
<http://www.mediaport.net/AfricArt/100CD/CD/059.html>
<http://www.mediaport.net/AfricArt/100CD/CD/062.html>
<http://www.cean.u-bordeaux.fr/gdr931.html>
<http://www.amazon.com/exec/obidos/ASIN/0864423381/inktombkasin-20/103-4179358-7107060>
<http://www.orient.uw.edu.pl/~afrykanistyka/ags.html>
<http://monarch.gsu.edu/nutrition/Amharic.htm>
Documentation (31)
<http://www.univ-orleans.fr/EXT/ASTEX/astex/doc/fr/web/html/astex012.htm>
<http://membres.lycos.fr/pedroiy/alphab/amharique.htm>
<http://www.afrik.com/journal/internet/?net-140-2.htm>
<http://members.aol.com/w4z5m4/wazema.html>
<http://www.abysiniacybergateway.net/fidel/unicode/new/references.html>
<http://geonames.nrcan.gc.ca/francais/resolutionA16.html>
<http://www.ethiobar.net/Amharic/download.htm>
<http://lweb.loc.gov/standards/iso639-2/englangn.html>
<http://www.rtt.org/ISO/TC37/SC2/WG1/N-doc/N066.doc>
<http://www.rtt.org/ISO/TC37/SC2/WG1/N-doc/N065.doc>
<http://www.color.org/unicode.html>
<http://babel.alis.com/langues/iso639.htm>
<http://mail.gnu.org/pipermail/commit-gnue/2001-July/000938.html>
<http://www.inalco.fr/pub/enseignements/langues/afrique/amharique/index.html>
<http://babel.uoregon.edu/yamada/guides/amharic.html>
<http://www.ethiopic.com/amharic.htm>
<http://www.africanlanguage.com/>
<http://www.encyclopedia.com/searchpool.asp?target=@DOCTITLE%20Amharic>
<http://www.tsculpitseries.org/amharic.html>
<http://www.iss.stthomas.edu/amharic/>
<http://www.geezsoft.com/>
<http://www.ethiolist.com/>
<http://www.culture.fr/culture/dglf/bpi/list-langues.html>
<http://www.ethnicarvest.org/bibles/amharic.htm>
<http://www.quartet.fr/amharique.html>
<http://www.amazon.com/exec/obidos/ASIN/2852650169/avsearch-bkasin-20/103-4179358-7107060>
<http://www.antiquariaten.com/rashi/catalogs/cat000000446.htm>
<http://www.amazon.com/exec/obidos/ASIN/0864423381/inktombkasin-20/103-4179358-7107060>
<http://www.gy.com/www/ac.htm>
<http://www.multilingualbooks.com/amharic.html>
http://www.pricegrabber.com/search_getprod.php/isbn=9966401466/ut=3e93589c80da16d2
Description (2)
<http://www.cyberethiopia.com/ethiopie/amharique.html>
<http://www.abysiniagateway.net/info/bender.html>
Échantillons (9)
http://monsite.wanadoo.fr/conscience_rasta/page5.html
<http://www.lodace.com/mabel/jtaime.htm>
<http://www.travlang.com/languages/cgi-bin/langchoice.cgi>

?lang1=english&lang2=amharic&page=main
<http://www.addishomes.com/>
<http://www.medhin.org/>
<http://www.cfee-fces.org/Francais/catalogueb.htm>
<http://www.ifrance.com/soleilrueethiopie/amharlexik.htm>
<http://info.uibk.ac.at/c/c7/c707/info/oecd-macroth/fr/4276.html>
http://www.yebbo.com/sydneyolympic/MARATHON/Yidnekachew_Tesema/Chat/Zero/zero.html
Texte (2)
<http://www.greatcom.org/laws/amharic/>
<http://www.immunize.org/catg.d/p4170am.pdf>
Cours (1)
http://ourworld.compuserve.com/homepages/GenX_jt_mtjr/GenXAmharic.html
Site (6)
<http://members.aol.com/w4z5m4/html/amargna.html>
<http://members.aol.com/w4z5m4/wazema.html>
<http://www.netsanet.com/>
<http://www.iss.stthomas.edu/amharic/>
 DW : <http://www.dwelle.de/amharic/Welcome.html>
 VOE : <http://www.medhin.com/Radio/RADIOAMH.HTM>
LC-F (18)
<http://www.quartet.fr/amharique.html>
<http://www.univ-orleans.fr/EXT/ASTEX/astex/doc/fr/web/html/astex012.htm>
<http://www.cyberethiopia.com/ethiopie/amharique.html>
<http://membres.lycos.fr/pedroiy/alphab/amharique.htm>
http://www.extense.com/bin/x2cgi_view.cgi?userID=64393794&view=on&query=amharique&url=http%3A%2F%2Fwww.quid.fr%2Fweb.php%3Fweb%3D%2FWEB%2FETATS%2FET%2FQ029450.HTM#marker
<http://www.univ-orleans.fr/EXT/ASTEX/astex/doc/fr/web/html/astex003.htm>
<http://www.afrik.com/journal/internet/?net-140-2.htm>
<http://www.afrik.com/journal/internet/?net-140-2.htm>
http://monsite.wanadoo.fr/conscience_rasta/page5.html
<http://www.mediaport.net/AfricArt/100CD/CD/059.html>
<http://www.ifrance.com/soleilrueethiopie/amharlexik.htm>
<http://www.inalco.fr/pub/enseignements/langues/afrique/amharique/index.html>
<http://geonames.nrcan.gc.ca/francais/resolutionA16.html>
<http://info.uibk.ac.at/c/c7/c707/info/oecd-macroth/fr/4276.html>
<http://www.mediaport.net/AfricArt/100CD/CD/062.html>
<http://www.lodace.com/mabel/jtaime.htm>
<http://www.cean.u-bordeaux.fr/gdr931.html>
<http://www.culture.fr/culture/dglf/bpi/list-langues.html>
LC-A (36)
Bambara / Bamanakan (27)
Mention (12)
<http://www.joshuaproject.net/Assets/Profiles/jpl208.htm>
<http://www.isnar.cgiar.org/publications/books/sahel/english/chap2-1.htm>
<http://www.musicavisual.com/30092000.htm>
http://www.citl.ohiou.edu/~mugane/critical/africa_index.htm
<http://www.maliwatch.org/SOCIO-ECO/Hadi/BiennaleV1.htm>
<http://createinternational.com/Top211List.html>

<http://www4.gvsu.edu/coler/GPY351/Readings/Cole/LUChangeMali.htm>
<http://gallica.bnf.fr/VoyagesEnAfrique/Geo/ListePeupl.htm>
<http://www.archaeology.org/online/features/mali/>
<http://www.malilink.net/archive/0157.html>
<http://www.ad2000.org/peoples/untarget.htm>
<http://www.soeursdelacharitestlouis.qc.ca/charitedestlouis.4/Magnanbugu.htm>
Documentation (3)
<http://www.goshen.edu/~megss/malibambara.html>
<http://www.eva.mpg.de/lingua/wals/atlasbib.html>
<http://arthistory.rutgers.edu/faculty/sbs.htm>
Description (10)
http://www.ethnologue.com/show_language.asp?code=BRA
<http://www.rosettaproject.org:8080/live/search/browsebyfamilyresult?searchtype=family&searchkey=Niger-Congo>
http://www.bethany.com/profiles/p_code2/1815.html
<http://www.christusrex.org/www3/ethno/Mali.html>
http://www.ethnologue.com/show_country.asp?name=CÃte+d'Ivoire
http://www.ethnologue.com/show_country.asp?name=Mali
http://www.ethnologue.com/show_country.asp?name=Burkina+Faso
<http://ikuska.com/Africa/Etnologia/Pueblos/Bamana/>
<http://www.christusrex.org/www3/ethno/Sene.html>
Échantillons (1)
<http://www.goshen.edu/~megss/bambara.html>
Textes (1)
 PN+AM <http://www.christusrex.org/www1/pater/JPN-bambara.html>
LC-F (4)
<http://www.maliwatch.org/SOCIO-ECO/Hadi/BiennaleV1.htm>
<http://gallica.bnf.fr/VoyagesEnAfrique/Geo/ListePeupl.htm>
<http://www.malilink.net/archive/0157.html>
<http://www.soeursdelacharitestlouis.qc.ca/charitedestlouis.4/Magnanbugu.htm>
LC-A (21)
<http://www.goshen.edu/~megss/malibambara.html>
<http://www.goshen.edu/~megss/bambara.html>
<http://www.christusrex.org/www1/pater/JPN-bambara.html>
<http://www.ad2000.org/peoples/jpl208.htm>
http://www.bethany.com/profiles/p_code2/1815.html
<http://www.isnar.cgiar.org/publications/books/sahel/english/chap2-1.htm>
http://www.ethnologue.com/show_language.asp?code=BRA
<http://www.christusrex.org/www3/ethno/Mali.html>
http://www.ethnologue.com/show_country.asp?name=CÃte+d'Ivoire
http://www.citl.ohiou.edu/~mugane/critical/africa_index.htm
http://www.ethnologue.com/show_country.asp?name=Mali
http://www.ethnologue.com/show_country.asp?name=Burkina+Faso
<http://createinternational.com/Top211List.html>
<http://www4.gvsu.edu/coler/GPY351/Readings/Cole/LUChangeMali.htm>
<http://www.archaeology.org/online/features/mali/>
<http://www.eva.mpg.de/lingua/wals/atlasbib.html>
<http://www.christusrex.org/www3/ethno/Sene.html>
<http://www.rosettaproject.org:8080/live/search/browsebyfamilyresult?searchtype=family&searchkey=Niger-Congo>
<http://arthistory.rutgers.edu/faculty/sbs.htm>
<http://www.ad2000.org/peoples/untarget.htm>
LC-O (2)

<http://www.musicavvisual.com/30092000.htm>
<http://ikuska.com/Africa/Etnologia/Pueblos/Bamana/>

Bangala (5)

Mention (1)

<http://www.christusrex.org/www1/pater/>

Texte (4)

AM: <http://www.christusrex.org/www1/pater/JPN-bangala.html>

PN: <http://www.christusrex.org/www1/pater/JPN-lingala-bangala.html>

Jean 1: <http://www.language-museum.com/b/bangala.htm>

4SL: <http://www.greatcom.org/spirit/bangala/>

LC-A (5)

Tous les sites trouvés sur cette langue sont en anglais.

Baulé (1)

Mention (1)

<http://www.ling.ohio-state.edu/research/jpcl/online/snotes/sn3.htm>

LC-A (1)

Bemba (16)

Mention (2)

<http://www.2001pray.org/PeopleGroups/Bemba.htm>

<http://www.anthro.mankato.msus.edu/cultural/oldworld/africa/bemba.html>

Documentation (7)

http://www.emory.edu/COLLEGE/LINGUISTICS/POLYGLOT/bemba_resources.html

<http://www.yourdictionary.com/languages/nigecong.html>

Bibliographie

<http://www.albany.edu/~lb527/Bemba.html>

<http://www.ethnicharvest.org/bibles/bemba.htm>

<http://www.foreignmedia.com/ilc-BJDC.html>

<http://www.worldlanguage.com/Languages/Bemba.htm>

http://www.walmart.com/catalog/product.gsp?product_id=348560&sourceid=0100000030390613102498

Description (6)

<http://www.emory.edu/COLLEGE/ANTHROPOLOGY/FACULTY/ANTDS/Bemba/profile.html>

<http://www.flw.com/languages/bemba.htm>

http://www.emory.edu/COLLEGE/ANTHROPOLOGY/FACULTY/ANTDS/Bemba/speech_profiles.html

<http://www.ikuska.com/Africa/Etnologia/Pueblos/Bemba/>

<http://www.linguistics.berkeley.edu/CBOLD/Data/LangInfoFiles/Bemba.M.info.html>

http://www.ethnologue.com/show_language.asp?code=BEM

Échantillons (1)

<http://www.emory.edu/COLLEGE/ANTHROPOLOGY/FACULTY/ANTDS/Bemba/>

LC-A (16)

Dinka (4)

Documentation (1)

<http://www.worldlanguage.com/Languages/Dinka.htm>

Description (2)

<http://www.flw.com/languages/dinka.htm>

<http://www.joshuaproject.net/Assets/Profiles/jpl394.htm>

Texte (1)

PN: <http://www.christusrex.org/www1/pater/JPN-dinka-padang.html>

LC-A (4)

Tous les sites trouvés sur cette langue sont en anglais.

Dyula / Dyoula / Dioula / Julia (42)

Mention (9)

<http://www.ksafe.com/profiles/clusters/8046a.html>

<http://www.rosettaproject.org:8080/live/search/browsebycountryresult?searchtype=countries&searchkey=Burkina%20Faso>

http://www.msds.ch/international/burkina/dvorb_welcome.htm

http://perso.club-internet.fr/galiotth/Cote_Ivoire/ci_index.html

<http://yoursay.imdb.com/Sections/Languages/Dioula/>

<http://www.xrefer.com/entry/571266>

<http://www.ilgradio.com/awr-monitor/awrafWAF.htm>

<http://www.inemanltd.com/blog/>

Documentation (14)

http://www.ethnologue.com/show_iso639.asp?code=dyu

<http://www.oasis-open.org/cover/langcodesEth.txt>

<http://www.w3.org/WAI/ER/IG/ert/iso639.htm>

<http://www.dsv.su.se/~jpalme/ietf/language-codes.html>

<http://www.oasis-open.org/cover/nisoLang3-1994.html>

<http://nativenet.uthscsa.edu/archive/ng/93/0182.html>

<http://www.indiana.edu/~librcsd/afrlg/data/0416.html>

<http://www.hum.gu.se/arkiv/BANTU-L/current/msg00111.html>

<http://www.unice.fr/CRI/ml/html/comparlingafric/1999/msg00020.html>

<http://languagelab.bh.indiana.edu/audio/dtapes.html>

<http://www.indiana.edu/~librcsd/afrlg/data/0149.html>

<http://www.amazon.com/exec/obidos/ASIN/2723605574/avsearch-bkasin-20/103-6819673-1337410>

<http://www.worldlanguage.com/Spanish/Languages/Dyula.htm>

<http://www.worldlanguage.com/Languages/Dyula.htm>

Description (16)

http://www.ethnologue.com/show_lang_family.asp?code=DGD

<http://www.ad2000.org/peoples/jpl208.htm>

<http://www.ad2000.org/peoples/jpl166.htm>

<http://www.ad2000.org/peoples/jpl172.htm>

<http://www.ad2000.org/peoples/jpl156.htm>

<http://www.ikuska.com/Africa/Etnologia/Pueblos/Julia/index.htm>

http://www.ethnologue.com/show_country.asp?name=Burkina+Faso

http://www.ethnologue.com/show_country.asp?name=Mali

http://www.ethnologue.com/show_country.asp?name=CÃ¢te+d'Ivoire

http://www.ethnologue.com/show_language.asp?code=DYO

<http://www.christusrex.org/www3/ethno/Burk.html>

<http://members.aol.com/grn4jean/bf.htm>

http://www.ethnologue.com/show_language.asp?code=DGD

<http://www.christusrex.org/www3/ethno/Sene.html>

http://www.ethnologue.com/show_country.asp?name=C%CC3%B4te+d%27Ivoire

<http://www.ad2000.org/peoples/jpl260.htm>

Échantillons (5)

<http://www.inforoutes-ardeche.fr/~l.bony/lexique.htm>

<http://home.earthlink.net/~riende9/dioula.html>

http://www.msds.ch/international/burkina/fprep_dioula.htm

<http://home.earthlink.net/~riende9/Africa.html>

<http://aliaspilote.com/bobodioula.htm>

LC-F (3)

<http://www.inforoutes-ardeche.fr/~l.bony/lexique.htm>

http://perso.club-internet.fr/galiotth/Cote_Ivoire/ci_index.html

<http://aliaspilote.com/bobodioula.htm>

LC-A (38)

LC-0

Espagnol

<http://www.worldlanguage.com/Spanish/Languages/Dyula.htm>

Fulfulde / Pulaar / Pular / Peul / Fulani (75)

Mention (2)

<http://news.uns.purdue.edu/UNS/expert.language/Pulaar.htm>

<http://fr.allafrica.com/stories/200205080842.html>

Documentation (38)

<http://www.languages-on-the-web.com/links/link-fulani.htm>

<http://africandl.org/research/pulaar-digitization.pdf>

<http://www.lmp.ucla.edu/>

<http://www.vada.nl/talenen.htm>

<http://languagelab.bh.indiana.edu/audio/fgtapes.html>

http://www.ethnologue.com/language_index.asp?letter=F

http://www.ethnologue.com/show_iso639.asp?code=ful

http://www.loc.gov/marc/languages/lang_o2s.html

<http://www.bisharat.net/A12N/FulUnicode.htm>

<http://laawol.pulaaku.net/langue/index.html>

<http://www.bpi.fr/9/peul.html>

<http://pages.ifrance.com/pages/langues/fulfulde/>

<http://www.people-park.com/l-FUV.htm>

<http://www.ceci.ca/w/i/n/m002335.htm>

<http://www.ceci.ca/w/i/mid11.htm>

http://www.nextag.com/Fulani_English_Practical_Dictionary~965758z2znzz1zmainz2-htm

<http://www.foreignmedia.com/ilc-PUGR.html>

<http://www.multilingualbooks.com/tlstore/pulaar.html>

http://www.ethnologue.com/show_language.asp?code=FUC

http://www.walmart.com/catalog/product.gsp?product_id=348665&sourceid=010000030390613102498

<http://www.multilingualbooks.com/pulaar.html>

<http://www.worldlanguage.com/Languages/PulaarFulani.htm>

<http://www.amazon.com/exec/obidos/ASIN/0761806113/avsearch-bkasin-20/103-4179358-7107060>

http://www.nextag.com/Pulaar_English_English_Pulaar~1553484z2znzz1zmainz2-htm

<http://www.worldlanguage.com/Italian/Languages/PulaarFulani.htm>

<http://www.101language.com/pulaar.html>

<http://www.amazon.com/exec/obidos/ASIN/3496025492/inktomi-bkasin-20/103-4179358-7107060>

<http://my.linkbaton.com/get?lbCC=q&nC=q&genre=book&item=0781804795>

http://www.maps2anywhere.com/Languages/languages_-_pulaar.htm

<http://www.worldlanguage.com/Products/31280.htm>

<http://www.language.com.au/pulaar.html>

<http://www.worldlanguage.com/Languages/PulaarFulani.htm>

<http://www.lmp.ucla.edu/>

<http://www.worldlanguage.com/Products/Peul/Tutorial-Learning/Page1.htm>

<http://www.imdb.com/Sections/Languages/Peul/>
<http://www.amazon.com/exec/obidos/ASIN/0320001830/avsearch-bkasin-20/103-4179358-7107060>
<http://www.amazon.com/exec/obidos/ASIN/3496025492/103-4179358-7107060>
<http://www.worldlanguage.com/Products/2722.htm>
<http://my.linkbaton.com/get?genre=book&item=0870133268>
http://www.walmart.com/catalog/product.gsp?product_id=1925316&sourceid=0100000030390613102498
Description (20)
http://www.ethnologue.com/show_lang_family.asp?code=FUC
http://www.ethnologue.com/show_lang_family.asp?code=FUH
http://www.ethnologue.com/show_lang_family.asp?code=FUQ
http://www.ethnologue.com/show_lang_family.asp?code=FUL
http://www.ethnologue.com/show_lang_family.asp?code=FUE
http://www.ethnologue.com/show_lang_family.asp?code=FUB
http://www.ethnologue.com/show_lang_family.asp?code=FUI
http://www.ethnologue.com/show_lang_family.asp?code=FUV
<http://www.uiowa.edu/~africart/toc/people/Fulani.html>
<http://www.joshuaproject.net/Assets/Profiles/jpl1807.htm>
http://www.gateway-africa.com/tribe/fulani_tribe.html
http://www.isp.msu.edu/AfrLang/Fula_root.htm
<http://www.shoalhaven.net.au/~kjpercival/transliteration.html>
<http://www.foreignmedia.com/ilc-main-PU.html>
http://www.ethnologue.com/show_language.asp?code=FUC
<http://www.flw.com/languages/fulfulde.htm>
http://www.ethnologue.com/show_language.asp?code=FUB
<http://www.flw.com/languages/fulfulde.htm>
<http://www.ikuska.com/Africa/Etnologia/Pueblos/Fulani/>
http://text.morgans-resources.com/Constraints_on_Pulaar_Phonology.html
Échantillons (2)
<http://www.itinerariafricani.net/Peul.htm>
<http://www.townbeacon.com/Nufulani/fulfulde.htm>
Texte (6)
 DH: <http://www.unhchr.ch/udhr/lang/fum1.htm>
 DH: <http://www.unhchr.ch/udhr/lang/fum.htm>
 DH: <http://www3.itu.int/udhr/lang/fum1.htm>
 Jean 1: <http://www.worldscriptures.org/pages/fulapulaartoucouleur.html>
 Jean 1: <http://www.worldscriptures.org/pages/fulfulde.html>
 PN: <http://www.christusrex.org/www1/pater/JPN-fulfulde-adamawa.html>
Cours (1)
<http://www.univie.ac.at/afrikanistik/african.fulcourse.html>
Site (3)
<http://www.pulaaku.net/>
<http://laawol.pulaaku.net/langue/index.html>
<http://www.entm.purdue.edu/entomology/research/cowpea/Extension%20bulletins/fulfulde.htm>
LC-F (3)
<http://www.bpi.fr/9/peul.html>
<http://fr.allafrica.com/stories/200205080842.html>
<http://www.ceci.ca/w/i/mid11.htm>
LC-A (71)
LC-O
 Italien

<http://www.itinerariafricani.net/Peul.htm>

Ewondo / Beti (29)

Mention (8)

<http://www.zompist.com/sources.htm>
<http://www.tlcdelivers.com/tlc/crs/lang0137.htm>
<http://www.flw.com/merry.htm>
<http://www.flw.com/languages/>
<http://www.sanctuaire-montligeon.com/priere/bas.html>
<http://www.language-museum.com/>
<http://membres.lycos.fr/elefantehiesel/humour/fables02.htm>
<http://www.tradeport.org/ts/countries/cameroon/bnotes.html>

Documentation (11)

<http://www.w3.org/WAI/ER/IG/ert/iso639.htm>
http://www.ethnologue.com/show_iso639.asp?code=ewo
<http://lcweb.loc.gov/standards/iso639-2/englangn.html>
<http://www.oasis-open.org/cover/nisoLang3-1994.html>
<http://www.dsv.su.se/~jpalme/ietf/language-codes.html>
<http://citeseer.nj.nec.com/context/112791/0>
<http://www.indiana.edu/~librcsd/afrlg/data/0165.html>
<http://cifdi.francophonie.org/Litter8.cfm>
<http://www.progiciels-bpi.ca/tcao/apercu-f.html>
<http://www.addall.com/Browse/Detail/1555674917.html>
<http://www.pucac.com/p33.html>

Description (5)

http://www.ethnologue.com/show_lang_family.asp?code=EWO
http://home.germany.net/101-108380/html/a_-_b.html
<http://www.flw.com/languages/ewondo.htm>
<http://www.culture.fr/culture/dglf/bpi/ewondo.html>
http://www.ethnologue.com/show_language.asp?code=EWO
Échantillon (2)
<http://www.elite.net/~runner/jennifers/hello.htm>
<http://www.namediffusion.net/talange/index.php>

Textes (2)

PN: <http://www.christusrex.org/www1/pater/JPN-ewondo.html>
 PN: <http://www.language-museum.com/e/ewondo.htm>
 PN: <http://www.christusrex.org/www1/pater/JPN-beti.html>
LC-F (6)
<http://www.indiana.edu/~librcsd/afrlg/data/0165.html>
<http://www.culture.fr/culture/dglf/bpi/ewondo.html>
<http://www.pucac.com/p33.html>
<http://membres.lycos.fr/elefantehiesel/humour/fables02.htm>
<http://cifdi.francophonie.org/Litter8.cfm>
<http://www.progiciels-bpi.ca/tcao/apercu-f.html>
LC-A (23)

Fon-gbe (30)

Mention (11)

<http://www.language-museum.com/>
<http://www.ikuska.com/Africa/Etnologia/Pueblos/Fon/>
<http://www.vada.nl/talenee.htm>
<http://www.vada.nl/talenen.htm>
http://www.ethnologue.com/show_language.asp?code=MXL
<http://www.campuscrusade.com/languages.htm>
<http://www.thejesusvideo.com/languages.htm>
<http://www.rosettaproject.org:8080/live/search/browsebylangname>
<http://www.izf.net/izf/Guide/Benin/pays.htm>

http://www.ethnologue.com/language_index.asp?letter=F
http://www.ethnologue.com/show_iso639.asp?code=ewe
Documentation (1)
http://www.ethnologue.com/show_iso639.asp?code=fon
Description (13)
http://www.ethnologue.com/show_lang_family.asp?code=FOA
http://www.ethnologue.com/show_family.asp?subid=70
http://www.ethnologue.com/show_family.asp?subid=44
http://www.ethnologue.com/show_family.asp?subid=42
http://www.ethnologue.com/show_family.asp?subid=43
<http://locke.ccil.org/~cowan/langtree.txt>
http://www.ethnologue.com/show_language.asp?code=FOA
http://www.ethnologue.com/show_country.asp?name=Benin
<http://www.christusrex.org/www3/ethno/Beni.html>
<http://www.ikuska.com/Africa/Etnologia/Pueblos/Fon/>
<http://members.aol.com/grn4jean/bj.htm>
http://www.ethnologue.com/show_country.asp?name=Togo
<http://www.christusrex.org/www3/ethno/Togo.html>
Échantillons (1)
<http://www.elite.net/~runner/jennifers/hello.htm>
Texte (2)
 PN: <http://www.christusrex.org/www1/pater/JPN-fon.html>
 DH: <http://www.language-museum.com/ffon-gbe.htm>
Cours (2)
http://www.geocities.com/fon_is_fun/index.html
http://www.geocities.com/fon_is_fun/French/fr_index.html
LC-F (2)
<http://www.izf.net/izf/Guide/Benin/pays.htm>
http://www.geocities.com/fon_is_fun/French/fr_index.html
LC-A (27)
LC-O (1)
 Espagnol
<http://www.ikuska.com/Africa/Etnologia/Pueblos/Fon/>
Gusii / Kisii (26)
Mention (7)
<http://bluegecko.crosswinds.net/kenya/tribes/gusii/gusii.htm>
<http://www.language-museum.com/>
<http://classweb.gmu.edu/accnt/gusii.html>
<http://classweb.gmu.edu/accnt/generalizations/gusii1/gusii1gen.html>
<http://classweb.gmu.edu/accnt/>
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=GUZ>
<http://www.kabalarians.com/female/kisii-f.htm>
Documentation (9)
<http://classweb.gmu.edu/accnt/>
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=GUZ>
<http://www.kisii.com/>
<http://www.kenyaweb.com/educ/kisii.html>
<http://www.indiana.edu/~librcsd/afri1g/data/0219.html>
<http://bluegecko.crosswinds.net/kenya/tribes/gusii/gusii.htm>
<http://www.amazon.com/exec/obidos/ASIN/9966884742/103-4179358-7107060>
<http://www.amazon.com/exec/obidos/ASIN/9966884742/inktomi-bkasin-20/103-4179358-7107060>
<http://www.worldlanguage.com/Languages/Kisii.htm>

Description (2)
http://www.ethnologue.com/show_lang_family.asp?code=GUZ
http://www.ethnologue.com/show_language.asp?code=GUZ
Échantillons (3)
<http://www.cam.ac.uk/societies/kenyap/>
<http://www.kabalarians.com/female/kisii-f.htm>
<http://www.cam.ac.uk/societies/kenyap/>
Texte (3)
 Jean 1 ; <http://www.language-museum.com/g/gusii.htm>
 PN: <http://www.christusrex.org/www1/pater/JPN-kisii.html>
<http://classweb.gmu.edu/accnt/gusii.html>
Cours
<http://classweb.gmu.edu/accnt/generalizations/gusii1/gusii1gen.html>
Site (1)
 Forum en ekegusi et en anglais
<http://www.kisii.com/>
LC-A (26)
Hausa (55)
Mention (17)
<http://www.lightlink.com/wysiati/LON/h/haousa.html>
<http://www.codecan.org/francais/activities/ghana.html>
<http://www.uni-bayreuth.de/afrikanistik/mega-tchad/Bulletin/bulletin2001/programmes/travaux.html>
<http://www.courrierint.com/pays/ng.htm>
<http://www.centraltranslations.co.uk/french/language.htm>
<http://www.sis.gov.eg/public/africanmag/issue02/html/frafr04.htm>
<http://www.africessor.org/Article%20-%20Autour%20du%20Swahili,%20page1.htm>
<http://www.hartford-hwp.com/archives/30/133.html>
<http://www.unesco.org/bpi/fre/unescopresse/99-72f.htm>
<http://www.alaska.net/~silenus/DUT3.HTM>
<http://home.t-online.de/home/Petra.Bode/wort.htm>
http://www.sas.upenn.edu/African_Studies/Hornet/irin_41499.html
<http://www.africaonline.co.ci/AfricaOnline/infos/lejour/1257CUL1.HTM>
<http://www.bbc.co.uk/cymru/radio/erthylgau/talwrn11.shtm>
<http://www.angelfire.com/az/ziyadah/>
Documentation (18)
<http://www.humnet.ucla.edu/humnet/aflang/Hausa/indexframe.html>
<http://www.humnet.ucla.edu/humnet/aflang/Hausa/hausa.html>
<http://www.language.com.au/hausa.html>
<http://www.humnet.ucla.edu/humnet/aflang/>
[Hausa%20Class%20Page/hausa_class_home.html](http://www.humnet.ucla.edu/humnet/aflang/Hausa%20Class%20Page/hausa_class_home.html)
<http://www.humnet.ucla.edu/humnet/aflang/Hausa/References/hrefs.html>
<http://www.univie.ac.at/afrikanistik/oracle/KofarHausaE.html>
http://languages.bigtome.com/big/page/HAUSA_
<http://polyglot.lss.wisc.edu/lss/lang/afrikan.html>
http://www3.aa.tufs.ac.jp/~P_aflang/TEXTS/sept96/philips.txt
http://www3.aa.tufs.ac.jp/~P_aflang/TEXTS/june97/philips.html
<http://www.sacred-texts.com/afri/hausa/index.htm>
<http://www.uni-bayreuth.de/afrikanistik/cnl/01/index.html>
<http://www.emich.edu/~linguist/ling-afri.html>
<http://www.linguistlist.org/pubs/africaine.html>
http://www.unesco.org/culture/africa/html_fr/comite.htm
<http://www.indiana.edu/~sal/titleindex.html>

<http://www.encyclopedia.com/searchpool.asp?target=@DOCTITLE%20Hausa>
<http://www.hausa.com/>
Description (13)
<http://www.berclo.net/page99/99fr-afr-languages.html>
http://www.ethnologue.com/show_country.asp?name=Cameroon
<http://ikuska.com/Africa/Etnologia/Pueblos/Hausa/>
http://lucy.ukc.ac.uk/EthnoAtlas/Hmar/Cult_dir/Culture.7844
<http://www.uiowa.edu/~africart/toc/people/Hausa.html>
<http://www.joshuaproject.net/Assets/Profiles/jpl1809.htm>
<http://wiem.onet.pl/wiem/007f76.html>
http://www.ethnologue.com/show_country.asp?name=Niger
http://www.ethnologue.com/show_country.asp?name=Nigeria
http://www.humnet.ucla.edu/humnet/aflang/Hausa/Hausa_online_grammar/grammar_frame.html
http://www.humnet.ucla.edu/humnet/aflang/Hausa/Hausa_online_grammar/Verb%20forms/verb_forms.html
<http://www.univie.ac.at/afrikanistik/SyntaxODL/HausaODL/H4.html>
<http://www.eirelink.com/alanking/modals/documents/do-g-hau.htm>
Échantillons (1)
<http://www.hausavoices.com/>
Site (6)
<http://www.gumel.com/>
<http://www.humnet.ucla.edu/humnet/aflang/Hausa/References/maillingList.html>
http://www.musavilari.org/html/12/index_d.htm
<http://www.univie.ac.at/afrikanistik/oracle/KofarHausa.html>
 BBC: <http://www.bbc.co.uk/hausaa/>
 DW: <http://www.dwelle.de/hausaa/Welcome.html>
LC-F (14)
<http://www.berclo.net/page99/99fr-afr-languages.html>
<http://www.codecan.org/francais/activities/ghana.html>
<http://www.uni-bayreuth.de/afrikanistik/mega-tchad/Bulletin/bulletin2001/programmes/travaux.html>
<http://www.courrierint.com/pays/ng.htm>
<http://www.centraltranslations.co.uk/french/language.htm>
<http://www.sis.gov.eg/public/africanmag/issue02/html/frafr04.htm>
<http://www.africessor.org/Article%20-%20Autour%20du%20Swahili,%20page1.htm>
<http://www.imam-khomeini.org/english/publication/publish-right.htm>
<http://www.emich.edu/~linguist/ling-afr.html>
<http://www.unesco.org/bpi/fre/unescopresse/99-72f.htm>
<http://www.africaonline.co.ci/AfricaOnline/infos/lejour/1257CUL1.HTM>
http://www.unesco.org/culture/africa/html_fr/comite.htm
LC-A (35)
LC-O (6)
<http://www.bbc.co.uk/cymru/radio/erthylgau/talwrn11.shtml>
<http://www.alaska.net/~silenus/DUT3.HTM>
 Allemand
<http://home.t-online.de/home/Petra.Bode/wort.htm>
<http://www.univie.ac.at/afrikanistik/SyntaxODL/HausaODL/H4.html>
 Espagnol
<http://ikuska.com/Africa/Etnologia/Pueblos/Hausa/>
 Polonais
<http://wiem.onet.pl/wiem/007f76.html>

Ibibio (4)

Description (3)

http://www.isp.msu.edu/AfrLang/efik_root.html
<http://www.uiowa.edu/~africart/toc/people/Ibibio.html>
<http://www.gateway-africa.com/tribe/ibibio-tribe.html>

Échantillons (1)

<http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/appendix/languages/ibibio/ibibio.html>
LC-A (4)

Tous les sites trouvés sur cette langue sont en anglais.

Igbo (10)

Mention (5)

http://www.igbo.co.uk/opilca_index.htm
<http://www.igbo.biz/igbo/>
<http://www.geocities.com/Athens/Acropolis/3629/>
<http://www.shef.ac.uk/~bsp98coi/seminar.html>
<http://www.umunna.org/igbos.htm>

Documentation (3)

<http://monarch.gsu.edu/nutrition/Igbo.htm>
<http://www.chiamaka.com/igbonames.htm>
<http://www.igbo.com/>

Description (1)

<http://www.uiowa.edu/~africart/toc/people/Igbo.html>
Cours (1)
<http://www.geocities.com/Athens/Acropolis/3629/igbo.html>
LC-F
<http://www.izf.net/izf/Guide/Benin/pays.htm>
LC-A (10)

Kabyle (4)

Mention (1)

<http://linguistlist.org/~ask-ling/archive-most-recent/msg00277.html>

Description (2)

<http://www.flw.com/languages/kabyle.htm>
<http://www.ad2000.org/peoples/jpl1176.htm>

Texte (2)

PN + AM: <http://www.christusrex.org/www1/pater/JPN-kabyle.html>
LC-A (4)

Tous les sites trouvés sur cette langue sont en anglais.

Kalenjin (13)

Mention (2)

http://www.mailafrica.net/credit_section_kalenjin.php
<http://www.christianvideos.org/kalenjinvideos.html>

Documentation (3)

<http://www.worldlanguage.com/Languages/Kalenjin.htm>
<http://www.worldlanguage.com/Languages/Kalenjin.htm>
<http://www.worldlanguage.com/Spanish/Languages/Kalenjin.htm>
Description (5)

http://www.ethnologue.com/show_family.asp?subid=2228
http://www.ethnologue.com/show_lang_family.asp?code=KLN
http://www.isp.msu.edu/AfrLang/kalenjin_root.html
<http://www.ikuska.com/Africa/Etnologia/Pueblos/Kalenjin/>
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=KLN>

Texte (3)

PN: <http://www.christusrex.org/www1/pater/JPN-kalenjin.html>

4SL: <http://www.greatcom.org/laws/kalenjin/>

4SL: <http://www.greatcom.org/spirit/kalenjin/>

LC-A (13)

Tous les sites trouvés sur cette langue sont en anglais.

Kamba (5)

Documentation (1)

<http://www.worldlanguage.com/Languages/Kamba.htm>

Description (3)

http://www.ethnologue.com/show_family.asp?subid=2262

<http://www.flw.com/languages/kamba.htm>

http://www.isp.msu.edu/AfrLang/kamba_root.html

Site (1)

<http://chumvi.tripod.com/mkizi.htm>

LC-A (4)

Kanuri (9)

Mention (1)

<http://www.bartleby.com/61/26/K0012650.html>

Documentation (5)

<http://www.amazon.com/exec/obidos/ASIN/3927620440/>

[avsearch-bkasin-20/002-3344447-9670423](http://www.amazon.com/exec/obidos/ASIN/0881332437/)

<http://www.amazon.com/exec/obidos/ASIN/0881332437/>

[avsearch-bkasin-20/002-3344447-9670423](http://www.amazon.com/exec/obidos/ASIN/0881332437/)

<http://www.worldlanguage.com/French/Languages/Kanuri.htm>

<http://www.worldlanguage.com/French/Languages/Kanuri.htm>

<http://www.worldlanguage.com/Languages/Kanuri.htm>

Description (2)

http://www.isp.msu.edu/AfrLang/kanuri_root.html

<http://www.ikuska.com/Africa/Etnologia/Pueblos/Kanuri/index.htm>

Texte (1)

<http://www.language-museum.com/k/kanuri.htm>

LC-F (1)

<http://www.worldlanguage.com/French/Languages/Kanuri.htm>

LC-A (8)

Kikoóngo / Kikongo (25)

Mention (8)

<http://www.kikongo.hpg.ig.com.br/index.html>

<http://www.vwv.cc/sprachen/kikongo.html>

<http://www.hum.gu.se/arkiv/BANTU-L/current/msg00049.html>

<http://www.hcjb.org.ec/Radio/schedB01esp.php3?sched=kikongo>

[http://www.moment-art.com/d/sander/show_](http://www.moment-art.com/d/sander/show_language.php3?lang_id=kikongo)

[language.php3?lang_id=kikongo](http://www.moment-art.com/d/sander/show_language.php3?lang_id=kikongo)

<http://us.imdb.com/Sections/Languages/Kikongo/links>

<http://www.worldscriptures.org/pages/photo-pages/kikongo.html>

<http://www.hyperdictionary.com/dictionary/Kikongo>

Documentation (5)

<http://uk.imdb.com/Sections/Languages/Kikongo/>

<http://us.imdb.com/Sections/Languages/Kikongo/sections>

<http://www.criabd.net/kikongo.htm>

<http://www.worldlanguage.com/Languages/Kikongo-SeeKongo.htm>

<http://www.attica->

[languages.com/recherche_avance.php/nuray/1335/rayon/afrique](http://www.attica-languages.com/recherche_avance.php/nuray/1335/rayon/afrique)

Description (3)

<http://members.aol.com/inquiceweb/ndinga.html>

<http://members.aol.com/inquiceweb/ndinga.html>

<http://www.ikuska.com/Africa/Lenguas/kikongo/morfologia.htm>

Échantillons (2)

<http://www.gospelweb.net/salvationindex.htm>

http://www.bantu.ovh.org/lari_greetings.html

Texte (3)

Jean 1: <http://www.worldscriptures.org/pages/kikongo.html>

Jean 1:

<http://www.worldscriptures.org/pages/kikongosansalvador.html>

Poème: <http://members.aol.com/inquiceweb/mambu.html>

Site (4)

<http://www.seghea.com/pat/bible/kikongo2.html>

Site kikongo - anglais

<http://www.gospelweb.net/kikongo.htm>

<http://www.ikuska.com/Africa/Lenguas/kikongo/>

<http://www.myhomepage.net/~jhdearmer/kikongo.htm>

LC-F (2)

<http://www.criabd.net/kikongo.htm>

<http://www.attica->

[languages.com/recherche_avance.php/nuray/1335/rayon/afrique](http://www.attica-languages.com/recherche_avance.php/nuray/1335/rayon/afrique)

LC-A (21)

LC-O (5)

Italien

<http://www.hcjb.org.ec/Radio/schedB01esp.php3?sched=kikongo>

Portugais

<http://www.kikongo.hpg.ig.com.br/index.html>

Allemand

<http://www.vwv.cc/sprachen/kikongo.html>

[http://www.moment-art.com/d/sander/show_](http://www.moment-art.com/d/sander/show_language.php3?lang_id=kikongo)

[language.php3?lang_id=kikongo](http://www.moment-art.com/d/sander/show_language.php3?lang_id=kikongo)

Espagnol

<http://www.ikuska.com/Africa/Lenguas/kikongo/morfologia.htm>

Kikuyu / Gikuyu (38)

Mention (16)

<http://www.ethnicarvest.org/bibles/kikuyu.htm>

<http://www.encyclopedia.com/articles/06957.html>

http://www.kikuyu.com/KMB/_mesboa/0000028b.htm

<http://www.kabarians.com/male/kikuyu-m.htm>

<http://www.bluegecko.org/kenya/tribes/kikuyu/>

<http://www.kabarians.com/female/kikuyu-f.htm>

<http://www.blissites.com/kenya/people.html>

<http://www.gy.com/language/ki.htm>

http://www.fontmenu.com/site/_Gikuyu.html

<http://www.ethnicarvest.org/bibles/gikuyu.htm>

<http://store.buying->

[behemoth.com/A_Paradigmatic_Grammar_of_Gikuyu.html](http://store.buying-behemoth.com/A_Paradigmatic_Grammar_of_Gikuyu.html)

<http://www.bartleby.com/61/31/G0123100.html>

<http://webhome.idirect.com/~mcsi/yr4life/gikuyu.htm>

<http://webhome.idirect.com/~mcsi/mcintosh/jgikuyu.htm>

http://www.unesco.org/courier/2000_04/uk/doss12.htm

<http://www.newafrica.com/culture/articles/article14.htm>

Documentation (9)

<http://www.gy.com/www/ki.htm>

<http://www.worldlanguage.com/Languages/Kikuyu.htm>

http://www.kikuyu.com/KMB/_mesboa/00000074.htm
<http://www.fontmenu.com/site/language.html>
<http://csli-publications.stanford.edu/site/1575860767.html>
<http://www.worldlanguage.com/Languages/Gikuyu.htm>
<http://www.amazon.com/exec/obidos/ASIN/9966884726/inktomi-bkasin-20/103-0050914-5311058>
<http://www.worldlanguage.com/German/Languages/Gikuyu.htm>
http://www.nextag.com/Paradigmatic_Grammar_of_Gikuyu~1627035z2znzzz1zmainz2-htm?nxtg=5c73d_8FA2CD731ACE200B&product=2383166
Description (4)
http://www.ethnologue.com/show_lang_family.asp?code=KIU
http://www.isp.msu.edu/AfrLang/kikuyu_root.html
<http://www.flw.com/languages/gikuyu.htm>
http://www.ethnologue.com/show_language.asp?code=KIU
Échantillons (4)
<http://users.erols.com/bccsbs/africa/gikuyu.htm>
<http://www.fb10.uni-bremen.de/anglistik/kerkhoff/AfricanLit/Ngugi/NgugiDorn.htm>
http://www.uni-bayreuth.de/Afrikanologie/publikationen/bass_37.html
<http://www.flw.com/merry.htm>
Texte (4)
 4SL: <http://www.greatcom.org/laws/kikuyu/>
 Message: http://www.kikuyu.com/KMB/_mesboa/00000074.htm
 Message: http://www.kikuyu.com/KMB/_mesboa/00000293.htm
 Message: http://www.kikuyu.com/KMB/_mesboa/00000234.htm
Site (1)
<http://www.greatcom.org/laws/kikuyu/>
LC-A (36)
LC-O (1)
 Allemand
<http://www.worldlanguage.com/German/Languages/Gikuyu.htm>

Kimbundu (32)

Mention (5)
<http://www.anthro.mankato.msus.edu/cultural/oldworld/africa/kimbundu.html>
<http://www.forumnow.com.br/vip/foruns.asp?forum=55337>
<http://www.yahooligans.com/reference/dictionary/entries/30/k0063000.html>
<http://arthur.ihs.ac.at/~isa/diplom/node63.html>
<http://www.bartleby.com/61/93/M0169350.html>
Documentation (11)
http://193.194.138.190/udhr/lang/mlo_email.htm
<http://www3.yahooligans.com/reference/dictionary/entries/30/k0063000.html>
http://www.explorelanguages.com/Framed_Pages/Main_Frame/World_Links/Languages/Kimbundu.htm
<http://odin.indstate.edu/level1.dir/cml/rbsc/cordell/kimbundu.html>
<http://www.indiana.edu/~librcsd/afrlg/data/0442.html>
<http://www.xrefer.com/entry/572191>
<http://www.worldlanguage.com/Languages/Kimbundu-SeeUmbundu.htm>
<http://www.worldlanguage.com/German/Languages/Kimbundu-SeeUmbundu.htm>

<http://www.worldlanguage.com/Polish/Languages/Kimbundu-SeeUmbundu.htm>
<http://www.worldlanguage.com/Languages/Mbundu.htm>
<http://www.amazon.com/exec/obidos/ASIN/0898755743/103-5377749-5369454>
Description (7)
http://www.ethnologue.com/show_family.asp?subid=27
http://www.isp.msu.edu/AfrLang/kimbundu_root.html
<http://www.infoplease.com/ipd/A0505628.html>
http://www.ethnologue.com/show_country.asp?name=Angola
http://www.angolanos.net/alberto_pinto_kimbundu.htm
http://www.angolanos.net/alberto_pinto_kimbundu_intro.htm
<http://www.rosettaproject.org:8080/live/search/detailedlanguage-record?ethnocode=MLO>
Échantillons (3)
http://www.angolanos.net/alberto_pinto_kimb_port_vocab.htm
<http://www.banjoist.de/>
<http://www.hyperdictionary.com/dictionary/Kimbundu>
Texte (6)
 4SL: <http://www.gospelweb.net/kikongo.htm>
 DH: <http://www.unhchr.ch/udhr/lang/mlo.htm>
 DHxt: <http://www.language-museum.com/m/mbundu-loanda.htm>
 Jean_1: <http://www.worldscriptures.org/pages/kimbunduluanda.html>
 AM: <http://www.christusrex.org/www1/pater/JPN-mbundu-loanda.html>
 Relig.: <http://www.seghea.com/pat/bible/mbundu.html>
LC-A (27)
LC-O (5)
 Portugais
http://www.angolanos.net/alberto_pinto_kimb_port_vocab.htm
http://www.angolanos.net/alberto_pinto_kimbundu.htm
<http://www.forumnow.com.br/vip/foruns.asp?forum=55337>
http://www.angolanos.net/alberto_pinto_kimbundu_intro.htm
<http://arthur.ihs.ac.at/~isa/diplom/node63.html>

Kinyarwanda (28)

Mention (6)
<http://www.rwandaforum.org/>
<http://www.rwanda.net/>
<http://www.emich.edu/~linguist/issues/9/9-1405.html>
<http://www.takingitglobal.org/other/sitesearch.html?searchonly=mem&LanguageID=96>
http://reisebuch.de/specials/kauderwelsch/kinyarwanda_fuer_ruanda_und_burundi-162.html
<http://linguistlist.org/~ask-ling/archive-1998.1/msg01551.html>
Documentation (6)
<http://languagelab.bh.indiana.edu/audio/ktapes.html>
http://www.explorelanguages.com/Framed_Pages/Main_Frame/World_Links/Languages/Kinyarwanda.htm
<http://www.african.gu.se/prj011.html>
<http://www.scar.utoronto.ca/~binnick/TENSE/OnKiny.html>
<http://www.worldlanguage.com/Languages/Kinyarwanda.htm>
<http://www.amazon.com/exec/obidos/ASIN/2738415415/avsearch-bkasin-20/103-6819673-1337410>
Description (3)
<http://www.flw.com/languages/kinyarwanda.htm>
<http://uts.cc.utexas.edu/~smyers/kinyarwanda.pdf>

<http://www.lowani.nl/kirundi.htm>

Échantillons (6)

<http://www.hirondelle.org/hirondelle.nsf/525125ad0c4e00a0c12564e500424876/37d28433340c2d30c125675f007214f3?OpenDocument>
<http://mgaetan.citeglobe.com/>

[http://store.allisons-](http://store.allisons-planet.com/Kinyarwanda_and_Kirundi_Names.html)

[planet.com/Kinyarwanda_and_Kirundi_Names.html](http://news.bbc.co.uk/1/hi/world/africa/2051526.stm)

<http://news.bbc.co.uk/1/hi/world/africa/2051526.stm>

<http://www.iwacu1.com/advertise/>

<http://www.harencity.nl/~harenkirambo/vocabulaire1>

Texte (4)

DH: <http://www.unhchr.ch/udhr/lang/rua1.htm>

DH <http://www3.itu.int/udhr/lang/rua1.htm>

PN: <http://www.christusrex.org/www1/pater/JPN-kinyarwanda.html>

4SL: <http://www.greatcom.org/laws/kinyarwanda/>

Site (3)

<http://www.bbc.co.uk/greatlakes/>

<http://www.iwacu1.com/indamutso.html>

<http://www.preventgenocide.org/rw/links.htm>

LC-F

<http://mgaetan.citeglobe.com/>

LC-A (23)

LC-O (1)

Allemand

http://reisebuch.de/specials/kauderwelsch/kinyarwanda_fuer_ruanda_und_burundi-162.html

Kirundi (23)

Mention (3)

<http://www.flw.com/languages/kinyarwanda.htm>

<http://linguistlist.org/issues/8/8-1693.html>

<http://www.icrc.org/icrcfre.nsf/4f16c583316c85944125673900241f30/b0968226a6f518f9c1256a7600438962?OpenDocument>

Documentation (11)

http://www.maps2anywhere.com/Languages/languages_-_kirundi.htm

<http://languagelab.bh.indiana.edu/audio/ktapes.html>

<http://www.indiana.edu/~librcsd/afrlg/data/0300.html>

<http://www.christianvideos.org/kirundi.html>

<http://www.multilingualbooks.com/kirundi.html>

<http://www.amazon.com/exec/obidos/ASIN/2738415415/>

[avsearch-bkasin-20/103-5377749-5369454](http://www.amazon.com/exec/obidos/ASIN/0884323765/avsearch-bkasin-20/103-5377749-5369454)

<http://www.amazon.com/exec/obidos/ASIN/0884323765/avsearch-bkasin-20/103-5377749-5369454>

<http://www.multilingualbooks.com/tlstore/kirundi.html>

<http://www.language.com.au/kirundi.html>

<http://www.worldlanguage.com/Languages/Kirundi.htm>

<http://www.worldlanguage.com/French/Languages/Kirundi.htm>

Description (3)

<http://www.flw.com/languages/kirundi.htm>

<http://www.101language.com/kirundi.html>

<http://www.lowani.nl/kirundi.htm>

Échantillons (2)

<http://www.copcity.com/anthems/burundi.html>

[http://store.allisons-](http://store.allisons-planet.com/Kinyarwanda_and_Kirundi_Names.html)

[planet.com/Kinyarwanda_and_Kirundi_Names.html](http://store.allisons-planet.com/Kinyarwanda_and_Kirundi_Names.html)

Texte (3)

4SL: <http://www.greatcom.org/laws/kirundi/default.htm>

PN: <http://www.christusrex.org/www1/pater/JPN-kirundi.html>

DH: <http://www3.itu.int/udhr/lang/rud1.htm>

Site (1)

<http://www.bbc.co.uk/greatlakes/>

LC-F (3)

<http://www.icrc.org/icrcfre.nsf/4f16c583316c85944125673900241f30/b0968226a6f518f9c1256a7600438962?OpenDocument>

<http://www.worldlanguage.com/French/Languages/Kirundi.htm>

LC-A (19)

Kiswahili (45)

Mention (9)

<http://www.moja.com/domain.html>

<http://www.yale.edu/swahili/faq.html>

<http://www.ilt.kuleuven.ac.be/ilt/index.htm>

<http://esperanto.online.fr/esperant97/swahili2.htm>

<http://www.esperanto.net/>

<http://www.yale.edu/swahili/>

<http://www.intratext.com/X/SWA0013.htm>

http://www.glcom.com/hassan/swahili_history.html

<http://www.swahili.dk/>

Documentation (6)

<http://www.unb.ca/web/civil/dccchair/dmm/swahili.html>

<http://www.kiswahili.de/>

<http://ccat.sas.upenn.edu/afli/penmoo.html>

http://www.csclub.uwaterloo.ca/u/mdmmrawi/Swahili_lit2.html

<http://www.geocities.com/africanwriters/Contemporaryswahili.html>

<http://www.yale.edu/swahili/sound/pronounce.htm>

Description (2)

<http://www.yale.edu/swahili/grammar/grammar.htm>

<http://www.iath.virginia.edu/swahili/swahili.html>

Échantillons (9)

<http://dmoz.org/World/Kiswahili/>

<http://www.caps.ucsf.edu/KFSIndex.html>

<http://www.uiowa.edu/~linguist/swahili/proverbs.html>

<http://www.glcom.com/hassan/swahili.html>

<http://www.yale.edu/swahili/>

<http://www.freedict.com/onldict/swa.html>

<http://www.glcom.com/hassan/swahili.html>

<http://www.yale.edu/swahili/dictionaries.html>

http://www.kent.wednet.edu/curriculum/soc_studies/kenya/

[swahili.html](http://www.kent.wednet.edu/curriculum/soc_studies/kenya/swahili.html)

Texte (4)

Droits de l'Homme:

<http://www.utamaduni.dk/swahili/human.htm>

Versets bibliques

<http://www.seghea.com/pat/bible/kiswahili.html>

Hymne national kenyan

<http://www.kenya.go.ke/anthem.html>

Swahili Bible concordance

<http://www.seghea.com/pat/bible/kiswahili.html>

Cours (3)

<http://www.ipl.org/div/kidspace/hello/swahili.html>

<http://ccat.sas.upenn.edu/plc/kiswahili/>

<http://ccat.sas.upenn.edu/plc/kiswahili/kiswahili.html>

Site (12)

<http://www.sas.upenn.edu/~tonyat/kiswahili.html>

<http://www0.bbc.co.uk/swahili/>
<http://www.sds.org/kiswahili/kiswahili.htm>
http://www.esperanto.net/info/index_sw.html
<http://www.irib.com/worldservice/swahili/>
<http://chumvi.tripod.com/>
<http://www.quranitukufu.com/>
 Site multilingue kiswahili et anglais (4)
<http://ccat.sas.upenn.edu/afl/kiswahili.html>
<http://ccat.sas.upenn.edu/plc/kiswahili/kiswahili.html>
<http://www.tanzania.go.tz/>
<http://www.kenyaconstitution.org/>
<http://www.itv.co.tz/>
LC-F (1)
<http://www.esperanto.net/>
LC-A (29)
LC-O (3)
 Allemand
<http://www.kiswahili.de/>
 Néerlandais
<http://www.ilt.kuleuven.ac.be/ilt/index.htm>
 Danois
<http://www.swahili.dk/>
Kituba (25)
Mention (5)
<http://www.mavica.ru/directory/spa/2446.html>
<http://www.mavicanet.org/directory/eng/2446.html>
<http://www.lifeway.com/international/kituba.htm>
<http://humanities.uchicago.edu/faculty/mufwene/>
<http://www.vada.nl/talenen.htm>
Documentation (5)
<http://www.101language.com/kituba.html>
<http://www.worldlanguage.com/Languages/Kituba.htm>
<http://www.worldlanguage.com/Products/8405.htm>
<http://www.christusrex.org/www3/ethno/Zair.html>
<http://www.worldlanguage.com/Products/Kituba/Tutorial-Learning/Page1.htm>
Description (4)
http://www.isp.msu.edu/AfrLang/Kongo_root.html
http://www.ethnologue.com/show_lang_family.asp?code=KTU
<http://ling.ucsd.edu/ling19/ling19langdis/kituba.htm>
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=KTU>
Echantillon (2)
<http://www.elite.net/~runner/jennifers/hello.htm#K>
<http://www.bantu.ovh.org/en/dico.html>
Texte (8)
 DHx : <http://www.language-museum.com/k/kituba.htm>
 DH : <http://www.unhchr.ch/udhr/lang/ktu.htm>
 DH : <http://www3.itu.int/udhr/lang/ktu.htm>
 PN : <http://www.christusrex.org/www1/pater/JPN-kituba-western.html>
 PN : <http://www.language-museum.com/k/kituba-western.htm>
 AM : <http://www.christusrex.org/www1/pater/JPN-kituba-eastern.html>
 AM : <http://www.language-museum.com/k/kituba-eastern.htm>
 EJ1x : <http://www.worldscriptures.org/pages/kituba.html>

LC-A (25)
 Tous les sites trouvées sur cette langue sont en anglais

Lingála (29)

Mention (3)
<http://www.imdb.com/Sections/Languages/Lingala/>
<http://www.verbix.com/languages/index.asp>
<http://linguistlist.org/issues/10/10-683.html>
Documentation (11)
<http://thor.prohosting.com/~linguist/lingala.htm>
<http://perso.wanadoo.fr/pascal.grouelle/apropos.htm>
<http://www.yourdictionary.com/languages/nigecong.html>
<http://www.worldlanguage.com/Languages/Lingala.htm>
<http://www.christianvideos.org/lingala.html>
<http://www.multilingualbooks.com/lingala.html>
<http://www.ethnicarvest.org/bibles/lingala.htm>
<http://www.101language.com/lingala.html>
<http://www.foreignmedia.com/ilc-LNDC.html>
<http://www.imdb.com/Sections/Languages/Lingala/>
<http://www1.buch24.de/shop/3-89416-260-0.html>
Description (2)
<http://www.isp.msu.edu/AfrLang/Language5.htm>
http://www.isp.msu.edu/AfrLang/Lingala_root.html
Échantillons (7)
<http://thor.prohosting.com/~linguist/media/lingalaphr.htm>
<http://www.lingala.de/>
http://www.bantu.ovh.org/lingala_greetings.html
<http://www.bantu.ovh.org/en/dico.html>
<http://pageperso.aol.fr/ashemtem/mapage/artiste.html>
<http://www.freelang.com/dictionnaire/lingala.html>
<http://www.yourdictionary.com/languages.html>
Texte (2)
 DH : <http://www.unhchr.ch/udhr/lang/lin.htm>
 Littérature. enfantine
<http://www.cherbearsden.com/story1AT.html>
Cours (4)
<http://perso.wanadoo.fr/pascal.grouelle/>
<http://www.linguistics.uiuc.edu/africanlang/elementarylingala.htm>
<http://lingalaenligne.iffrance.com/lingalaenligne/>
<http://www.vwv.cc/sprachen/Lingala.html>
LC-F (5)
<http://perso.wanadoo.fr/pascal.grouelle/>
<http://www.freelang.com/dictionnaire/lingala.html>
<http://lingalaenligne.iffrance.com/lingalaenligne/>
<http://pageperso.aol.fr/ashemtem/mapage/artiste.html>
<http://perso.wanadoo.fr/pascal.grouelle/apropos.htm>
LC-A (22)
LC-O (2)
 Allemand
<http://www.lingala.de/>
<http://www.vwv.cc/sprachen/Lingala.html>
Lomnwe / chilomwe / elomwe (15)
Mention (5)
<http://www.imb.org/southern-africa/prayerletters/Brownfields%20Lomwe%20team.htm>
http://www.ethnologue.com/show_language.asp?code=KZN

<http://www.imb.org/southern-africa/archive/Mission%20Stories/Lomwe%20story.HTM>
<http://members.aol.com/grn4jean/mw.htm>
http://www.imb.org/southern-africa/site_map.htm
Documentation (3)
<http://www.worldlanguage.com/Languages/Lomwe.htm>
<http://www.worldlanguage.com/French/Languages/Lomwe.htm>
<http://www.worldlanguage.com/Spanish/Languages/Lomwe.htm>
Description (5)
http://www.ethnologue.com/show_lang_family.asp?code=NGL
http://www.isp.msu.edu/AfrLang/makua_root.html
<http://www.ikuska.com/Africa/Etnologia/Pueblos/lomwe/>
http://www.ethnologue.com/show_language.asp?code=NGL
<http://www.imb.org/southern-africa/peoplegroups/lomwe.htm>
Échantillons (1)
<http://www.kabalarians.com/female/lomwe-f.htm>
Texte (1)
 Jean_1 : <http://www.language-museum.com/l/lomwe.htm>
LC-F
<http://www.worldlanguage.com/French/Languages/Lomwe.htm>
LC-A (14)

Luganda (19)

Mention (3)
<http://www.luganda.com/>
<http://www.travlang.com/languages/cgi-bin/langchoice.cgi?page=main&lang1=english&flags.x=291&flags.y=379>
<http://www.linglist.org/issues/6/6-420.html>
Documentation (4)
<http://www.buganda.com/language.htm>
<http://www.buganda.com/bitabop1.htm>
<http://www.multilingualbooks.com/luganda.html>
<http://www.worldlanguage.com/Languages/Luganda.htm>
Description (3)
<http://www.flw.com/languages/luganda.htm>
http://www.maps2anywhere.com/Languages/languages_-_luganda.htm
<http://www.buganda.com/ggulama.htm>
Échantillons (5)
<http://www.buganda.com/pride.htm>
<http://www.buganda.com/phrasbk.htm>
<http://www.uebersetzung.at/twister/lug.htm>
<http://www.kabalarians.com/male/lugandam.htm>
<http://luganda.canadianwebs.com/Research-3a.html>
Texte (2)
 Telecenter : http://www.nakaseke.or.ug/html/luganda_version.html
 AM : <http://www.christusrex.org/www1/pater/JPN-luganda.html>
Cours (1)
<http://www.buganda.com/luganda.htm>
Site (1)
 Anglais - Luganda
<http://www.iss.stthomas.edu/luganda/>
LC-A

Luo (5)

Description (2)
<http://www.flw.com/languages/luo.htm>

http://www.ethnologue.com/show_language.asp?code=LUO
Échantillons (1)
<http://www.jaluo.com/>
Texte (1)
 4SL : <http://www.greatcom.org/laws/luo/>
Site (1)
<http://www.jaluo.com/>
LC-A (4)

Luyia (22)

Mention (8)
<http://www.mavica.ru/directory/ukr/16230.html>
<http://www.christusrex.org/www1/pater/>
<http://members.aol.com/grn4jean/ke.htm>
<http://www.language-museum.com/>
<http://www.tlcdelivers.com/tlc/crs/lang0144.htm>
<http://www.britannica.com/seo/l/luhya/>
<http://www.vada.nl/talenen.htm>
<http://www.sim.org/xq/asp/id.28/qx/countryFAQ.asp>
Documentation (5)
<http://www.indiana.edu/~librcsd/afrlg/data/0394.html>
<http://www.indiana.edu/~sal/titleindex.html>
<http://www.christusrex.org/www3/ethno/Keny.html>
<http://www.albany.edu/~lb527/Luano.html>
<http://www.albany.edu/~lb527/Mashi.html>
Description (6)
http://www.ethnologue.com/show_lang_family.asp?code=LUY
http://www.ethnologue.com/show_family.asp?subid=2238
http://www.isp.msu.edu/AfrLang/luyia_root.html
http://www.ethnologue.com/show_language.asp?code=LUY
http://www.ethnologue.com/show_country.asp?name=Uganda
http://www.ethnologue.com/show_language.asp?code=RAG
Texte (3)
 PN : <http://www.christusrex.org/www1/pater/JPN-luyia.html>
 Jean_1 : <http://www.language-museum.com/l/luyia.htm>
 Gospel : <http://gospelgo.com/fluyiajoh.htm>
LC-A (21)
LC-O
 Russe
<http://www.mavica.ru/directory/ukr/16230.html>

Makhuwa / makuwa (15)

Mention (7)
<http://www.byhisgrace.com/makhuwa/>
<http://www.imb.org/southern-africa/prayerletters/Makhuwa.htm>
<http://www.byhisgrace.com/makhuwa/makhuwa.htm>
<http://www.2001pray.org/Africa.htm>
<http://www.language-museum.com/>
<http://www3.itu.int/udhr/navigate/alpha.htm>
<http://www.joshuaproject.net/Assets/Profiles/peoplxx.htm>
Documentation (1)
http://www.imb.org/Southern-Africa/links_to_web_sites.htm
Description (6)
http://www.ethnologue.com/show_lang_family.asp?code=VMK
<http://www.imb.org/Southern-Africa/peoplegroups/MakhuwaofMozambique.htm>
<http://www.joshuaproject.net/Assets/Profiles/jpl62.htm>

http://www.ethnologue.com/show_country.asp?name=Mozambique
<http://www.ikuska.com/Africa/Etnologia/Pueblos/makua/>
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=MAK>

Texte (1)

Jean_1 : <http://www.language-museum.com/m/makhuwa-metto.htm>

LC-A (14)

LC-O (1)

Espagnol

<http://www.ikuska.com/Africa/Etnologia/Pueblos/makua/>

Mende (4)

Documentation (2)

http://www.library.cornell.edu/africana/Writing_Systems/Mende.html

<http://www.worldlanguage.com/Languages/Mende.htm>

Description (2)

<http://www.uiowa.edu/~africart/toc/people/Mende.html>

http://www.ethnologue.com/show_language.asp?code=MFY

LC-A (4)

Tous les sites trouvés sur cette langue sont en anglais.

Mongo-Nkundu (20)

Mention (1)

<http://www.language-museum.com/>

Documentation (5)

<http://www.tlcdelivers.com/tlc/crs/lang0187.htm>

<http://www.rba.ru:8101/rusmarc/rusmarc/supa/17.htm>

http://www.ethnologue.com/show_iso639.asp?code=lol

http://www.loc.gov/marc/languages/lang_12n.html

http://www.loc.gov/marc/languages/lang_pt2.html

Description (6)

http://www.ethnologue.com/show_lang_family.asp?code=MOM

http://www.ethnologue.com/show_family.asp?subid=2807

<http://www.christusrex.org/www3/ethno/Zair.html>

<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=MOM>

<http://www.ikuska.com/Africa/Etnologia/Pueblos/mongo/index.htm>

http://www.ethnologue.com/show_language.asp?code=MOM

Échantillons (5)

http://www.loc.gov/marc/languages/lang_12n.html

<http://www.cs.wm.edu/~radu/peace.html>

<http://www.mindmills.net/articles/peace.shtml>

<http://www.giftofpeace.net/PeaceLanguages.htm>

<http://users.bestweb.net/~om/yoga/peace.html>

Texte (3)

PN : <http://www.language-museum.com/m/mongo-nkundu-mongo.htm>

AM_Angola : <http://www.christusrex.org/www1/pater/JPN-lomongo-s-nkundu.html>

AM_Eaiïre : <http://www.christusrex.org/www1/pater/JPN-longo.html>

LC-A (18)

LC-O (2)

Espagnol

<http://www.ikuska.com/Africa/Etnologia/Pueblos/mongo/index.htm>

Russe

<http://www.rba.ru:8101/rusmarc/rusmarc/supa/17.htm>

Mooré (19)

Mention (11)

http://www.glc.org/jcolman/moore_primer.html

<http://www.worldchristians.org/>

<http://www3.itu.int/udhr/navigate/alpha.htm>

<http://www.globaldialogues.org/FirstFilms.htm>

http://www.biblesociety.org/wr_362/362_12.htm

<http://www.film.de/film/kino/3585.shtml>

<http://research.umbc.edu/~fischett/ml1310.html>

http://www.univie.ac.at/linguistics/gur/qui_m-n.htm

<http://www.syfia.com/premiere/radiosom03.asp>

<http://www.worldbank.org/afr/ik/french/frdbref.htm>

<http://membres.lycos.fr/eauvive/ev-025.htm>

Documentation (1)

<http://www.lmp.ucla.edu/lmd/cals.htm>

Échantillons (5)

<http://www.aliaspilote.com/ouagamoore.htm>

<http://www.fraternet.org/lafibeme/intro.htm>

<http://www.aqoci.qc.ca/qsfs/qsfs2/csisag.html>

<http://www.inforoutes-ardeche.fr/~1.bony/burkina.htm>

http://www.mediaport.net/AeS/5/5_95_01.fr.html

Textes (1)

DH : <http://www.unhchr.ch/udhr/lang/mhm.htm>

Cours (1)

<http://www.glc.org/jcolman/primer/>

LC-F (9)

<http://www.aliaspilote.com/ouagamoore.htm>

<http://www.aqoci.qc.ca/qsfs/qsfs2/csisag.html>

<http://www.fraternet.org/lafibeme/intro.htm>

<http://www.inforoutes-ardeche.fr/~1.bony/burkina.htm>

http://www.mediaport.net/AeS/5/5_95_01.fr.html

<http://www.syfia.com/premiere/radiosom03.asp>

http://www.francophonie.org/oif/francophonie/membres/ARTICLES/cl_748.html

<http://www.worldbank.org/afr/ik/french/frdbref.htm>

<http://membres.lycos.fr/eauvive/ev-025.htm>

LC-A (9)

LC-O (1)

Allemand

<http://www.film.de/film/kino/3585.shtml>

Ndebele / isiNdebele (15)

Mention (5)

<http://www.soulcity.org.za/materials%20index.htm>

<http://www.gov.za/documents/2000/mdda/>

<http://cgi.gates96.com/cam/africa/South%20Africa/media.html>

<http://www.letsgo.com/SAF/01-teaser-01>

<http://www.suedafrika.net/Norden/g1mpuma2.htm>

Documentation (2)

<http://www.linguistics.berkeley.edu/cgi-bin/CBOLD/jgspltsrnm.cgi?simpledbq=Ndebele.Pelling.1971.txt>

http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm

Description (3)

http://www.ethnologue.com/show_language.asp?code=NDF

<http://www.icon.co.za/~filizana/ndebele.htm>

<http://www.isp.msu.edu/AfrLang/Language7.htm>

Echantillons (1)

<http://www.kabalarians.com/female/ndebelef.htm>

Texte (4)

<http://www.gov.za/documents/2000/mdda/ndebele.pdf>

Bible: <http://www.gospelcom.net/ibs/bibles/ndebele/>

Edit Translation

<http://ndebele.edit.fr/>

<http://www.edit.fr/ndebele/index.shtml>

LC-A (6)

LC-0

Allemand

<http://www.suedafrika.net/Norden/g1mpuma2.htm>

Oromo (6)

Mention (4)

http://www.sas.upenn.edu/African_Studies/Hornet/Afaan_Oromo_19777.html

http://www.sas.upenn.edu/African_Studies/Articles_Gen/Oromo.html

<http://www.oromo.org/>

<http://www.voa.gov/allsked.cfm>

Documentation (1)

<http://monarch.gsu.edu/nutrition/Oroma.htm>

Échantillons (1)

<http://www.oromo.org/>

LC-A (6)

Tous les sites trouvés sur cette langue sont en anglais.

Sängö / Sango (19)

Mention (2)

<http://www.beafrica.org/>

<http://www.sangonet.com/1Csituation.html>

Documentation (7)

<http://www.osil.ch/gtf-rifal/bulletinRifal2000/page3.html>

<http://www.acelf.ca/bottins/bottin8.html>

http://www.lib.msu.edu/limb/a-z/az_bg2.html

<http://w3.olf.gouv.qc.ca/inventaire/fiche/608498.htm>

<http://www.fiu.edu/~linguist/Vol26-28.htm>

<http://www.ling.lu.se/research/profileareas/Typology/TypDiss.html>

<http://www.u-paris10.fr/bibethno/kintz.html>

Description (5)

<http://babel.ling.upenn.edu/~nagy/nwav/WWWabs/Samarin.html>

<http://home.earthlink.net/~ddweston/car/sango/>

http://www.isp.msu.edu/AfrLang/Sango_root.html

<http://www.sangonet.com/6Cformation.html>

<http://www.unesco.org/comnat/france/colloquemars3.htm>

Echantillons (3)

<http://www.beafrica-opinions.com/reflexions/reflexions9.htm>

<http://www.sangonet.com/Formulefin2000SNW.html>

<http://www.sozoala.com/pagedegarde.htm>

Texte (1)

<http://www.osil.ch/gtf-rifal/bulletinRifal2000/>

Site (1)

<http://sango.free.fr/>

LC-F (11)

<http://www.beafrica.org/>

<http://www.sangonet.com/Formulefin2000SNW.html>

<http://www.sangonet.com/6Cformation.html>

<http://w3.olf.gouv.qc.ca/inventaire/fiche/608498.htm>

<http://www.osil.ch/gtf-rifal/bulletinRifal2000/page3.html>

<http://www.u-paris10.fr/bibethno/kintz.html>

<http://www.beafrica-opinions.com/reflexions/reflexions9.htm>

<http://www.osil.ch/gtf-rifal/bulletinRifal2000/>

<http://www.acelf.ca/bottins/bottin8.html>

<http://www.unesco.org/comnat/france/colloquemars3.htm>

<http://www.sozoala.com/pagedegarde.htm>

LC-A (8)

Sesotho Nord /Pedi (14)

Mention (4)

<http://www.soulcity.org.za/materials%20index.htm>

<http://www.gov.za/documents/2000/mdda/>

<http://www.ncrf.org.za/mpumalanga.htm>

<http://www.letsgo.com/SAF/01-teaser-01>

Documentation (6)

<http://www.101language.com/sotho.html>

<http://www.language.com.au/sotho.html>

<http://www.amazon.com/exec/obidos/ASIN/0627015867/>

<http://www.flw.com/languages/sotho.htm>

<http://www.walmart.com/catalog/product.jsp>

[;jsessionid=6101860123fb4bae61c54a9d83fe795af8707fb63_](http://www.walmart.com/catalog/product.jsp)

[10-15-20-185.2951.1034693579283.:0f](http://www.walmart.com/catalog/product.jsp)

[?product_id=1623134&sourceid=0100000030390613102498](http://www.walmart.com/catalog/product.jsp)

http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm

Description (1)

<http://www.flw.com/languages/sotho.htm>

Texte (3)

REL: <http://www.greatcom.org/laws/pedi/>

LO <http://www.geocities.com/Athens/9479/zulu.html>

<http://www.gov.za/documents/2000/mdda/nsotho.pdf>

LC-A (14)

Tous les sites trouvés sur cette langue sont en anglais.

Sesotho / Sud (29)

Mention (5)

<http://www.jazz-pop-classic.de/9814901.htm>

<http://www.soulcity.org.za/materials%20index.htm>

<http://www.gov.za/documents/2000/mdda/>

<http://www.ncrf.org.za/mpumalanga.htm>

<http://www.letsgo.com/SAF/01-teaser-01>

Documentation (11)

<http://www.cyberserv.co.za/users/~jako/sesotho.htm>

<http://www.cyberserv.co.za/users/~jako/books.htm>

<http://www.worldlanguage.com/Languages/Sotho.htm>

<http://www.language.com.au/sotho.html>

<http://www.up.ac.za/academic/libarts/afriLang/nshome.htm>

<http://www.worldlanguage.com/Indonesian/Languages/Sotho.htm>

<http://www.flw.com/languages/sotho.htm>

http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm

<http://www.worldlanguage.com/Indonesian/Languages/Sotho.htm>

<http://www.worldlanguage.com/Products/Sotho/Games/Page1.htm>

<http://www.foreignmedia.com/ilc-STDC.html>

Description (3)

http://www.ethnologue.com/show_language.asp?code=SSO

<http://www.flw.com/languages/sothosouth.htm>
<http://www.flw.com/languages/sotho.htm>
Echantillon (7)
<http://www.cyberserv.co.za/users/~jako/phrases.htm>
<http://www.cyberserv.co.za/users/~jako/sentence.htm>
<http://www.cyberserv.co.za/users/~jako/ref.htm>
<http://www.knet.co.za/lesedi/sotho.htm>
<http://www.cyberserv.co.za/users/~jako/greet.htm>
<http://vraptor.jpl.nasa.gov/voyager/languages/sotho.html>
<http://users.skynet.be/afrika/ZuSo.html>

Textes (3)

DH : <http://www.unhchr.ch/udhr/lang/sso.htm>
 LO : <http://www.geocities.com/Athens/9479/zulu.html>
<http://www.gov.za/documents/2000/mdda/ssotho.pdf>
 LC-A (29)

Tous les sites trouvés sur cette langue sont en anglais.

Seswati (25)

Mention (17)

<http://www.lowani.nl/swati.htm>
http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm
<http://www.bday.co.za/bday/content/direct/1,3523,556632-6078-0,00.html>
<http://www.soulcity.org.za/materials%20index.htm>
<http://www.soulcity.org.za/materials%20index.htm>
<http://www.doj.gov.za/trc/amntrans/2000/0007310804JHB/200802jb.htm>
<http://www.crwflags.com/fotw/flags/sz-ngwan.html>
<http://www.gov.za/documents/2000/mdda/>
<http://www.crwflags.com/fotw/flags/za-ec-nm.html>
<http://members.aol.com/grn4jean/sz.htm>
http://www.pienternet.be/zuidafrika/zafrika_cijfers.html
<http://www.ncrf.org.za/mpumalanga.htm>
<http://www.ncrf.org.za/mpumalanga.htm>
<http://www.limpopo.gov.za/people/Magadzi.htm>
<http://cgi.gates96.com/cam/africa/South%20Africa/media.html>
<http://www.letsgo.com/SAF/01-teaser-01>
<http://www.mozambique.mz/namaacha/agerais.htm>

Documentation (2)

<http://www.missionsalive.org/gospeltract/tracts.html>
<http://www.sabc.co.za/TV/tvindex.html>

Texte (2)

<http://www.gov.za/documents/2000/mdda/seswati.pdf>
<http://www.dwaf.gov.za/Docs/Internet/NWRS/PCM/SESWATI%20REGISTR%20PCM.doc>

Site (1)

<http://www.action.co.zw/Magazines/venacular/aids/seswati/html/14-15.html>

LC-A (21)

LC-O (3)

<http://www.lowani.nl/swati.htm>
http://www.pienternet.be/zuidafrika/zafrika_cijfers.html
<http://www.mozambique.mz/namaacha/agerais.htm>

Setswana (6)

Mention (6)

<http://www.soulcity.org.za/materials%20index.htm>

<http://www.gov.za/documents/2000/mdda/>
<http://www.ncrf.org.za/mpumalanga.htm>
<http://www.letsgo.com/SAF/01-teaser-01>
<http://www.african.gu.se/prj001.html>
<http://linguistlist.org/issues/10/10-1210.html>
<http://www.nflrc.hawaii.edu/aboutus/ithompson/flmedia/languages/setswana.htm>

Documentation (13)

<http://www.worldlanguage.com/Products/Setswana/ESL-EnglishasSecondLanguage/Page1.htm>
<http://www.worldlanguage.com/Products/Setswana/Kids/Page1.htm>
<http://www.worldlanguage.com/Products/Setswana/Tutorial-Learning/Page1.htm>
<http://www.worldlanguage.com/German/Languages/Setswana.htm>
<http://www.worldlanguage.com/French/Languages/Setswana.htm>
<http://www.bartleby.com/61/29/S0292900.html>
http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm
<http://www.worldlanguage.com/Languages/Setswana.htm>
<http://www.amazon.com/exec/obidos/ASIN/0799408166/avsearch-bkasin-20/102-4162701-8993710>
<http://www.amazon.com/exec/obidos/ASIN/0796006393/avsearch-bkasin-20/102-4162701-8993710>
<http://www.worldlanguage.com/Products/Setswana/System/Page1.htm>
<http://www.worldlanguage.com/Products/Setswana/Reference/Page1.htm>
<http://www.amazon.com/exec/obidos/ASIN/0627022472/inktomibkasin-20/102-4162701-8993710>

Echantillons (1)

<http://www.info.bw/~msc/setswana.html>

Texte (3)

<http://www.gov.za/documents/2000/mdda/tswana.pdf>
 DH : <http://www3.itu.int/udhr/lang/tsw.htm>
 DH : <http://www.unhchr.ch/udhr/lang/tsw.htm>

Cours (4)

<http://botswana.free.fr/setswana.htm>
http://exploringtheplanet.com/botswana_setswana.htm
http://willow.cats.ohiou.edu/~mugane/setswana/step/step_content.html
http://willow.cats.ohiou.edu/~mugane/setswana/exp_files/01_exp.htm

Site (5)

<http://www.survival-international.org/old%20site/setswanaub9912.htm>
<http://www.survival-international.org/old%20site/setswana.htm>
<http://www.cmn.co.za/tswana/>
http://www.uniwest.ac.za/hp_35.htm
http://www.gov.za/symbols/coa_setswana.htm

LC-F

<http://www.worldlanguage.com/French/Languages/Setswana.htm>

LC-A (33)

LC-O

Allemand

<http://www.worldlanguage.com/German/Languages/Setswana.htm>

Shona (3)

Documentation (1)

<http://www.worldlanguage.com/Languages/Shona.htm>

Description (1)

<http://www.uiowa.edu/~africart/toc/people/Shona.html>

Cours (1)

<http://www.shonalanguage.fsnet.co.uk/>

LC-A (3)

Tous les sites trouvés sur cette langue sont en anglais.

Somali (10)

Mention (1)

<http://www.bbc.co.uk/somali/inenglish.shtml>

Documentation (1)

<http://www.geocities.com/Paris/Louvre/2521/somali.html>

Échantillons (3)

<http://www.somalia-rebirth.dj/>

http://www.dm.unipi.it/~jama/alif/qaamuus/qaamuus_en.html

<http://www.redsea-online.com/modules.php?name=dictionary>

Cours (1)

<http://iteslj.org/v/so/>

Site (4)

<http://www.somalitalk.com/>

Anglais et somali

<http://somalitalk.com/chat/>

<http://www.somalipress.com/sp/Default.asp>

<http://www.somaliweyn.com/>

LC-A (6)

Tamajaq / Tamasheq (42)

Mention (22)

http://www.ethnologue.com/show_language.asp?code=TDA

<http://issco-www.unige.ch/staff/andrei/formRIFAL2000/>

<http://home.tiscali.be/kwia/tijdschrift/dos51.html>

<http://users.antrasite.be/ppoisie/Documents/saharatouareg.htm>

<http://www.joshua.sim.org/contents.html>

<http://www.sil.org/silesr/1999/008/nsonghay.html>

<http://www.linguistlist.org/issues/11/11-379.html>

<http://members.aol.com/grn4jean/ne.htm>

<http://www.mediaport.net/AfricArt/100CD/CD/015.html>

<http://membres.lycos.fr/temoust/disque2.htm>

<http://www.medecinsdumonde.org/2missions/niger/jourJ.htm>

<http://www.bartleby.com/65/ha/HamitoSe.html>

<http://www.arzouni.com/niger.html>

http://www.mediaport.net/AfricArt/100CD/index_langue.html

<http://perso.wanadoo.fr/theotokos/foucauld.html>

http://membres.lycos.fr/djilalibenamrane/ntic_uk.htm

<http://www.nativeweb.org/resources/organizations/>

[aboriginal_indigenous_nations/](http://www.nativeweb.org/resources/organizations/aboriginal_indigenous_nations/)

<http://membres.lycos.fr/temoust/tradition2.htm>

<http://www.manntaylor.com/foucauld.html>

<http://www.isoc.org/oti/articles/0401/gallagher.html>

<http://www.albright.edu/reporter/wint02/sahara01.html>

<http://membres.lycos.fr/djilalibenamrane/ntic.htm>

Documentation (6)

<http://www.freefind.com/find.html>

[?id=8889810&map=0&page=0&lang=fr](http://www.freefind.com/find.html?id=8889810&map=0&page=0&lang=fr)

<http://www.rosettaproject.org:8080/live/search/>

[browsebycountryresult?searchtype=countries&searchkey=Niger](http://www.rosettaproject.org:8080/live/search/browsebycountryresult?searchtype=countries&searchkey=Niger)

http://www.ethnologue.com/show_iso639.asp?code=tml

http://www.lepetitprince.net/sub_Touareg/Tamachek-E.html

<http://www.ee.umd.edu/~sellami/ref.html>

<http://www.amazon.com/exec/obidos/ASIN/>

[9060223594/104-3675279-9639150](http://www.amazon.com/exec/obidos/ASIN/9060223594/104-3675279-9639150)

Description (10)

http://www.ethnologue.com/show_lang_family.asp?code=TTQ

<http://www.sim.org/PGFAQ.asp?id=34>

<http://www.joshuaproject.net/Assets/Profiles/jpl938.htm>

http://www.ethnologue.com/show_language.asp?code=TTQ

<http://www.ad2000.org/peoples/jpl938.htm>

<http://www.joshuaproject.net/Assets/Profiles/jpl938.htm>

http://www.ethnologue.com/show_country.asp?name=Niger

http://www.ethnologue.com/show_language.asp?code=THV

<http://www.christusrex.org/www3/ethno/Nige.html>

http://www.ethnologue.com/show_country.asp?name=Libya

Échantillons (2)

<http://home.t-online.de/home/Petra.Bode/wort.htm>

<http://www.manntaylor.com/tenere.html>

Texte (2)

Bible : <http://pages.ifrance.com/pages/langues/tamajaq/>

<http://www.ifrance.com/pages/infos/touareg.htm>

LC-F (8)

<http://issco-www.unige.ch/staff/andrei/formRIFAL2000/>

http://www.sas.upenn.edu/African_Studies/Country_Specific/

[niger_info.html](http://www.sas.upenn.edu/African_Studies/Country_Specific/niger_info.html)

<http://users.antrasite.be/ppoisie/Documents/saharatouareg.htm>

<http://membres.lycos.fr/temoust/disque2.htm>

<http://www.medecinsdumonde.org/2missions/niger/jourJ.htm>

<http://www.mediaport.net/AfricArt/100CD/CD/015.html>

<http://membres.lycos.fr/temoust/tradition2.htm>

<http://membres.lycos.fr/djilalibenamrane/ntic.htm>

LC-A (32)

LC-O (2)

Néerlandais

<http://home.tiscali.be/kwia/tijdschrift/dos51.html>

Allemand

<http://home.t-online.de/home/Petra.Bode/wort.htm>

Tigrinya (23)

Mention (10)

<http://www.ethnicarvest.org/bibles/tigrinya.htm>

<http://www.africanlanguage.com/>

http://store.browsing-empire.com/Tigrinya_Phrase_Book.html

<http://home.wxs.nl/~hans.mebrat/eritrea-languages.htm>

http://search.samanthasdomain.com/Tigrinya_Phrase_Book.html

http://www.usatoday.com/news/opinion/2002-07-18-ncguest1_x.htm

<http://book.netstoreusa.com/index/bkbl350.shtml>

<http://www.thing.net/~grist/ld/haile/rh-ref.htm>

<http://www.centrelink.gov.au/>

<http://www.africanlanguage.com/demo/>

Documentation (6)

http://www.nextag.com/Sewasiw_Tigrinya_B_Sefihu~

[1806332z2znzz1zmainz2-htm](http://www.nextag.com/Sewasiw_Tigrinya_B_Sefihu~1806332z2znzz1zmainz2-htm)

[?nextg=10596f_5A350D9CE963ABF8&product=2617118](http://www.nextag.com/Sewasiw_Tigrinya_B_Sefihu~?nextg=10596f_5A350D9CE963ABF8&product=2617118)

<http://www.africanlanguage.com/products/eztigrinya.html>

<http://www.walmart.com/catalog/product.gsp>

[?product_id=575023&sourceid=0100000030390613102498](http://www.walmart.com/catalog/product.gsp?product_id=575023&sourceid=0100000030390613102498)

<http://my.linkbaton.com/get?lbCC=q&nC=q&genre=book&item=0001474243>
<http://my.linkbaton.com/get?genre=book&item=1569020833>
<http://www.amazon.com/exec/obidos/ASIN/0932415210/avsearch-bkasin-20/102-4162701-8993710>
<http://www.amazon.com/exec/obidos/ASIN/1569020833/inktomi-bkasin-20/102-4162701-8993710>
Description (1)
<http://www.isp.msu.edu/AfrLang/Language6.htm>
Texte (3)
<http://www.centrelink.gov.au/internet/internet.nsf/languages/ti.htm>
<http://www.immunize.org/catg.d/p4170ti.pdf>
 PN + AN: <http://www.christusrex.org/www1/pater/JPN-tigrinya.html>
Cours (1)
<http://www.eritrea.de/forum/messages/411.html>
Site (2)
<http://members.asmarino.com/DrReesom/>
<http://www.calpoison.org/lang/tig/home.html>
LC-A (20)
LC-O (1)
 Allemand
<http://www.eritrea.de/forum/messages/411.html>

Tiv (4)

Mention (1)
http://lucy.ukc.ac.uk/EthnoAtlas/Hmar/Cult_dir/Culture.7874
<http://www.muta.org/>
Documentation (1)
<http://www.indiana.edu/~librcsd/afrlg/data/0642.html>
Texte (1)
 4SL: <http://www.greatcom.org/laws/tiv/>
LC-A (4)
 Tous les sites trouvés sur cette langue sont en anglais.

Tsonga (33)

Mention (10)
<http://www.soulcity.org.za/materials%20index.htm>
<http://www.gov.za/documents/2000/mdda/>
<http://www.ncrf.org.za/mpumalanga.htm>
<http://www.letsgo.com/SAF/01-teaser-01>
http://members.tripod.com/lingua_consultants/
<http://123hostnow.com/articles/LCID.asp?LCID=1073>
<http://www.nationallottery.co.za/IE/HTPL/HTPT.htm>
http://www.wordtheque.com/owa-wt/new_wordtheque.wcom_literature.literaturea_page?lang=TS&letter=B&source=search&page=1
<http://encyklopedia.pwn.pl/1.php?arg=76576>
<http://www.greatcom.org/laws/languages.html>
Documentation (5)
<http://www.oasis-open.org/cover/iso639a.html>
http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm
<http://www.indiana.edu/~librcsd/afrlg/data/0651.html>
<http://www.amazon.com/exec/obidos/ASIN/0949981370/avsearch-bkasin-20/102-4162701-8993710>
<http://www.amazon.com/exec/obidos/ASIN/0869814753/avsearch-bkasin-20/102-4162701-8993710>
Description (6)
http://www.ethnologue.com/show_lang_family.asp?code=TSO

<http://www.isp.msu.edu/AfrLang/language9.htm>
http://www.isp.msu.edu/AfrLang/tsonga_root.html
<http://www.ikuska.com/Africa/Etnologia/Pueblos/tsonga/>
http://www.ethnologue.com/show_language.asp?code=TSO
<http://www.rosettaproject.org:8080/live/search/detailedlanguage-record?ethnocode=TSO>
Echantillon (1)
<http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/appendix/languages/tsonga/tsonga.html>
Texte (8)
<http://www.gov.za/documents/2000/mdda/tsonga.pdf>
 4SL: <http://www.greatcom.org/laws/changana/>
 4SL: <http://www.greatcom.org/laws/tsonga/>
<http://www.doh.gov.za/docs/legislation/patientsright/tsonga.pdf>
<http://www.nda.org.za/publications/brochures/tsonga.pdf>
 PN: <http://www.christusrex.org/www1/pater/JPN-tsonga.html>
<http://www.anc.org.za/about/anc-tsonga.html>
<http://www.iba.org.za/tsonga.htm>
Site (3)
<http://www.edit.fr/tsonga/index.shtml>
http://education.pwv.gov.za/DoE_Sites/Curriculum/Final%20curriculum/policy/Xitsonga/tsonga/Tsonga-home%20lang.pdf
http://www.gov.za/symbols/orders_nsotho.htm
LC-A (29)
LC-O (1)
 Polonais
<http://encyklopedia.pwn.pl/1.php?arg=76576>

Tumbuka / Chitumbuka (12)

Mention (2)
<http://www.britannica.com/seo/t/tumbuka/>
<http://www.rosettaproject.org:8080/live/search/addressresourceform?ethnocode=TUW&langname=Tumbuka>
Documentation (2)
<http://www.albany.edu/~lb527/Tumbuka.html>
<http://www.worldlanguage.com/Languages/Tumbuka.htm>
Description (5)
http://www.ethnologue.com/show_lang_family.asp?code=TUW
http://www.ethnologue.com/show_family.asp?subid=2268
<http://www.flw.com/languages/tumbuka.htm>
http://www.isp.msu.edu/AfrLang/tumbuka_root.html
<http://www.imb.org/Southern-Africa/peoplegroups/tumbuka.htm>
<http://www.kidlink.org/kie/africa/angola/languages.html>
<http://www.rosettaproject.org:8080/live/search/detailedlanguage-record?ethnocode=TUW>
Échantillons (1)
<http://www.flw.com/merry.htm>
Texte (2)
 PN: <http://www.christusrex.org/www1/pater/JPN-tumbuka.html>
 Relig.: <http://www.jceal.org/tracts/tum9915t.htm>
LC-A (12)
 Tous les sites trouvés sur cette langue sont en anglais.

Umbundu (20)

Mention (8)
<http://www.worldlanguage.com/Languages/Kimbundu-SeeUmbundu.htm>

<http://www.kidlink.org/kie/africa/angola/protect.htm>
<http://www.linguaportuguesa.ufrn.br/>
<http://www.library.yale.edu/div/fa/181.htm>
<http://www.bantu.ovh.org/tones.html>
<http://www.christusrex.org/www1/pater/>
<http://www.kidlink.org/kie/africa/angola/specifics.html>
<http://www.tlcdelivers.com/tlc/crs/lang0153.htm>
Documentation (12)
<http://www.amazon.com/exec/obidos/ASIN/3927620157/avsearch-bkasin-20/102-4162701-8993710>
<http://www.worldlanguage.com/Languages/Umbundu.htm>
Description (5)
http://www.ethnologue.com/show_lang_family.asp?code=MNF
<http://www.flw.com/languages/umbundu.htm>
http://www.isp.msu.edu/AfrLang/umbundu_root.html
http://www.ethnologue.com/show_language.asp?code=MNF
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=MNF>
Échantillons (2)
<http://www.flw.com/merry.htm>
<http://www.bartleby.net/61/63/U0016375.html>
Texte (3)
 PN : <http://www.christusrex.org/www1/pater/JPN-umbundu.html>
 DH : <http://193.194.138.190/udhr/lang/mnf.htm>
 Jean_1 : <http://www.worldscriptures.org/pages/umbundu.html>
LC-A (19)
LC-0 (1)
[http://www.linguaportuguesa.ufrn.br/Venda/Venda\(8\)](http://www.linguaportuguesa.ufrn.br/Venda/Venda(8))
Mention (6)
<http://www.berntsen.com/u-z.html>
<http://www.gov.za/gazette/notices/2001/22223a.pdf>
<http://www.soulcity.org.za/materials%20index.htm>
<http://www.gov.za/documents/2000/mdda/>
<http://www.ncrf.org.za/mpumalanga.htm>
<http://www.letsgo.com/SAF/01-teaser-01>
Documentation (1)
http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm
Texte (1)
<http://www.gov.za/documents/2000/mdda/venda.pdf>
LC-A (8)
 Tous les sites trouvés sur cette langue son en anglais

Wolayita (29)

Mention (14)
http://www.ethnologue.com/show_language.asp?code=OYD
<http://www.towerofbabel.com/translators/>
http://www.ethnologue.com/show_language.asp?code=BSW
<http://www.campuscrusade.com/languages.htm>
http://www.ethnologue.com/show_language.asp?code=CRA
<http://www.biblesociety.org/bs-eth.htm>
<http://www.thejesusvideo.com/languages.htm>
<http://www.ufc.se/lagen/j-lag00/09.htm>
http://www.ethnologue.com/show_language.asp?code=GMO
http://www.ethnologue.com/show_language.asp?code=BST
http://bamse.ling.su.se/~ljuba/maps/all_langs_mac.txt
<http://www.ti.ch/DIC/DC/CenD/sottile.html>

http://www.ethnologue.com/show_language.asp?code=KCX
http://www.gengo.l.u-tokyo.ac.jp/enshu98-01_e.html
Documentation (4)
http://www.koeppe.de/html/d_slao.htm
http://www.koeppe.de/html/d_aethio.htm
<http://www.fu-berlin.de/semiarab/weinf/bibliog.html>
<http://www.fu-berlin.de/semiarab/voigt.html>
Description (10)
<http://www.tlfq.ulaval.ca/axl/monde/famarabe.htm>
<http://www.freelang.com/freelang/apprendre/familles.html>
<http://www.koeppe.de/katalogD/3-89645-040-9.html>
http://www.ethnologue.com/show_language.asp?code=WBC
<http://www.koeppe.de/katalogE/3-89645-040-9.html>
<http://www.christusrex.org/www3/ethno/Ethi.html>
http://www.ethnologue.com/show_country.asp?name=Ethiopia
http://ngilegacy.com/language_families.htm
<http://www.rosettaproject.org:8080/live/search/detailedlanguagerecord?ethnocode=WBC>
<http://theoryx5.uwinnipeg.ca/CPAN/data/perl/I18N/LangTags/List.html>
Texte (1)
 PN : <http://www.christusrex.org/www1/pater/JPN-wolayta.html>
LC-F (1)
<http://www.freelang.com/freelang/apprendre/familles.html>
LC-A (24)
LC-0 (4)
 Allemand
<http://www.koeppe.de/katalogD/3-89645-040-9.html>
<http://www.fu-berlin.de/semiarab/weinf/bibliog.html>
 Suédois
<http://www.ufc.se/lagen/j-lag00/09.htm>
 Italien
<http://www.ti.ch/DIC/DC/CenD/sottile.html>

Wolof (26)

Mention (6)
<http://users.mildura.net.au/users/mjackson/>
<http://www.afrika-erleben.de/senegal/radtour.htm>
<http://users.mildura.net.au/users/mjackson/Music.htm>
<http://www.afrika-erleben.de/>
<http://languagelab.bh.indiana.edu/audio/wtapes.html>
<http://www.kingdomofbiffeche.net/people.htm>
Documentation (6)
<http://www.gy.com/www/wo.htm>
<http://us.imdb.com/Sections/Languages/Wolof/>
<http://www.dejudicibus.it/dizionario/wolof/wolof.html>
<http://www.uwasa.fi/comm/termino/collect/diction/wolof.html>
<http://www.africanculture.dk/gambia/langabot.htm>
<http://www.worldlanguage.com/Languages/Wolof.htm>
Description (6)
<http://www.uiowa.edu/~africart/toc/people/Wolof.html>
http://lucy.ukc.ac.uk/EthnoAtlas/Hmar/Cult_dir/Culture.7882
<http://www.humnet.ucla.edu/humnet/aflang/Wolof/>
<http://www.ku.edu/~wolof/>
<http://www.ad2000.org/peoples/jpl455.htm>
<http://www.africanculture.dk/gambia/ftp/wolofgram.pdf>

Échantillons (1)

<http://www.sf.airnet.ne.jp/~ts/language/number/wolof.html>

Cours (4)

http://ourworld.compuserve.com/homepages/GenX_jt_mtjr/GenXWolof.html

<http://www.wolofonline.com/>

<http://www.linguistics.uiuc.edu/wolof/>

<http://www.humnet.ucla.edu/humnet/aflang/Wolof/wolofRes.html>

Site (3)

<http://www.ixpres.com/twolff/>

[senegalcommunicationsresearchwolof.htm](http://www.senegalcommunicationsresearchwolof.htm)

<http://www.bok.net/pajol/index.wo.html>

<http://www.worldbank.org/afr/ik/wolof/wdefault.htm>

LC-A (20)

LC-O (3)

Allemand

<http://www.afrika-erleben.de/senegal/radtour.htm>

<http://www.afrika-erleben.de/>

Italien

<http://www.dejudicibus.it/dizionario/wolof/wolof.html>

Xhosa (23)

Mentionnée (10)

<http://www.soulcity.org.za/materials%20index.htm>

<http://www.gov.za/documents/2000/mdda/>

<http://www.ncrf.org.za/mpumalanga.htm>

<http://cgi.gates96.com/cam/africa/South%20Africa/media.html>

<http://www.letsgo.com/SAF/01-teaser-01>

http://members.tripod.com/lingua_consultants/

<http://library.thinkquest.org/22868/?tqskip1=1&tqtime=0221>

<http://www.isp.msu.edu/AfrLang/Language7.htm>

<http://www.sacred-texts.com/afr/>

<http://www.lmp.ucla.edu/profiles/profx01.htm>

Documentation (7)

http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm

<http://isixhosa.co.za/>

http://www.databank.oxydex.com/compendium_bibliographium/world_cultures_arts/Xhosa.html

<http://www.sacred-texts.com/afr/xft/index.htm>

http://www.pricegrabber.com/search_getprod.php/masterid=541659001/ut=c3535cb249592eb8

<http://www.worldlanguage.com/Languages/Xhosa.htm>

<http://www.amazon.com/exec/obidos/ASIN/0844201774/>

[inktomi-bkasin-20/104-8587239-9621508](http://www.inktomi-bkasin-20/104-8587239-9621508)

Description (2)

<http://www.cyberserv.co.za/users/~jako/lang/xho.htm>

<http://www.lmp.ucla.edu/lmpweb/profiles/profx01.htm>

Échantillon (7)

<http://www.knet.co.za/lesedi/xhosa.htm>

<http://library.thinkquest.org/22868/introduce.htm>

<http://hctv.humnet.ucla.edu/departments/linguistics/>

[VowelsandConsonants/course/chapter6/xhosa/xhosa.html](http://www.vowelsandconsonants.com/course/chapter6/xhosa/xhosa.html)

Phonétique Phrases et traduction

<http://mokennon.albion.edu/>

<http://mokennon.albion.edu/xhosa.htm>

<http://www.wcape.school.za/subject/xhosa/xhoshome.htm>

http://ourworld.compuserve.com/homepages/GenX_jt_mtjr/GenXXhosa.html

Textes (4)

Litt. Orale <http://www.geocities.com/Athens/9479/zulu.html>

<http://www.gov.za/documents/2000/mdda/xhosa.pdf>

Chant : <http://web.uct.ac.za/projects/poetry/xhosa.htm>

HN : <http://www.polity.org.za/html/misc/nkosi.html?rebookmark=1#classic>

Sites (2)

<http://xhosa.dianetics.org/>

Xhosa - anglais

<http://www.edit.fr/xhosa/index.shtml>

LC-A (37)

Yoruba (24)

Mention (5)

<http://www.aladura.de/english.htm>

<http://www.cean.u-bordeaux.fr/gdr931.html>

<http://profiles.iagora.com/gaijinjanai/bio.html>

<http://membres.lycos.fr/houles/Porto.html>

<http://www.nuffic.nl/prisma/pers/pp008526.htm>

Documentation (6)

<http://65.107.211.206/post/nigeria/yorubaov.html>

<http://www.eit.bj/golfefm.htm>

<http://www.languages-on-the-web.com/links/link-yoruba.htm>

<http://babel.alis.com/langues/iso639.htm>

<http://www.culture.fr/culture/dglf/bpi/list-langues.html>

<http://www.fiu.edu/~africana/conferences/yoruba/program.htm>

Description (2)

<http://www.tlfq.ulaval.ca/axl/monde/famnigero-congolaise.htm>

<http://www.culture.fr/culture/dglf/bpi/yoruba.html>

Échantillon (4)

<http://agora.legratos.org/?rubrique=religion>

<http://www.yoruba.org/education.htm>

<http://www.anet.net/~ifa/>

<http://www.religioustolerance.org/voodoo.htm>

<http://sites.uol.com.br/odetumbi/>

Textes (2)

<http://monarch.gsu.edu/nutrition/Yoruba.htm>

<http://www.greatcom.org/laws/yoruba/>

Cours (3)

<http://african.lss.wisc.edu/yoruba/>

<http://www.learnyoruba.com/>

<http://african.lss.wisc.edu/yoruba/YLTF/>

LC-F (9)

<http://www.eit.bj/golfefm.htm>

<http://www.aladura.de/english.htm>

<http://babel.alis.com/langues/iso639.htm>

<http://www.culture.fr/culture/dglf/bpi/yoruba.html>

<http://agora.legratos.org/?rubrique=religion>

<http://www.cean.u-bordeaux.fr/gdr931.html>

<http://www.tlfq.ulaval.ca/axl/monde/famnigero-congolaise.htm>

<http://www.culture.fr/culture/dglf/bpi/list-langues.html>

<http://membres.lycos.fr/houles/Porto.html>

LC-A (14)

LC-O (1)

http://sites.uol.com.br/odetumbi/index_pt.htm

Zarma / Dyerma (15)

Documentation (2)

<http://www.moonmountainpub.com/>

<http://www.cilf.org/pub/74-contb.fr.html>

Description (8)

http://www.ethnologue.com/show_lang_family.asp?code=DJE

<http://www.ad2000.org/peoples/jpl1153.htm>

<http://www.joshuaproject.net/Assets/Profiles/jpl1153.htm>

<http://www.ad2000.org/peoples/jpl206.htm>

<http://www.joshuaproject.net/Assets/Profiles/jpl206.htm>

http://www.ethnologue.com/show_language.asp?code=DJE

<http://www.flw.com/languages/Zarma.htm>

http://www.isp.msu.edu/AfrLang/songhai_root.html

Échantillons (21)

<http://www.afrocom.org/zarma.htm>

<http://www.flw.com/merry.htm#S>

Texte (2)

<http://www.christusrex.org/www1/pater/JPN-zarma.html>

<http://www.language-museum.com/z/zarma.htm>

Cours (1)

<http://perso.wanadoo.fr/denis.niamey/html/zarma.htm#ancre1>

LC-F (2)

<http://www.afrocom.org/zarma.htm>

<http://perso.wanadoo.fr/denis.niamey/html/zarma.htm#ancre1>

LC-A (13)

Zulu (9)

Mention (5)

<http://www.soulcity.org.za/materials%20index.htm>

<http://www.doj.gov.za/trc/amntrans/2000/>

[0007310804JHB/200802jb.htm](http://www.doj.gov.za/trc/amntrans/2000/0007310804JHB/200802jb.htm)

<http://www.gov.za/documents/2000/mdda/>

<http://www.ncrf.org.za/mpumalanga.htm>

<http://cgi.gates96.com/cam/africa/South%20Africa/media.html>

<http://www.letsgo.com/SAF/01-teaser-01>

Documentation (1)

http://www.dwaf.gov.za/dir_ws/content/lids/sanitation.htm

Description (1)

<http://www.lmp.ucla.edu/profiles/profz01.htm>

Textes (2)

LO <http://www.geocities.com/Athens/9479/zulu.html>

<http://www.gov.za/documents/2000/mdda/zulu.pdf>

LC-A (9)

Les langues africaines sur la Toile

Étude des cas haoussa, somali, lingala et isixhosa

Les langues africaines sont présentes sur la Toile. Cette présence facilite la création de corpus linguistiques. Pour illustrer notre propos et notre méthodologie, nous nous concentrons sur quelques langues des quatre coins de l'Afrique: le haoussa, le somali, le lingala et l'isixhosa. La taille et la composition des corpus respectifs donnent une idée de ce qui est disponible sur la Toile pour ces langues et permettent aussi de comparer ces langues entre elles. Afin d'illustrer le potentiel de ces corpus linguistiques nous fabriquerons, à titre d'exemple, des logiciels de grande utilité: à savoir, des correcteurs d'orthographe pour chacune de ces langues africaines.

Termes-clés:

la Toile; correcteurs d'orthographe; langues africaines; haoussa; somali; lingala; isixhosa.

1 Introduction

ON ENTEND souvent dire que les langues africaines ne sont pas présentes sur la Toile. Cette vision doit être sérieusement révisée. Plusieurs centaines de langues africaines ont déjà réussi à conquérir une place dans le cyberspace. Le nombre de sites et donc de textes par contre, dépend de la langue en question et n'est pas toujours en correspondance avec le nombre de locuteurs ni avec l'importance supposée de la langue.

Chanard et Popescu-Belis (2001) ont déjà détaillé les trois phases nécessaires pour qu'on puisse parler de l'informatisation d'une langue. Dans leur article ils ont traité la toute première étape, à savoir le stockage des documents « *sous forme alphabétique, syllabique ou idéographique* sur un support informatique, c'est-à-dire sous la forme d'une suite de caractères qui seule permet l'édition, la recherche, bref la manipulation *analytique* du contenu linguistique » (Chanard et Popescu-Belis 2001: 33).

La deuxième et la troisième étape sont d'après eux respectivement les outils informatiques adaptés à la langue et la diffusion de la langue sur la Toile.

Dans la présente contribution par contre, nous renversons leur séquence en choisissant la Toile comme point de départ. En effet, nous nous basons sur des documents disponibles dans le cyberspace, pour créer des

corpus linguistiques, et ensuite des logiciels comme par exemple des correcteurs d'orthographe. Il va de soi que l'existence de documents en langues africaines sur la Toile présuppose le stockage de ces textes sur un support informatique – la première phase de Chanard et Popescu-Belis.

2 La Toile et les langues africaines

Quoique l'anglais reste de loin la langue dominante sur la Toile, Grefenstette (2002) montre que beaucoup d'autres langues de par le monde gagnent du terrain. Ceci est aussi le cas pour les langues africaines, puisque la proportion de celles-ci ne cesse de croître. Pour une esquisse traitant des langues africaines dans le cyberspace dans ce contexte, voir De Schryver (2002). Néanmoins, une question valable reste la suivante: pourquoi est-ce que la présence des langues africaines sur la Toile reste aussi précaire bien que celles-ci constituent 30% de toutes les langues du monde?

2.1 La situation précaire des langues africaines dans le cyberspace

Est-ce un problème d'existence de caractères? Non, de nos jours toute écriture peut être utilisée sur la Toile, aussi bien les systèmes alphabétiques (comme par exemple le kiswahili¹ ou l'arabe²), que les systèmes syllabiques (comme par exemple l'amharique³), que les systèmes idéographiques (comme pour par exemple le *han* chinois ou le *kanji* japonais). En ce qui concerne les langues africaines, il est bien connu que la plus grande partie (de celles qui sont déjà écrites) est basée sur l'alphabet latin. De la même façon que pour plusieurs langues indo-européennes qui emploient l'alphabet latin comme base, des signes diacritiques peuvent avoir été ajoutés ou des caractères latins de base peuvent avoir été adaptés en ajoutant des glyphes (pseudo-)phonétiques. Pour pouvoir « voir » l'orthographe correcte, ou du moins l'orthographe du

1 p. ex. ippmedia.com/alasiri.htm

2 p. ex. news.bbc.co.uk/1/1/arabic/news

3 p. ex. www2.dw-world.de/ambaric/

document, il suffit d'employer un navigateur plutôt récent (par exemple *Explorer 4+* ou *Netscape 4+*), et/ou de télécharger le jeu de caractères qui a été employé pour la création du document.

Est-ce un problème de standard orthographique? Pour le haoussa par exemple, il existe une orthographe officielle. Par contre sur la Toile, comme on le verra plus loin, on trouve au moins six types d'orthographe. Cette situation ne facilite pas nécessairement l'échange de documents et peut même être un obstacle dans la création et le téléchargement depuis les serveurs de documents.

Est-ce un problème d'accès? Les difficultés d'accès aux technologies de l'information et de la communication (TIC) ne sont en effet pas à sous-estimer en Afrique. Les cybercafés ne se trouvent que dans les grandes villes et ils ne sont pas accessibles à tout le monde. En plus, l'analphabétisme constitue un obstacle à la fréquentation de textes en l'occurrence sur ordinateur.

Est-ce un problème de motivation? Le manque de motivation parmi les Africains à écrire dans leur propre langue est une des raisons que l'on peut citer pour expliquer le relatif insuccès des langues africaines sur la Toile. Le cybernaute qui s'exprime sur la Toile veut être lu et compris, il va donc écrire dans une langue connue par le plus grand nombre de gens. En effet, une grande partie des textes en langues africaines trouvés sur la Toile n'a pas été écrit par des Africains, comme nombre de documents religieux ou de textes destinés à l'enseignement. Des forums où des Africains communiquent avec d'autres Africains, en langues africaines, sont l'exception et non la règle.

2.2 Estimer le nombre de mots en langues africaines dans le cyberspace

Grefenstette et Nioche (2000) ont proposé une méthode pour estimer le nombre de mots, pour une langue donnée, dans le cyberspace. Même si le nombre de mots en langues africaines est, jusqu'à ce jour, en fait toujours trop faible pour ce genre de calcul, on peut déjà employer la méthode de Grefenstette et Nioche avec succès. L'idée de base est toute simple: il suffit de diviser le nombre de fois qu'une sélection de mots uniques d'une langue donnée a été indexée par un moteur de recherche, par la fréquence

relative de ces mêmes mots dans un corpus. La moyenne des résultats offre alors une estimation du nombre de mots sur la Toile dans la langue étudiée.

On peut brièvement illustrer la méthode pour le sepedi. La colonne 1 dans [1] liste les dix mots qui ont été utilisés, avec leurs équivalents français dans la colonne 2. La colonne 3 montre les fréquences relatives dans un corpus du sepedi (pour ce corpus, voir De Schryver & Lepota 2001 : 3). Les dix mots de la colonne 1 ont été soumis au moteur de recherche *Google*⁴ (mars 2003) et le nombre de « pages vues » pour chacun de ces mots est montré dans l'avant-dernière colonne. Les estimations de la dernière colonne sont basées sur les chiffres des deux colonnes précédentes.

[1] Estimation du nombre de mots sepedi sur la Toile

Mot en sepedi	Équivalent en français	Fréquence dans un corpus (en %)	Résultat <i>Google</i>	# mots sepedi sur la Toile
<i>kgopela</i>	demander	0,03961934	46	116 105
<i>latelago</i>	suivre (+relatif)	0,03302768	57	172 582
<i>kgauswi</i>	prêt (de)	0,02695642	31	115 000
<i>bolelago</i>	parler (+relatif)	0,02260245	21	92 910
<i>mafelelong</i>	à la fin	0,01974029	26	131 710
<i>tsebago</i>	savoir (+relatif)	0,01925458	18	93 484
<i>mangwalo</i>	lettres	0,01075481	20	185 963
<i>blogong</i>	dans la tête	0,01030380	8	77 641
<i>sengwalwa</i>	manuscrit	0,00995687	9	90 390
<i>phapano</i>	différence	0,00988749	19	192 162
				126 795

En considérant la moyenne, l'on peut conclure qu'il y a au moins cent vingt-cinq mille mots sepedi sur la Toile. Nous remarquons aussi que même si l'on n'a pas accès à un corpus d'une certaine langue (pour en extraire les chiffres de la colonne 3), il est évidemment possible de télécharger d'abord quelques textes (dans la langue à examiner) de la Toile. Les fréquences relatives peuvent alors être dérivées de l'ensemble de ces textes.

3 Quatre études de cas: haoussa, somali, lingala et isixhosa

Pour illustrer l'existence des langues africaines sur la Toile, nous avons choisi quatre langues représentant quatre régions géographiques de l'Afrique sub-saharienne: le haoussa pour l'Ouest, le somali pour l'Est, le lingala pour la région centrale et l'isixhosa pour la région australe. Notons en passant que le Nord de l'Afrique est évidemment bien représenté par l'arabe, et que la langue sub-saharienne dominante sur la Toile est le kiswahili. Ces deux dernières langues sont d'ailleurs tellement présentes, avec plusieurs millions de mots chacune, qu'il est possible de rechercher uniquement des pages écrites dans ces langues avec le moteur de recherche *All the Web*⁴.

Rechercher des sites et/ou des pages dans une langue africaine donnée peut facilement se faire à l'aide d'un des nombreux moteurs de recherche. Il suffit d'introduire un ou plusieurs mots de la langue recherchée dans le champ de recherche, d'appuyer sur la touche *Entrée* du clavier, et – à condition que les mots n'existent que dans cette langue – tous les liens cliquables sur la page avec les résultats mèneront à des sites et/ou des pages dans la langue recherchée.

C'est exactement ce procédé-là que nous avons suivi pour le haoussa, le somali, le lingala et l'isixhosa. Pour chacune de ces langues, nous avons navigué sur la Toile pendant environ quatre jours (mars 2003). À partir des textes trouvés, nous avons alors construit un corpus pour chaque langue. La taille et le contenu de ces différents corpus dépendent bien sûr de ce qui est disponible sur la Toile. Ensuite nous avons utilisé ces corpus pour créer plusieurs correcteurs d'orthographe dans les quatre langues.

Dans ce qui suit, nous présenterons les difficultés que nous avons rencontrées, les résultats, puis nous expliquerons comment on peut fabriquer un correcteur d'orthographe. Nous mettrons aussi les correcteurs d'orthographe à l'épreuve à l'aide d'un texte qui existe dans les quatre langues concernées, c'est-à-dire la *Déclaration universelle des droits de l'homme*⁵.

⁴ www.google.com

⁵ www.alltheweb.com

⁶ www.umbcbr.cb/udbr

⁷ www.etbnologie.com

3.1 L'Afrique de l'Ouest: haoussa

Pour l'Afrique de l'Ouest, le choix s'est facilement porté sur le haoussa qui est, avec ses trente-neuf millions de locuteurs dans huit pays différents⁷, la plus grande *lingua franca* de cette région géographique. Le haoussa est une langue de la famille chadique qui fait partie de la grande unité des langues afro-asiatiques. Ce grand nombre de locuteurs du haoussa comme première ou deuxième langue et sa fonction comme langue véhiculaire ont malheureusement pour conséquence que les textes haoussa sur la Toile se présentent sous des formes très diverses. Nous avons aussi remarqué que la plupart des sites sont d'origine nigériane.

Le grand nombre d'orthographes différentes, utilisées dans des textes haoussa trouvés sur la Toile, pose problème. Historiquement, le haoussa connaît deux types d'écritures, l'ajami et le boko. L'écriture ajami, étant basée sur des caractères arabes, était utilisée dès le XVII^e siècle. À présent par contre, l'écriture basée sur l'alphabet latin, le boko, a gagné beaucoup de terrain. Cette écriture boko, utilisée par les Européens depuis le début du XIX^e siècle, est la seule trouvée sur la Toile. Dans l'étude présente nous ne nous arrêterons pas sur la question de l'écriture ajami.

Malgré l'imposition d'une orthographe standard pour le haoussa en janvier 1980 pendant une conférence à Niamey (Wolff 1991), au moins six types d'orthographe sont utilisés dans le cyberspace. L'alphabet latin « pur » ne répond pas aux besoins d'une écriture haoussa, puisque des caractères symbolisant des coups de glottes, des implosifs et des éjectifs sont absents. Dans l'orthographe de 1980, le « b implosif » est représenté par le caractère phonétique **ɓ** et le « d implosif » par le caractère phonétique **ɗ**. Le « k éjectif » correspond au caractère phonétique **ƙ**. Le « coup de glotte » est représenté par l'apostrophe ' et la combinaison du « coup de glotte et y » s'écrit **y**. On peut considérer les autres types d'orthographe (trouvés sur la Toile et ailleurs) comme des adaptations graduelles de cette orthographe standard. Le **y** peut être remplacé par 'y. Une deuxième simplification est le remplacement de **ɓ** et **ɗ** par 'b et 'd et le remplacement de **ƙ** par k'. Ici on voit encore bien la différence entre les consonnes implosives et la consonne éjective. Dans une troisième étape, par contre, cette distinction est omise et on trouve b', d' et k'. Une simplification qui va encore plus loin est l'omission du caractère implosif ou éjectif des

consonnes, on garde seulement l’apostrophe pour le coup de glotte et avant y. Dans la dernière étape, seul le coup de glotte est présent. Ces six types d’écriture haoussa sont non seulement tous présents sur la Toile, mais parfois plusieurs orthographes sont utilisées sur un seul site ou même dans un seul texte. Bien que la simplification ne semble pas poser trop de problèmes pour les locuteurs maternels du haoussa – dans les différents forums, les cybernautes ont tendance à utiliser l’orthographe la plus simplifiée – les orthographes simplifiées ne contribuent pas à la compréhension de la langue pour ceux qui étudient le haoussa.

Bien que le haoussa soit une langue tonale, les tons ne sont pas marqués dans la plupart des textes haoussa disponibles sur la Toile. Nous avons trouvé seulement cinquante-six mots sur lesquels des tons sont marqués; il s’agit des tons hauts (´), des tons bas (̀) et des tons descendants (ˆ). Il n’y a que neuf textes sur trois sites différents qui utilisent ces mots avec tons.

Surtout pour le traitement informatique efficace d’une langue, l’utilisation d’un seul type d’orthographe est nécessaire. Le meilleur choix pour le haoussa est sans aucun doute l’orthographe standard. L’utilisation des caractères phonétiques ne pose pas de problèmes, puisque ceux-ci sont présents dans le jeu Unicode. Même pour l’élaboration d’un

correcteur d’orthographe, ces caractères phonétiques (Unicode) peuvent être utilisés sans aucun problème.

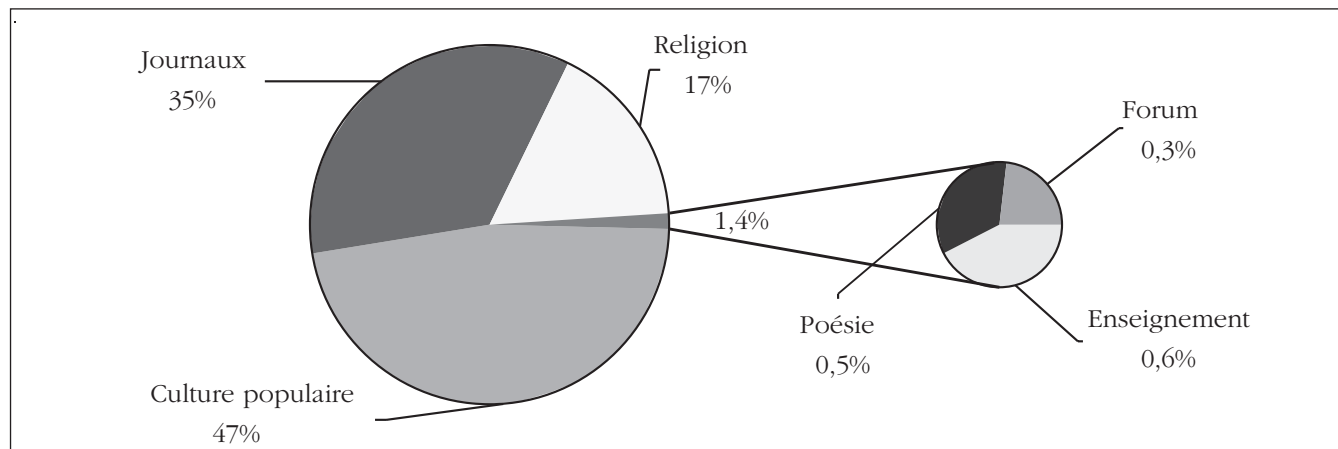
3.1.1 Distribution du haoussa sur la Toile selon le contenu

De la navigation sur la Toile pendant quatre jours à la recherche de documents haoussa résulte un corpus linguistique contenant 858 734 mots au total (les « tokens »), dont 30 996 mots différents (les « types »). Comme on peut le voir dans [2], les textes haoussa peuvent être divisés en trois grandes catégories: culture populaire (47%), journaux (35%) et religion (17%). Le pourcentage de textes produits sur des forums par des cybernautes est minuscule (0,3% du total des mots).

3.1.2 Correcteurs d’orthographe haoussa

Il y a plusieurs façons de construire un correcteur d’orthographe. Pour un aperçu avec l’accent sur les langues africaines, voir Prinsloo et De Schryver (2003). Une des pistes est tellement simple qu’on pourrait parler de « correcteur d’orthographe maison ». En effet, tous les logiciels populaires de traitement de textes, comme *Microsoft Word* ou *Corel WordPerfect*, ont une fonction qui

[2] Distribution du haoussa sur la Toile selon le contenu



permet d'ajouter un ou plusieurs « dictionnaires personnels ». Il suffit donc de composer une liste des mots fréquents dans un corpus, ou même une liste de tous les mots différents dans un corpus, et d'en faire un ou plusieurs correcteurs d'orthographe.

À titre d'exemple, [3] montre une partie d'un texte du site *Bisbarat*⁸: ce texte ne se trouve *pas* dans le corpus, et il est vérifié avec un correcteur d'orthographe dérivé du corpus contenant 858 734 *tokens* et 30 996 *types*. On voit clairement qu'il n'y a aucun problème à traiter l'orthographe standard du haoussa (avec les caractères phonétiques) dans un outil d'édition, même au niveau du logiciel de vérification.

Dès qu'on dispose d'un correcteur d'orthographe utilisant une certaine orthographe, il est généralement trivial d'employer la fonction « rechercher et remplacer » pour en faire un vérificateur capable de reconnaître une autre orthographe. Puisque nous avons décidé de contrôler l'efficacité de tous nos correcteurs d'orthographe à l'aide du même texte disponible dans les quatre langues – la *Déclaration universelle des droits de l'homme*, un texte qui n'est évidemment pas inclus dans les corpus respectifs – nous avons produit un correcteur d'orthographe haoussa qui utilise l'écriture de ce texte, et dans le cas présent, l'écriture la plus simplifiée. Les résultats des tests sont résumés dans [4].

Comme on peut le voir, un total de cinq correcteurs d'orthographe a été produit, chacun correspondant à une

[3] Texte haoussa dans l'orthographe standard, vérifié à l'aide d'un correcteur d'orthographe dérivé de la Toile (2 mots corrects ne sont pas reconnus)

	<p>Les changements technologiques et socio-économiques se déroulent rapidement. Cependant, en Afrique, aussi bien qu'ailleurs dans le Sud, il y a peu de ressources disponibles pour traiter des aspects de TIC (et en même temps, il y a beaucoup d'autres besoins de base exigeant ces ressources).</p>
<p>Ga albarkacin ilimin na'urar zamani, halayen al'amuran duniya na <u>cuɗayyar</u> jama'a, sai ci gaba da tasowa suke yi da sauri. Amma duk da haka, a faɗin Afirka da sauran wurare na kudu a bisa doron duniya, akwai karancin albarkatun na'urar zamani waɗanda za'a fuskanci hanyoyin labarai da sadarwa na zamani da su (kuma ga buƙatu iri-iri na sahan gaba sun fuskanto a gaggauce).</p>	<p>Des opportunités peuvent être perdues dans la mesure où des actions opportunes ne sont pas prises. En reconnaissant ce fait, des agences et des organismes extérieurs cherchent à aider l'Afrique dans le domaine des TIC, se concentrant d'abord sur les questions essentielles de la connectivité de base et des aspects de « l'accès physique ». Cependant, peu abordent encore les questions également importantes de « l'accès significatif » et du contenu.</p>
<p>Ga halin haka kuwa, hanyoyin dacewa da dama sukan iya ɓacewa in ba'a ɗauki wasu matakai na aikata abubuwan da wuri ba. Bayan dogon nazari bisa ga dukan al 'amuran nan ne wasu ma'aikata da kungiyoyi masu zaman kansu, suke aiki domin su taimaki Afirka ta fannin sababbin hanyoyin labarai da sadarwa na zamani, ta a maido hankali wajen sawa a gane da yadda hanyoyin nan suke, da kuma yadda ake iya cin moriyarsu a aikace.</p>	

[4] Construction d'un correcteur d'orthographe haoussa

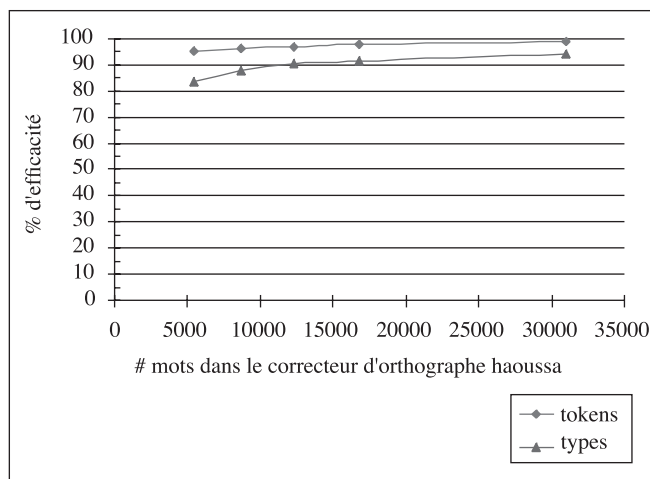
Correcteur d'orthographe haoussa (dérivé de 858 734 <i>tokens</i>)			<i>Déclaration universelle des droits de l'homme</i> haoussa (2 751 <i>tokens</i> ; 556 <i>types</i>)				
mots dans chaque niveau			pas reconnu		% d'efficacité		
fréquence	nombre	%	<i>tokens</i>	<i>types</i>	<i>tokens</i>	<i>types</i>	utilisateurs
10 ou plus	5 418	17,48	130	90	95,27	83,81	96,73
5 jusqu'à 9	3 218	10,38	100	67	96,36	87,95	97,56
3 et 4	3 657	11,80	81	53	97,06	90,47	98,07
2	4 519	14,58	64	46	97,67	91,73	98,33
1 (<i>hapax</i>)	14 184	45,76	35	31	98,73	94,42	98,87
	30 996	100,00					

différente section ou « couche » du corpus: (i) tous les mots avec une fréquence d'au moins 10; (ii) tous les mots avec une fréquence de cinq à neuf; (iii) tous les mots avec une fréquence de trois et quatre; (iv) tous les mots avec une fréquence de deux; et (v) tous les mots qui n'apparaissent qu'une fois dans le corpus (les « hapax », bons pour 45,76% du total). La *Déclaration universelle des droits de l'homme* contient 2 751 *tokens* et 556 *types*. Nous présumons qu'il n'y a pas de fautes dans ce texte, et donc qu'un vérificateur « parfait » devrait pouvoir reconnaître tous les mots. Autrement dit, l'efficacité (ou le « taux de rappel ») devrait être de 100%. En utilisant seulement le premier niveau, l'efficacité du point de vue des *tokens* est déjà de 95,27%. En ajoutant le second niveau au premier, l'efficacité grimpe jusqu'à 96,36%; en ajoutant le troisième niveau, l'efficacité est de 97,06%; en ajoutant ensuite le quatrième niveau, l'on atteint 97,67%; finalement avec tous les niveaux ensemble, l'efficacité grimpe jusqu'à 98,73%. Du point de vue des *types*, l'efficacité est moins bonne, mais du point de vue des utilisateurs, l'efficacité grimpe même jusqu'à 98,87%. Pour calculer ce dernier, le nombre de *types* non reconnu est comparé au nombre de *tokens* dans le texte. Le raisonnement est que, du point de vue de l'utilisateur, il suffit d'ajouter chaque *type* une seule fois avec la fonction « Ajouter » afin que le correcteur d'orthographe reconnaisse le mot en question. Les graphiques dans [5] résument l'efficacité.

3.2 L'Afrique de l'Est: somali

Le somali – aussi une langue afro-asiatique, mais faisant partie de la famille cushitique – est parlé en Afrique de l'Est, plus précisément au Somali et dans les pays limitrophes: le Djibouti, l'Éthiopie et le Kenya. La population totale de locuteurs du somali peut être chiffrée à

[5] Efficacité d'un correcteur d'orthographe haoussa

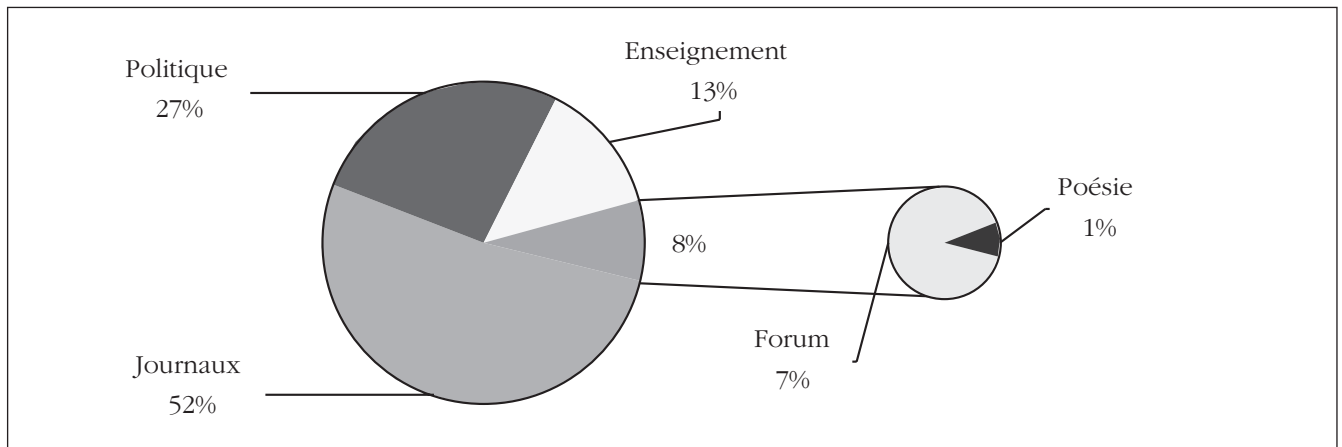


environ dix millions⁹. Ce nombre de locuteurs indique que cette langue joue un rôle important parmi les langues africaines. L'écriture du somali se fait à l'aide de l'alphabet latin.

3.2.1 Distribution du somali sur la Toile selon le contenu

Un corpus de 304 361 *tokens* et 40 251 *types* a été rassemblé, dont la distribution est montrée dans [6].

[6] Distribution du somali sur la Toile selon le contenu



Comparé à la distribution pour le haoussa, on voit l'apparition d'un grand bloc de politique (27%) et d'enseignement (13%). Aussi les forums grandissent substantiellement (7%). Notons toutefois que malgré le fait qu'il n'y a pas de textes religieux dans le corpus, il y en a qui sont disponibles sur la Toile. En ce qui concerne les documents politiques, il est étonnant qu'un grand nombre provienne de sites officiels de gouvernements non-africains (australiens, finlandais, suédois, etc.). La présence de locuteurs du somali dans ces pays peut expliquer cela. Une autre partie des documents politiques est originaire de la « République de Somaliland ». Environ la moitié (52%) du matériel somali sur la Toile est composé d'articles de journaux.

3.2.2 Correcteurs d'orthographe somali

Pour le somali aussi, cinq correcteurs d'orthographe ont été produits selon la même procédure que pour le haoussa. Malgré le fait que le corpus somali soit trois fois plus petit que celui du haoussa, il contient un quart de *types* de plus, dont 60,09% d'hapax. Le nombre de *types* en soi ne dit évidemment pas tout, et l'efficacité des cinq niveaux du correcteur d'orthographe est en effet moins bonne que pour le haoussa. Les chiffres et graphiques sont résumés dans [7] et [8].

Pour le somali, la compagnie *SomiTek* distribue un correcteur d'orthographe appelé « Hikaadiye » qui peut être téléchargé depuis la Toile¹⁰. Après une comparaison entre notre correcteur d'orthographe et le correcteur offert par *SomiTek*, nous avons constaté que beaucoup plus de mots n'étaient pas reconnus par Hikaadiye. Il est possible que *SomiTek* se soit basée sur un autre dialecte du somali pour l'élaboration de son logiciel de vérification. Dans [9] les deux correcteurs d'orthographe sont appliqués sur le même texte, notamment l'Article 21 de la *Déclaration universelle des droits de l'homme*. Les mots qui ne sont pas reconnus sont en gras.

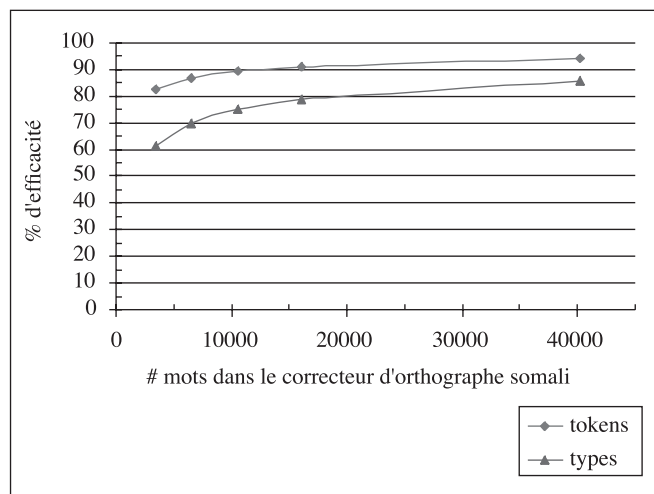
⁹ www.ethnologue.com

¹⁰ www.somitek.com

[7] Construction d'un correcteur d'orthographe somali

Correcteur d'orthographe somali (dérivé de 304 361 <i>tokens</i>)			<i>Déclaration universelle des droits de l'homme</i> somali (1 919 <i>tokens</i> ; 710 <i>types</i>)				
mots dans chaque niveau			pas reconnu		% d'efficacité		
fréquence	nombre	%	<i>tokens</i>	<i>types</i>	<i>tokens</i>	<i>types</i>	utilisateurs
10 ou plus	3 433	8,53	334	276	82,60	61,13	85,62
5 jusqu'à 9	3 026	7,52	255	215	86,71	69,72	88,80
3 et 4	4 056	10,08	207	175	89,21	75,35	90,88
2	5 551	13,79	176	151	90,83	78,73	92,13
1 (<i>hapax</i>)	24 185	60,09	107	100	94,42	85,92	94,79
	40 251	100,00					

[8] Efficacité d'un correcteur d'orthographe somali



Pour ce petit extrait, le nombre de mots qui ne sont pas reconnus par le correcteur d'orthographe somali Hikaadiye est, comparé au nôtre, trois fois plus grand.

[9] Comparaison entre le correcteur d'orthographe somali Hikaadiye et le nôtre

Correcteur d'orthographe somali Hikaadiye
Qod XXI

1. Qof kastaa wuxuu xaq u leeyahay inuu ka qayb galo maamulka Dalkiisa si toos ah ama isagoo si xornimo ah u dooranaya cid wakiil ka ah.
2. Dadka oo dhami way u **simanyihiin** inay ka qayb galaan maamulka sare ee dalka ka jira.
3. Rabitaanka dadweynuhu waa saldhigga awoodda Dawladda; **rabitaankaas** waxaa lagu muujinyaa dooroshooyin run ah oo xilliyo joogta ah dhacaya, laga wada qayb-galaya, waxayna u dhacayaan si qarsoodi ah ama hab kale oo wax lagu doorto oo xor ah.

Notre correcteur d'orthographe pour le somali
Qod XXI

1. Qof kastaa wuxuu xaq u leeyahay inuu ka qayb galo maamulka Dalkiisa si toos ah ama isagoo si xornimo ah u dooranaya cid wakiil ka ah.
2. Dadka oo dhami way u **simanyihiin** inay ka qayb galaan maamulka sare ee dalka ka jira.
3. Rabitaanka dadweynuhu waa saldhigga awoodda Dawladda; **rabitaankaas** waxaa lagu muujinyaa dooroshooyin run ah oo xilliyo joogta ah dhacaya, laga wada qayb-galaya, waxayna u dhacayaan si qarsoodi ah ama hab kale oo wax lagu doorto oo xor ah.

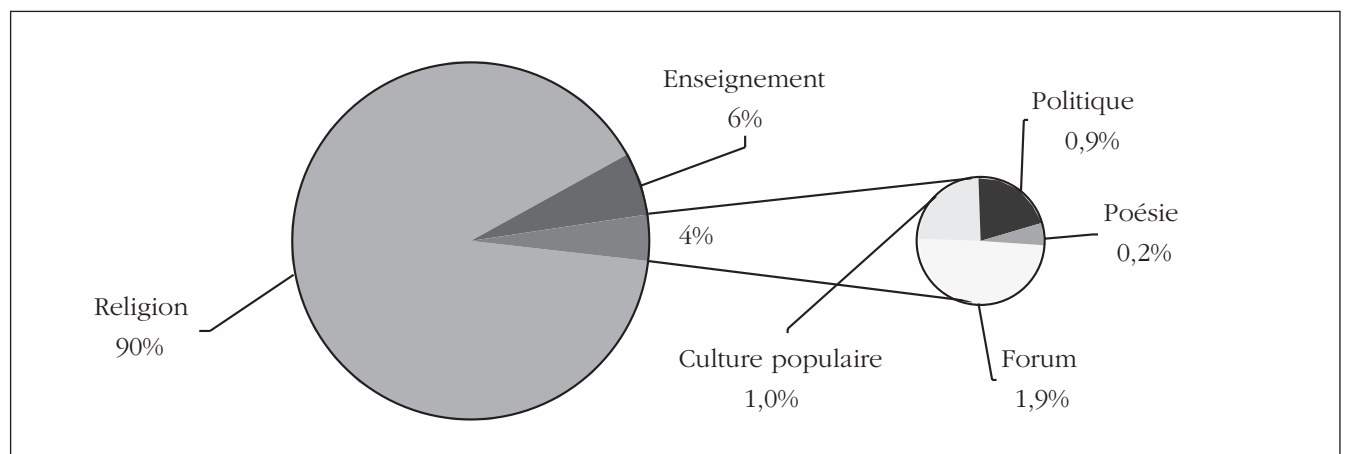
3.3 L'Afrique centrale: lingala

Le lingala – une langue bantoue de la zone C de Guthrie (C36d), parlée dans la République démocratique du Congo (Kinshasa), dans la République du Congo (Brazzaville) et en Angola – peut être considérée comme une *lingua franca* de l'Afrique centrale. Le nombre de locuteurs tourne autour de sept millions¹¹. Malgré ce nombre de locuteurs et le statut de la langue, il n'y a pas d'orthographe fixe. Seules les lettres de l'alphabet latin sont utilisées, du moins sur la Toile, mais l'indication des tons est plutôt sporadique et capricieuse. Les trois systèmes plus ou moins « systématiques » utilisés sur la Toile sont: (i) aucun ton est indiqué; (ii) les tons sont seulement indiqués pour les « paires minimales »; et (iii) tous les tons sont indiqués.

3.3.1 Distribution du lingala sur la Toile selon le contenu

La première chose que l'on remarque pour le lingala sur la Toile est que les sites ne sont pas nombreux. Des quatre langues, le corpus téléchargé pour le lingala est aussi le plus petit avec 193 772 *tokens* et 11 557 *types*. En plus, la distribution selon le contenu est très différente comme on peut le voir dans [10].

[10] Distribution du lingala sur la Toile selon le contenu



La très grande majorité des textes lingala se trouve dans la catégorie religion (90%). La deuxième catégorie, l'enseignement (6%), est beaucoup moins répandue. Le pourcentage des forums (1,9%) est plus haut que pour le haoussa, mais n'atteint pas du tout le pourcentage du somali.

3.3.2 Correcteurs d'orthographe lingala

Puisque l'orthographe de la *Déclaration universelle des droits de l'homme* est sans tons, nous avons fabriqué un correcteur « lingala sans tons » à partir de tous les textes dans le corpus. Ceci est facile, puisqu'il suffit d'utiliser la fonction « rechercher et remplacer » pour se débarrasser des tons avant de faire la liste des mots pour le correcteur d'orthographe. Les résultats du correcteur « lingala sans tons » sont résumés dans [11] et [12].

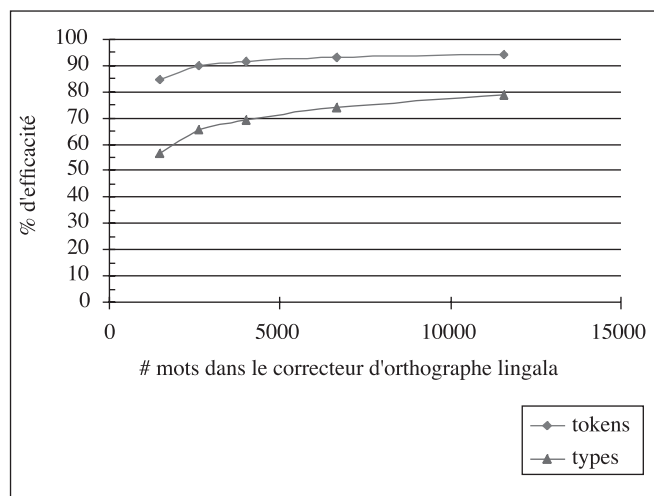
Comparée aux correcteurs haoussa et somali, l'efficacité du correcteur lingala se trouve entre les deux, et cela avec beaucoup moins de mots. Notons tout de même qu'en plus de ce correcteur « lingala sans tons », nous avons aussi fabriqué un « lingala paires minimales » et un « lingala avec tons ».

11 www.ethnologue.com

[11] Construction d'un correcteur d'orthographe lingala

Correcteur d'orthographe lingala (dérivé de 193 772 <i>tokens</i>)			<i>Déclaration universelle des droits de l'homme</i> lingala (1 854 <i>tokens</i> ; 361 <i>types</i>)				
mots dans chaque niveau			pas reconnu		% d'efficacité		
fréquence	nombre	%	<i>tokens</i>	<i>types</i>	<i>tokens</i>	<i>types</i>	utilisateurs
10 ou plus	1 469	12,71	282	156	84,79	56,79	91,59
5 jusqu'à 9	1 130	9,78	182	125	90,18	65,37	93,26
3 et 4	1 388	12,01	154	110	91,69	69,53	94,07
2	2 668	23,09	132	94	92,88	73,96	94,93
1 (hapax)	4 902	42,42	104	77	94,39	78,67	95,85
	11 557	100,00					

[12] Efficacité d'un correcteur d'orthographe lingala



3.4 L'Afrique australe : isixhosa

Pour l'Afrique australe finalement, nous avons décidé d'étudier l'isixhosa, une langue de l'Afrique du Sud mais aussi parlée au Botswana et au Lesotho. Comme le lingala, l'isixhosa compte environ sept millions de locuteurs¹². Cette langue bantoue de la zone S de Guthrie (S41) a été

influencée par les langues khoisans. L'alphabet latin est utilisé pour l'écriture, et même les « clics » sont représentés par des caractères de l'alphabet latin. Les tons ne sont jamais indiqués.

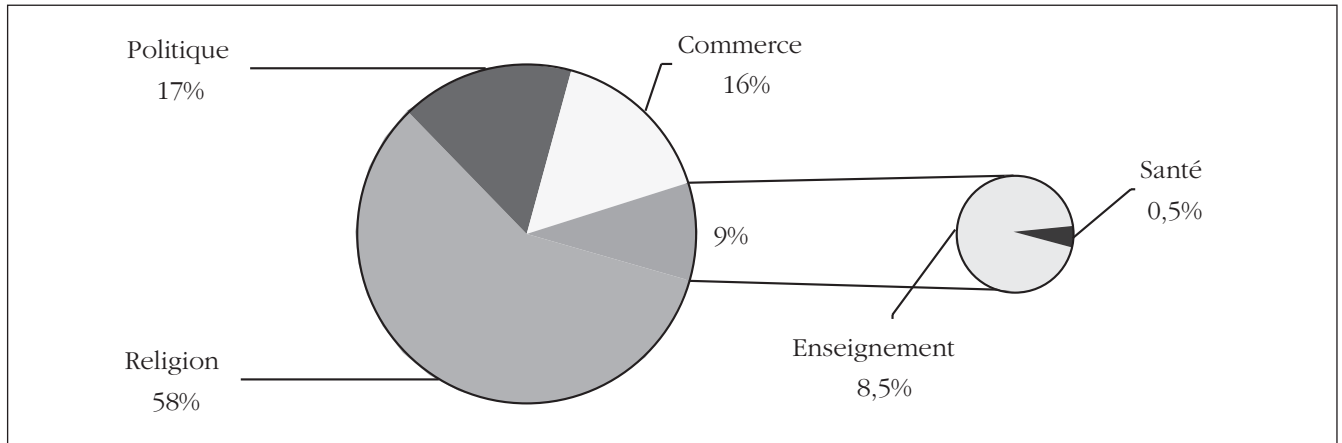
3.4.1 Distribution de l'isixhosa sur la Toile selon le contenu

Des neuf langues bantoues officielles de l'Afrique du Sud, l'isixhosa est certainement celle pour laquelle il y a « beaucoup » de matériel sur la Toile. Un corpus de 943 772 *tokens* et 149 553 *types* a été rassemblé, dont la distribution est montrée dans [13].

Les textes isixhosa peuvent être divisés en trois grandes catégories. Comme pour le lingala, la religion forme le plus grand bloc (58%). La deuxième catégorie est constituée, comme pour le somali, de documents politiques (17%). La troisième catégorie, le commerce (16%), n'existe pas pour les autres langues de notre échantillon. Puisqu'il nous a semblé qu'il n'y a pas de forums isixhosa sérieux ou substantiels sur la Toile, cette catégorie n'est pas représentée dans notre corpus. Un bloc enseignement (8,5%) par contre, est présent comme pour les autres langues.

12 www.ethnologue.com

[13] Distribution de l'isixhosa sur la Toile selon le contenu



3.4.2 Correcteurs d'orthographe isixhosa

L'isixhosa est écrit de manière dite « conjonctive » (voir Prinsloo et De Schryver 2002). En conséquence, le nombre de mots différents, ou donc le nombre de *types*, est très élevé. Ceci ne facilite pas la tâche d'un correcteur d'orthographe, comme on peut le voir dans [14] et [15].

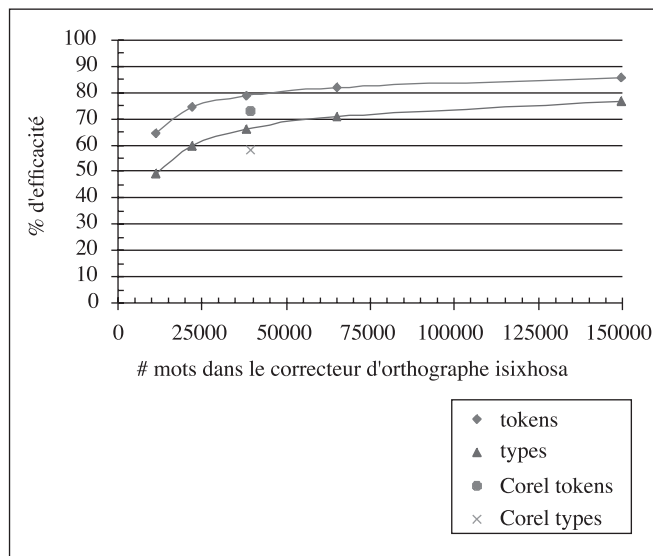
Même avec un total de 150 000 mots, l'efficacité n'atteint qu'un peu plus de 85 %.

Pour l'isixhosa, un correcteur d'orthographe dit « de première génération » (voir Prinsloo et De Schryver 2001 : 129) est inclus dans le logiciel de traitement de textes *Corel WordPerfect 9* distribué en Afrique du Sud depuis l'année 2000. Ce correcteur contient 39 192 mots. L'efficacité de ce correcteur a été calculée sur le même texte, à savoir la *Déclaration universelle des droits de l'homme*, et on a trouvé que la reconnaissance était un peu moins bonne, comme on peut le voir dans [15].

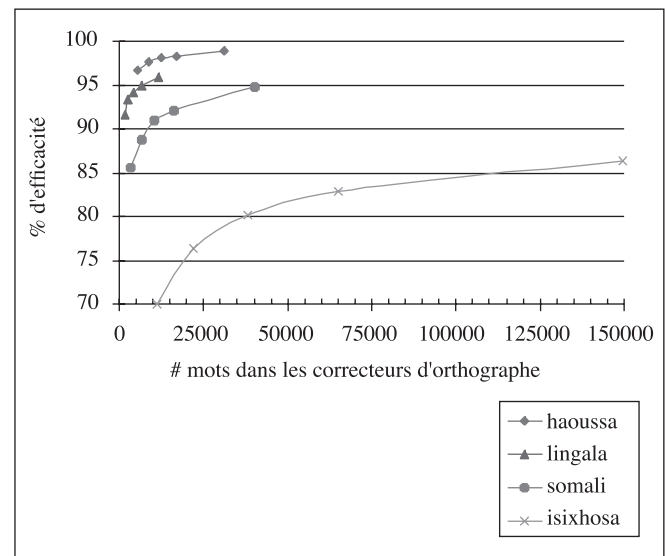
[14] Construction d'un correcteur d'orthographe isixhosa

Correcteur d'orthographe isixhosa (dérivé de 943 772 <i>tokens</i>)			<i>Déclaration universelle des droits de l'homme</i> isixhosa (1 196 <i>tokens</i> ; 705 <i>types</i>)				
mots dans chaque niveau			pas reconnu		% d'efficacité		
fréquence	nombre	%	<i>tokens</i>	<i>types</i>	<i>tokens</i>	<i>types</i>	utilisateurs
10 ou plus	11 026	7,37	422	358	64,72	49,22	70,07
5 jusqu'à 9	11 123	7,44	302	282	74,75	60,00	76,42
3 et 4	16 069	10,74	252	238	78,93	66,24	80,10
2	26 902	17,99	216	206	81,94	70,78	82,78
1 (hapax)	84 433	56,46	172	164	85,62	76,74	86,29
	149 553	100,00					

[15] Efficacité d'un correcteur d'orthographe isixhosa



[16] Comparaison de l'efficacité des correcteurs d'orthographe pour les quatre langues



4 Conclusion

Dans cette contribution nous avons démontré que les langues africaines sont non seulement présentes sur la Toile – et certaines même en force – mais aussi qu'il n'y a aucune raison technique qui devrait empêcher le téléchargement d'encore plus de documents et l'augmentation du nombre de langues utilisées, tout cela en utilisant les orthographes correctes. L'étude du contenu de sites haoussa, somali, lingala et isixhosa semble indiquer qu'à l'Ouest et qu'à l'Est du continent les actualités sont favorisées, tandis que pour la zone bantoue on remarque une grande attention pour la religion. Des échantillons d'autres langues africaines confirment cette tendance.

En naviguant sur la Toile le même nombre de jours (quatre) pour le haoussa, le somali, le lingala et l'isixhosa, des corpus linguistiques d'environ 850 000, 300 000, 200 000 et 950 000 mots (*tokens*) ont pu être rassemblés. Dû à des différences entre les structures de ces langues, et pour les langues bantoues à des différences entre les degrés de conjonction, le nombre de mots différents (*types*) est

très varié: respectivement 30 000, 40 000, 10 000 et 150 000.

Cette dernière série de mots a ensuite été employée pour la fabrication de correcteurs d'orthographe. Une comparaison de l'efficacité de ces correcteurs d'orthographe, du point de vue de l'utilisateur, en employant la *Déclaration universelle des droits de l'homme* pour les quatre langues de l'échantillon, est présenté dans [16].

Il est clair que pour des langues comme le haoussa et le somali, ainsi que pour des langues bantoues avec une orthographe disjonctive, des logiciels de vérification basés sur des « mots orthographiques » sont tout à fait réalisables. Par contre, pour des langues bantoues avec une orthographe conjonctive, comme l'isixhosa (mais aussi l'isizulu, l'isindebele, le siswati, etc.), un logiciel de vérification ne sera satisfaisant qu'en employant des approches qui tiennent compte de la morphologie de ces langues. Une piste qui semble prometteuse est d'utiliser des machines à états finis (comme cela a été fait pour par exemple le finlandais). Des projets actuels et similaires traitent des langues comme l'amharique, l'arabe et l'isizulu.

En conclusion il est manifeste que les langues africaines méritent et prennent déjà leur place sur la Toile et il est réjouissant que le traitement automatique du langage naturel (TALN) pour les langues africaines soit bien en cours.

Anneleen Van der Veken,
Linguistique africaine, Université libre de Bruxelles, Bruxelles,
Belgique.
avdveken@ulb.ac.be

Gilles-Maurice de Schryver,
Département de langues et cultures africaines, Université de Gand,
Gand, Belgique.
gillesmaurice.deschryver@ugent.be

Bibliographie

- BBC, 2003: *News in 43 languages, Arabic*, news.bbc.co.uk/hi/arabic/news.
- Bisharat, 2003: *Initiative langues – technologie – développement*, www.bisharat.net.
- Chanard (Chr.) et Popescu-Belis (A.), 2001: « Encodage informatique multilingue: application au contexte du Niger », dans *Les Cahiers du Rifal*, n° 22, p. 33-45.
- De Schryver (G.-M.), 2002: « Web for/as Corpus: A Perspective for the African Languages », dans *Nordic Journal of African Studies*, n° 11 (2), p. 266-282.
- De Schryver (G.-M.) et Lepota (B.), 2001: « The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails », dans *Lexikos*, n° 11 (Afrilex-reeks/series 11: 2001), p. 1-37.
- Deutsche Welle, 2003: *News in 31 languages, Amharic*, www.dwelle.de/amharic/Welcome.html.
- Fast Search & Transfer, 2003: *moteur de recherche « All the Web »*, www.alltheweb.com.
- Google, 2003: *moteur de recherche « Google »*, www.google.com.
- Grefenstette (Gr.), 2002: « The WWW as a Resource for Lexicography », dans Corréard (Marie-Hélène), éd., *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins*, Euralex, p. 199-215.
- Grefenstette (Gr.) et Nioche (J.), 2000: « Estimation of English and non-English Language Use on the WWW », dans *Proceedings of Riao 2000*, Paris, 12-14 Avril 2000.
- Grimes (B. F.) et Grimes (J. E.), éd., 2000: *Ethnologue. Languages of the World, 14th Edition*, Dallas: SIL International. Voir aussi www.ethnologue.com
- IPPmedia, 2003: *Alasiri, Afternoon news from Dar es Salaam, Tanzania*, ippmedia.com/alasiri.htm.
- Prinsloo (D. J.) et De Schryver (G.-M.), 2001: « Corpus applications for the African languages, with special reference to research, teaching, learning and software », dans *Southern African Linguistics and Applied Language Studies*, n° 19 (1-2), p. 111-131.
- Prinsloo (D. J.) et De Schryver (G.-M.), 2002: « Towards an 11 x 11 Array for the Degree of Conjunction / Disjunction of the South African Languages », dans *Nordic Journal of African Studies*, n° 11 (2), p. 249-265.
- Prinsloo (D. J.) et De Schryver (G.-M.), 2003: « Towards Second-Generation Spellcheckers for the South African Languages », dans De Schryver (Gilles-Maurice), éd., *Tama 2003 South Africa: Conference Proceedings*, Pretoria: (SF)² Press, pp. 135-141.
- SomiTek (Somali Information Technology), 1999: *Hikaadiye, a complete wordprocessor with Somali and English spell checking capability*, www.somitek.com.
- United Nations Organisation, 1948: « Universal Declaration of Human Rights », Adopted and proclaimed by General Assembly resolution 217 A (III) of 10 December 1948, www.unhchr.ch/udhr.
- Wolff (E. H.), 1991: « Standardization and varieties of written Hausa (West Africa) », dans Von Gleich (U.) et Wolff (E. H.), éd., *Standardization of national languages. Symposium on language standardization, 2-3 February 1991*, Hamburg: UEI reports, Unesco Institute for Education, p. 21-32.

Flore: un site coopératif pour recueillir et diffuser les noms des plantes dans les langues africaines

Les travaux que nous présentons s'inscrivent dans une problématique de sauvegarde des langues pour lesquelles il existe peu de ressources linguistiques comme les dictionnaires ou les lexiques. La rareté de ces ouvrages de référence fragilise ces langues car elle favorise la multiplication d'écrits incorrects, c'est-à-dire non conformes aux décrets de transcription officiels en vigueur. Nous considérons que la constitution de ressources lexicales électroniques via le réseau Internet représente une solution viable. D'une part elle permet de fédérer les travaux de chercheurs géographiquement éloignés. D'autre part elle facilite la diffusion des connaissances puisque celles-ci sont disponibles pour n'importe quel internaute.

Le site Flore s'inscrit dans ce cadre. Il est spécialisé dans les noms des plantes. La définition précise des informations stockées sur ce site par les partenaires du projet (informaticiens, linguistes et ethnobotanistes) nous a amené à dépasser l'objectif initial et à ajouter la description botanique des plantes ainsi que leurs utilisations. Cette extension a abouti à la définition de stratégies de coopération entre les personnes de différentes compétences susceptibles de fournir de nouvelles informations au site (linguistes, botanistes, nutritionnistes, etc.). Nous présentons les problèmes techniques rencontrés lors du développement qui font obstacle à l'informatisation des langues africaines: il s'agit des caractères spéciaux spécifiques à ces langues et inexistantes sur les claviers courants et dans les polices de caractères courantes. Enfin nous explorons les multiples extensions envisageables, telles la mise en service d'un forum de discussion réservé aux contributeurs du site ou l'édition de son contenu dans un format facilitant le partage des données: l'XML.

Termes-clés:

langue africaine; nom de plante; flore; site web; site coopératif; transcription; décret de transcription.

Introduction

LES NOMS des plantes dans les langues africaines sont généralement présentés à la fin des ouvrages de botanique, mais ces noms sont souvent erronés. Nous présentons le site électronique *Flore* qui est destiné à recueillir ces noms, les faire corriger par des linguistes, puis présenter les versions correctes afin qu'elles soient diffusées et employées.

Nous montrons que cette démarche s'inscrit dans un contexte plus général de sauvegarde des langues, et nous passons en revue les motivations pour développer ce site

électronique. Ensuite, nous détaillons les informations qu'il contient et son mode de fonctionnement pour l'apport de nouvelles informations et leur validation. Nous insistons sur la nécessaire collaboration entre spécialistes de compétences différentes (linguistes et botanistes) et présentons les modes de coopération que nous avons mis en place. Enfin, nous ouvrons la liste des développements futurs qui feront évoluer le site *Flore*¹.

1 Motivation

Dans de nombreux domaines (santé, éducation, économie) l'Afrique est dans une situation catastrophique, le domaine linguistique n'échappe pas à cette situation. Bien que ce continent soit riche d'un millier de langues, celles-ci sont en grande majorité peu valorisées, voire abandonnées. La transmission de connaissances est réalisée dans des langues que les populations ne maîtrisent pas, ce qui accroît encore les difficultés. Cet abandon des langues utilisées par les populations a une seconde conséquence difficilement perceptible *a priori*: une diminution de la capacité globale de l'espèce humaine à innover. En effet, chaque langue est plus qu'un simple code mais reflète une conceptualisation particulière de la réalité. Edward Sapir (Sapir [1921] cité par Durand [2002: chap. 2]) souligne ainsi que le langage est une traduction symbolique de la réalité mais que la réalité, telle que nous la percevons, est pour une large part inconsciemment fondée sur nos habitudes linguistiques. Chaque langue présente des rapprochements sémantiques, des collisions phonétiques, qui lui sont propres et peuvent donner lieu à des découvertes qui ne seraient pas concevables dans une autre langue. En permettant la construction de perceptions différentes de la réalité, la diversité des langues entraîne le progrès car elle favorise la multiplicité des langues et du vécu.

Cet abandon des langues est favorisé par le faible nombre de cadres qui sont formés en Afrique, y compris dans le domaine très particulier de la linguistique. Dans

¹ Nous remercions le Fonds francophone des inforoutes de l'Agence de la francophonie qui a financé la conception de ce site (août 1999 - juin 2001).

une langue donnée, il est très difficile d'obtenir des compétences langagières du niveau d'un locuteur maternel si l'on a grandi dans une autre langue, or le nombre de linguistes travaillant sur leurs langues maternelles est extrêmement faible (si l'on compare, par exemple, avec le nombre de linguistes français travaillant sur la langue française). À cause, précisément, de ce manque de linguistes formés, de nombreuses questions linguistiques ne sont pas résolues.

– Souvent il n'existe aucune ressource lexicale, en l'occurrence des dictionnaires, qui permette de fixer la graphie, la phonétique et les différents sens des lexèmes. Ainsi le premier dictionnaire monolingue en zarma a vu le jour en 2000 (Oumarou 2000)² alors qu'il existait auparavant un dictionnaire bilingue zarma-français (Bernard 1994). Des ressources bilingues zarma-anglais sont également apparues sur la Toile (<http://www.bisharat.net/Zarma/>).

– Les ouvrages rassemblant les règles syntaxiques sont très rares, peu connus et peu utilisés.

Ce manque chronique de ressources lexicales freine largement la production de textes écrits alors que ces langues de tradition orale sont largement pratiquées par les populations. Lorsque des transcriptions voient le jour, c'est souvent le fait de sociologues ou d'ethnologues sans les compétences linguistiques nécessaires pour les réaliser dans les règles de l'art. La multiplication de ces textes mal transcrits contribue à dégrader encore davantage les capacités des populations de lire et d'écrire ces langues. Cette méconnaissance de leurs propres langues les pénalise d'autant plus qu'elle leur interdit l'accès aux nouvelles technologies, et en particulier à Internet, dans leur langue.

2 Signalons, cependant, les projets dimo-lexis (dictionnaires monolingues et lexiques spécialisés) financés par l'Agence de coopération culturelle et technique (ACCT) de 1984 à 1989 dans le cadre de ses programmes de coopération linguistique.

organisé par la chaire Unesco en aménagement linguistique et didactique des langues dans les systèmes éducatifs et l'Institut de linguistique de l'Université de Mons-Hainaut.

3 Séminaire tenu à Mons en Belgique, du 16 au 22 mars 2002. Ce séminaire a été

Pourtant Internet pourrait contribuer à relier ces populations à la diaspora, encourager la production et la diffusion d'écrits, stimuler l'activité économique. Le fait que ces activités doivent toutes être réalisées dans une langue étrangère (généralement le français en Afrique de l'Ouest) aggrave encore la situation linguistique.

Cette pauvreté linguistique a des répercussions sur le développement économique des populations car la transmission des connaissances techniques et scientifiques est souvent portée par une langue occidentale (généralement celle d'un des pays anciens colonisateurs). Non seulement les populations doivent acquérir de nouveaux savoirs, mais elles doivent le faire (à supposer qu'elles en aient la possibilité et l'opportunité) dans une langue étrangère pas ou peu maîtrisée, ce qui ne facilite pas l'apprentissage.

Il apparaît donc que la production et la diffusion d'informations linguistiques concernant les dénominations scientifiques et techniques constituent une condition nécessaire (mais non suffisante) au développement économique et un facteur essentiel de la renaissance africaine.

Et comme le font remarquer les experts réunis à Mons à l'occasion du séminaire³ sur le thème « Respect de la diversité linguistique et accès universel au cyberspace » : « Nul doute (...) qu'à court terme, un nombre substantiel d'activités humaines ne pourront plus être envisagées sans moyens de télécommunications informatisés » (Harmegnies et Renard 2002: 12).

2 Choix du médium

La diffusion des connaissances a été longtemps monopolisée par les ouvrages imprimés (revues ou livres). Mais les ouvrages imprimés restent chers et, de ce fait, peu diffusés. On peut évoquer ici les nombreuses collectes de terminologies (vocabulaires ou lexiques) effectuées par des institutions du Sud et qui dorment, pour ainsi dire, dans les tiroirs des chercheurs, faute de moyens financiers pour en assurer l'édition. Et en tout état de cause, celles qui parviennent à être publiées n'atteignent pas forcément tous les publics visés qui, de toute manière, ne bénéficieraient pas de l'interactivité qu'offre une base de données sur la Toile doublée d'un forum de discussions.

Internet a donc bouleversé ce schéma, non seulement en donnant la possibilité de mettre fin à cette discrimination et à cette injustice dont l'Afrique est la principale victime, mais également par l'apport d'une série de fonctionnalités inédites : il est possible de stocker, manipuler et diffuser les données sous forme électronique.

- le stockage dans des bases de données exige une définition conceptuelle des données et systématise le recueil des informations ;
- le partage de grandes quantités de données électroniques est facilité par l'utilisation de formats comme l'XML ;
- les données peuvent être accessibles via un site Internet à partir de n'importe quel ordinateur connecté à la Toile ;
- l'ajout de données est aisé et ces nouvelles données sont immédiatement accessibles en consultation .

Cette technologie présente par ailleurs quelques inconvénients facilement maîtrisables :

- il faut être équipé d'un ordinateur connecté à la Toile pour accéder aux informations ;
- les données électroniques sont très volatiles et peuvent être irrémédiablement perdues en cas de panne ou d'erreur de manipulation. Toutefois le stockage des informations dans des bases de données garantit la sauvegarde des données.

Plusieurs stratégies permettent de minimiser ces inconvénients. Tout d'abord, la production régulière d'ouvrages imprimés constitue une diffusion supplémentaire des connaissances qu'il ne faut pas oublier : elle touche les populations rurales qui ne sont pas connectées et elle permet une consultation fréquente des informations. Par ailleurs, ces ouvrages imprimés constituent une forme de pérennisation et de sauvegarde des informations qu'il ne faut pas négliger. Toutefois, la sauvegarde régulière des informations sur support électronique doit évidemment être réalisée scrupuleusement.

L'ancien vecteur de diffusion des connaissances qu'est le livre réapparaît donc comme un complément primordial de la technologie Internet.

3 Les noms de plantes

Les objets répertoriés dans les ontologies naturalistes (zoologie ou botanique) sont désignés par un unique nom

savant de forme latinisante, indépendant de la langue. Il s'agit d'une dénomination d'un concept.

Dans chacune des langues, ces objets peuvent avoir une ou plusieurs dénominations, ou parfois n'en avoir aucune. Ces dénominations sont les étiquettes linguistiques des objets qui constituent les ontologies naturelles, il s'agit donc de termes (Sager, 1990).

Les ouvrages de botanique et de zoologie présentent (généralement en fin d'ouvrage) des listes des noms de ces objets dans les langues locales mais ces noms recueillis par des naturalistes démunis des compétences linguistiques nécessaires sont entachés d'erreurs.

Voici quelques exemples :

Le nom du *Cassia Occidentalis* L. en wolof cité dans Fortin *et al.* (1997 : 92) est *bentemaré* alors que dans von Maydell (1983 : 447) est énoncé *bantamare*. En réalité ces deux dénominations coexistent effectivement mais dans des dialectes différents.

Deux noms en wolof sont cités dans Fortin *et al.* (1997 : 92) pour le *Moringa oleifera* Lam : *nevöday* et *nüböday* alors que dans Von Maydell (1983 : 447) est énoncé *benaille*. Ces trois dénominations sont grossièrement erronées : les caractères *ö* et *v* n'existent pas dans la langue wolof, tandis que *benaille* est une transcription défectueuse du français *ben ailé* ! En fait, le nom wolof devrait être transcrit *nebedaa* (ce nom serait issu de l'expression anglaise *never die*).

Le nom du *Tamarindus Indica* L. en wolof cité dans Fortin *et al.* (1997 : 219) est *dakkar* alors que dans von Maydell (1983 : 447) sont énoncés *dakak* et *dakkar* : *dakak* est erroné tandis que *dakkar* est mal transcrit. La dénomination correcte en wolof établie en respectant l'alphabet et les règles de transcription officiels est : *daqaar*.

D'une manière générale, nous constatons que les règles de transcription ne sont pas respectées. Ainsi, bien que de nombreuses langues africaines s'écrivent à l'aide de caractères non compris dans l'alphabet latin (comme le b crossé : ḅ ; ou le n palatal : ɲ), nous n'en avons trouvé aucune trace dans ces listes.

Les ouvrages dans lesquels nous avons relevé de nombreux noms mal transcrits sont largement diffusés, et même admis comme des références dans le domaine de la botanique. Comme ils sont souvent cités, ils présentent l'inconvénient majeur de propager des informations linguistiques fausses allant par exemple jusqu'à implicitement remettre en cause les caractères

alphabétiques adéquats pour une langue comme dans le cas du *nevöday*. Ils constituent de véritables virus linguistiques (par analogie avec les virus informatiques) susceptibles de pervertir la langue.

Il existe peu de sites répertoriant les noms savants dans les langues nationales. Ecoport, un site américain, présente des listes de noms dans différentes langues. Mais, en français les noms présentent des erreurs dues à l'utilisation exclusive des caractères latins non accentués (*arbre a beurre* au lieu de *arbre à beurre*). En fait, ce site présente l'inconvénient majeur de ne pas identifier les contributeurs du site qui ajoutent de nouvelles informations. Celles-ci sont alors présentées sans vérification d'aucune sorte.

Les noms fournis au site *Flore* sont transcrits ou retranscrits par des linguistes locuteurs des langues concernées conformément aux décrets de transcription officiels en vigueur dans les pays où ces langues sont parlées. Il serait d'ailleurs souhaitable que les ministères chargés des langues nationales rendent accessibles ces décrets en les mettant à disposition sur la Toile.

Au Sénégal, plusieurs décrets ont été promulgués. Concernant le wolof, on peut citer le décret n° 72-702 du 16 juin 1972 modifiant le décret n° 71-566 du 21 mai 1971 relatif à la transcription des langues nationales et le décret n° 85-1232 du 20 novembre 1985 modifiant le décret n° 75-1026 du 10 octobre 1975 relatif à l'orthographe et à la séparation des mots. Le décret n° 75-1025 du 10 octobre 1975 vise le sereer. L'orthographe du pulaar est régie par le décret n° 1049 du 14 octobre 1980, celle du soninké, du manding et du jóola repose sur des textes de 1980 non encore promulgués. Récemment, la Direction de la promotion des langues nationales (D.P.L.N.) a procédé à la codification de l'écriture de plusieurs langues parlées au Sénégal dont le balante, le mankañ et le noon.

4 Réalisation: le site *Flore*

L'objectif principal du site *Flore* est de collecter et de diffuser les noms des plantes dans les langues du Sahel, tous ces noms ayant été au préalable établis par des linguistes en respectant les décrets de transcription. Ce sont la qualité et la fiabilité de ces informations qui font la valeur de ce site.

Le nombre de langues concernées n'est pas limité. Pour

l'instant, il comprend le bambara, le haoussa, le jóola, le peul⁴, le sereer, le wolof et le zarma, auxquels nous avons joint le français et l'anglais.

Toutefois, nommer les plantes suppose de savoir les reconnaître, il nous est donc apparu indispensable de présenter la description botanique des plantes. Enfin, de nombreuses plantes offrent des possibilités d'utilisation qui peuvent être vitales dans des pays démunis, qu'il s'agisse d'utilisations culinaires (on peut se nourrir de certains fruits, de certaines racines), domestiques (comme la confection de clôtures), ou médicales. Nous avons donc décidé de répertorier également ces utilisations.

La langue pivot, utilisée pour l'interface du site et pour renseigner les informations autres que les noms des plantes, est le français.

4.1 Modèle de données

Le modèle de données comprend trois parties, chacune pouvant être identifiée à un domaine des savoirs que nous avons choisi de modéliser. Ainsi, il y a une partie linguistique, une partie botanique, et une partie utilisation. Chacune de ces parties a été conçue en étroite collaboration avec des professionnels du domaine concerné: deux linguistes, Boureima Oumarou (docteur en linguistique) et Chérif Mbodj (docteur en linguistique) ont contribué à la conceptualisation de la partie linguistique, tandis qu'Anne Luxereau (docteur en ethnobotanique) intervenait sur les parties botanique et utilisations.

Le modèle de données réalisé est centré sur la notion d'espèce (celle-ci est repérée par son nom latin et son inventeur). Pour chaque espèce, il existe une partie linguistique et une partie utilisations.

4.1.1 Partie linguistique

Chaque espèce peut avoir un ou plusieurs noms dans chaque langue.

Chaque nom est assorti de sa graphie, de sa phonétique, d'informations linguistiques et d'un commentaire libre.

Les informations linguistiques diffèrent selon les

⁴ Appelé également *pula(a)r*, *ful* ou *fulfulde*.

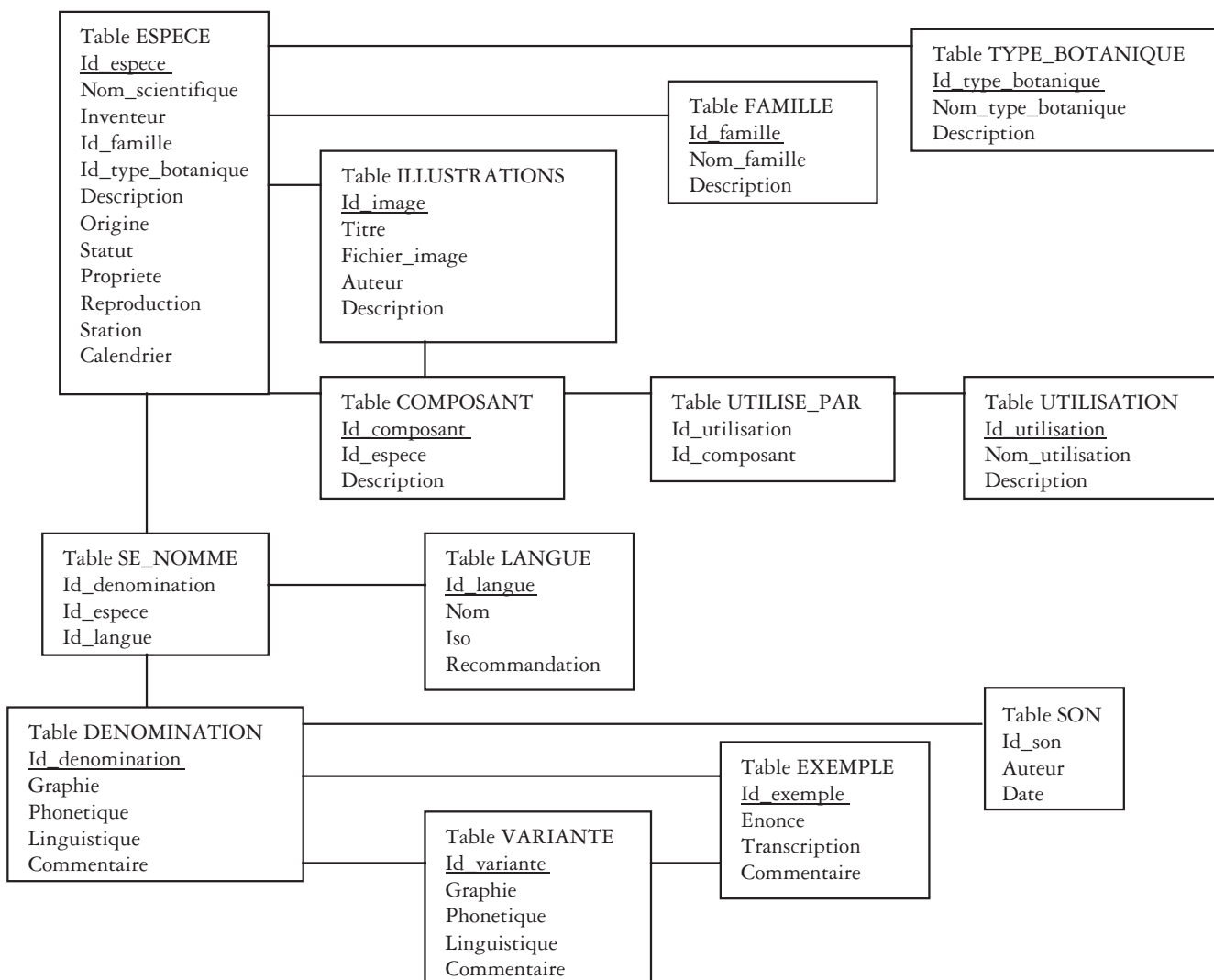
langues. En français, par exemple, il s'agit de fixer le genre du nom (masculin ou féminin) alors qu'en wolof il faut préciser sa classe nominale (les suffixes de classe les plus fréquents étant g- ou b- dans le domaine de la flore). Les informations linguistiques sont primordiales pour l'éventuelle intégration de ces données au sein de ressources linguistiques utilisables en traitement automatique des langues.

Des exemples d'usage accompagnent cette description.

Chaque exemple comprend :

- son énoncé dans la langue locale;
- sa transcription en français (la langue pivot);
- un éventuel commentaire;
- un ou plusieurs enregistrements audio de l'énoncé prononcés par un locuteur natif de la langue.

Modélisation des informations (extrait)



Certains noms peuvent présenter de légères variations (appelées variantes) selon la région. Ainsi, le fruit du *Solanum aethiopicum* se nomme *jaxatu* en wolof, mais dans la région de Saint-Louis, au nord du Sénégal, il est appelé *xuluñe*.

Une variante est attachée à un nom. Elle est décrite par sa graphie, sa phonétique, les informations linguistiques et un commentaire libre qui précise notamment la région géographique concernée. Des exemples d'usage complètent cette description.

4.1.2 Partie botanique

Chaque espèce est repérée par son nom scientifique, sa famille botanique et son inventeur. Elle est décrite par:

- son type botanique (arbre, arbuste, graminée, etc.);
- sa description botanique;
- son origine, mentionnée si l'espèce a été importée d'une autre partie du monde;
- son statut mentionne le caractère spontané ou cultivé de l'espèce, et, dans ce dernier cas, précise le mode de culture (bouturable, transplantable, etc.);
- ses propriétés (exemple: fixation de l'azote dans le sol);
- son mode de reproduction (rhizome ou graine ou clonage, etc.);
- son calendrier (exemple: période de floraisons);
- sa station décrit les lieux où peuvent se rencontrer des individus de l'espèce.

Les composants de l'espèce⁵ sont également décrits et illustrés de photographies ou de dessins.

Les éventuels synonymes du nom latin sont signalés: il s'agit d'anciens noms que des révisions de la classification ont évincés.

4.1.3 Partie utilisations

Enfin, la connaissance des plantes en Afrique étant souvent liée à leurs utilisations, nous avons ajouté à cette description botanique une partie utilisation. Nous distinguons une liste de plusieurs catégories d'utilisations (artisanale, cosmétique, culinaire, mystique, domestique,

⁵ Bourgeon, branche, cendre, écorce, feuille, fleur, fruit, gomme, épine, graine, inflorescence, racine, tronc.

fourragère, phyto-sanitaire, autre) qui peut être étendue.

L'utilisation médicale est traitée spécialement puisqu'il faut distinguer les différentes affections visées par les traitements. Cette partie a été conçue en collaboration avec Marcelle Talbot (directrice de la coopérative Banituri) et Anne Luxereau.

Ici encore, il s'agit de ne répertorier que des utilisations effectivement attestées, ou des soins scientifiquement fondés.

4.2 Fonctionnalités du site

Le site présente plusieurs caractéristiques qui conditionnent son mode de fonctionnement.

1) Les informations sont visibles par tous sur la Toile.

Le site est donc librement accessible en consultation à tout internaute.

2) Les informations présentées dans le site sont de grande qualité.

Cette condition exige la collaboration de différents contributeurs responsables de la qualité des informations selon leur domaine de compétence. Ainsi, les informations linguistiques sont contrôlées par des linguistes (chacun étant qualifié pour une langue), tandis que des botanistes se chargent des informations botaniques. Les utilisations, du fait de leur diversité, peuvent être issues de différents métiers: ethnologues, nutritionnistes, médecins, personnel de santé, etc. Techniquement, cette exigence de qualité demande donc la définition de différents types de contributeurs et leur identification à l'aide d'un nom d'utilisateur et d'un mot de passe.

3) Des contributeurs de différentes compétences renseignent le site.

Énoncer le nom d'une espèce dans une langue particulière suppose des compétences dans plusieurs domaines: il faut connaître cette langue (et donc avoir des compétences linguistiques pour établir correctement sa graphie), et savoir identifier l'espèce (et donc disposer des compétences botaniques nécessaires). Il est illusoire de penser trouver des contributeurs qualifiés à la fois en linguistique et en botanique, il faut donc définir un mode de fonctionnement permettant la collaboration entre les différentes compétences.

Cet objectif peut être rempli en définissant une procédure en deux passes (saisie, puis validation) classique en terminologie. Tout contributeur peut saisir des informations dans tous les domaines. Ainsi, un linguiste peut entrer une nouvelle espèce, un botaniste peut proposer un nom dans une langue ou une utilisation, etc. Les informations seront ensuite corrigées lors de l'étape de validation, celle-ci étant strictement limitée à certaines informations selon les domaines de compétence de la personne qualifiée: les botanistes qualifiés dans une langue particulière passent en revue, corrigent puis valident les informations linguistiques préalablement saisies dans cette langue; de la même façon les informations botaniques sont validées par des botanistes. Les informations sur les utilisations font intervenir des domaines différents et sont validées selon les compétences des valideurs.

Cette caractéristique a pour conséquence la définition précise des domaines de compétence des utilisateurs chargés de la validation des informations. Pour chacun de ces utilisateurs, sont donc mémorisés (en plus du nom d'utilisateur et du mot de passe) un ou plusieurs domaines de compétence (linguistique, botanique ou utilisation). Pour la linguistique, il faut également noter la langue de compétence et pour les utilisations le ou les types d'utilisation concernés. Lors de l'étape de validation, seules les informations correspondant à ses compétences sont proposées à l'utilisateur qualifié en validation.

Ces exigences aboutissent à la définition de trois modes de fonctionnement du site:

4.2.1 Consultation

Tout internaute peut consulter le site. Il accède aux informations pour toutes les espèces. Il peut consulter les informations botaniques et /ou linguistiques et /ou sur les utilisations.

4.2.2 Saisie

L'utilisateur s'identifie à l'aide d'un nom d'utilisateur et d'un mot de passe préalablement délivrés par l'administrateur du site.

L'utilisateur peut :

- ajouter une espèce;

- compléter la description d'une espèce;
- ajouter des composants à des espèces existantes;
- ajouter ou compléter des utilisations;
- ajouter un nom à une espèce;
- ajouter un exemple d'usage à un nom;
- ajouter un enregistrement audio;
- compléter la description d'un nom;
- ajouter une variante;
- etc.

4.2.3 Validation

L'utilisateur s'identifie à l'aide d'un nom d'utilisateur et d'un mot de passe préalablement délivrés par l'administrateur du site.

L'utilisateur est qualifié en botanique ou en linguistique ou pour les utilisations. Dans le cas de la linguistique, il est qualifié pour une seule langue. Dans le cas des utilisations, il est qualifié pour une ou plusieurs utilisations.

Ces utilisateurs valident selon leurs compétences les informations préalablement saisies par des utilisateurs qualifiés en saisie.

Une fois validées ces informations sont présentées en consultation.

4.3 Difficultés techniques

La technologie des autoroutes de l'information a été originellement développée aux États-Unis, pays anglophone. Comme la langue anglaise n'utilise que les 26 caractères non accentués et sans signe diacritique de l'alphabet latin, cette technologie n'est pas adaptée aux autres alphabets. Ainsi, la plupart des utilisateurs francophones du courrier électronique ont-ils déjà été confrontés à des messages dans lesquels les caractères accentués sont dénaturés et ne s'affichent pas correctement. Ces caractères ne sont pas codés de la même manière par tous les systèmes, ce qui entraîne ces difficultés (Chanard *et al.*, 2002). Le récent standard Unicode représente une avancée considérable dans le domaine de la transmission électronique des informations puisqu'il fournit un codage unique pour chacun des caractères utilisés dans le monde (y compris les idéogrammes chinois par exemple).

Cependant, ce standard est récent, et nous sommes encore confrontés à de nombreuses difficultés techniques dues à l'utilisation de caractères spéciaux (comme le b crossé: ɓ; ou le n palatal: ŋ) dans certaines langues africaines:

- ces caractères ne sont pas supportés par la plupart des polices de caractères. Nous tenons donc à disposition des utilisateurs une police de caractères qu'ils peuvent télécharger en ligne. Cette police de caractères est suffisante pour la plupart des langues africaines et permet d'avoir des affichages sans défaut lors de la consultation du site.
- les caractères spéciaux sont mal représentés dans les formulaires utilisés lors de la saisie ou de la validation des données. Nous sommes actuellement à la recherche d'une solution satisfaisante.
- les caractères spéciaux ne sont pas présents sur les claviers des ordinateurs. L'utilisateur peut saisir le code du caractère (par exemple ŋ pour le n palatal N), mais cette solution n'est pas ergonomiquement satisfaisante. Prochainement il sera possible de saisir un caractère spécial en utilisant une combinaison de touches appropriée.

4.4 État du site

Le site *Flore* est visible sur la Toile à l'adresse <http://www.sciences.univ-nantes.fr/irin/flore/>.

Pour l'instant il contient 30 plantes (uniquement des arbres et des arbustes), 9 illustrations, 165 noms de plantes dans 4 langues (français, joola, pulaar, wolof, zarma), 17 références bibliographiques et 4 utilisations. Les noms sont accompagnés de 30 exemples d'usage (en zarma).

Vous pouvez consulter, à titre d'exemple, la fiche décrivant l'espèce *Andansonia Digitata* L. annexée à cet article.

5 Perspectives

Les perspectives de développement de ce site sont multiples et concernent l'extension de la base de données et des fonctionnalités.

Concernant la base de données, les contacts avec des nutritionnistes nous ont convaincus de rajouter les noms de

certaines composantes comme les fruits pour les arbres fruitiers, ou la racine lorsqu'il s'agit de plantes dont la racine est comestible (comme la carotte).

Par ailleurs, nous envisageons d'étendre ce site à l'ensemble des plantes et à toutes les langues souhaitées (sans nous limiter aux langues du Sahel) tout en privilégiant cependant le public africain.

Concernant les fonctionnalités, les premiers mois de fonctionnement du site ont montré la nécessité d'un espace d'échange et de discussion entre les contributeurs du site (en saisie ou en validation). Cet espace de discussion se matérialisera par l'adjonction d'un forum de discussion.

Les informations saisies peuvent être modifiées, voire rejetées lors de la validation, ces décisions seront notifiées, voire expliquées, aux auteurs de la saisie afin de permettre aux contributeurs de suivre le devenir de leur contribution et les raisons des décisions finalement prises. Cette communication entre les contributeurs nous paraît essentielle pour fidéliser leur participation et étendre le contenu du site.

Nous ajouterons au site la possibilité d'éditer son contenu (en totalité ou partiellement) sous la forme d'un fichier au format XML pour intégrer facilement nos données à d'autres bases de données, et sous la forme d'un fichier mis en forme qui, imprimé, permettra la diffusion de ces données hors du monde électronique.

Conclusion

Même si le site *Flore* est principalement dédié à la linguistique, l'élaboration de ce site nous a donné l'occasion de travailler dans un cadre pluridisciplinaire en faisant collaborer informaticiens, linguistes de plusieurs langues, ethnologues, botanistes, nutritionnistes, personnels de santé. Cette collaboration a considérablement enrichi le site car nous avons essayé de respecter chacune de ces disciplines en modélisant la représentation des données après des discussions poussées avec chacun des interlocuteurs. Il est apparu que chacun est compétent dans son domaine mais devient très incompetent dans le domaine des autres: typiquement, botanistes et nutritionnistes ne peuvent jouer le rôle des linguistes (c'est-à-dire recueillir les noms des plantes dans les langues nationales), tandis que les

linguistes sont absolument incompetents pour décrire finement une espèce botanique. Pourtant les informations intéressantes sont par essence pluridisciplinaires: il n'y a aucun sens à connaître le nom d'une plante sans savoir la reconnaître, sans connaître ses utilisations, sans savoir la transcrire correctement.

Par ailleurs, le travail pluridisciplinaire, en faisant communiquer des professionnels de compétences différentes, fait émerger de nouvelles questions et exige des réponses claires pour lesquelles le spécialiste du domaine concerné doit faire un réel effort didactique vis-à-vis de ses collègues d'autres disciplines.

Il est, peut-être, permis d'espérer que le site sur la flore puisse contribuer efficacement à la préservation du patrimoine culturel et matériel de l'Afrique, à la protection de ses langues dont les ressources ainsi recueillies sont mises à la disposition des usagers et, enfin, à la diversité culturelle de la Francophonie.

En vérité, il reste que l'appel de Dakar de mai 1989 sur les « enjeux du multilinguisme... » doit être entendu si l'on veut que l'Afrique, par le virage culturel que permet de réaliser Internet, puisse s'appropriier les sciences et les techniques.

*Chantal Enguehard,
Institut de recherche en informatique de Nantes, Nantes, France.
chantal.inguehard@irin.univ-nantes.fr*

*Chérif Mbodj,
Centre de linguistique appliquée de Dakar, Dakar, Sénégal.
chembodj@ucad.refer.sn*

Bibliographie

Bernard (Y.), White-Kaba (M.), 1994: *Dictionnaire zarma-français (République du Niger)*, Paris: Agence de coopération culturelle et technique.

Chanard (C.), Popescu-Belis (A.), 2001: « Encodage informatique multilingue: application au contexte du Niger », dans *Cahiers du Rifal*, n° 22, p. 33-45.

Durand (C.-X.), 2002: *La nouvelle guerre contre l'intelligence II. La manipulation mentale par la destruction des langues*, Paris: F.X. de Guibert.

Fortin (D.), Lô (M.), Maynard (G.), 1997, *Plantes médicinales du Sabel*, Dakar: ENDA (Tiers-Monde, série Études et recherches).

Maydell (H.-J. von), 1983: *Arbres et arbustes du Sabel, leurs caractéristiques et leurs utilisations*, Eschborn: GTZ.

Harmegnies (B.) et Renard (R.), 2002: *Rapport d'évaluation du séminaire sur le respect de la diversité linguistique et l'aspect au cybercopie*, Mons: Université de Mons - Hainaut.

Mbodj (Ch.), 1994: « L'activité terminologique au Sénégal », dans *Terminologies nouvelles*, n° 11, p. 3-9.

Oumarou (I. A.), 2000: *Zarmaciine kaamuusu kayna*, Niamey: ACCT, Éditions Alpha.

Sager (J.), 1990: *A practical course in terminology processing*, Amsterdam: John Benjamins Publishing Company.

Sapir (E.), 1921: *Le langage. Introduction à l'étude de la parole*: http://www.uqac.quebec.ca/zone30/Classiques_des_sciences_sociales/livres/Sapir_edward/langage/langage.html.
www.unicode.org

Annexe

Fiche de l'espèce *Adansonia digitata* L.

Partie botanique

Espèce : *Adansonia digitata* L.
 Famille : Bombacaceae
 Synonyme : *Adansonia sphaerocarpa* A. Chev.
 Type botanique : Arbre
 Description : Arbre haut de 10 à 15 m, à tronc énorme atteignant plusieurs mètres de diamètre, écorce lisse et grisâtre. Les branches sont très robustes, peu nombreuses et étalées.
 Station : Terres légères et sablonneuses ou terrains calcaires
 Description des composants :
 Feuille : Les feuilles sont composées, digitées, alternes, ayant un nombre variable (6 à 8) de folioles obovales longues de 10 à 15 cm et larges de 4 à 6 cm. Le pétiole est long de 10 à 20 cm ou davantage.
 Fleur : Les grandes fleurs blanches, larges de 15 à 20 cm, pendent à l'extrémité d'un long pédoncule. Elles s'épanouissent le soir et sont fécondées par les chauve-souris.
 Fruit : Les fruits subsphériques ou ovoïdes de 15 à 35 cm de long sont ligneux, verdâtre bronzé et veloutés. Ils contiennent de nombreuses graines noires arrondies, noyées dans une pulpe farineuse blanche entremêlée de fibres rougeâtres. Le fruit mûrit de janvier à avril au Sahel. Sa pulpe, généralement blanchâtre, mais pouvant être jaune ou rosée, appelée « pain de singe », est très riche en acide ascorbique (vitamine C; de 169 à 270 mg pour 100 g de matière fraîche ou 73 mg vit. C. 100 g MS/pulpe) et en thiamine (vitamine B1 : 0,38 mg/100 g M.V.) ainsi qu'en potassium et en glucose : on dit qu'à volume égal elle est plus riche que l'orange en vitamine C.
 Utilisation : culinaire : Quand on est fatigué, mâcher la pulpe du

fruit du baobab redonne vite des forces. On peut mâcher la pulpe et l'avalier, ou bien la dissoudre dans de l'eau pour faire une boisson rafraîchissante qui est employée quelquefois comme un succédané du lait. On peut aussi dissoudre la poudre dans du lait. Cette boisson est quelquefois mélangée à la « mérisa », sorte de bière de sorgho fermenté très commune au Soudan.

Partie linguistique

Français

Nom : baobab
 Phonétique : [baɔbab]
 Références bibliographiques :
 Fortin (Daniel), Lô (Modou), Maynard (Guy), *Plantes médicinales du Sabel*, CECI 1997, ENDA 1997. ISBN 2-920114-0-2, 1990.
 von Maydell (H.-J.), *Arbres et arbustes du Sabel, leurs caractéristiques et leurs utilisations*, GTZ, 1983.

Jóola-éssulaalu?

Nom : bubak
 Référence bibliographique :
 Sambou (Pierre-Marie), 2001.

Pulaar

Nom : bowdi
 Référence bibliographique :
 Fal (Arame), Santos (Rosine), Doneux (Jean Léonce), *Dictionnaire wolof-français*, Karthala, 1990.

Wolof

Nom : guy
 Phonétique : [guj]
 Références bibliographiques :
 Berhaut Jean, *Flore du Sénégal*, Dakar, Éditions ClairAfrique, 1967.
 Diaw Abdoul Aziz, *Vocabulaire de la Flore*, Dakar, CLAD, 1981.
 Saar Àllaaji Taalam, *Gëstu ci wàllu garab*, tomes 1, 2, 3, Dakar, Sénégalaise de l'imprimerie, 1998.

Zarma

Nom : koo
 Phonétique : [ko:]
 Référence bibliographique :
 Oumarou Boureima, Document interne, 2001.
 Exemple d'usage :
 Énoncé : Hano kaŋ koo ga wura hay, talka si duu goobu kaŋna a catu.

Constitution de banques de textes multilingues : un mécanisme fondé sur le standard XML

Nous présentons dans cet article une méthodologie pour la réalisation de ressources linguistiques réutilisables, à savoir des documents munis de balises, en vue de leur intégration dans des banques de données textuelles. Après avoir présenté les mérites du standard de balisage XML, nous proposons l'utilisation de la spécification XCES, application de XML, pour le balisage des ressources documentaires, et nous la décrivons brièvement. Enfin, nous traçons les principales étapes du processus de conversion des ressources vers ce format, et suggérons quelques repères pédagogiques pouvant aider les chercheurs qui abordent ces tâches de conversion.

Termes-clés : banques de textes ; balisage de textes ; XML ; XCES ; conversion des ressources.

1 Introduction : motivation et objectifs

L'UTILITÉ des banques de textes dans le domaine du traitement informatique des documents, dans le contexte de la société d'information, est désormais établie. Les banques de textes sont utilisées, par exemple, pour la construction de ressources lexicographiques et terminologiques, pour les études linguistiques, en particulier de corpus, et pour la mise au point d'outils de traitement de la langue. Les banques de textes multilingues fournissent un matériel appréciable aux études comparatives et, lorsque les textes dans plusieurs langues sont alignés avec une précision raisonnable, ces banques permettent la construction de lexiques multilingues et de ressources pour les outils d'aide à la traduction, tels les mémoires de traduction. Toutefois, pour que les banques de textes puissent jouer ce rôle, il est nécessaire qu'elles soient convenablement *structurées*. En d'autres mots, le contenu des banques doit être décrit de façon interprétable et uniforme, et, dans la mesure du possible, la structure des textes qui constituent la banque doit être également mise en évidence. C'est donc la structuration dans un format explicite et transparent qui distingue une banque de textes d'une simple collection de documents.

Dans cet article, nous nous proposons de décrire une méthode, des outils et un format permettant la structuration des banques de textes, à partir de ressources documentaires préexistantes, dans différentes langues. Le fondement en est le standard XML pour le marquage des

documents, sur lequel se fonde le standard d'annotation de corpus XCES, que nous proposons d'utiliser. Ce dernier conçoit les documents comme formés d'un en-tête, contenant l'information à *propos du* document, et d'un corps, contenant le document lui-même, tous deux structurés selon des spécifications précises. Une procédure pour l'application de ces standards, ainsi que des éléments pédagogiques, sera également décrite.

Le cadre proposé ici, développé en liaison avec le programme de formation du Rifal pour 2002-2003, est tout particulièrement destiné aux institutions et aux chercheurs qui souhaitent augmenter le potentiel d'une collection de documents électroniques disponibles, en particulier dans des langues encore peu représentées dans la société de l'information, par exemple sur Internet. En effet, un des objectifs du Rifal est d'aider ses membres du Sud à se doter de banques de données terminologiques et textuelles multilingues capables de répondre aux besoins locaux tout en constituant une ressource globalement distribuée. Alors que les formations antérieures du Rifal se sont concentrées sur les techniques terminologiques et la conversion des données existantes, les formations à venir porteront sur la construction de banques de données textuelles et terminologiques multilingues. Le logiciel BTML (Yang 2001), paramétré selon les besoins du Rifal, permettra la gestion de ces banques multilingues distribuées, accessibles *via* Internet.

2 Notions théoriques

Nous présentons en premier lieu dans cette section les préliminaires théoriques aux méthodes que nous proposons. Il s'agit d'abord du standard XML, qui définit une syntaxe abstraite pour le marquage de documents à l'aide de balises ainsi que des outils associés, tels les DTD (définition du type de document) et le langage XSLT (langage extensible pour les feuilles de style). Puis nous décrivons brièvement et montrerons l'intérêt des spécifications XCES (standard pour l'encodage des corpus en XML) par rapport à nos objectifs¹.

¹ Nous employons le mot « standard » pour désigner ces spécifications, bien qu'aucun organisme normatif ne les ait adoptées jusqu'à présent (sauf

pour SGML, cf. section 2.1.1). Il s'agit plutôt de standards auxquels la communauté choisit de se conformer pour des raisons de compatibilité et d'échange.

2.1 Le standard de marquage XML et les outils associés

2.1.1 Origines et intérêt

Le langage XML (en anglais, *eXtended Markup Language*, ou langage de marquage étendu) permet de marquer les documents grâce à des *balises*, afin de les structurer. L'idée centrale est de marquer le contenu d'un document plutôt que sa forme, par exemple le sens de ses différentes parties. Cette façon de rendre explicite le contenu des documents est flexible et expressive, ce qui fait que les applications d'XML sont très nombreuses. Les documents structurés en XML s'échangent facilement entre les humains ou entre des programmes de traitement conçus pour utiliser tel ou tel jeu de balises. XML est très utilisé pour les applications multilingues, grâce aussi à ses liens avec Unicode, jeu de caractères universel (*cf.* section 3.2 ci-dessous) : un document XML est un simple document texte, muni de balises et écrit avec un jeu de caractères précis, par défaut Unicode/UTF-8. Outre les premières lignes qui déclarent le document comme étant écrit « en XML », un document XML contient du texte, enchâssé dans une série hiérarchique de balises. Par ailleurs, XML, standard libre de droits pour la définition des balises, ne doit pas être confondu avec un langage de programmation qui définit des instructions.

Signalons, d'un point de vue historique, que l'ancêtre de XML et de HTML est le langage SGML (*Standard Generalized Markup Language*, ou langage général standard pour le marquage), conçu au début des années 1980, et devenu le standard ISO 8879 en 1986. HTML, le langage de formatage de la plupart des documents accessibles sur Internet, fut à l'origine conçu comme une application de SGML en vue de l'affichage, à savoir un jeu de balises particulières respectant la syntaxe SGML. Quant à lui, XML dérive d'une initiative visant à alléger SGML, mais garde toute sa généralité. La spécification XML 1.0, parue en 1998, est en vigueur encore aujourd'hui (*cf.* www.w3.org).

XML présente quelques points communs mais aussi d'importantes différences par rapport à HTML. Ce dernier ne permet qu'un balisage « graphique » des textes, en vue de l'affichage, alors que XML permet de définir des jeux de balises plus riches, plus nuancés, permettant d'exprimer des

informations plus détaillées sur la structure et la provenance des textes. Dans XML, les noms et la syntaxe des balises sont à définir selon les besoins, alors qu'en HTML ils sont fixés une fois pour toutes. Enfin, XHTML est une application de XML très proche de HTML, appelée à le remplacer.

2.1.2 La DTD, une « grammaire des balises »

Un langage à part permet de définir les balises autorisées dans une application XML et leur ordre, en spécifiant ainsi une « grammaire des balises » nommée DTD (définition du type du document). Une solution équivalente plus récente porte le nom de Schéma XML, et au contraire de la DTD, est écrite en XML ; toutefois, l'utilisation des DTD reste la plus répandue. Une DTD permet de définir et communiquer un jeu de balises à utiliser, en spécifiant notamment les noms de balises autorisés, leur enchaînement, les attributs autorisés, etc.

Ce sont les DTD qui, avec les spécifications écrites en langue naturelle, définissent les « applications d'XML ». Chaque application conçoit ses outils de traitement en fonction du sens des balises décrit dans sa spécification et de leur syntaxe précise décrite dans la DTD. Parmi les applications XML dans le domaine du traitement multilingue, citons plusieurs formats d'échange de ressources entre outils déjà existants. Ces formats permettent de structurer des données de telle sorte qu'elles soient lisibles indépendamment du logiciel utilisé. Ainsi, on a défini le format XLIFF² pour la « localisation », *i.e.* l'adaptation de documents ou programmes à une langue et à une culture données ; le format OLIF³ pour les échanges de données lexicales et terminologiques ; le format TMX⁴ pour l'échange, entre différents logiciels propriétaires, de mémoires de traduction (paires de syntagmes dans la langue source et la langue cible).

2 XLIFF : *XML Localization Interchange File Format*, www.xliff.org, format d'échange XML pour les fichiers à localiser.

4 TMX : *Translation Memory Exchange*, www.lisa.org/tmx, échange des mémoires de traduction.

3 OLIF : *Open Lexicon Interchange Format*, www.olif.net, format ouvert pour l'échange de lexiques.

2.1.3 La conversion des formats et l’affichage: XSLT

Les textes balisés en XML n’ont pas de format d’affichage intrinsèque. Pour les visualiser sous une forme quelconque, qui dépend des objectifs de l’application envisagée, il faut par exemple leur appliquer une feuille de style écrite en XSL⁵, autre standard basé sur XML. Cette feuille de style, appliquée à travers un logiciel de conversion, changera les balises du fichier initial en un autre jeu de balises XML, ou en des balises HTML, ou même en un fichier texte. Des méthodes plus complexes existent également, permettant de produire des fichiers au format imprimable (PDF ou *PostScript*). On peut alors visualiser le résultat dans un navigateur classique – dans ce cas la feuille de style XSL joue le rôle d’une feuille de style CSS, utilisée pour HTML.

L’intérêt principal du mécanisme de feuilles de style XSL est que l’on peut définir de nombreux affichages différents (chaque feuille de style correspondant à un affichage) pour un même document en XML, selon la demande de celui ou de celle qui consulte la banque de données. On peut ainsi afficher tout ou partie d’un document (vue abrégée ou vue détaillée), afficher certaines explications, varier la langue dans laquelle on les affiche, et varier bien sûr l’aspect: couleurs, taille et définition de l’écran, etc.

2.1.4 La vérification et la validation

On dit d’un document balisé en XML qu’il est *bien formé* lorsque ses balises respectent les principes généraux de la syntaxe XML, tel l’emboîtement. On dit d’un document qu’il est *valide* si ses balises respectent *en plus* les règles définies dans une DTD. Le test de validité est donc plus contraignant que le test de bonne formation. On comprend en tout cas que pour qu’un document stocké dans une banque de textes puisse être utilisé par la suite – qu’il soit transformé pour être affiché selon divers formats, ou traité par différents outils logiciels – il est nécessaire qu’il soit valide.

⁵ XSL: *eXtended Stylesheet Language*, www.w3.org/Style/XSL/, langage de feuilles de style étendu; plus spécifiquement, ici, XSLT, *i.e.* transformations XSL.

De nombreux outils, plus ou moins simples, gratuits ou non, permettent de faire ces vérifications. Certains sont de simples vérificateurs autonomes, prenant peu de place en mémoire et offrant peu d’options, alors que d’autres font partie de boîtes à outils logicielles, dont ils constituent les éléments premiers. Les versions les plus récentes des navigateurs Internet courants effectuent également les tests de bonne formation: en ouvrant le document XML à tester dans un navigateur, le programme s’arrêtera sur la première erreur de bonne formation trouvée, s’il y en a une, ou affichera l’intégralité d’un document correct. Naturellement, lorsque le document est affiché, on voit en général l’ensemble des balises, sans formatage particulier, puisqu’il n’y en a pas par défaut (sans feuille de style). Plus récemment, les navigateurs peuvent aussi appliquer une feuille de style donnée à *n* fichier XML. Ces outils ne semblent pas, dans leur dernière version, vérifier la validité d’un document.

2.2 L’encodage des textes selon la spécification XCES

Le stockage des textes dans une banque de données nécessite en général l’utilisation d’un format spécifique, souvent fondé sur des *balises* insérées dans un texte, telles que définies initialement dans le langage SGML. La TEI (*Text Encoding Initiative*, initiative pour le codage des textes) a défini un tel jeu de balises pour le marquage des textes au format électronique, jeu spécifié initialement en SGML. Une simplification de la TEI est proposée dans le standard CES (*Corpus Encoding Standard*, standard de codage des corpus), qui s’inspire également des recommandations du projet européen EAGLES (*cf.* www.ilc.cnr.it/EAGLES96/home.html). CES a été récemment mis à jour et changé en XCES (*cf.* www.cs.vassar.edu/XCES/) pour être conforme au standard XML, en définissant grâce à une DTD des balises XML pour marquer les principales structures d’un texte.

Le modèle de données CES (donc XCES) définit deux principales classes d’annotation. D’abord, chaque document de la banque de textes contient un ensemble de méta-données, en format texte, structuré par des balises spécifiques à XCES. On peut indiquer ainsi un grand nombre d’informations utiles à propos du texte, dont un

certain nombre sont obligatoires : par exemple, l'auteur, la date, un titre, la version, etc. Les balises fournies ici par la spécification XCES sont très nombreuses et permettent de structurer finement ces méta-données, pour inclusion plus tard dans un catalogue électronique, par exemple. Après cet en-tête, le corps du document contient le texte proprement dit, muni lui aussi de nombreuses annotations possibles. Celles-ci sont organisées en plusieurs niveaux : section ou chapitre, paragraphe, phrase ; à chaque niveau, un ensemble de phénomènes linguistiques et discursifs peuvent être annotés.

Nous proposons d'utiliser ici le standard XCES pour l'annotation des textes dans une banque. En effet, cette annotation est répandue dans le monde du traitement des corpus, et son respect de la syntaxe XML fait qu'elle peut être convertie facilement avec une feuille de style XSLT vers d'autres formats d'annotation, le cas échéant. De plus, le groupe de travail XCES met gratuitement à la disposition de la communauté une DTD (assez complexe), et des feuilles de style, pour transformer les ressources XCES. Pour commencer, on pourra utiliser seulement un sous-ensemble restreint de balises extraites du standard XCES⁶.

3 La conversion de données textuelles vers le format XCES

Nous proposons dans ce qui suit un procédé de conversion vers XML et XCES des données textuelles qui peuvent être disponibles dans une institution ou une équipe.

3.1 Un procédé en trois étapes

Les principales étapes peuvent se résumer ainsi : (1) conversion des données existantes au format HTML, à l'aide du logiciel d'édition ayant servi à les créer ; (2) conversion des caractères dits « spéciaux » vers un jeu de caractères compatible avec XML (par exemple ISO-LATIN-1 ou UTF-8/Unicode) ; (3) conversion des balises dans les documents depuis HTML vers les conventions XCES ; (4) validation du document produit. Les documents ainsi obtenus peuvent être stockés dans une banque de textes ; en

fonction du logiciel de gestion de la banque, une conversion des documents par des feuilles de style XSLT peut être nécessaire. Aussi, si l'affichage au format HTML est souhaité, les feuilles de style « XCES vers HTML » fournies avec XCES pourront être utilisées.

Parmi les étapes ci-dessus, la première est la moins coûteuse en temps et en ressources. La deuxième requiert un certain savoir-faire dans le domaine des jeux de caractères et nous présenterons ci-après (3.2) les principales conclusions d'un article paru antérieurement dans la présente revue (Chanard et Popescu-Belis 2001).

La troisième étape est la plus délicate, puisqu'il faut assigner les balises XCES aux différentes parties des documents, en fonction de leur « sens » – cette étape semble donc très difficile à automatiser. Les balises XCES étant plus riches (ou diverses) que les balises de formatage HTML, on ne peut savoir à l'avance comment encoder un texte donné dans le format XCES. On pourra commencer par appliquer un logiciel de « nettoyage » du code HTML (par exemple *HTML Tidy*⁷), et choisir d'abord des textes courts, ayant une structure assez simple. Parmi les opérations à accomplir, l'ajout de la déclaration XML est essentiel, et tient en quelques lignes seulement (prendre modèle sur des exemples fournis par XCES). Par ailleurs, il faut saisir un en-tête pour chaque document, dans le format XCES, ce qui suppose une bonne connaissance de la spécification XCES. Cet en-tête peut toutefois être simplifié pour des documents courts. On passera ensuite à un balisage simple et clair du corps du texte, que l'on pourra complexifier par la suite.

Une fois le document ainsi codé, la validation consiste à vérifier que le balisage suit celui spécifié dans la DTD de XCES. Pour ce faire, des outils gratuits de validation peuvent être téléchargés ; la section 3.3 décrit un ensemble d'éléments utiles pour la manipulation des fichiers XML et XCES.

⁶ Un manuel de balisage XCES, qui explique l'utilisation d'un sous-ensemble de balises XCES, a été utilisé dans la formation Rifal 2002-2003. Ce document a été élaboré par Marc Van Campenhoudt (ISTI, Bruxelles).

⁷ *HTML Tidy* : écrit par Dave Raggett, disponible gratuitement sur le site tidy.sourceforge.net.

3.2 L'encodage des caractères spéciaux : Unicode et XML

Si la conversion en HTML des ressources documentaires est souvent possible grâce aux logiciels d'édition, le problème des caractères dits « spéciaux » n'est pas toujours correctement résolu. Un caractère « spécial » peut être, par exemple, tout caractère qui ne figure pas dans l'ensemble de 26 lettres non accentuées de l'alphabet latin, ou plus souvent un caractère qui n'appartient pas à l'ensemble de caractères « courants » d'une région, tels le *o avec barre* (ø) ou le *s avec cédille* (ç) pour le monde francophone. On voit donc ce que cette notion comporte de relatif; précisément pour cette raison, le standard Unicode met sur un pied d'égalité les différents alphabets grâce à un codage plus riche, contenant 65536 caractères.

Afin de réaliser des ressources interopérables, l'utilisation d'Unicode, prévue par défaut dans les documents XML, est vivement conseillée (Chanard et Popescu-Belis, 2001). Unicode permet en effet la représentation informatique aisée de multiples alphabets. Toutefois, le format d'affichage HTML généré par les logiciels ne suit pas forcément le codage Unicode, mais peut faire appel à des polices de caractères « locales » (balisage du type: *Mot en langue locale*< /font >). Or, dans les documents XML, tous les caractères doivent se conformer au jeu de caractères déclaré au début du document XML, sans pouvoir passer par une information de police. Si l'on opte pour Unicode, on peut saisir directement les codes sur deux octets de chaque caractère (ou les générer grâce à un logiciel de conversion), ou bien insérer des entités avec le code hexadécimal, par exemple, en Unicode, ə pour le « *schwa* » (code hexadécimal 0259) ou bien ə pour le même caractère (code décimal 601).

3.3 Outils pédagogiques pour la conversion

Avant d'entreprendre la conversion d'un nombre important de ressources selon notre procédé, il est utile de le tester sur des documents relativement courts, et se familiariser avec les notions de bonne formation, validité, encodage, etc. Un outil simple de vérification, disponible gratuitement, est *RXP*⁸; un outil gratuit pour l'encodage est *Recode*⁹; un outil pour appliquer les feuilles de style est

*Saxon*¹⁰; rappelons également que les principaux navigateurs contiennent aussi des fonctionnalités pour XML. Dans la formation Rifaal 2002-2003 que nous avons définie, un paquet pédagogique regroupait un certain nombre d'exercices utiles, ainsi que des logiciels et des exemples. À titre d'exercice, une application XML simple était proposée dans un premier temps: l'idée était de coder des fiches bibliographiques en XML, et de les transformer en HTML pour visualisation. Dans un second temps, des modèles de documents XCES (extraits du corpus Multext), la DTD de XCES et une feuille de style étaient introduits. Pour tous ces documents, on peut ainsi tester la bonne formation, la validité, et appliquer des feuilles de style. Notons toutefois ici que le jeu complet de feuilles de style XCES, conçu par les auteurs de XCES, nécessite un programme particulier de transformation XSLT, nommé *XT*¹¹.

4 Conclusion

L'alimentation des banques textuelles par la méthode décrite dans cet article se fonde sur plusieurs techniques élémentaires de manipulation de documents: production de documents XML munis de balises de type XCES, vérification de leur validité, application de feuilles de style XSLT données afin de générer des fichiers pour le transfert ou la visualisation.

L'ensemble de la méthode de conversion présentée ici peut paraître complexe et coûteuse. Il s'agit néanmoins de l'une des possibilités les plus directes pour la constitution de banques de textes multilingues à partir de documents existants rédigés dans des formats non contrôlés. Dans le cas de documents suffisamment importants, le coût de leur encodage correct en XCES est faible par rapport à la valeur de la ressource en elle-même, et les avantages apportés par

8 *RXP*, écrit par Richard Tobin, disponible à: www.cogsci.ed.ac.uk/~richard/rxp.html.

9 *Recode*, écrit par François Pinard, disponible à: www.iro.umontreal.ca/contrib/recode/HTML/.

10 *Saxon*, écrit par Michael Kay, disponible à: saxon.sourceforge.net.

11 *XT*, écrit par James Clark, disponible à: www.blnz.com/xt/.

l'existence d'une banque de textes, véritable bibliothèque minutieusement structurée, compense largement les efforts de conversion.

Andrei Popescu-Belis
Institut pour les études sémantiques et cognitives (ISSCO),
Unité de traitement informatique multilingue (TIM),
École de traduction et d'interprétation, Université de Genève,
Genève, Suisse.
andrei.popescu-belis@issco.unige.ch

Remerciements

Le contenu pédagogique de la formation Rifal 2002-2003 a été préparé par Andrei Popescu-Belis (ISSCO/TIM/ETI, Genève) pour l'Observatoire suisse des industries de la langue. L'auteur souhaite remercier pour leur aide Marc Van Campenhoudt (Isti, Bruxelles) Christian Chanard (LLACAN-CNRS, Villejuif), Marcel Diki-Kidiri (LLACAN, Villejuif), responsable du Groupe de travail formation du Rifal, ainsi que Marcel Grangier (Chancellerie fédérale suisse, Berne).

Bibliographie

Amann (B.) et Rigaux (P.), 2002 : *Comprendre XSLT*, Paris : Éditions O'Reilly.

Chanard (C.) et Popescu-Belis (A.), 2001 : « Encodage informatique multilingue : application au contexte du Niger », *Cahiers du RIFAL*, n° 22, p. 33-45.

Consortium Unicode, 2000 : *The Unicode Standard Version 3.0*, Reading : Addison Wesley. Voir aussi www.unicode.org.

Harold (E. R.) et Scott Means (W. S.), 2001 : *XML in a Nutsbell: a Desktop Quick Reference*, Sebastopol : O'Reilly & Associates.

Ide (N.), Bonhomme (P.) et Romary (L.), 2000 : « XCES: An XML-based Encoding Standard for Linguistic Corpora », Actes de LREC 2000 (2^e Conférence internationale sur les ressources linguistiques et l'évaluation), Athènes, p. 825-830. Voir aussi www.cs.vassar.edu/XCES/.

ISO/IEC 10646-1, 2000 : *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane*, Genève : Organisation internationale de normalisation. Voir aussi www.iso.ch.

Popescu-Belis (A.), 2002 : « Apports d'Unicode à l'édition numérique multilingue », *Document numérique*, vol. 6, n° 3-4, p. 139-153.

Yang (J.), 2001 : « Comment construire une banque de terminologie véritablement multilingue », Actes du colloque *L'impact des nouvelles technologies sur la gestion terminologique*, Toronto.

Tri informatique pour le lingala et le hausa dans le projet BTML

Le projet BTML de l'Office québécois de la langue française vise à construire une banque de terminologie multilingue capable de soutenir sur le plan technologique les caractéristiques spécifiques de diverses langues. Le présent article présente un compte rendu des expériences du traitement de langues africaines dans le cadre de ce projet, notamment pour ce qui concerne le tri multilingue à l'aide de la méthode de classement à multiples clés recommandée par la norme ISO/CEI-14651. Des exemples concrets de mise en œuvre de cette méthode pour le lingala et le haussa montrent ses avantages et, aussi, sa limite.

Termes-clés:

banque de terminologie multilingue; traitement informatique de langues africaines; classement à multiples clés; ISO/CEI-14651; Unicode.

Introduction

LE PROJET BTML (banque de terminologie multilingue) de l'Office québécois de la langue française vise à concevoir et à créer une banque de terminologie qui facilite le partenariat international pour la production et le partage de terminologies multilingues et qui profite pleinement des nouvelles technologies de l'information et des communications telles que l'Internet, le multimédia, le traitement informatique des langues naturelles et le support des langues nationales. Selon les principes du véritable multilinguisme, une telle banque doit être capable non seulement de supporter les différentes langues mais en plus de respecter leurs caractéristiques propres et de permettre une grande souplesse aux partenaires. Pour mettre en œuvre ces principes, il a fallu résoudre des problèmes théoriques, tels que la définition d'une banque de terminologie multilingue, la relation entre les langues dans une telle banque, et des problèmes techniques, notamment ceux concernant le codage et les polices de caractères, le tri selon les particularités des langues, l'aide à la saisie des caractères spéciaux, la neutralisation de signes et de caractères non essentiels pour faciliter la recherche, la localisation des interfaces, etc., sans oublier la prise en compte des exigences politico-législatives propres à l'environnement où le système sera utilisé (par exemple, chaque administrateur du système doit pouvoir définir facilement une langue

permanente tout en assurant une égalité d'accès et de fonctions à toutes les langues traitées). Dans les lignes qui suivent, nous nous en tiendrons à une discussion sur quelques aspects des tris informatiques en différentes langues et nous présenterons notamment un bilan de nos tests pour le lingala et le haussa basés sur la norme ISO/CEI-14651 (2001).

Problématique du tri et principes de notre solution

Quand on parle du tri, le plus souvent, on pense à une simple mise en ordre des lettres d'un alphabet, comme s'il existait un ordre préétabli et transcendant. Mais, en réalité, le tri est beaucoup plus complexe que l'expression « de A à Z » ne le suggère. S'agissant du tri en informatique, la question se complique davantage en raison de la diversité des finalités, des destinataires et des environnements du traitement où s'inscrit l'opération du tri. Ainsi, le tri peut viser simplement à classer des éléments dans un ordre *quelconque*, à l'aide d'un algorithme déterministe pour que des opérations informatiques comme la comparaison et la recherche puissent s'effectuer de façon fiable, c'est-à-dire de façon à aboutir toujours aux mêmes résultats étant donné les mêmes conditions d'exécution. Ce genre de tri est destiné à la machine et l'aspect linguistique et culturel n'y compte guère. Tous ceux qui ont utilisé MS-DOS se rappellent ce programme de tri, *SORT*, qui classe les lettres et les autres signes purement et simplement selon la valeur numérique de leur code. Pour les 26 lettres de base de l'alphabet latin, les minuscules sont donc rangées après « Z » majuscule; quant aux caractères accentués, quand ils sont disponibles, ils sont tous placés après les lettres de base, sans aucune logique apparente quant à leur ordre. Ce tri fonctionne sans problème pour la machine, d'autant plus que la distribution des codes est plus ou moins faite pour optimiser le traitement informatique. Par exemple, pour les 26 lettres de base, un écart de 32 (en décimal) sépare les lettres en majuscule et leur correspondant en minuscule, ce qui permet de les convertir d'une casse à l'autre à l'aide d'une opération facile et rapide de OR ou XOR. Toutefois, si l'on doit utiliser le résultat de ce genre de tri pour afficher les données destinées à l'humain, l'absence totale du

respect des habitudes linguistiques et culturelles devient alors inconvenant, frustrant, voire révoltant. En effet, l'ordre dans lequel les locuteurs d'une communauté linguistique ont l'habitude de classer et de chercher les éléments de leur langue fait partie de leur manière d'utiliser cette langue, au point qu'une déviation peut être perçue comme une anomalie ou même un barbarisme, et peut devenir un obstacle qui empêche un usage efficace, cohérent et correct des résultats du traitement informatique. Ainsi, si quelqu'un cherche un mot dans une longue liste chevauchant plusieurs pages et s'attend à trouver classés de façon consécutive tous les mots commencés par «a», toutes casses confondues, avec ou sans signe diacritique, le résultat d'un tri selon la valeur décimale des codes risque de l'induire en erreur en le poussant à abandonner prématurément la recherche dès qu'il voit une entrée commencée par «B».

Or, effectuer les tris informatiques en tenant compte de la diversité linguistique et culturelle des destinataires humains n'est pas une tâche facile, du fait que le tri informatique se doit d'être absolument prévisible, déterministe et fiable alors que les habitudes humaines dans la réalité des langues naturelles ne sont pas toujours de cette rigueur. Pour l'alphabet latin, si la convention est plus ou moins bien établie quant à l'ordre du classement des caractères de base de A à Z, beaucoup de langues peuvent avoir des caractères particuliers qui ne se trouvent pas dans cet alphabet de base. Le français a ses caractères accentués et ses bigammes soudés, sans oublier d'autres signes de toutes sortes qui peuvent être utilisés dans son écriture. Même s'il existe des conventions non écrites suivies de façon plus ou moins rigoureuse par les grands dictionnaires, il n'est pas certain que les règles du tri soient connues et partagées par tous. Si l'on demande à des locuteurs francophones comment il faut classer des mots qui ne se distinguent que par des signes diacritiques à différentes positions dans les mots, ou comment il faut classer les mots précédés par des chiffres ou par des lettres grecques comme «α» et «β», il y a fort à parier qu'on obtiendra des réponses différentes. Le chinois pose un problème d'une autre nature, parce qu'il utilise un grand nombre de caractères idéographiques au lieu d'un alphabet ayant un nombre bien restreint de lettres. Force est de combiner plusieurs niveaux de critères. Il existe plusieurs méthodes différentes à cet effet : on peut commencer par un classement selon les radicaux pour ensuite ordonner les caractères d'un même radical selon le

nombre de traits composants, puis la prononciation, etc. ; on peut aussi ordonner les caractères d'abord selon les codes représentant les graphèmes des quatre coins du caractère pour ensuite considérer d'autres critères ; les dictionnaires modernes utilisent le plus souvent un classement qui considère d'abord la prononciation, puis le ton, ensuite le nombre de traits et, s'il y a lieu, le radical. Toutefois, quelle que soit la méthode combinatoire utilisée, il peut toujours y avoir une zone floue où des éléments ne se distinguent plus selon les critères appliqués. L'arbitraire et la variation trouvent alors leur place. Dans le cas d'une langue comme le coréen, qui a pourtant un alphabet restreint appelé *hangül*, les conventions du tri peuvent varier au niveau des lettres consonantiques ou vocaliques dites doubles, par exemple «ㄱㄱ» (kk) par rapport à «ㄱ» (k), ou «ㅇ» (ye) par rapport à «ㅇ» (e). Certains préfèrent les classer comme des éléments de tri indépendants l'un de l'autre, d'autres préfèrent les regrouper. On pourrait donc voir, à titre d'illustration schématique, des séquences de tri du genre «계 계 계 계» si on préfère séparer les consonnes simples et doubles mais regrouper les voyelles simples et doubles (ce qui correspond à la pratique la plus courante), ou «계 계 계 계» si on préfère l'inverse.

Adapter les tris informatiques aux destinataires humains, c'est aussi tenir compte de la diversité des besoins selon les contextes d'utilisation. On peut s'en tenir à des éléments de tri qui se distinguent par leur forme graphique seulement, ou opérer sur des éléments de tri qui sont des regroupements logiques basés sur des critères sémantiques, phonétiques, thématiques, typographiques, etc. Par exemple, selon les cas, on peut préférer classer «ph» comme «f» en se basant sur un critère phonétique, ou comme deux caractères séparés selon un critère purement orthographique. Pour le français, on peut vouloir classer ensemble les paires de mots comme «clé» et «clef», à titre de variantes d'un même élément, si on opère sur un niveau lexical, tout comme on peut regrouper «e, é, è, ê...» en les considérant comme des variantes d'un même caractère de base. Selon que l'on tienne compte ou non des conventions typographiques, on peut traiter «œ» et «æ» comme des séquences de caractères séparés, quand c'est écrit ainsi, ou appliquer systématiquement une analyse pour déterminer s'il s'agit de bigammes soudées «œ» et «æ» (cet exemple, présenté dans ce sens, concerne notamment la comparaison et la recherche, car au niveau du tri proprement dit, la question se pose le plus souvent dans l'autre sens : classer les

bigammes soudées comme deux caractères séparés ou comme un seul caractère à part).

Enfin, l'adaptation des tris informatiques aux destinataires humains doit faire face à un obstacle de taille : la diversité parfois déroutante des « cultures informatiques ». En effet, la grande variété des systèmes d'exploitation, les multiples jeux de caractères, les langages de programmation, les familles de SGBD (système de gestion de bases de données), les langages de script pour la publication sur le web (ASP, PHP, JavaScript, VBScript), les logiciels ou les progiciels d'application spécifique... ce sont autant de « fiefs » où il peut y avoir une « culture » spécifique ayant une influence favorable ou défavorable sur les tris informatiques. Ainsi, le monde humain, qui a créé le monde informatique, cherche maintenant à adapter celui-ci à ses besoins, sous peine de se faire aliéner de sa particularité linguistique et culturelle. À en croire Marc Wilhelm Küster (2000), la difficulté pour classer correctement en informatique les « ch », « ll » et « ñ » de l'espagnol en tant qu'éléments à part après les « c », « l » et « n » respectivement aurait même amené à réviser la norme traditionnelle du classement en espagnol pour l'adapter à la tyrannie informatique.

Ce tour d'horizon de la problématique du tri est de nature à nous indiquer quelques principes pragmatiques pour guider notre recherche de solution en matière de tri informatique adaptable aux besoins de diverses langues dans notre projet BTML.

– Le tri étant variable selon son utilité, son destinataire et son contexte d'application, il convient de bien délimiter les besoins du tri dans différentes parties de notre système afin de diminuer l'ampleur des difficultés et de conserver les caractéristiques particulières propres à chaque langue là où l'effet vaut la peine de notre effort. Plus précisément, il convient de concentrer notre effort sur l'adaptation du tri à différentes langues là où les destinataires humains doivent parcourir la liste des éléments linguistiques dont on suppose normalement qu'ils sont classés selon l'habitude de la langue en question. En revanche, dans des parties du système où la comparaison et la recherche à l'aide d'un tri informatique concernent seulement l'ordinateur, il est alors superflu de vouloir imposer la culture humaine. Par exemple, à un moment donné il nous a fallu exporter une longue liste de mots d'exclusion à partir du SGBD vers un programme en VBScript du côté du serveur web. Des comparaisons et des recherches très rapides doivent être

effectuées par l'ordinateur entre les mots clés entrés par les utilisateurs et les mots de cette liste. Pour cette raison, la liste doit être triée. Au niveau du SGBD, le programme du tri est adapté à l'aspect linguistique et culturel de chaque langue ; mais au niveau du VBScript, le tri par défaut suit des règles qui ne respectent pas les caractéristiques des langues humaines. Dans un cas pareil, vouloir imposer le respect des cultures humaines à VBScript semblerait une lutte don-quistotique.

– Pour une banque de terminologie multilingue, qui relève d'applications spécifiques, il convient de déterminer une procédure commune à toutes les langues et facile à mettre en œuvre pour adapter les règles du tri à divers besoins. Par contre, lorsqu'il existe plusieurs conventions de tri pour une même langue, il n'est pas nécessaire pour nous de les intégrer toutes dans le système. En fait, le tri n'est pas une fin en soi, mais un moyen pour faciliter la comparaison et la recherche. Or, les moyens de recherche exacte ou floue à l'aide de l'ordinateur deviennent de nos jours tellement évolués que si, en appliquant une méthode du tri parmi d'autres selon les particularités d'une langue, on ne peut toujours pas permettre aux utilisateurs de consulter facilement la liste des résultats, ce sera alors au niveau conceptuel du processus d'interrogation et de recherche qu'il y a lieu de mettre un peu plus d'effort.

Solution de tri basée sur la norme ISO/CEI-14651

Suivant l'analyse qui précède, nous avons décidé d'adopter les directives et le modèle de classement de la norme ISO/CEI-14651 comme cadre méthodologique pour adapter et mettre en œuvre les règles du tri selon les langues.

Pour résumer, cette norme remonte aux travaux d'Alain LaBonté¹, informaticien-conseil au Secrétariat du Conseil du trésor du Gouvernement du Québec. Au cours des

1 L'auteur de cet article voudrait exprimer son remerciement à Alain LaBonté pour avoir accepté de relire et de corriger les paragraphes concernant ses travaux et la norme ISO/CEI-14651 dont il a été le rédacteur. Les lecteurs intéressés

par la question du tri informatique en contexte multilingue sont invités à lire le texte original de la norme et les publications d'Alain pour obtenir des informations techniques plus détaillées et plus précises.

années 80, dans sa recherche d'une méthode de tri informatique absolument prévisible quant aux résultats et respectueuse de l'aspect linguistique et culturel du français, il détermina qu'il était nécessaire d'effectuer une décomposition à plusieurs niveaux de l'information pour en rendre le classement prévisible de manière absolue et, aussi, permettre d'établir un tri dont la précision pouvait varier selon les besoins. Pour intéresser les producteurs de logiciels, il essaya de démontrer qu'une telle méthode pourrait s'appliquer à plusieurs autres langues et il fit un parallèle avec le traitement à plusieurs niveaux utilisé dans les dictionnaires chinois. Il fit aussi un parallèle entre la structure informatique des nombres à virgule flottante, où l'on peut, en effectuant une comparaison uniquement sur l'exposant exprimé, classer les nombres selon leur ordre de grandeur, la mantisse venant au besoin ajouter plus ou moins de finesse au résultat des comparaisons...

Ainsi il a mis au point une méthode de classement qui permet de trier le français avec haute précision et grande souplesse. Sa méthode s'appuie sur l'utilisation de clés de comparaison à quatre niveaux définis selon leur degré de précision, à savoir, le niveau des caractères de base, celui des signes diacritiques, celui de la casse et enfin celui des autres signes distinctifs. Sauf le premier niveau, les autres sont optionnels et leurs clés peuvent faire l'objet de réduction selon certaines techniques, ce qui réduit la longueur des clés et le poids de stockage.

Ses travaux ont ensuite conduit à une norme canadienne du classement, CAN-CSA Z243.4.1, qui permet de traiter adéquatement à la fois l'anglais, le français, le portugais, l'allemand, l'italien et le néerlandais. Sous son instigation, le projet ISO/CEI-14651 fut créé en 1992 et il en a été le rédacteur. Après de longues années de travail ardu, souvent contre vents et marées, il a enfin réussi à faire émerger un consensus dans les pays membres de l'ISO et chez tous les géants incontournables de l'industrie informatique, autour de la nécessité d'une norme internationale du classement et autour de la pertinence de sa méthode de base. Celle-ci, après des années de révision, s'est améliorée et s'est élargie pour devenir une méthode universelle d'adaptation des règles de comparaison et de classement des données informatiques. La norme a finalement été adoptée en 2001. Selon une étude, cette norme est la douzième norme internationale de l'informatique la plus citée dans le monde (la première étant la norme du jeu universel de caractères

codés sur plusieurs octets (JUC), soit la norme ISO/CEI-10646, avec laquelle le développement d'Unicode est harmonisé).

Inutile de préciser qu'il ne s'agit point pour cette norme de fournir des règles du tri prêtes à appliquer à toutes les langues, mais bien d'une méthode d'adaptation uniforme, cohérente et facile d'emploi des règles du tri pour toutes les langues. La norme comporte en annexe quatre tables de symboles de tri (déjà classés) relevant des quatre niveaux sus-mentionnés, plus une table commune de classement, basé sur ces symboles, de tous les caractères du JUC. Ces symboles et leur classement tiennent compte des particularités des différentes écritures utilisées par les langues du monde et conviennent déjà à beaucoup de contextes courants pour des destinataires humains. Pour toute langue qui a des besoins spécifiques en tri, la norme fournit les directives pour élaborer un fichier delta qui répertorie les symboles, éléments et classement faisant l'objet d'adaptation à un niveau ou à un autre. L'adaptation, qui porte souvent sur un nombre limité de symboles et d'éléments pour une langue donnée (excepté le cas du chinois), peut naturellement bénéficier du modèle des tables offertes par la norme, de sorte que, lorsqu'on utilise Unicode pour construire une application à portée multilingue, il est possible de réaliser des tris informatiques qui soient à la fois adaptés à une langue en particulier et compatibles avec d'autres langues en général.

Dès lors, il nous a paru essentiel de trouver une façon d'appliquer la méthode d'adaptation de cette norme à notre système de façon à ce que par la suite, quand le besoin se ferait sentir, l'adaptation particulière pour une langue puisse s'effectuer isolément et facilement. Sur ce point, le travail se trouve grandement facilité par l'architecture globale de notre système dans laquelle chaque langue est traitée dans une base à part, tandis qu'une base centrale coordonne ce qui touche à toutes les bases. Ainsi, nous pouvons facilement instaurer soit des règles communes à plusieurs bases à l'aide de la base centrale, ou des règles spécifiques à une langue dans la base de celle-ci. L'application d'une méthode universelle et cohérente basée sur une norme internationale conduit dès le début à créer des programmes et fonctions partageables par toutes les langues concernées même si elles nécessitent des règles de tri différentes.

Ce qui suit a été relativement simple. D'une part, nous avons analysé les données et repéré les catégories de données qui peuvent être affichées aux utilisateurs sous forme de liste triée. Pour toutes ces catégories de données, nous avons créé des champs en double avec une longueur trois fois plus grande que la longueur initiale (selon nos tests, les clés ne dépassent pas cette longueur) pour y stocker les clés de tri. D'autre part, avec plus ou moins d'adaptation selon les cas, nous avons créé les tables des symboles de tri pour les quatre niveaux, fixé une méthode pour attribuer les poids aux symboles de tri de façon à ce que les valeurs des poids de chaque niveau ne chevauchent pas celles des autres niveaux, créé les tables (adaptées) du classement où chaque élément de tri possède quatre clés générées selon les poids des symboles de tri utilisés. Parallèlement à ces tables de classement, nous avons créé une fonction qui peut être appelée avec comme paramètres une chaîne de caractères et une indication du niveau et qui retourne la clé complète du niveau indiqué pour la chaîne en question. Si nous voulons un tri à deux niveaux, il suffit de concaténer les clés des deux niveaux ainsi obtenues, et ainsi de suite. Le cas échéant, nous pouvons même changer l'ordre des niveaux ou sauter certains niveaux intermédiaires pour des besoins spécifiques. S'il faut générer les clés d'un certain niveau avec un balayage à rebours, comme c'est le cas pour le français au deuxième niveau selon la norme canadienne du classement, il suffit d'inverser la chaîne de caractères en intrant ou la clé générée en extrant.

Une fois que nous avons établi la procédure d'adaptation des règles du tri selon la norme ISO/CEI-14651 et créé les programmes et fonctions nécessaires à la fabrication des clés de tri pour chacun des quatre niveaux, nous avons mené des tests sur le lingala et le hausa en collaboration avec les membres du LLACAN (Edema Atibakwa, Christian Chanard, Marcel Diki-Kidiri), pour fins de démonstration dans le cadre d'un projet du Rifal. Tous les tests ont été réalisés dans un environnement de Microsoft, avec Windows 2000 serveur, SQL 2000 et IIS 5. Ce choix s'explique par le simple fait que cette plate-forme supporte Unicode au niveau natif, ce qui simplifie beaucoup notre travail pour développer une application à portée multilingue (ce n'est pas pour rien que la norme JUC est la plus citée, comme nous l'avons mentionné plus haut).

Tests du tri pour le lingala et le hausa

Les informations linguistiques de base concernant les règles du tri en lingala et en hausa nous ont été fournies par Christian et Edema. En guise de résumé, voici la liste des caractères de ces deux langues, avec toutes les combinaisons possibles des signes diacritiques, que nous devons parvenir à classer dans l'ordre indiqué :

Pour le lingala

a à á â ã ä å b c d e é ê ë ÷ ε è é ê ÷ f i ï ï ÿ k l m mb n nd ŋ
ng j ny nz o ò ó ô õ ö ÷ ÷ ÷ ÷ ÷ p s t ts u ù

Pour le hausa

a à á â ã ä å b b̄ c d d̄ e ē ê ê̄ ÷ f fy g gw gy h i ï ï̄ ÿ j k
kw ky k̄ kw̄ ky l m n o ò ó ô õ ö ÷ ÷ ÷ ÷ ÷ p r r̄ s h t ts u ù û ü ù
w y 'y z

Ces caractères peuvent varier en casse, bien entendu, le majuscule devant précéder le minuscule selon l'information fournie.

Il existe plusieurs écoles de pensée en matière de conventions régissant le classement des caractères dans ces deux langues. Les uns préfèrent classer les caractères selon leur graphie, les autres selon leur valeur phonologique. Pour fins de test, nous avons choisi de suivre l'école qui classe les caractères selon leur valeur phonologique, car ce classement est plus exigeant pour le traitement informatique et convient mieux à nos tests. Notons que pour l'objectif de nos tests et dans l'état actuel du développement des normes de base concernant l'orthographe et les règles du tri de ces langues africaines, la question de savoir dans quelle mesure ces informations ont été validées par un consensus plus ou moins large n'est pas pertinente pour nous. Nous prenons pour acquis que lorsque vient le moment pour nous d'intégrer officiellement les règles du tri pour ces langues, la question de la validité des informations linguistiques de base aura été réglée en amont.

Ainsi, pour le lingala, les groupes de caractères « mb », « nd », « ny », « nz » et « ts », qui correspondent à des unités phonologiques à part, doivent être traités comme des éléments de tri simples et, dans un classement, doivent suivre respectivement « m* », « n* », « nd* », « ny* » et « t* ». Autrement dit, par exemple, « mba » doit être classé après « mza », ce que, pour le moment, aucun logiciel commercialisé sur le marché ne peut réaliser par défaut,

bien entendu. Il en est de même pour le hausa, dont les groupes de caractères «fy», «gw», «gy», «kw», «ky», «ƙw», «sh», «ts» et «'y» doivent être traités comme des éléments de tri à part et classés dans l'ordre indiqué par la liste montrée plus haut. Nous allons appeler ces groupes de caractères «bigammes» dans le reste du texte.

Mais les particularités des tris informatiques pour ces deux langues ne se limitent pas aux groupes de caractères qui doivent être traités à la manière de bigammes soudées et qui, naturellement, n'existent pas encore comme tels dans le répertoire d'Unicode. En effet, les nombreux signes diacritiques nécessitent aussi un traitement particulier. Dans certains cas, il s'agit de caractères composés à l'aide de signes combinatoires prévus à cet effet dans Unicode; dans d'autres, on voit des combinaisons de deux signes diacritiques sur un caractère de base.

Bigammes ou caractères composés, leur point commun pour nous, c'est que c'est à nous qu'il incombe de voir à ce qu'ils soient traités comme des éléments de tri soudés. À cet effet, dans un premier temps, il a fallu établir un fichier delta pour documenter les adaptations à effectuer à partir des tables modèles fournies par la norme.

En premier lieu, il a fallu définir autant de nouveaux éléments de tri qu'il y a de bigammes et de caractères composés à classer qui n'existent pas déjà dans le répertoire d'Unicode. Un élément de tri peut être simple ou complexe, c'est-à-dire composés de plusieurs caractères simples. Les caractères «Ä» et «ã» du hausa, par exemple, suivant les directives de la norme, peuvent être définis comme étant la combinaison des caractères de base «A» et «ã» avec le signe diacritique grave «`». On peut les noter comme suite dans le fichier delta :

```
% Lettre majuscule hausa a avec macron et grave
collating-element <U0100_0300> from "<U0100><U0300>"
% Lettre minuscule hausa a avec macron et grave
collating-element <U0101_0300> from "<U0101><U0300>"
```

Ensuite, il faut adapter les symboles de tri du premier niveau. Un symbole de tri peut être associé à plusieurs éléments de tri. Par exemple, le symbole de tri <S0061> est associé à tous les éléments de tri qui comportent le caractère de base «a», peu importe s'ils portent ou non des signes diacritiques et peu importe s'ils sont au minuscule ou au majuscule. C'est de cette façon qu'on peut obtenir un classement qui est indifférent aux signes diacritiques et aux casses. En l'occurrence, pour le lingala et le hausa, deux

approches sont possibles selon qu'on veut ou non pouvoir par la suite faire varier le degré de précision en jouant avec les symboles de tri du deuxième niveau qui concernent normalement les signes diacritiques. D'abord, on peut définir de nouveaux symboles de tri complexes pour les bigammes et les caractères composés, et on leur attribue des poids. Ainsi traités, ces bigammes et ces caractères composés propres à ces deux langues seront comme des caractères de base simples au point de vue du classement. Pour l'exemple de «A avec macron et grave» en hausa, suivant cette approche, on peut alors définir un nouveau symbole de tri minuscule qui va être utilisé pour les deux casses de ce caractère composé :

```
% Lettre minuscule hausa a avec macron et grave
collating-symbol <S0101_0300>
```

En revanche, suivant une deuxième approche, on peut ne pas définir de nouveaux symboles de tri au premier niveau mais, plutôt, on y associe des symboles de tri existants qui correspondent aux caractères de base qu'on peut extraire des caractères composés à traiter. Toujours pour l'exemple de

«A avec macron et grave», en hausa, on peut ainsi utiliser <S0061> («a») comme symbole de tri au premier niveau. Lors de la définition des clés complètes du classement pour ces caractères composés, on définit au deuxième niveau une clé qui est composée de deux symboles de tri correspondant aux deux signes diacritiques superposés :

```
% Réordonner après lettre majuscule latine a macron
% Lettre a macron et grave
<U0101_0300> <S0061>;" <BASE><MACRO><GRAVE>";
" <MIN><MIN><MIN>";<U0101_0300>
% Lettre A macron et grave
<U0100_0300> <S0061>;" <BASE><MACRO><GRAVE>";
" <CAP><MIN><MIN>";<U0100_0300>
```

Cette seconde approche est préférable à notre sens, car elle permet de tirer profit d'une méthode de tri à plusieurs niveaux. Par exemple, si on utilise seulement les clés du premier niveau, on réalisera un tri indifférent aux signes diacritiques; ajoutant les clés du deuxième niveau, et le tri devient alors sensible aux signes diacritiques. Pour nos tests sur le lingala et le hausa, par commodité, nous avons considéré les seconds caractères des bigammes au même titre que des signes diacritiques des caractères composés au deuxième niveau. Cette seconde approche peut donc nous donner le même type de souplesse pour ces bigammes que

pour les caractères accentués, à l'aide des clés du deuxième niveau.

Quant aux clés des deux derniers niveaux, celles du troisième niveau permettent de faire varier la précision du tri en rapport avec les casses, tandis que celles du quatrième niveau, elles, ont une utilité particulière: elles permettent de réaliser un tri informatique de façon absolument déterministe et prévisible peu importe comment on aura défini les clés aux trois premiers niveaux, car elles s'appuient sur des valeurs uniques correspondant aux positions respectives des éléments de tri dans la table du classement.

Pour les symboles de tri utilisés aux deuxième et troisième niveaux, ainsi que les tables de classement de ces symboles, nous avons simplement modifié l'ordre du classement de certains signes diacritiques (placer l'accent aigu après l'accent grave) et des deux casses (mettre la majuscule avant la minuscule), en suivant littéralement les informations fournies.

Dans le cas des bigammes qui doivent se placer en bloc après certains caractères, il s'agit d'insérer dans la table du classement, à la position voulue, les éléments de tri définis pour ces bigammes. Voici, à titre d'exemple, la structure complète du fichier delta pour le lingala, avec toutes les définitions concernant la bigamme «mb» à divers niveaux:

```
% Fichier DELTA pour le lingala
reorder-after <SFFFF>
order_start forward;forward;forward;forward
% copy ISO14651_2000_TABLE1
% Définition de symboles de tri ajoutés pour le lingala
%
% Digraphe minuscule lingala mb
collating-symbol <S006D_0062>
[...]
% Niveau 1: réordonner et/ou ajouter des symboles de tri
reorder-after <S006D>
% Digraphe minuscule lingala mb
<S006D_0062>
[...]
% Niveau 2: inverser l'ordre de <GRAVE> et <AIGUT>
reorder-after <GRAVE>
<AIGUT>
% Niveau 3: inverser l'ordre de <CAP> et <MIN>
reorder-after <CIRCLECAP>
<MIN>
<WIDE>
<COMPAT>
<FONT>
<CIRCLE>
% Définition d'éléments de tri ajoutés pour le lingala
[...]
% Digraphe minuscule lingala mb
```

```
collating-element <U006D_0062> from "<U006D><U0062>"
% Digraphe majuscule lingala mb
collating-element <U004D_0062> from "<U004D><U0062>"
[...]
% Ajouter les nouveaux éléments de tri dans la table de tri
[...]
reorder-after <U004D> <S006D>;<BASE>;<CAP>;<U004D>
% Digraphe minuscule lingala mb
<U006D_0062> <S006D_0062>;<BASE>;<MIN>;<U006D_0062>
% Digraphe majuscule lingala mb
<U004D_0062> <S006D_0062>;<BASE>;<CAP>;<U004D_0062>
[...]
reorder-end
% Fin du fichier DELTA pour le lingala
```

Ci-dessous sont présentés des exemples de résultat des tris réalisés après ces adaptations.

Exemple 1: Comparaison des tris avec ou sans adaptation

Adaptation	Classement des éléments	Éléments comptés
Aucune	mbayu	5
	muyu	4
	pingu	5
	pinu	4
Pour le lingala	muyu	4
	mbayu	4
	pinu	4
	pingu	4
Aucune	shuku	5
	sosho	4
	soso	4
	suku	4
Pour le hausa	soso	4
	sosho	4
	suku	4
	shuku	4

Exemple 2: Tri du lingala (interface BTML)

Rifal Mode de recherche

Tri alphabétique |<-| a| b| c| d| e| f| g|

Tri numérique |<| <<| >>| >|

196 [mpô ya mibalé](#) (1)

197 [mpô ya moyikinyí](#) (1)

198 [mundimami](#) (1)

199 [mwăna ya boténgú](#) (1)

200 [mbéba na momekano](#) (1)

201 [mbuta](#) (1)

202 [nkóla elongóbání](#) (1)

203 [nkóla ya etéyelo](#) (1)

204 [nkóla ya litéya lya engébéné](#) (1)

205 [nkóla ya liyékoli](#) (1)

206 [nkúta ya bontíndi](#) (1)

207 [nkúta ya kelási](#) (1)


208 [ntéi ya lóndo](#) (1)

209 [ndingisa](#) (1)

210 [ndingisa ya mayékoli libándá lya etéyelo](#) (1)

Exemple 3: Tri du hausa (interface GDT)

**Office québécois
de la langue
française**

Québec 

1. RECHERCHE

Langue Langue
d'interrogation de l'équivalent

haoussa français

Interrogation

s

par terme dans la définition

2. Index

- 📁 sājēwar aikin mak ařantā hōrō (1)
- 📁 sāmùn izinin kārātū waje da mak arantarsa
- 📁 sāmùn shìga
- 📁 sāti fik àt
- 📁 sōkewā
- 📁 shāgō hōrō (1)
- 📁 shāgòn hōrō
- 📁 shèk arār hōrō
- 📁 shèk arār kāmā aikī
- 📁 shèk arār kārātū
- 📁 shèk arār kārātū dà ak à mällak ā
- 📁 shèk àrun kārātū
- 📁 shūgāban mak ařantā

Remarques sur la limite de notre approche

À première vue, la mise en œuvre de la norme ISO/CEI-14651 semble donner des résultats assez satisfaisants. Par exemple, pour le français, notre programme du tri range correctement « école-clinique » entre « école centrale » et « école commerciale », ce que le tri de MS-SQL2000 ne permet de faire en aucun des modes prévus pour le français. Pour le lingala et le hausa, les bigammes sont classées correctement selon les règles indiquées. En plus, elles sont comptées comme des caractères simples par le programme du tri.

Toutefois, nous avons également rencontré des difficultés dues essentiellement à deux types de raisons.

Limite d'une méthode de tri « étrangère » aux systèmes natifs

D'abord, nous avons constaté la limite d'une méthode de tri dont la mise en œuvre demeure « étrangère » aux systèmes natifs (que ce soit le système d'exploitation ou le système de bases de données) sur lesquels l'application est développée et doit s'appuyer pour son fonctionnement. C'est que la réalisation des tris selon les règles spécifiques du classement pour une langue nécessite non seulement la mise en ordre de tous ses éléments, mais aussi la définition de ceux-ci. Les systèmes d'exploitation ou de gestion de bases de données supportent de façon native certaines définitions de certains éléments, permettant ainsi le choix, par exemple, entre plusieurs jeux de caractères, ou entre plusieurs types de classement pour une langue donnée. Lorsqu'un choix est fait, le fonctionnement tout entier de ces systèmes sous-jacents bascule vers les programmes et les ressources adéquats, garantissant ainsi que toutes les opérations du traitement informatique s'effectuent de façon cohérente en fonction du choix. Or, le traitement des langues comme le lingala et le hausa implique la définition *ad hoc* de certains éléments que les systèmes sous-jacents ne reconnaissent que dans la limite des programmes et fonctions que nous avons développés explicitement à cet effet. Hors de cette limite, point de salut.

Dans notre cas, nous pouvons certes afficher les résultats de recherche selon les règles du tri particulières des deux langues africaines testées, mais ce support

demeure imparfait, restreint et fragile. Par exemple, quand on voit « sh* » classé correctement après « so » pour le hausa dans l'index des résultats de recherche, il y a en fait déjà une petite anomalie (anodine, faut-il dire, pour notre objectif): la séquence « sh » n'étant définie comme une bigamme que dans la limite de nos programmes, les deux caractères composants « s » et « h » demeurent toujours deux entités séparées pour les systèmes sous-jacents et sont donc traités comme tels aussitôt qu'on sort de nos programmes et fonctions spécifiques du tri. C'est pour cette raison que lorsqu'on cherche tout ce qui commence par « s », on peut toujours trouver les séquences « sh* », ce qui ne devrait pas être le cas si « sh » était reconnu comme une entité à part entière et différente de « s ». Une autre version de cette anomalie peut être observée lorsqu'on cherche tout ce qui se trouve après « so »: cette fois-ci, les séquences « sh* » ne reviennent plus, ce qui représente une inconsistance par rapport à ce qu'on voit avec la recherche par « s ». Techniquement, et théoriquement, on sait comment on peut éliminer ce genre d'anomalies: il s'agit, d'une part, de remplacer systématiquement et partout les caractères à comparer et à chercher avec leurs clés de tri dûment générées selon les règles spécifiques et, d'autre part, d'effectuer systématiquement et partout la comparaison et la recherche non pas dans le champ original mais dans le champ correspondant qui contient les clés de tri. Toutefois, dans la pratique, cette solution n'est guère réalisable en dehors des systèmes natifs, parce qu'elle risque d'alourdir excessivement le traitement et est trop coûteuse en termes de ressources informatiques requises. Dans le fonctionnement d'une application, il y a beaucoup d'endroits où les opérations de comparaison et de recherche sont nécessaires; les systèmes sous-jacents mettent souvent en place des moyens d'optimisation et ce, au cœur même des systèmes, au niveau des codes machine, pour accélérer ces opérations. Appliquer systématiquement et partout une solution *ad hoc* de comparaison et de recherche signifie ni plus ni moins substituer un système étranger à un système natif. Plus une application est complexe, plus cette substitution sera difficile. Dans notre cas, pour le champ des entrées de la table principale d'interrogation, nous avons créé plusieurs champs de recherche où nous stockons les données ayant subi des traitements spécifiques afin d'offrir des méthodes de recherche optionnelles et performantes: neutralisation et substitution de certains

signes et caractères, inversion de la chaîne, nettoyage de balises HTML, etc. Pour garantir une performance satisfaisante lors de ces traitements, nous devons profiter pleinement des moyens d'optimisation offerts par les systèmes natifs. Si nous devons effectuer systématiquement la comparaison et la recherche avec les clés de tri spécifiques de chaque langue, nous devrions alors augmenter énormément le volume de données à traiter et alourdir considérablement chaque opération. Force nous est alors de trouver un équilibre entre des systèmes natifs performants mais incompatibles avec les besoins spécifiques, et une solution idéale pour ceux-ci mais impraticable tant que nous ne pouvons nous passer de ces systèmes natifs.

Complexité de la réalité

La norme ISO/CEI-14651 propose une méthode de classement basée sur une décomposition des données à plusieurs niveaux, allant du moins précis au plus précis. Par défaut, ces niveaux sont définis comme concernant les caractères de base, les signes diacritiques, les casses et tous les autres signes distinctifs. Au cours de nos tests, nous nous sommes rendu compte que dans la vie réelle, les problèmes à résoudre pour un traitement adéquat des données dans le cas d'une banque de terminologie d'envergure alimentée par de nombreuses personnes dans des conditions variées au cours de longues années, posent parfois des « colles » auxquelles ne peut suffire cette décomposition hiérarchique nette et claire en ces quatre niveaux. Plus précisément, il nous a paru que le principe véritablement essentiel et pertinent de la méthode, c'est la décomposition des données selon leur état et selon le degré de précision requis; en revanche, la nature de ces quatre niveaux et l'ordre dans lesquels ils sont proposés importent parfois peu. L'adaptation de la méthode sur ce point est d'autant plus nécessaire que nous devons surmonter les difficultés dues au manque de support natif à l'égard des règles spécifiques (voir plus haut). Un exemple concret que nous avons vécu fournit une illustration de ce point.

Pour l'anglais et le français, nous avons dans un premier temps utilisé des clés à quatre niveaux, en vue de pouvoir classer adéquatement les termes qui ne diffèrent que par la présence des signes de ponctuation ou des espaces intercalés mais qui doivent tout de même être traités,

comptés et affichés séparément en raison de leur association avec les différents domaines et en raison des fiches notionnelles dans lesquelles ils sont inscrits. En plus, il faut harmoniser les comptages effectués au niveau du serveur de données à l'aide des clés de tri et au niveau du programme d'interface qui n'a pas d'accès à ces clés pour des raisons d'optimisation. Citons, par exemple, « plate-forme, plate forme, plateforme », « fire-wall, fire wall, firewall », « ac, a.c. », etc. Par la suite, nous avons remarqué des problèmes liés à la casse: le fait d'inclure la clé du troisième niveau entraîne une distinction systématique dans le traitement des termes qui ne diffèrent que par la casse, partiellement ou entièrement. Cette distinction de casses a une conséquence au niveau du comptage et au niveau du produit de jointure des tables. L'analyse des cas montre qu'il y a trois situations possibles: 1) La différence des casses n'est pas pertinente, étant due à une erreur de saisie; 2) La différence est pertinente, parce qu'on veut consigner les variations orthographiques, mais les termes renvoient à la même notion (« Cegep » et « CEGEP »); 3) La différence est pertinente et les termes renvoient à des notions différentes (« bac » et « BAC »). Dans cette situation, nous ne pouvons pas simplement utiliser une fonction de conversion pour traiter toutes les données en minuscule ou en majuscule; en outre, nous ne pouvons pas non plus compter sur la possibilité de faire effectuer une correction massive et systématique dans les données à la source. Il ne nous reste plus qu'à essayer d'enlever la clé du troisième niveau tout en conservant celle du quatrième niveau. L'adaptation semble fonctionner, puisque l'indifférence aux casses convient aux situations 1) et 2), et que cette indifférence est compensée par l'unicité des identifiants des fiches notionnelles qui entrent en ligne de compte à un certain point de notre traitement. Suite à cela, nous avons remarqué qu'en fait, il n'y a qu'un tout petit nombre de signes qu'il nous faut prendre en considération au quatrième niveau pour satisfaire à nos besoins. Ce sont l'« espace », le « trait d'union-signes moins », le « point » et la « barre oblique », pour ne citer que ce que nous avons noté jusqu'ici. Inclure le traitement des clés du quatrième niveau juste pour ces signes ne nous paraît pas une solution économique. Nous avons donc décidé de définir des symboles de tri pour ces signes et de les prendre en compte au deuxième niveau. Étant donné que dans la table de classement du quatrième niveau proposée comme modèle

par la norme ISO/CEI-14651, ces signes sont classés avant les signes diacritiques, nous les avons donc insérés immédiatement après le symbole <BASE> au deuxième niveau. Cette adaptation fonctionne pour le moment à notre satisfaction.

En fait, l'adaptabilité de la démarche et des niveaux de la décomposition des données en vue de concevoir les clés adéquates selon les langues et selon les besoins est sous-entendue par la norme ISO/CEI-14651. Universelle, cette norme aurait du mal à se justifier comme telle si elle ne pouvait pas s'adapter et s'appliquer aux besoins d'une langue comme le chinois, par exemple. Or, la décomposition pour les caractères chinois doit se faire sur des niveaux complètement différents, tant par leur nature que par leur nombre, de ceux proposés par la norme pour les autres langues, le chinois n'ayant ni signes diacritiques ni casses. L'éclaircissement de ce point dans notre compréhension de la norme ISO/CEI-14651 en ce qui concerne l'adaptabilité de sa méthode de tri doit, à notre avis, permettre de mieux concevoir et adapter nos solutions selon les besoins et selon les langues.

Conclusion

Les difficultés pour adapter le tri informatique à l'humain s'expliquent par la diversité des besoins, la diversité linguistique et culturelle, la diversité des systèmes informatiques et, bien évidemment, la différence entre le mode de fonctionnement des ordinateurs et celui des êtres humains. Comprendre objectivement tous ces aspects de la problématique du tri constitue déjà un pas en avant dans la recherche d'une solution adéquate et efficace. Tout en maintenant notre principe du respect des caractéristiques propres aux langues traitées, nous pensons également qu'il importe de bien cerner les fonctions et les données pour lesquelles il y a lieu de mettre en œuvre spécifiquement les règles du tri selon les langues, car pour le moment nous ne pouvons le faire qu'en recourant à des méthodes «étrangères» aux systèmes natifs sous-jacents à notre application. Dans cette perspective, la norme ISO/CEI-14651 nous fournit une méthode d'adaptation universelle, cohérente et facile à implanter. Le résultat des tests sur le lingala et sur le hausa nous a confirmé la possibilité de

l'utilisation de cette norme dans notre projet BTML et, en même temps, nous a permis de prendre conscience de certaines limites. Sur le plan technique, en dernière analyse, le mécanisme le plus important d'un tri informatique adaptable à la diversité linguistique et culturelle, c'est de décomposer adéquatement les données à classer pour construire des clés en plusieurs niveaux allant du moins précis au plus précis. Le contexte de développement d'une application peut imposer parfois des difficultés imprévisibles; suivre une approche pragmatique nous semble donc de mise pour pouvoir réaliser des tris informatiques à la fois prévisibles, fiables et adaptables à diverses langues. Il va de soi que toute réalisation du tri informatique doit s'appuyer sur des informations précises en amont quant à la description linguistique des éléments à classer.

Jian Yang,
Chef de projet BTML, Office québécois de la langue française,
Québec.
jian@oqlf.gouv.qc.ca

Bibliographie

LaBonté (A.), 1988:
– Règles du classement alphabétique en langue française et procédure informatisée pour le tri, Québec: Publications du Québec.

LaBonté (A.), 1989:
Technique de réduction – Tris informatiques à quatre clés, Québec: Publications du Québec.
Note: Ces deux publications ont été revues et corrigées par l'auteur en 1998 pour leur version électronique publiée à <http://www.tresor.gouv.qc.ca/doc/classm.htm> et à <http://www.tresor.gouv.qc.ca/doc/techtri.htm> respectivement.

ISO/IEC 14651, 2001:
Technologies de l'information – Classement international et comparaison de chaînes de caractères – Méthode de comparaison de chaînes de caractères et description du modèle commun et adaptable de classement

Küster (M. W.), 2000:
«Developing European Ordering Rules – The basics of a new sorting standard from A to Ω to å, æ and beyond», dans *MultiLingual computing & Technology*, Volume 11 Issue 5, July/August 2000, p. 33-36.

Les caractères spéciaux dans l'orthographe courante : le cas du malgache

En décrivant le rôle que jouent les caractères spéciaux tant au niveau sémantique que syntaxique, et les problèmes connexes, nous avons voulu montrer tout simplement qu'une simplification excessive de l'écriture de la langue, due à des problèmes d'informatisation de celle-ci, ne peut que l'appauvrir.

Termes-clés :

caractères spéciaux ; orthographes ; informatisation de la langue ; Madagascar.

LE MALGACHE, langue nationale de Madagascar, écrit actuellement en caractères latins, présente plusieurs variétés régionales. Nous n'utilisons, en tout, que sept caractères spéciaux. Et c'est pourquoi le malgache fait partie de ces langues qui ont assimilé ses caractères accentués spéciaux aux formes standard les plus proches. En malgache, on utilise, par exemple :

- le [ÿ] i grec accent grave en position finale, qui prend la forme d'un [i] i minuscule accent grave en position médiane. Il est assimilé à [y] i grec sans accent en position finale, et à un [i] i minuscule sans accent en position médiane.
- le [ñ] vélaire et [n̄] palatale sont assimilés à [n] dentale.
- le [ŝ] s accent circonflexe qui note une palatale est assimilé au [s] de la dentale.
- la glottale [ʔ] est assimilée au point d'interrogation [?].

Dès lors, on peut se demander si le fait d'ignorer ainsi les caractères spéciaux constitue un avantage ou un handicap pour la promotion du malgache ?

D'une part, cela constitue un avantage, car à cause de cela, le malgache fait partie de ces langues d'Afrique qui n'ont aucun problème de mise en ligne quant à l'encodage de la langue, puisqu'il suffit de choisir le français. C'est pourquoi, depuis longtemps déjà, communiquer en courrier électronique entièrement en malgache ne pose pas de problème. Créer un site bilingue ou entièrement en malgache ne pose pas de problème non plus dès lors que l'on a acquis les connaissances techniques requises. Ceci explique le nombre assez élevé des sites bilingues français-malgache à la disposition des internautes, des sites tels que :

- <http://serasera.org>, site et forum très utilisés par la diaspora malgache de France ;
- <http://www.encyclopediemalgache.org> ;
- <http://dico.malgache.free.fr>, un site destiné à apprendre le malgache aux étrangers ;

– <http://www.teny.org>, notre site bilingue français-malgache ;
et beaucoup d'autres.

Cette option est un avantage car l'Internet facilite l'acquisition des connaissances, permet une communication rapide et l'évolution de la langue, les termes se créant au fur et à mesure des utilisations qui en sont faites. Actuellement, converser en ligne en malgache pour nos étudiants est devenu une distraction ; le développement des inforoutes, surtout dans le domaine de l'éducation et de la formation, nous rend optimistes quant à l'avenir, si l'on admet que le développement des nouvelles technologies de l'information et de la communication facilite le développement technologique tout court.

Mais d'autre part, ignorer ces caractères spéciaux constitue un handicap considérable, et nous en avons fait l'expérience. Comme on le sait, l'histoire de l'écriture du malgache est liée à la pénétration, sur la côte Est de Madagascar, des Arabes islamisés qui y ont établi des comptoirs, favorisant un syncrétisme communicationnel, où l'écriture du malgache était une adaptation de l'écriture arabe. Cette toute première écriture du malgache fut appelée l'arabico-malgache. Mais comme cette forme d'écriture n'a été utilisée que par les communautés vivant sur les côtes Sud et Est, et seulement dans le but de retracer les événements de la vie quotidienne que l'on appelle « tantaran-drazana », ce domaine d'utilisation restreint, avec les croyances y afférentes, l'a rendu difficile d'accès et l'a condamnée à n'être utilisée que par les « Katibo », c'est à dire les gardiens de la tradition. Ainsi, vers 1810, le Roi Radama I^{er}, avec le concours des missionnaires français et anglais, procéda à la première réforme de l'orthographe du malgache en mettant au point l'orthographe du parler merina. Il prit au français et à l'anglais le matériau nécessaire pour écrire ce qu'il institua comme étant le malgache officiel. Mais bien qu'inspirée du français et de l'anglais, cette orthographe du malgache officiel, écrit désormais en caractères latins, utilise encore quelques caractères accentués spéciaux qui jouent un rôle très important tant sur le plan sémantique que syntaxique :

- L'y accent grave [ÿ] par exemple joue un rôle discriminatoire pour deux mots ou deux énoncés de sens différents qui s'écrivent de la même façon à l'accent près, comme illustré ci-après.

– *atỳ* avec un [ỳ] accent grave signifie : « ici » ; par contre, *aty* sans accent signifie « le foie »
 – *arỳ* avec [ỳ] accent grave est un substitut locatif signifiant « là-bas » ; par contre *ary* sans accent est une conjonction de coordination équivalent à « et » ;
 – *misỳ barika ao amin'io tanàna io* « il y a une épidémie dans ce village » et
 – *misy barika ao amin'io tanàna io* « il y a un tonneau dans ce village ».

Sans le caractère accentué, il y a une ambiguïté au niveau sémantique, parfois même, le terme perd son sens ; c'est ce qu'on a dans l'exemple suivant : *lalambỳ* « chemin de fer » ; mais, écrit sans le caractère accentué, le mot *lalambly* ne veut rien dire. En outre, comme il n'y a que deux modes en malgache, l'indicatif et l'impératif, d'une façon générale, à chaque forme indicative correspond une forme impérative marquée par le suffixe « -a ». Or, il arrive souvent que, dans les énoncés, ce suffixe tombe en finale et ne se prononce pas (ou ne soit pas audible), de sorte que l'impératif ne se distingue de l'indicatif que par le déplacement de l'accent tonique provoqué par la suffixation du morphème d'impératif. Ainsi, sans les caractères accentués [ì] / [ỳ] et [ò] on ne saurait faire la différence entre l'énoncé suivant au mode indicatif :

– *mitsimbina ny mahantra ianao* : « vous prenez soin des pauvres »
 et l'énoncé au mode impératif :
 – *mitsimbina ny mahantra ianao* : « prenez soin des pauvres »
 de même pour l'énoncé de mode indicatif :
avalony ny akanjony : « il range ses vêtements »,
 et l'impératif :
avalòny ny akanjony : « rangez ses vêtements ».

Au début, comme tout se faisait manuellement, l'écriture des accents ne posait pas de problème, les supports écrits étant encore limités. Mais avec l'utilisation de la machine à écrire sur laquelle ces caractères spéciaux ne figurent pas, ils ont été assimilés aux caractères standard non accentués les plus proches.

D'autre part ce malgache officiel n'était pas en mesure de mettre en relief la différence entre les énoncés du type :
Manasa ny antsiny eo amoron-drano izy : « il lave son couteau au bord de la rivière » et
Manasa ny antsiny eo amoron-drano izy : « il aigüise son couteau au bord de la rivière »

et les mots construits comme :

manembana « voler » et *manembana* « déranger »

alors que d'autres variétés du malgache le font aisément en utilisant les caractères spéciaux, en tant que variantes graphiques :

– [ñ] n vélaire devant un radical commençant par [e] ou [i]
 – [ñ] n palatal devant un radical commençant par un [a] ou un [o], à côté de la dentale [n] que l'on utilise devant un radical commençant par une consonne ;

comme dans l'énoncé qui suit :

mañ-asa ny antsiny eo amoron-drano izy : « il aigüise son couteau au bord de la rivière » ;

man- {s}asa ny antsiny eo amoron-drano izy : « il lave son couteau au bord de la rivière » ;

et les termes construits comme :

mañ-embana « voler » et *man- {s}embana* « déranger »

Par ces lacunes sur le plan orthographique, qui ne lui permettent pas de décrire la réalité comme d'autres parlars peuvent le faire, ce malgache officiel n'a pas été très apprécié, si bien qu'après le soulèvement de 1972, qui réclamait la revalorisation de la langue malgache, la société a opté pour une seconde réforme orthographique avec la mise en place d'un « malgache commun », défini comme une forme standard élaborée à partir de tous les autres parlars, et qui doit être l'aboutissement d'un processus d'emprunt à tous les parlars, donc utilisant tous les caractères spéciaux, y compris la glottale [ʔ] jusque-là utilisée uniquement dans le parler antandroy, une communauté vivant dans le Sud et le Sud-Ouest de Madagascar. Cette glottale qui introduit toujours soit un complément secondaire soit un complément primaire ne peut être absente de l'énoncé où il devait être. Par exemple :
 – *Nibabe ? i Goa ty Kepeke* : « Kepeke porte Goa sur son dos », comparé à
 – **Nibabe i Goa ty Kepeke* (énoncé mal formé).

Mais bien qu'inscrits à l'examen officiel du BEPC depuis 1975, ces caractères spéciaux ont une difficulté à s'implanter du fait de leur absence sur le clavier, il fallait toujours les rajouter manuellement après la saisie. Par la suite, comme ce « malgache commun » était le résultat d'une réforme orthographique consécutive à la malgachisation, celle-ci ayant échoué, le malgache commun a été abandonné et les caractères spéciaux aussi. On est donc revenu rapidement à situation antérieure où les caractères spéciaux sont assimilés aux caractères standard les plus

proches. Mais cette fois, il s'en est suivi un laisser-aller qui a facilité un emprunt massif au français, ce qui ne peut que faire chuter la performance tant en malgache et qu'en français.

Face à cela, comme l'objectif du Centre des langues de l'Académie malgache est plus particulièrement le développement parallèle du malgache et du français, pour vraiment œuvrer dans ce sens, nous avons puisé dans les différents parlers les termes nécessaires pour enrichir le malgache.

Ainsi, dans l'élaboration des lexiques bilingues français-malgache sur l'environnement, pour prendre un exemple, il nous a fallu faire l'inventaire de toutes les espèces botaniques, et chaque espèce a déjà un nom dans les parlers régionaux que sont les différentes variétés du malgache. Il nous paraît donc erroné d'en inventer d'autres juste pour éviter les caractères spéciaux largement utilisés dans ces parlers. De même pour le dictionnaire monolingue malgache auquel le Centre s'attelle maintenant, en parallèle avec les autres lexiques bilingues, nous avons procédé à des collectes de termes dans ces parlers avant de recourir à des néologismes afin de lutter contre les emprunts massifs. Tout cela nous oblige à utiliser à nouveau ces caractères spéciaux. Dans cette perspective, le Centre compte donner l'exemple d'utilisation et amener petit à petit les locuteurs à les utiliser de nouveau. Il reste à bien préparer la diffusion sur tous les supports (papier, CD, Toile) pour répondre à tous les besoins.

Enfin, pour être complet, il faut rappeler qu'actuellement on n'utilise plus l'arabico-malgache si ce n'est pour faire de la recherche sur des documents anciens. Toutefois elle a laissée sa trace sur le système d'énumération qui se lit et s'écrit de droite à gauche (sauf dans le parler antandroy qui fonctionne toujours de gauche à droite). Avec l'informatisation de la langue ce système d'énumération pose des problèmes surtout quand il s'agit, par exemple, de faire entrer dans un téléphone portable le numéro d'un correspondant dicté par un tiers ou de saisir sur une machine à calculer ou sur un ordinateur, un nombre élevé comme: 12 353. Suivant l'ordre de diction des chiffres qui composent ce nombre, on aura 35 321 au lieu de 12 353. Il est vrai que des logiciels ont été déjà conçus pour surmonter ce type de problème, mais le problème dans le cas du malgache, le système d'écriture et le système d'énumération fonctionnent différemment, ce qui pousse les

locuteurs à faire une gymnastique intellectuelle avant de trouver le nombre juste.

Pour tenter de résoudre ce problème, on a décidé en 1982 d'étendre à tout le pays le système d'énumération du parler antandroy en l'introduisant dans le programme scolaire au primaire. Mais puisqu'il s'agit d'une habitude, d'une culture et d'une tradition, presque la totalité des établissements scolaires fonctionnaient de fait toujours avec l'ancien système d'énumération hérité de l'arabico-malgache. Ce qui fait qu'après deux ans d'expérimentation, on a dû abandonner la tentative de réforme. La solution pragmatique que les gens adoptent dans la vie quotidienne est de dicter systématiquement les nombres en malgache puis en français pour faciliter la saisie.

Tout cela nous amène à croire qu'une simplification excessive de l'écriture de la langue appauvrit la langue.

Oliva Ramavonirina,

Centre des langues de l'Académie malgache, Antananarivo, Madagascar.

olivaramavonirina.r@free.fr

Bibliographie

Bulletin de l'Académie malgache, 1977: numéro spécial du 75^e anniversaire (1902-1977)

Fugier (H.), 1999: *Syntaxe Malgache*. Peeters, Louvain-la-Neuve: p. 257.

Ludvig (M.), 1982: *La tradition arabico-malgache vue à travers les manuscrits A-6 d'Oslo et d'autres manuscrits disponibles*, Antananarivo: TPFLM.

Rakotosaona, 1972: *Rakibolana Antandroy*, Tome I, dactylographié.

Rajaona (S.), 1978: *Structure du malgache*, Fianarantsoa: Ambozontany.

Rajaona (S.); 1987: *Problème de morphologie malgache*, Fianarantsoa: Ambozontany.

Renard (R.), 2000: *Une éthique pour la francophonie: questions de politique linguistique*, Paris: Didier érudition.

Yacoub (J.), 2001: « Afrique noire: l'éternelle question linguistique », dans *Terminogramme: La protection internationale des minorités linguistiques*, n° 95-96, p. 237-245.

L'orthographe des langues de la République démocratique du Congo : entre usages et norme

Depuis l'époque coloniale jusqu'en 1974, beaucoup de consignes ont été promulguées pour asseoir un système orthographique harmonisé des langues congolaises qui en respecte les particularités. Seulement, cette norme reste d'un usage très limité faute d'avoir été enseignée et rendue obligatoire par une législation officielle, ce qui conduit à l'existence d'au moins deux usages orthographiques de fait parfois à l'intérieur d'une même langue. Cet article appelle à décréter une orthographe de droit.

Termes-clés :

langues congolaises ; orthographe ; législation linguistique ; norme ; usages.

Introduction

Il existe beaucoup de textes à caractère législatif qui font allusion à la politique linguistique depuis l'époque coloniale (1886, *cf.* annexe 1), jusqu'aux résolutions de la Conférence nationale souveraine en 1992. Ces textes ne parlent, de façon générale, que du *statut* et de l'usage officiel des langues parlées au Congo. Très peu se penchent sur leur normalisation, leur standardisation et la nécessité de leur instrumentalisation. Tout se passe comme si la chose était déjà réglée, la question de l'orthographe, par exemple, résolue.

C'est peut-être parce que l'orthographe peut paraître d'une banalité affligeante pour beaucoup. La chose n'est cependant pas si simple. En effet, le silence du législateur congolais en cette matière laisse croire, à tort, que l'orthographe est facile ou que tout le monde est d'accord sur sa forme.

Dans ce petit propos, nous allons montrer que sans un texte fort, à caractère coercitif, sans vulgarisation (enseignement, affiches), sans son officialisation, l'existence d'un système orthographique uniformisé par les linguistes (*cf.* annexe 3) peut donner naissance, non pas à une mais à plusieurs orthographe, les anciennes habitudes restant toujours vivaces (pire encore, des noms d'une même langue pouvant s'écrire différemment pour des raisons purement ethnocentriques ou confessionnelles ; par exemple ; *Tshimanga vs Cimanga ; Ngoy vs Ngoie*). Ce qui explique qu'au Congo l'orthographe va dans tous les sens et souvent pas dans le bon. Parfois on se trouve devant des graphies

parfaitement fantaisistes du genre *Z'abidi*, nom issu du swahili et qui, en tant que lexie, ne s'écrit nullement ainsi dans un dictionnaire swahili.

À la lumière de trois documents relatifs à l'orthographe des langues congolaises (*cf.* annexes 2 et 3), les seuls auxquels nous avons pu accéder, nous allons montrer que le silence du législateur en matière de réglementation de l'orthographe des langues congolaises n'est pas dû à l'absence de norme mais bien plus à la non-publicité et l'inapplication de celle-ci.

Entre l'homme politique et le linguiste, qui n'agit pas pour asseoir un système officiel et uniforme utilisable par tous ? Il semble que ce manque de collaboration « résulte dans certains cas de l'incapacité du linguiste à faire du politique, et dans d'autres, de celle du politique à faire du linguistique » (Halaoui 1996 : 3). Il serait plus que temps que linguiste et politique se mettent autour de la même table pour arrêter la fluctuation dont souffrent l'orthographe des langues congolaises.

Circulaire du 19 avril 1921¹

Le premier texte officiel fixant les règles orthographiques propres aux langues congolaises est la circulaire du 19 avril 1921 destinée sans doute aux Belges qui écrivent sur le Congo-belge car Bevel (1950 : 8) note que « bien des auteurs éprouvent des hésitations quant à l'orthographe à appliquer aux mots géographiques congolais ».

Cette circulaire, tout comme la suivante, ne concerne pas, à proprement parler, l'orthographe en langues congolaises dans tous leurs mécanismes comme l'a fait la commission chargée de la standardisation et de l'uniformisation de l'orthographe des langues congolaises (*cf.* annexe 3). Il serait donc inutile d'en critiquer certaines dispositions d'autant plus que la circulaire suivante (*cf.* annexe 2) et les recommandations de la commission en ont déjà corrigé les erreurs.

On soulignera juste que, malgré cette correction, (la révision de la consigne n° 2, *cf.* annexe 2) certaines erreurs dues au départ à une connaissance incomplète et insuffisante de la chose linguistique par les premiers administrateurs ou missionnaires (les langues congolaises ne sont pas encore

¹ *Cf.* Bevel (1950 : 7-8).

suffisamment étudiées) ont été maintenues. Fort de ce nouvelles connaissances, l'État (colonial ou indépendant) aurait pu revenir aux graphies correctes de certains toponymes comme :

- *Buta* issu, de *Mbota* (langue boa; espèce de serpent);
- *Gombe*, issu de *Ngombe* (langue teke; promontoire, butte);
- *Doruma*, issu de *Ndolomo* (langue zande; anthroponyme), etc.²

Circulaire n° 21/30 du 12 août 1953³

Il n'y a rien à dire sur cette circulaire.

Recommandation au premier séminaire des linguistes du Zaïre

Sortant de querelles souvent plus idéologiques que véritablement scientifiques, le premier séminaire des linguistes du Zaïre, tenu à Lubumbashi en 1974, après avoir entendu la Commission III, *chargée de la standardisation et de l'uniformisation de l'orthographe des quatre langues nationales zaïroises*⁴, avait proposé une orthographe commune et harmonisée des langues congolaises qui reprend en grande partie la tradition des missionnaires tout en l'adaptant aux données et connaissances linguistiques nouvelles.

2 Pour ne citer qu'un exemple, les Malgaches sont passés de l'ancienne Tananarive à l'actuelle Antananarivo.

3 Dans *Kongo Overzee*, Vol. XX, n° 1, 1954, p. 71-73.

4 La Commission note en introduction ce qui suit : « si les principes proposés s'appliquent avant tout aux quatre langues nationales cela ne signifie nullement que leur champ d'application se limite à elles seules. Chaque fois que cela a été possible, il a été tenu compte de faits particuliers existant en dehors de ces quatre langues

nationales de telle sorte que la transcription proposée ici puisse répondre à la plupart des situations qui se rencontrent dans les autres langues zaïroises » (CELTA 1974 : 148).

5 Composée d'éminents linguistes africanistes : Kadima Kamuleta, Jan Daeleman, Bokula Moiso, Bunduki, Rwakazina K., Tingbo nyi Zonga, Nkiko Munya Rugero, Nzakimwena.

6 Il suffit de parcourir par exemple les revues *Æquatoria*, *Kongo Overzee*, ou *Kongo ya Sika*.

Les propositions de la commission (travail colossal de 122 pages que nous ne pouvons reproduire entièrement ici), basées sur des descriptions complètes, phonétiques, phonologiques, morphologiques et syntaxiques d'une bonne partie des langues congolaises, sont restées méconnues du grand public, l'État congolais n'en ayant jamais tiré les conséquences pratiques pour asseoir un système orthographique officiel unifié.

Les membres de la commission⁵ constatent d'abord qu'il « existe à l'heure actuelle des transcriptions, pour plusieurs langues zaïroises, ayant servi dans l'enseignement. Ces transcriptions varient selon les langues, selon régions et selon qu'elles avaient été élaborées par des sociétés missionnaires catholiques ou protestantes.

Dans l'ensemble, elles peuvent être considérées comme assez heureuses dans la mesure où elles représentent directement l'état synchronique de la langue et où elles n'ont pas été trop influencées par des orthographes des langues étrangères, notamment du français.

Ces transcriptions traditionnelles restent néanmoins insuffisantes, incapables de représenter tous les faits de nos langues. On n'y trouve pas en effet les moyens de notation pour plusieurs traits fonctionnels, voyelles ouvertes dans les langues à 7 voyelles, certaines consonnes distinctives, tons, quantité etc. » (CELTA 1974 : 148)

Constat sans doute un peu injuste car dans de nombreuses publications de l'époque coloniale, les missionnaires, notamment les Pères Hulstaert et Everbrook, notent systématiquement les tons, les voyelles ouvertes ou longues, pour toutes les langues qu'ils étudient⁶.

Dans tous les cas, l'ensemble des problèmes de l'orthographe des langues congolaises dans tous leurs mécanismes est examiné et les questions sont résolues. Les principes de base sont assis et des propositions concrètes sont faites (cf. annexe 3).

Étant en pleine période de politique d'authenticité, les langues nationales entrent dans l'enseignement plus qu'avant et le CELTA produit des manuels sur l'enseignement des et en langues congolaises (uniquement les quatre langues nationales) qui tiennent compte de ces recommandations.

Absence de législation linguistique et variantes orthographiques des noms des personnes physiques

Malheureusement, du point de vue de l'usage administratif et officiel, ces recommandations sont restées lettre morte car elles ne constituent pas, jusqu'à ce jour, une référence pour tous les Congolais.

Quand on parcourt par exemple le texte de la loi 73-022 du 20 juillet 1973 relative au port du nom des personnes physiques, adoptée à l'époque par le Parlement, aucune mention n'est faite, de façon explicite, sur l'orthographe de ces noms et « post-noms » « authentiques » dont il est simplement dit qu'ils « doivent être trouvés exclusivement dans le patrimoine culturel » du pays (article 4 de la loi précitée.)

L'article 10 interdit « de changer de nom, d'en modifier l'orthographe » sauf autorisation du « tribunal de première instance du ressort de la dernière résidence du demandeur ». De quelle orthographe s'agit-il ? De celle donnée par le porteur ou de celle du législateur ?

L'article 12 autorise le titulaire du nom à user de l'action en justice « pour obliger les tiers à le respecter et à ne pas en user irrégulièrement ». D'accord, mais quelles sont les règles de cet usage ? Si j'écris *Ngoma*, au lieu de *N'goma*, quel texte de loi m'oblige à écrire le premier plutôt que le second ou vice versa ?

Quelle différence y a-t-il entre *Ngoy* et *Ngoie* ? Aucune. Le premier est de la province du Kasááyi (Kasaï), le second de la province du Katanga. Les deux dérivent de la même ethnie luba. Toutes ces orthographes du même nom sont enregistrées dans les documents officiels au Congo. D'où de nombreuses variantes graphiques des mots issus pourtant d'une même langue.

Conclusion

Avec tant de systèmes fonctionnant de façon parallèle, il n'est donc pas surprenant de constater que les lexies venant de langues locales présentent aujourd'hui encore autant de variantes orthographiques. On a ainsi *Kasai vs Kasaï vs Kasaayi vs Kasáayi*; *Tshitenge vs Citenge*, etc.

Si les caractères spéciaux sont d'ordinaire rigoureusement respectés dans les écrits scientifiques, dans les dictionnaires usuels du lingala et dans les manuels d'enseignement, ils demeurent totalement inconnus des usagers congolais. Il manque donc une officialisation, une publicité, une vulgarisation et un enseignement de ces notations.

Il est vrai, par ailleurs, que « *des réticences se manifestent souvent lorsque des innovations sont proposées dans le domaine de l'orthographe des langues zairoises. La raison avancée: il ne faut pas compliquer l'orthographe. Il y a cependant un fait qu'on doit reconnaître, c'est que les transcriptions de nos langues sont au départ très simples et très faciles et qu'elles doivent être innovées pour une plus grande efficacité. Le degré de difficultés qu'entraîneront ces innovations n'atteindra jamais celui de certaines autres langues comme le français et l'anglais. Des expériences menées dans des écoles secondaires ont montré, par exemple, que les élèves parvenaient vite à noter correctement les tons, considérés pourtant comme les traits les plus difficiles* » (CELTA 1974:149).

En résumé, il existe en République démocratique du Congo trois orthographes de fait; une première, héritée de l'époque coloniale avec ses erreurs, une deuxième du fait des usagers seuls, et une troisième conforme aux recommandations des linguistes. Toutes sont reconnues par l'État. Il est temps qu'une orthographe de droit soit décrétée.

*Edema Atibakwa-Baboya,
Cellule des langues congolaises, CELTA / LLACAN-CNRS,
Kinshasa - Villejuif, République démocratique du Congo - France.
edema_atibakwa@yahoo.fr*

Bibliographie

Bevel (M.-L.), 1950: *Le dictionnaire colonial (Encyclopédie)*, Bruxelles: Guyot.

Bursdens (A.), 1972: *La notation des langues négro-africaines. Signes typographiques à utiliser*, Bruxelles: Académie royale des sciences d'Outre-Mer.

CELTA, 1974: *Actes du 1^{er} Séminaire des linguistes du Zaïre*, Lubumbashi.

Edema (A.B.) & Nduku (K.), 1993: « Les variantes graphiques de l'IFA: un point de vue zaïrois », dans Latin (D.) & Tabi-Manga (J.), (éds), *Inventaires des usages de la francophonie. Nomenclature et méthodologies*, Paris: Aupelf-Uref et John Libbey, (Actualité scientifique), p. 355-372.

Halaoui (Nazam), 1996: *Politique et aménagement linguistiques*, Montréal: GRESLET.

Annexes

Annexe 1⁷

Répertoire des textes législatifs et réglementaires sur l'usage des langues en République démocratique du Congo.

- ordonnance du 14 mai 1886 de l'Administrateur général de l'État indépendant du Congo;
- décret du 4 mai 1881 (art. 21);
- circulaire du 6 août 1887 réglementant l'usage des langues dans la Force publique;
- circulaire n°41 du 01 juillet 1895;
- circulaire n°62 du 6 août 1895;
- convention entre l'E.I.C. et le Saint Siège du 26 mai 1906 (art. 6);
- conventions particulières entre la colonie et les congrégations (art. 3);
- charte coloniale du 18/10/1908 (art. 3);
- circulaire n°65 du 08/03/1910 relative au degré de connaissance des dialectes indigènes que devait posséder le personnel de la colonie;
- circulaire n°68 du 24 mai 1912;
- arrêté royal du 10 janvier 1922;
- texte créant la commission Franck;
- loi du 28 juin 1936;
- circulaire du 23 mars 1951 du ministre des Colonies;
- décret royal créant la Commission de linguistique africaine du 21-08-1950;
- décret du février 1957 (emploi simultané du français et du néerlandais);
- loi du 17/06/1960 sur les enquêtes parlementaires (art. 10 et 11);
- loi fondamentale du 17 à 1960.;
- constitution du 1^{er} août 1964 (art. 89);
- ordonnance n°174 du 17.10.1962 (suppression des langues nationales);
- résolutions du 1^{er} Congrès ordinaire du M.P.R. (1972);
- résolutions du 1^{er} séminaire des linguistes du Zaïre de 1974;
- circulaire du 31 août 1974.;
- 1982 Résolution du 3^e congrès ordinaire du M.P.R.;
- résolutions du colloque sur l'utilisation des langues nationales dans l'éducation et dans la vie socioculturelle de 1985;
- loi-cadre de l'enseignement de 1986 N° 86005 du 22-09-1986 (art. 120);
- constitution de la C.N.S. (art. 2);
- acte constitutionnel de la transition de 1994 (art. 1^{er});
- charte de l'éducation de la C.N.S. 51992 (art. 41 et 42, 43);
- décisions de la C.N.S. ; Acte sur la réhabilitation des langues nationales et leur utilisation dans le système éducatif.

⁷ Répertoire réuni par le Professeur Matumele Maliya du CELTA.

Actes portant intégration des langues nationale dans le système éducatif.

Différents programmes scolaires :

- programme de 1925 ;
- programme de 1929 ;
- programme de 1938 ;
- programme de 1948 ;
- réglementation scolaire de 1952 ;
- programme de 1958 ;
- programme de 1963 ;
- programme de 1974 ;
- programme de 1980 ;
- programme de 1984.
- Code de procédure pénale (art. 2 et 11) :
- circulaire de 1977 invitant les présidents des chambres pénales à résumer dans la langue nationale l'essentiel de la décision de justice rendue ;
- Circulaire de 1975 du 1^{er} Président de la cour suprême de justice créant un service d'accueil et d'information.
- Décret du 5/2/1957 introduit par [...]et néerlandais en matière judiciaire.
- Décret royal du 05/05/1881 relatif à la procédure devant le conseil Supérieur.

Annexe 2

a) Circulaire du 19 avril 1921, relative à l'orthographe des noms géographiques au Congo-belge

Principe de base

Le son véritable de chaque nom, tel qu'il est prononcé par les indigènes, sera pris comme base de l'orthographe.

Étant donné que la représentation phonétique parfaite de chaque son ne peut être obtenue au moyen des lettres de notre alphabet et des signes orthographiques, on s'appliquera à indiquer, le plus exactement possible, la prononciation locale avec les caractères ci-après :

1. les voyelles *a, e, i, o* et les consonnes *b, d, f, j, k, l, m, n, p, r, t, v, t, z* se prononceront comme en français ;

2. les voyelles *a* et *o* auront toujours le son bref comme dans « cas » et « coke ».

Exemples : Palabala, Lukolela. L'allongement d'une voyelle sera indiqué par un accent circonflexe. Exemples : Kitâmo, Malêla ;

3. *c* et *q* disparaîtront, comme faisant double emploi et seront remplacés par *k* ;

4. *e* (sans accent) aura toujours le son de l'*é* fermé en français. On l'emploiera également pour représenter le son *ai*, tel qu'il se prononce dans le mot « gai ». On écrira donc Bakange, Mokoange et non Bakanghe, Mokoangai ;

5. *g* aura toujours le son dur, comme dans *gare*, quelle que soit la voyelle qui suit immédiatement. On écrira Isangila, Giri, Bakange, Agenge, Zange, Kenge et l'on prononcera comme si l'on écrivait : Issanghui, Ghiri, Bakangai, Aguenghe, Zangai, Kenghé ;

6. *b* ne sera employé que dans l'articulation *sb* ;

7. *i* semi-voyelle, sera représenté par *y*, comme dans *yard*.

Exemples : Yambuya, Yalundi, Yakoya. Les sons *ya, ye, etc.*, ne seront jamais représentés par *ja, je, etc.* On écrira donc Bayanda, Bayeye, Bapeye, et non Bajanda, Bajeje, Bapeje. On ne terminera jamais un mot par *y* ; la voyelle *i* sera seule employée dans ce cas. Exemples : Noki, Dri ;

8. *j* sera employé uniquement pour représenter le son qu'il a dans le mot *jour*. Exemples : Bunji, Djabir ;

9. *ô* (avec un accent circonflexe) servira à représenter le son *au* tel qu'il se prononce dans *aube*. On écrira : Yôrembe, Bôra et non Yaurembe, Baura ;

10. *s* aura toujours le son sifflant comme dans *sinistre*. On écrira : Ysangi, Kasongo, Yakusu et l'on prononcera comme si on écrivait Issangui, Kassongo Yakussu ;

11. *u* représentera toujours le son *ou* français. Exemples : Rubi, Ubangi ;

12. *w*, semi-voyelle, se prononcera comme dans le mot anglais William. On aura soin de ne pas confondre les *wa, we, wi, etc.*, avec *ua, ue, ui*. En écrivant *wa, we, wi*, l'accent tombe entièrement sur les voyelles *a, e, i*, tandis qu'en écrivant *ua, ue, ui*, on représentera deux sons distincts dans chacune de ces diphtongues, et d'égale importance au point de vue de l'émission (*ua* prononcez *oua*, *ue* prononcez *ou-é*, *ui* prononcez *ou-i*). Exemples : Kalugwa, Niangwe, Kwilu, Muala, Kuela, Duizi ;

13. *x* ne sera jamais employé ;

14. l'articulation représentée en français par *ch* s'écrira *sh*. Exemples : Shonzo, Tshuapa ; prononcez comme si l'on écrivait Chonzo, Tchuapa ;

15. les sons doubles sont figurés par les lettres représentant les sons qui les composent. Exemples : Tshumbiri, Budja ;

16. deux voyelles juxtaposées se prononceront séparément. Ainsi, *ai* se prononcera *a-i* comme dans *mais* ; *ao* se prononcera *a-o* comme dans *cacao* ; *au* se prononcera *a-u* (*u* ayant le son *ou*) comme dans *raût* ; *ei* se prononcera *e-i* comme dans la dernière syllabe de *soleil*. On écrira Rafai, Kasai, et non Rafai, Kasai ;

17. *ph* ne sera jamais employé ;

18. toutes les lettres se prononceront ;

19. les voyelles ne seront doublées que lorsqu'il y aura deux sons distincts à représenter. Exemples : *Zuulu*, prononcez *Zou-oulou* ; *Oosila*, prononcez *O-osila* ;

20. les consonnes ne seront jamais doubles. Exemples : *Mangbetu*, *Bangaso*, *Kasai*, *Kobo* ;

21. les lettres *m* et *n*, que font sonner souvent les indigènes devant les mots commençant par des consonnes, seront négligées. On écrira, par conséquent, *Zobe*, au lieu de *N'Zobe*, *Doruma* au lieu de *N'Doruma*, etc. ;

22. l'emploi des accents et du tréma sera absolument banni, sauf en ce qui concerne l'accent circonflexe dans le cas indiqué au 2^o et au 9^o ;

23. à titre exceptionnel, on conservera l'orthographe de certains noms lorsqu'elle a été sacrée déjà par un long usage. Exemple : *Congo* ;

24. on n'ajoutera jamais de *s* au nom de peuplades comme marque de pluriel. On écrira donc, les Mombutu, les Bangala, les Ngombe.

b) Circulaire n° 21/30 du 12 août 1953, fixant les règles à suivre pour l'orthographe des noms géographiques du Congo

[Nous ne reproduisons que les dispositions modifiées, les autres étant restées les mêmes]

La Commission de linguistique africaine chargée par le département d'élaborer les règles concernant l'orthographe des noms géographiques congolais a édicté les dispositions suivantes qu'il y a lieu d'appliquer dorénavant à l'exclusion de toutes les autres.

Le son véritable de chaque nom, tel qu'il est prononcé par les indigènes, sera pris comme base de l'orthographe.

Étant donné que la représentation phonétique parfaite de chaque son ne peut être obtenue au moyen des lettres de notre alphabet et des signes orthographiques, on s'appliquera à indiquer, le plus exactement possible, la prononciation locale avec les caractères ci-après :

2. la différence entre voyelle brève et voyelle longue ne sera pas indiquée;
3. *c* et *q* ne seront pas employés pour noter le son *k*;
4. la différence entre *e* fermé (par exemple été, nez) et *e* ouvert (par exemple très, mais, net, même) ne sera pas indiquée;
6. la lettre *b* ne sera écrite que dans le groupe *sb* et dans les noms qui comportent une véritable aspiration (Yahuma, Kahemba, Kalehe);
9. la différence entre *o* fermé (par exemple chose, aube, beau) et *o* ouvert (par exemple tort, robe) ne sera pas indiquée;
10. *s* aura toujours le son sifflant de *s*, *c*, *ç*, dans servir, acier, garçon. On écrira: Ysangi, Kasongo, Yakusu et l'on prononcera comme si on écrivait Issangui, Kassongo, Yakussu;
11. *u* représentera toujours le son français *ou*. Exemples: Rubi, Ubangi, Uele;
15. les consonnes doubles seront figurées par les sons qui les composent. Exemples: Tshubiri, Budja;
16. deux voyelles juxtaposées se prononceront séparément. Ainsi *ai* se prononcera *a-i* comme dans maïs; *ao* se prononcera *a-o* comme dans cacao; *au* se prononcera *a-u* (*u* ayant le son *ou*) comme dans raôût; *ei* se prononcera *e-i* comme dans pléistocène, réitéré;
20. les consonnes ne seront pas doublées; on écrira donc pas *tt*, *ss*, *bb*. Exemples: Mangbetu, Bangaso, Kasai, Kobo;
21. des mots indigènes commencent souvent par *m* et *n* suivis d'une autre consonne; dans ce cas *m* et *n* font partie intégrante du mot et ils seront donc écrits sans employer l'apostrophe. Exemples: Mbao, Mpozo, Ndekeshu, Nko, Nsontin, (Nyanza, Mwangi);
22. l'emploi des accents et du tréma sera absolument banni;
23. a titre exceptionnel, on continuera à écrire «Congo» avec un *c*; toutefois, en néerlandais la graphie Kongo, avec *k*, sera adoptée;
24. on n'ajoutera jamais de *s* au nom de peuplades comme marque de pluriel. On écrira donc, les Mombutu, les Bangala, les Ngombe.

En outre, en vue d'éviter l'inconvénient des noms de localité qui peuvent être traduits dans les deux langues nationales et dans le but de maintenir l'intégrité du patrimoine culturel des populations autochtones auquel les noms géographiques appartient, on

s'abstiendra à l'avenir d'introduire des noms européens dans la toponymie congolaise.

Je rappelle enfin que par la dénomination néerlandaise de la Colonie, il y a lieu d'employer exclusivement l'orthographe «Belgisch-Kongo» la seule qui soit conforme aux règles orthographiques de la langue néerlandaise.

Les prescriptions de la circulaire n° 13 du 19 avril 1921 sont rapportées.

Annexe 3

Rapport de la commission III, chargée de la standardisation et de l'uniformisation de l'orthographe des quatre langues nationales zaïroises

(Extrait, CELTA, 1974, p. 263, *passim*)

La Commission pose un préalable sur les principes de base qui doivent orienter les propositions concrètes à émettre. Ces principes de base sont les suivants :

1. L'orthographe (écriture phonologique) doit être distinguée de la transcription (notation phonétique).
2. Le principe fondamental de toute orthographe qui se veut fonctionnelle est celui qui commande de représenter tout son distinctif par un seul signe.
3. Par réalisme cependant, certains traits phonétiques peuvent être représentés dans l'écriture. C'est ainsi que des allophones d'un même phonème peuvent être notés. Cela doit dépendre de la distance phonétique qui les sépare ou de la nature des contextes qui les déterminent.
4. Parce que l'orthographe est une convention qui doit viser avant tout l'efficacité dans la perspective de son apprentissage et de son utilisation par les locuteurs de la langue, la Commission a tenu compte des situations préexistantes; nous ne sommes donc pas partis de zéro. Nous avons, en partie, été d'accord avec ce qui existait déjà. Nous avons cependant apporté des modifications lorsque l'uniformisation l'imposait.
5. L'orthographe peut avoir une fonction morphologique; cependant, le niveau morphologique ne peut être représenté dans l'écriture que dans la mesure où cette représentation n'entraîne aucun inconvénient.
6. La transcription doit éviter de s'embarasser des considérations d'ordre diachronique; elle doit plutôt s'appliquer à l'état actuel de la langue.

Après avoir posé un préalable sur les principes de l'orthographe tels qu'énumérés ci-dessus, la commission a adopté les propositions concrètes suivantes :

I Les voyelles

1 Les voyelles orales

La commission maintient la graphie traditionnelle des cinq voyelles / i e a o u / pour les trois langues nationales (kiswahili, ciluba

et kikongo). Elle introduit deux nouveaux signes ε et \circ en lingala⁸ où il existe des oppositions distinctives entre e et ε et entre o et \circ . Autrement dit, le système vocalique définitif adopté par la commission III se présente comme suit :

i ε a \circ o u

où, dans les langues à cinq voyelles, ε et \circ ne seront pas mentionnés.

2 Les voyelles nasalisées

Les membres de la commission proposent que l'usage du tilde (~) soit adopté pour indiquer les voyelles nasales dans les langues où elles sont attestées.

II Les semi-voyelles

La commission propose que soit généralisée la notation directe par y, la semi-voyelle correspondant aux voyelles antérieures (i, e) et par w, la semi-voyelle correspondant aux voyelles postérieures (o, u).

III. Les consonnes

1 symbolisées par un signe simple

Pour la nasale bilabiale [m], la nasale dentale [n], les deux bilabiales [b] et [p], les deux dentales [d] et [t], les deux vélares [g] et [k], la latérale [l], la vibrante [r] les fricatives labio-dentales [v] et [f], les fricatives dentales [z] et [s], la commission propose le maintien des signes préexistants.

2 symbolisées par un digraphe

Etant donné que la Commission n'est pas partie de zéro, et s'est référée aux faits préexistants et que par ailleurs l'orthographe est une convention, la Commission propose :

a) que les palatales [ɲ] et [ʃ] restent respectivement représentées par les digraphes ny et sh.

b) qu'à la liste des consonnes soient ajoutées les labio-vélares [kp] et [gb], qui seront représentées par les mêmes signes dans l'orthographe courante (kp et gb).

c) la nasale vélaire [ŋ] jusqu'ici symbolisée par / ng / demeure symbolisée par ng dans les cas où il n'y a pas d'opposition distinctive entre cette nasale vélaire [ŋ] et le complexe [ng]. Par ailleurs, si cette opposition est attestée, la nasale vélaire sera symbolisée par ng', tandis que la séquence nasale vélaire (n-g) sera symbolisée par ng.

d) l'affriquée palatale sourde qui a été diversement présentée par *tsh*, *tch*, *ch*, sera dorénavant représentée par l'unique signe *c*⁹. La correspondance sonore reste symbolisée par le digraphe *dj*, lorsque dans une langue déterminée, il y a opposition entre [dä] et [ä]. Par contre, dans les langues dans lesquelles cette opposition n'est pas attestée, seul le symbole [j] est gardé. En conclusion la commission propose la graphie ci-après :

8 Nous rappelons que dans de nombreux écrits que nous avons consultés et datant au moins de 1940, les voyelles ouvertes et les tons sont déjà notés en lingala et en lomongo.

9 En raison de sa qualité de véhiculaire transnationale et d'une longue habitude, on conserve « ch » pour le kiswahili.

Voyelles et semi-voyelles

i	u
e	o
ε	\circ
a	
y	w

Consonnes simples

	Bil.	Lab.dent.	Dent.	Palat.	Vélaire.	Laryng.	Lab.vél.
Nasales	m		n	ny	ng(')		
Occlus.	b, p		d, t		g, k		gb, kp
Fricat.		v, f	z, s	j, sh		h	
Affr.			dj, c				
Latér.			l				
Vibrante			r				

IV La quantité vocalique

La commission recommande la notation de la quantité vocalique. Le redoublement de la voyelle reste la notation la plus pratique pour les voyelles longues. Toutefois, là où l'opposition de quantité vocalique est neutralisée, le voyelle longue sera représentée par le signe non-redoublé.

La quantité ultra-longue (notamment dans les idéophones et les onomatopées) est à représenter par le triplement du signe.

V Les tons

La commission recommande la notation des tons, vu leur valeur distinctive dans les langues africaines. Leur apprentissage au niveau primaire ne posera pas de problème, la méthode actuelle de l'enseignement étant plutôt globale qu'analytique.

1) Principe de notation

Le ton haut est symbolisé par l'accent aigu [´] et le ton bas par l'accent grave [˘]. Par contre, pour ne pas surcharger l'écriture, le ton le plus fréquent sera représenté par l'absence du signe. Toutefois, dans le cas de la nasale syllabique le ton (haut ou bas) sera toujours indiqué.

2) Tons complexes

La commission propose que soit adoptée la notation par l'accent antiflexe [˘˘] pour le ton montant, par l'accent circonflexe [ˆ] pour le ton descendant. Toutefois, sur les voyelles longues, non-neutralisées, les tons descendant et montant seront représentés par un accent aigu respectivement sur la première et la deuxième partie de la voyelle longue.

3) Tons moyens

La commission propose que le ton moyen dans les langues où il est attesté soit symbolisé par un trait vertical [˘˘] au-dessus de la voyelle.

VI La syllabation

1 La syllabe, dans les langues bantu, compte généralement un sommet vocalique, portant un ton dans les langues tonales. Elle se présente sous une des structures suivantes :

V	voyelle	a
CV	consonne-voyelle	na
NCV	nasale-consonne-voyelle	mba
CSV	consonne-sémi-voyelle-voyelle	bwa
NCSV	nasale-consonne-sémi-voyelle-voyelle	mbwa

Le découpage des mots, lors du passage d'une ligne à une autre, doit tenir compte de cette structure syllabique.

2. Par fidélité au réalisme phonétique, la nasale syllabique portera toujours son ton (haut ou bas), dans les langues tonales. Dans les langues non-tonales, une apostrophe placée, soit au-dessus de la nasale minuscule (ex. Mtu), soit après la nasale majuscule (ex. M'tu) servira à distinguer cette nasale syllabique de la nasale homorganique non syllabique

3. Les séquences vocaliques sont en nombre limité suite à l'importance de la contraction vocalique. La commission rappelle que les diphtongues n'existent pas dans les langues bantou; mais qu'on a plutôt affaire à des séquences de voyelles appartenant des syllabes différentes. En conséquence, en vertu de ce qui précède au point VI.1. relatif à la structure syllabique, la Commission propose :

1. les séquences vocaliques de deux ou trois voyelles (ex. en swahili: nimeiona « je l'ai vue (la chose) ») doivent être considérées comme respectivement de deux ou trois syllabes;

2. l'usage du tréma est à rejeter.

4 Séquences consonantiques

Quatre¹⁰ types de séquences sont courants dans les langues bantu.

C		m	n	ny	ng	b	p	d	t	g	k	gb	kp	v	f	z	s	j	sh	c	dj	l	h
NC		mm	nn	nny	nng	mb	mp	nd	nt	ng	nk	ngb	nkp	mv	mf	nz	ns	nj	nsh	nc	ndj		
CS	y	my	ny		ng'y	by	py	dy	ty	-? ¹¹	ky			vy	fy	zy	sy		-	cy	-	-?	-
	w	mw	nw	nyw	ng'w	bw	pw	dw	tw	-?	kw			vw	fw	zw	sw			cw		-	-
NCS	y	mmy*	-	-	-	mby	mpy	ndy	nty	ngy	nky	-	-	mv	mfy	nzy	nsy	-	-	-	-	-	-
	w	mmw	nnw	-	-	mbw	mpw	ndw	ntw	ngw	nkw	-	-	mvw	mfw	nzw	nsw	-	-	-	-	-	-

10 L'ouvrage note trois types mais il s'agit sans doute d'une erreur.

11 Nous mettons un point d'interrogation pour les séquences pour lesquelles nous pensons qu'elles existent bien.

Le kiswahili à l'épreuve de la modernité

Le lexique du kiswahili est le domaine qui reflète tous les efforts d'accès de cette langue à la modernité. Celle-ci, après avoir été symbolisée pendant des siècles par la culture orientale essentiellement arabe, est incarnée depuis plus d'un siècle dans le monde occidental. Recourant à différents procédés néologiques pour enrichir ce lexique, les acteurs coloniaux et plus tard africains ont réussi à transformer cette langue en un véritable moyen d'information. Même si aujourd'hui le kiswahili est apte à répondre aux défis que constituent l'usage des NTIC et les effets de la mondialisation, il est néanmoins confronté à un manque d'harmonisation. Celui-ci est issu principalement de l'étendue de son aire de diffusion et de l'hétérogénéité de ses divers paysages linguistiques et idéologiques.

Termes-clés :

kiswahili; néologie; terminologie; harmonisation; NTIC.

Introduction

Le kiswahili¹, comme toute langue vivante, a toujours puisé sa vitalité dans sa capacité de s'adapter aux nouvelles réalités chaque fois qu'il en était besoin. Ainsi la langue swahili, dans sa forme actuelle, est le résultat d'un long cheminement que se propose de revisiter cette étude, en remontant à son apparition sur la côte de l'Afrique orientale jusqu'à son actuelle diffusion internationale. Elle dégagera les différents processus d'enrichissement de son lexique, l'accueil des locuteurs, les problèmes rencontrés par les différents acteurs de cette planification, variables selon les époques et les contrées. Nous esquisserons sa situation actuelle à la lumière des NTIC et du phénomène de la mondialisation.

1 Le terme « kiswahili » sera exclusivement réservé à la langue tandis que « swahili » sera utilisé comme adjectif.

2 Famille à laquelle appartient la très grande majorité des langues parlées dans la partie comprise entre le Cameroun au nord, le Kenya à l'est et la république Sud Africaine au sud.

3 De l'arabe *sabil*, pluriel sawahil côte, rivage.

Bref rappel

Le kiswahili – faut-il le rappeler – est une langue africaine appartenant à la famille des langues bantu². Apparue sur la côte est-africaine au début du premier millénaire, elle est longtemps restée ancrée sur ces rives de l'océan Indien (actuellement Kenya et Tanzanie). Langue maternelle de populations locales dont l'activité principale était le commerce et la pêche, il facilitait également le contact entre ces dernières et les étrangers qui venaient y écouler leurs produits et s'approvisionner par la même occasion. Sa diffusion à l'intérieur du continent, assez récente, s'est effectuée dès la deuxième moitié du XIX^e siècle. Elle était liée dans un premier temps à la traite négrière et à l'essor économique de Zanzibar, véritable puissance régionale à l'époque. Son rayonnement ultérieur sera tributaire de l'activité coloniale et missionnaire, ainsi que des gouvernements nationaux.

Deux civilisations, l'orientale et l'occidentale, auront largement contribué à l'enrichissement tant culturel que social du monde swahili. Chacune servira de référence et de pont vers ce qui est considéré par son époque comme moderne.

1 Le contact avec la modernité

1.1 L'influence orientale

Les échanges entre la côte et les contrées du Moyen et Extrême-Orient (Arabie, Perse, Péninsule indienne et Chine) remontent au début du premier millénaire et n'ont cessé de se développer depuis. De toutes ces cultures, celle qui exercera une influence prépondérante et laissera le plus de traces dans la vie des populations swahili sera l'arabomusulmane. L'explication en est entre autres la proximité géographique.

Un coup d'œil au lexique du kiswahili témoigne de l'ampleur de cet apport, à commencer par le nom de la langue lui-même³. Cette influence, assez faible jusqu'au XV^e siècle, s'intensifie dès la fin du XVII^e siècle avec l'avènement de l'autorité omanaise qui venait d'évincer les Portugais de la côte.

L'emprunt est le seul procédé employé pour enrichir la langue et jusque-là il concernait surtout les domaines nouveaux comme la religion islamique et le commerce. Ce tournant annonce une transformation radicale de la société swahili; elle sera opérée au niveau social, administratif, culturel et aura pour seul modèle la civilisation arabe. C'est l'ère de l'*ustaarab*⁴. À l'époque, le kiswahili est déjà écrit en caractères arabes⁵.

Au niveau du lexique, l'emprunt connaît deux formes, celui fait par nécessité et un autre par « snobisme ». Si l'entrée dans la langue de termes en rapport avec le commerce comme *duka*, magasin, *deni*, dette, *sarafu*, monnaie, *bidhaa*, marchandise... paraît évident, l'usage de certains autres se superpose à ceux qui devaient exister. L'exemple est donné par le mot *samaki*, poisson, *bahari*, mer... Il serait étonnant pour un peuple riverain de l'océan de ne pas avoir de termes pour désigner ces deux notions.

Les effets de cette influence arabe se retrouvent également dans la littérature, le mode d'habillement et différents autres secteurs de la vie des habitants de la côte swahili.

Cette arabisation accélérée connaîtra son déclin avec la concurrence que constitueront l'ouverture de la région au monde occidental au milieu du XIX^e siècle et l'arrivée des premiers explorateurs et missionnaires. La page orientale s'effacera progressivement pour laisser place à une civilisation plus conquérante et de surcroît technologiquement supérieure.

1.2 L'influence occidentale

1.2.1 L'avènement de la colonisation

Bien avant l'arrivée des explorateurs et des missionnaires au XIX^e siècle, les Portugais avaient occupé la côte de 1498 à 1752. De cette période marquée par des conflits incessants avec les Arabo-Swahili, la langue n'aura gardé qu'une centaine de mots (Knappert 1989 : 49-52).

Contrairement aux Portugais qui se contentaient de la gestion de quelques comptoirs et de la sécurisation de la route des Indes, les Omanais, eux, allèrent plus loin dans la conquête de l'intérieur du continent. Une telle entreprise fut facilitée par le kiswahili. Celui-ci passera du stade de simple moyen d'expression à celui de véritable langue véhiculaire régionale.

Au fur et à mesure que l'œuvre coloniale et missionnaire prenait forme, le besoin de transformer cette langue en un véritable moyen d'information devenait plus qu'urgent. Pour la première fois, la question de sa planification se posait véritablement. Trois actions s'avéraient prioritaires pour améliorer le rendement qu'on pouvait tirer du kiswahili : le passage de la graphie arabe aux caractères latins, la modernisation de son lexique et sa standardisation qui serait basée sur le choix d'un dialecte⁶ à aménager.

Le premier point fut vite réglé et à la fin du XIX^e siècle, quelques écrits (dictionnaires, grammaires, journaux, catéchèses...) étaient déjà disponibles⁷, quoiqu'en distribution limitée, en Afrique orientale. Au Congo belge, la langue swahili était à la fois sujet et moyen d'enseignement comme l'attestent les manuels de latin utilisés dans les séminaires tenus par la Congrégation des Pères blancs⁸. Ils avaient élaboré une terminologie pour traduire les notions grammaticales latines⁹, leur lexique puisant dans les ressources déjà disponibles en kiswahili auxquelles s'ajoutaient quelques emprunts à la langue arabe surtout. L'influence de cette langue ne pouvait pas s'effacer du jour au lendemain, une certaine transition entre le monde oriental et occidental s'imposait.

Après la première guerre mondiale et le départ des Allemands, dans la partie sous tutelle allemande (actuels Tanganyika et Zanzibar), l'autorité coloniale laissera au protectorat britannique une situation linguistique assez

4 Civilisation à la manière arabe. (globe), du soleil, de la lune et des étoiles), *Jiolojia*, UMCA, Zanzibar 1901 (Géologie),

5 Le plus ancien manuscrit swahili attesté, le *Hamziyya*, date de 1728.

6 Le kiswahili compte une quinzaine de dialectes disséminés du sud de la Somalie au nord du Mozambique.

7 En guise d'exemples les journaux *Msimulizi* (1888), *L'Informateur*, *Habari za Mwezi* (1894) *Les Nouvelles du Mois*, *Puwani na Bara* (1910), À la Côte et sur le Continent et les manuels *Stronomia – Elimu ya Dunia na Jua na Mwezi na Nyota*, UMCA, Zanzibar 1901 (Astronomie, Science de la terre

Zanzibar 1901 (Géologie), *Lojiki; Mlango wa Filosofia 1*, UMCA, Zanzibar 1901 (La logique; Porte de la Philosophie).

8 Mission des Pères blancs, 1925, *Lexicographia-Kanuni za Maneno mbalimbali 1*, Albertville (Congo Belge) et *Grammatica Latina-Sarufi ya Kilatini, Syntaxis Kanuni za Maneno mbalimbali 2*, Albertville (Congo Belge).

9 *geuzo*, déclinaison, *kiungo*, conjonction, *kilio*, interjection, *tamko*, syllabe, *vocativo*, vocatif, *daraja ya ukuu*, superlatif...

normalisée. Le kiswahili avait le statut de langue officielle dans toute la *Deutsch Ost Afrika*. Il ne leur restait qu'à s'occuper principalement de la standardisation. Celle-ci ne fut pas aussi aisée que pour les deux actions précédentes. Deux dialectes, le kimvita, au nord, (Mombasa) et le kiunguja, au sud, (Zanzibar) se livreront une concurrence acharnée à travers leurs spécialistes et amateurs respectifs. Chacun des deux parlars avait ses atouts propres. Si une riche production littéraire et culturelle était associée au kimvita, il lui manquait cependant le rayonnement dont jouissait le kiunguja. Le trafic des matières premières et la traite avait porté celui-ci jusque dans la partie orientale de l'actuelle République démocratique du Congo. En revanche il ne bénéficiait d'aucun prestige culturel. Après plusieurs concertations, ce n'est que lors de la conférence de 1928 que le kiunguja l'emporta et fut sélectionné comme le dialecte à aménager, devant être utilisé comme langue de référence dans toute la région swahiliphone. *L'East African InterTerritorial Language Committee* (EAITLC)¹⁰ regroupant les représentants des quatre principales colonies swahiliphones (Kenya, Ouganda, Tanganyika et Zanzibar) fut créé et devint effectif dès le 1^{er} janvier 1930 avec comme premier siège Dar es Salaam.

Son objectif principal était de « promouvoir la standardisation et le développement de la langue swahili » à l'intention des quatre territoires (Whiteley 1969: 82). Un rapide parcours des treize objectifs que s'était fixé ce comité montre que la priorité était donnée au travail d'édition de livres scolaires et d'ouvrages à caractère littéraire. Les manuels en vigueur en Grande-Bretagne et la littérature populaire de l'époque en constituaient la seule référence. Il s'attela, avec une grande rapidité, à leur traduction en kiswahili. Cependant la tâche n'était pas facile dans la mesure où le swahili n'était pas préparé lexicographiquement. Et ceci constituait un handicap auquel ses membres n'avaient pas pensé car des treize objectifs, seulement un seul, le second, mentionne de façon allusive le travail lexicographique : « *Securing as far as possible uniformity in the use of existing and new words of the*

exercise of control over the publication of school and other dictionaries ». Pressés par le temps, les concepteurs de manuels se livreront à une traduction servile d'ouvrages britanniques, allant jusqu'à rendre littéralement en kiswahili des proverbes et maximes anglais. Le résultat fut une européanisation du kiswahili et de sa culture.

Si la future intelligentsia était familière à cette nouvelle langue, le décalage n'en était pas moins prononcé avec le commun des locuteurs swahili. Ils ne pouvaient s'identifier à cette nouvelle langue qui, à leurs oreilles, était artificielle. Ils ne tardèrent pas à l'affubler des noms de *kizungu*, la langue des Blancs¹¹, *kiserkali*, la langue du gouvernement ou tout simplement *kiswahili cha kiskuli*, le kiswahili scolaire. Même les locuteurs du kiunguja, pourtant noyau de cette « langue standard », ne suivaient plus et pour les nostalgiques de l'*ustaarab*, « *Kiswahili kilichovuliwa na kiarabu si kiswahili tena, ni kisbenzi* », le kiswahili dépouillé de l'arabe n'est plus du kiswahili, c'est un barbarisme (Ohly 1982: 73 et *sq.*). Même les responsables de cette planification reconnaîtront les défaillances de leur « bébé », comme en témoigne un mémorandum de l'époque cité par Whiteley (1969: 87): « *There is one other general point which must be noticed. Most of the Swahili contained in these books is correct grammatically, and may be defended on that ground. Grammatical accuracy, however, does not of itself constitute a language, and it is perhaps this very exaggerated application of grammatical rules that has led away from the real Swahili language, and made us substitute something which is at its best lifeless, though intelligible, at its worst both lifeless and unintelligible* ».

Le débat sur le kiswahili standard était lancé et ne s'arrêtera pas tout de suite comme l'indique la parution d'un livre pamphlet (Khalid 1977) près d'un demi-siècle plus tard.

Malgré les faiblesses méthodologiques, le comité aura eu le mérite d'avoir mené pendant les vingt premières années une vraie standardisation de la langue, en fixant les règles d'orthographe et de grammaire encore en vigueur de nos jours.

Pour pallier les insuffisances terminologiques du kiswahili¹², une rubrique consacrée aux termes techniques fut ajoutée au bulletin du comité à partir du n°13 de 1939. Le comité procédait de deux façons :

– soit en proposant des mots anglais avec leurs équivalents swahili sur lesquels devaient se prononcer les lecteurs comme on peut le lire dans une correspondance de Akida

10 Comité dans ce texte.

12 Surtout dans les domaines comme le commerce, l'administration coloniale et les sciences.

11 La participation effective des Africains aux réunions du comité n'aura lieu que seize ans après son instauration (Whiteley 1969: 82).

a) Dérivation :		
connection	<i>kiungo</i> (<- <i>unga</i> , unir)	
croissance	<i>ukuaji</i> (<- <i>kua</i> , grandir, croître)	
immigration	<i>uhamiaji</i> (<- <i>hama</i> , démenager)	
table de matières	<i>yaliyomo</i> (< <i>li+mo</i> , être à l'intérieur, litt. ce qui est dedans)	
b) Composition		
politicien	<i>mtu wa siasa</i> , (< <i>mtu+siasa</i> , homme+politique)	* <i>mwanasiasa</i>
supporter	<i>mwunga mkono</i> (< <i>kuunga+mkono</i> , unir+main)	* <i>mfuasi</i>
opposant	<i>mtu wa mjadala</i> (< <i>mtu+mjadala</i> , homme+discussion)	* <i>mpinzani</i>
c) Emprunt		
ampère	<i>ampere</i> (ampere)	* <i>ampea</i>
banque	<i>benki</i> (<bank)	
avance	<i>advansi</i> (<advance)	
bouchon	<i>plagi</i> (<plug)	
électricité	<i>stimu</i> (<steam)	
d) Transposition		
Chambre des Lords	<i>Baraza la Wakubwa</i> (<House of Lords)	* <i>inusicé</i>
citoyens	<i>wenyeji wa mji</i> (<habitants d'une ville)	* <i>wananchi</i>
régime républicain	<i>hukumu ya watu wengi</i> (<jugement de plusieurs personnes)	* <i>utawala wa kijamburi</i>
électorat	<i>Wenye haki ya kuchagua</i> (ceux qui ont le droit de voter)	* <i>wachaguzi</i>
circuit	<i>mzunguko kamili</i> (tour complet)	* <i>saketi</i>

(1958) qui réfute les termes proposés en s'appuyant sur des arguments linguistiques et culturels.

– soit en proposant une liste de mots anglais pour lesquels les lecteurs devaient créer eux-mêmes des équivalents swahili.

La procédure, quoique démocratique et animée de bonne volonté, baignait dans un empirisme absolu. Cette création était libre et ne suivait aucune recommandation ou règle précise. Et en fin de compte, seul le comité, sans beaucoup d'explications, détenait le dernier mot quant à l'approbation ou au refus du néologisme proposé.

Quelques exemples de l'époque permettent de dégager quatre principaux procédés de création auxquels il recourait. Les mots marqués d'un astérisque désignent le terme actuel (voir ci-dessus).

À partir des années 50, le travail du comité connut un net déclin en comparaison des résultats obtenus durant les

deux premières décennies. Ceci était dû au manque de moyens et à un désintérêt croissant du gouvernement colonial qui était confronté à un mouvement inéluctable de décolonisation. Son autorité fut également affaiblie par l'*East African Literature Bureau*¹³, pourtant son émanation, qui se mettra à éditer et diffuser des manuscrits sans son imprimatur (Mutahi 1986: 105).

1.2.2 Après les indépendances

L'avènement des indépendances correspond au déménagement du comité de Mombasa à Dar es Salaam en 1961. Cet événement aura l'une ou l'autre conséquence pour le kiswahili. Il annonce entre autres la disparition du comité tel qu'il avait été conçu initialement et sonne un coup d'arrêt à la standardisation. Le sort du kiswahili revenait entièrement aux mains des autochtones et chaque pays modèlera la langue selon ses aspirations politiques et socio-économiques.

¹³ Maison d'édition créée en 1948 et qui était chargée de diffuser des documents en langue swahili.

1.2.2.1 La Tanzanie

Suite à diverses raisons, la Tanzanie se retrouvera en première ligne dans le développement du kiswahili. En 1964, le comité fut transformé en *Tuki*¹⁴ et fut rattaché à l'université de Dar es Salaam. Cette institution avait pour mission de mener des recherches littéraires et linguistiques, dont la lexicographie et la terminologie. Ceci fut en soi une heureuse coïncidence dans la mesure où le kiswahili y était déjà enraciné et y jouait un rôle politique, économique et social de premier plan; la situation n'est pas pareille dans les autres pays de la région. Ainsi elle se retrouvait, pour ainsi dire, aux commandes de tout ce qui se rapportait au développement du kiswahili.

Déjà le 8 décembre 1962, jour de l'indépendance, le président Julius K. Nyerere avait prononcé son discours en kiswahili. Ce geste était comme un signe clair quant au rôle qu'il entendait assigner à cette langue dans le jeune état indépendant. En 1964, le poste de promoteur du kiswahili fut créé au sein du ministère du Développement communautaire et de la culture nationale. Un de ses rôles devait être la création de clubs de langue en milieu urbain (secteur public et privé) et rural. Au sein de différents services, des responsables furent nommés et chargés de veiller au bon usage de la langue.

L'année 1967 marque l'adoption de l'*Ujamaa*¹⁵ comme politique qui devait régir la Tanzanie. Une telle orientation prônant l'égalité entre tous les citoyens ne pouvait réussir que si elle était expliquée et pratiquée dans la langue comprise par tous et qui n'était autre que le kiswahili. Le gouvernement et les autorités déployèrent tous les moyens possibles pour lui permettre de jouer pleinement ce rôle. À cet effet, *Bakita (Baraza la Kiswahili Tanzania)*¹⁶ fut créé en août 1967 avec comme objectifs de :

– promouvoir le développement et l'usage du kiswahili à travers la République Unie de Tanzanie;

- coopérer avec les autres organismes dans le pays qui s'occupent de la promotion de la langue swahili et tenter de coordonner leurs activités;
- encourager l'usage de la langue swahili dans la gestion des affaires officielles et de la vie publique en général;
- encourager la réalisation de niveaux élevés (high standards) dans l'usage de la langue swahili et en décourager les mauvais usages;
- coopérer avec les institutions concernées pour mettre en place la traduction de termes techniques en kiswahili standard;
- publier un journal swahili consacré à la langue et la littérature swahili;
- offrir des services linguistiques au gouvernement, aux organismes publics et aux auteurs indépendants qui écrivent en kiswahili.

Mais à cette époque le kiswahili s'était déjà imposé dans toutes les activités officielles (correspondance, discours) et le recours à l'anglais ne se faisait qu'en cas de stricte nécessité. Une vraie culture nationale était née et une forme spécifique de kiswahili qui devait en servir de véhicule prenait progressivement forme.

Le lancement de l'*Ujamaa*¹⁷ permit au kiswahili d'enrichir son lexique surtout dans le domaine politique. Cependant tout ce travail ne suivait aucune règle établie ni ne se référait aux recommandations en la matière édictées par l'Unesco. Cette création, plutôt libre, se faisait dans un cadre volontariste et révolutionnaire. La contribution du président Nyerere fut capitale dans la définition de l'*Ujamaa* (Philippson 1970) et l'élaboration d'une vraie culture politique (Abdulaziz 1971 et Martin 1988). Son exemple fut suivi par ses collaborateurs qui ne cessaient d'apporter de nouveaux termes à la langue. *Bakita* et *Tuki* n'étaient pas en reste. Leurs bulletins sortaient de façon régulière des termes prêts à l'emploi et qui touchaient les domaines aussi variés que la politique, la biologie, l'agriculture, la psychologie (Akida 1971-1976; Gurnah 1974; Madalla 1972; Whiteley 1971 et 1974). Toutes ces initiatives visaient à équiper le kiswahili appelé à remplacer progressivement l'anglais comme langue d'enseignement et ce à tous les niveaux.

Les chercheurs de *Bakita* et *Tuki* ne changèrent pas sensiblement la méthodologie en vigueur du temps du comité et recoururent aux mêmes procédés que leurs prédécesseurs. Cet enthousiasme des années 70 fera place à

14 *Taasisi ya Uchunguzi wa Kiswahili*, Institut de recherches sur le kiswahili.

15 Famille, parenté en kiswahili. Nyerere proposera ce terme pour désigner le socialisme à l'africaine, différent d'*usobalisti*, socialisme de type occidental.

16 Conseil national du kiswahili.

17 Famille, parenté en kiswahili. Nyerere proposera ce terme pour désigner le socialisme à l'africaine, différent d'*usobalisti*, socialisme de type occidental.

une approche structurée et pluridisciplinaire dans les années 80. *Tuki*, avec le concours de l'Unesco, organisera des sessions où seront invités des spécialistes de disciplines concernées originaires de toute la région swahiliphone, appelés à donner leur avis sur les termes proposés. *Bakita*

avait recommandé une certaine hiérarchie des langues devant intervenir dans la création terminologique: le kiswahili, les langues bantu, les autres langues africaines et les langues non africaines. Les procédés de création les plus courants étaient les suivants :

a) Équivalence

Si le mot existe en kiswahili, il est repris pour traduire le terme technique proposé :

quantité	<i>kiasi</i>	indemnité	<i>fidia</i>
énergie	<i>nishati</i>	habitat	<i>makazi</i>
cœur	<i>moyo</i>	commerce	<i>biashara</i>
humus	<i>mboji</i>	nectar	<i>mbochi</i>

b) Composition

Partant de deux notions, ce procédé permettra de traduire ou de transposer le terme anglais en un seul terme swahili. Les combinaisons les plus exploitées sont les suivantes :

(i) nom + nom

<i>mwanamaji</i>	(<i>mwana maji</i> , enfant eau)	marin
<i>mwanasoka</i>	(<i>mwana soka</i> , enfant football)	footballeur
<i>mwanaanga</i>	(<i>mwana anga</i> , enfant ciel)	astronaute
<i>elimumaadili</i>	(<i>elimu maadili</i> , étude morale)	éthique
<i>elimumagonjwa</i>	(<i>elimu magonjwa</i> , étude maladie)	pathologie
<i>elimumsamiati</i>	(<i>elimu msamiati</i> , étude vocabulaire)	lexicographie
<i>nyumbataa</i>	(<i>nyumba taa</i> , maison lampe)	phare
<i>halijoto</i>	(<i>hali joto</i> , état chaleur)	température
<i>mcheduara</i>	(<i>mche duara</i> , prisme cercle)	cylindre
<i>msanifujenzi</i>	(<i>msanifu ujenzi</i> , designer construction)	architecte
<i>mchangukonuru</i>	(<i>mchanguko nuru</i> , désintégration lumière)	photodésintégration

(ii) nom + adjectif/numéral ou adjectif + nom

pembetatu	(<i>pembe tatu</i> , angles trois)	triangle
simbadume	(<i>simba dume</i> , lion mâle)	lion

(iii) verbe + nom

<i>mpandafarasi</i>	(<i>-panda farasi</i> , monter cheval)	cavalier
<i>kinzamimba</i>	(<i>-kinza mimba</i> , obstruer grossesse)	contraceptif
<i>pimajoto</i>	(<i>-pima joto</i> , mesurer chaleur)	thermomètre
<i>pimamvua</i>	(<i>-pima mvua</i> , mesurer pluie)	pluviomètre
<i>mfanyakazi</i>	(<i>-fanya kazi</i> , faire travail)	travailleur

c) Dérivation

Ce procédé est le plus productif dans la mesure où il a recours aux capacités internes du kiswahili, une langue flexionnelle où un mot peut donner plusieurs dérivés, par préfixation ou suffixation.

<i>kigawanyi</i>	(-gawanya, diviser)	diviseur
<i>kionyeshi</i>	(-onyesha, monter)	démonstratif
<i>kifaa</i>	(-faa, convenir)	matériel
<i>kielezo</i>	(-eleza, expliquer)	index
<i>kitenzi</i>	(-tenda, agir)	verbe
<i>ubaguzi</i>	(-bagua, séparer)	discrimination
<i>uapaji</i>	(-apa, jurer)	prestation de serment
<i>ubalozi</i>	(balozi, ambassadeur)	ambassade, consulat
<i>urais</i>	(rais, président)	présidence
<i>uburu</i>	(buru, libre)	liberté, indépendance
<i>ukubwa</i>	(kubwa, grand)	volume, capacité
<i>maendeleo</i>	(-endelea, continuer)	développement
<i>mawasiliano</i>	(-wasiliana, communiquer)	communication
<i>kizidisbo</i>	(-zidisba, multiplier)	multiple
<i>uzaliano</i>	(-zaa, engendrer)	reproduction
<i>kano</i>	(-kana, nier)	nihilisme
<i>shinikizo</i>	(-shinikiza)	pression
<i>ueneaji</i>	(-enea, se répandre)	distribution
<i>ubamaji</i>	(-bama, déménager)	migration
<i>upimaji</i>	(-pima, mesurer)	test
<i>upokezi</i>	(-pokea, recevoir)	réception
<i>utafiti</i>	(-tafuta, rechercher)	recherche

d) Apposition

Elle consiste à juxtaposer deux termes simples et parfois des propositions. Elle sert surtout dans la création des équivalents des adjectifs qualificatifs:

<i>ukaaji uchi</i>	(fait de vivre nudité)	naturisme
<i>ubusiano usawe</i>	(relation égalité)	relation équivalente
<i>fundi bomba</i>	(technicien tuyau)	plombier
<i>mfumo wa metriki</i>		système métrique
- a bali ya juu	(état de dessus)	supérieur, raffiné
- enye kasoro	(ayant un défaut)	défectueux
- sio ya serikali	(n'étant pas du gouvernement)	non gouvernemental
- iliyo ya ukweli	(étant de la vérité, réalité)	véritable
- silaha za sumu	(armes de poison)	armes chimiques

e) Extension sémantique

Ce procédé consiste à élargir le champ sémantique de mots existants :

	sens premier	néologisme
<i>kifaru</i>	rhinocéros	char d'assaut
<i>beberu</i>	bouc	impérialiste
<i>ujamaa</i>	parenté	socialisme
<i>bepari</i>	commerçant	capitaliste
<i>kupe</i>	tique	parasite
<i>tanzia</i>	annonce de décès	tragédie (théâtre)
<i>mzinga</i>	ruche	canon
<i>mfuko</i>	sac, poche	fonds
<i>tawi</i>	branche de l'arbre	agence, succursale

f) Contraction

Peu utilisée car ne découlant d'aucune règle précise et tendant à basculer parfois dans la fantaisie, cette méthode n'aura donné que quelques mots à la langue swahili.

<i>kidubini</i>	<i>KIjiDUdu kinachoonekana kwa msaada wa baduBINI</i> , petit insecte qu'on voit à l'aide du microscope, micro-organisme
<i>chajio</i>	<i>CHAkula cha JIOni</i> , repas du soir, dîner ou souper
<i>kasorobo</i>	<i>KASOro ROBO</i> , défaut quart, moins le quart
<i>bataza</i>	<i>HATi ya kukaTAZA kuiga uvumbuzi</i> , certificat pour interdire d'imiter une invention, brevet d'invention

La plupart de ces mots n'auront pas franchi le stade de proposition, le public ayant préféré les formes plus longues comme :

Mtu aSIye na KWAO, une personne sans chez lui, vagabond au lieu de *msikwao*
Vifaa vya kuLIA chakula, ustensiles pour manger la nourriture, couvert au lieu de *vilia*
DAwa ya waDUDU, médicament pour les insectes, insecticide au lieu de *dadudu*.
HUduma HAdi BANdarini, Free Of Board (FOB), au lieu de Huhaba.

18 Plus de cent langues bantu sont dénombrées en Tanzanie et au Kenya.

g) L'emprunt

Deux groupes de langues constituent la source des emprunts, d'une part les langues africaines et d'autre part les langues « étrangères ».

(i) Les langues africaines

Même si celles-ci sont vivement recommandées comme langues sources, elles n'auront grandement contribué à la terminologie swahili que de façon sporadique et marginale. Deux raisons expliquent cela : *primo* l'absence d'études sur ces langues et une tradition lexicographique inexistante ; *secundo* il ne serait pas facile de fixer les critères de la (des) langue(s) source(s) sans heurter la susceptibilité de la multitude des locuteurs dont les langues auraient été oubliées¹⁸.

Néanmoins quelques termes font désormais partie du vocabulaire courant swahili et les swahiliphones les ont assimilés sans aucun problème. La plupart du temps l'acceptabilité d'un terme est liée à l'influence de ceux qui les ont suggérés ou utilisés pour la première fois.

	Sens	Langue d'origine
<i>kabwela</i>	prolétaire	(kizanaki)
<i>Ikulu</i>	palais présidentiel	(kinyamwezi)
<i>Bunge</i>	Parlement	(kihaya)
<i>ngeli</i>	classe nominale	(")
<i>rovu</i>	goitre	(kipemba)
<i>uga</i>	court, terrain	(kizigua)
<i>fuwele</i>	cristal	(")
<i>kitivo</i>	faculté	(kipare)
<i>wawe</i>	chant d'agriculteurs au champ	(kiamu/kipate)
<i>ikari</i>	barbecue	(kimeru)
<i>ndigano</i>	sorte de fièvre côtière	(masai)
<i>mbuti</i>	duodénum	(")

Des tentatives pour adopter des mots issus de langues parlées dans d'autres régions d'Afrique n'ont pas abouti. Sur la liste d'attente se retrouvent :

<i>rara</i>	ballade	(yoruba)
<i>mkola</i>	kola	(hausa)

(ii) Les langues étrangères

– L'arabe

Ayant été pendant longtemps la référence par excellence pour la langue swahili, il garde une place de choix et a ses traces gravées dans son lexique. Cependant de nos jours, et comme naguère, il n'est parlé que par une poignée de lettrés musulmans et ne figure pas sur la liste des matières scolaires. Cette faible diffusion explique le peu d'engouement qu'il suscite auprès du grand public en général et des terminologues en particulier. L'art et la littérature restent les principaux domaines où il s'épanouit. Il intervient surtout dans les notions très académiques réservées à quelques initiés et dépassant rarement leur cercle.

<i>fasibi</i>	littérature	<i>sanaa</i>	art
<i>nabau</i>	expression idiomatique	<i>tamthilia</i>	théâtre
<i>tabakiki</i>	critiquet	<i>aswira</i>	image, imagerie
<i>sitiari</i>	métaphore	<i>riwaya</i>	roman

– L'anglais

Après l'indépendance et surtout durant la période de l'*Ujamaa*, l'emprunt à l'anglais était vécu comme une « trahison », un déni des capacités de la langue swahili de traduire les réalités modernes. Cette fierté et ce nationalisme linguistique s'opposaient au *kasumba*¹⁹ et favorisaient l'accès à la science pour tous dans un système qui préconisait l'égalité de tous ses citoyens. La fin de l'*Ujamaa*, peu après 1983, voit fléchir cette ardeur et consacre le retour de la langue de l'ancien colonisateur. Le rêve des partisans du kiswahili de le voir supplanter un jour l'anglais avait fait long feu. La diffusion et le développement de ce dernier sont soutenus par la puissance britannique et une distribution linguistique le consacre *de facto* comme langue de l'enseignement secondaire et

19 Aliénation culturelle.

20 *Elimu*, (éducation, connaissance) est un emprunt à l'arabe intégré et connu de tous.

supérieur, des hautes juridictions, de la science et de la technologie et des relations internationales. Le kiswahili se retrouve relégué dans l'enseignement primaire, l'alphabétisation, le domaine politique et social ainsi que la juridiction inférieure.

Au niveau terminologique, on assiste à un phénomène de déswahilisation :

Le kiswahili avait été privilégié pour la création de termes techniques et scientifiques. Ils seront abandonnés au profit d'emprunts à l'anglais. En effet, l'acceptabilité d'un nouveau terme par le public dépend de son sens et de sa dénotation. Quand le sens du terme ne correspond pas au concept qu'il dénote, le public le rejette (Mdee, 1980 : 94). Le terme doit s'inscrire dans l'imaginaire collectif des Tanzaniens. Ceci explique l'impopularité de la plupart des termes issus de la contraction (voir *supra*) et le succès dont jouissaient les termes suivants :

dynamique	<i>elimuwendo</i>	(étude marche)
hygiène	<i>elimusiba</i>	(étude santé)
sémantique	<i>elimaana/elimumaana</i> ²⁰	(étude sens)
sociologie	<i>elimujamii</i>	(étude société)
psychologie	<i>elimunafsi</i>	(étude âme)
gynécologie	<i>elimuuzazi</i>	(étude procréation)
cylindre	<i>mcheduara</i>	(prisme cercle)
condom	<i>mpira wa uume</i>	(caoutchouc pour masculinité)

Le retour de l'anglais entraîne une certaine déswahilisation et au grand dam des locuteurs swahiliphones, ces termes sont progressivement remplacés par *semantiksi*, *sosholojia*, *saikolojia*, *jinakolojia*, *silinda*, *kondomu*. Ce retour à l'anglais trouve son explication dans le fait que le jargon scientifique est quasi identique dans toutes les langues, les racines grecques ou latines étant sauvegardées.

<i>ikolojia</i>	écologie
<i>biolojia</i>	biologie
<i>heksgramu</i>	hexagramme
<i>herbaria</i>	herbarium
<i>hepatitisi</i>	hépatite
<i>asidi hidrokloria</i>	acide hydrochlorique
<i>kompyuta analojia</i>	ordinateur analogique
<i>biokemia</i>	biochimie

<i>oksijeni</i>	oxygène
<i>kondensa</i>	condenseur
<i>theolojia</i>	théologie
<i>mwentomolojia</i>	entomologiste

Malgré cela le kiswahili n'est pas totalement abandonné et reste incontournable dans certains néologismes hybrides anglais-kiswahili ou vice versa :

<i>Saketi wazi</i>	circuit ouvert
<i>namba mchanganyiko</i>	numéral composé
<i>numerali za kiarabu</i>	numéraux arabes
<i>elementi za kemikali</i>	éléments chimiques
<i>mbembeo themodainamia</i>	cycle thermodynamique
<i>mwelekeo kibarometa</i>	tendance barométrique
<i>mwatuko nyuklia</i>	fission nucléaire

viii) Calque et transposition

L'anglais sert de langue de départ pour les calques et transpositions.

UKIMWI: Ukosefu wa KInga MWilini, manque d'immunités dans le corps
AIDS: Acquired Immuno-Deficiency Syndrome
KUSOTU: KUMbukumbu ya SOma TU, ROM (Read Only Memory)
KUFINA: KUMbukumbu ya FIkiO NAsibu, RAM (Random Access Memory)
msihakamano wa umemetuli, adhésion électrostatique
mfumo wa mwasbo, système d'allumage
nguvufarasi, borse power

21 *Tafsiri sanifu*, traduction standard. Série de 5 fascicules contenant des termes nouveaux d'économie et commerce, administration, biologie... ; *Lugha yetu*, Notre langue, *Kiswahili Fasaha*, Le kiswahili correct, ouvrages sur le bon usage de la langue et *Lugha ya Taifa*, la langue nationale, un programme radiophonique hebdomadaire.

ala za matamsbi, organes de prononciation
mfereji wa chakula, canal alimentaire *i.e.* appareil digestif
nchi zisizoegemea upande wowote, les pays qui ne penchent d'aucun côté *i.e.* les pays non alignés
silaba za nyuklia, armes nucléaires
silaba za sumu, armes chimiques

Cette inspiration de l'anglais se reflète également dans les définitions scientifiques comme celle proposée pour le terme *iode*: *iodini: elementi simetali halojeni, ya rangi kijivunyeusi ambayo ni mango fuwelia inayovukamanga kuwa mvuke urujuani mzito. Hutumiwa kwenye mmumunyo wa alkoboli kama kiuaviini. Alama yake ni: I; uzani atomia: 126.904; namba atomia: 53; graviti mabususi (mango): 4.93 katika 20°C* (Irira 1995: 90).

Officiellement *Bakita* et *Tuki* continuent de s'occuper de la création terminologique et maintiennent leur collaboration même si elle n'est pas parfois totale comme le montrent les exemples issus pourtant de deux dictionnaires contemporains :

	Irira (1995)	Tuki (1996)
accumulateur	<i>betrimzo</i>	<i>kilimbikizi</i>
neutron	<i>nutroni</i>	<i>nyutroni</i>
ignition	<i>mwasbo</i>	<i>mwako</i>
hydrologie	<i>hidrolojia</i>	<i>baidrolojia</i>
hydrochlorique	<i>hidrokloria</i>	<i>-a baidrokloriki</i>
hyperbole	<i>hipabola</i>	<i>mbalagha</i>
power station	<i>kituo cha nguvu</i>	<i>kituo cha umeme</i>
incubateur	<i>tamikanza</i>	<i>kitamizi/kitangulio</i>

Bakita, après une décennie (1974-1985) relativement riche en publications²¹, semble aujourd'hui survivre difficilement à la disparition du système *Ujamaa* qui l'avait créé. Il ne s'est jamais départi de ce cachet politique qui l'a toujours caractérisé et qui ne manque pas d'incidence dans la gestion quotidienne du Conseil et qui l'a toujours caractérisé. Son activité se trouve ralentie par le manque de moyens alloués à la publication des nouvelles terminologies dont les nombreux manuscrits sommeillent sur les étagères du Conseil. En revanche, *Tuki* a repris le flambeau et publié de façon assez régulière dictionnaires (*Tuki* 1981, 1996, 2001) et lexiques kiswahili-anglais de langue et linguistique (*Tuki* 1990a), biologie, physique et chimie (*Tuki* 1990b), droit (Mlacha 1999) et kiswahili du parlement (Mwansoko et Tumbo-Masabo 1996).

Cependant ils subissent la concurrence d'autres institutions issues de la conjoncture nationale et internationale. En effet, l'effondrement de l'*Ujamaa* et l'avènement du multipartisme ont entraîné bien des mutations politico-économiques accompagnées de disparités sociales naguère discrètes. La libération des mentalités, la libéralisation de l'économie se sont traduites par la naissance de sociétés et d'organismes privés. Toutes ces nouvelles formes de rapports sociaux seront d'ailleurs très vite exprimées en kiswahili²². Des holdings investissent dans les médias²³. La plupart des journalistes et éditeurs ont été scolarisés dans le système en vigueur du temps de l'*Ujamaa* et écrivent très bien le kiswahili. Ils s'efforcent de maintenir le niveau requis du kiswahili standard. Cependant quand il est question de traduire certaines réalités, ils n'attendent pas les recommandations de *Bakita* et *Tuki* et n'hésitent pas à proposer des termes nouveaux. Cette création spontanée ne facilite pas la tâche de ces deux organes qui se retrouvent parfois mis devant le fait accompli.

1.2.2.2 Le Kenya

À son indépendance, le Kenya optera pour un système politique et économique hérité de la Grande-Bretagne. Sa politique linguistique ne se départira pas de cette allégeance et l'anglais primera sur les langues africaines dont le kiswahili. Si celui-ci y survit, c'est surtout grâce à la persévérance et à l'astuce du père de l'indépendance kenyane, Jomo Kenyatta. Il s'est arrangé pour l'imposer en douceur ne fût-ce que dans l'assemblée nationale, aux côtés de l'anglais. Sa promotion relève plus de la pression du peuple qui l'utilise dans sa vie quotidienne que de la volonté des autorités qui, elles, mettent un point d'honneur à parler la langue de la Reine. Ceci explique l'infime contribution du Kenya, pourtant berceau du kiswahili, au développement de cette langue.

22 *Ujamaa wa kipindi cha mpito*, socialisme de la période de transition, *sera ya uwekezaji wa rasilimali*, politique d'investissement du capital, *ulegezaji wa masharti ya biasbara*, libéralisation commerciale, *mpango wa ufufuaji wa uchumi*, plan de relance de l'économie... Pour une analyse détaillée du langage politique tanzanien, cf. Crozon (1998).

23 Journaux, radio, TV... Cf. Issa (2000).

Des initiatives de promotion et de développement du kiswahili n'apparaîtront que dans le milieu des années 70. Cependant, toutes sont minées par un manque de méthode et de coordination entre les différents acteurs.

Au niveau national, deux formes de kiswahili se démarquent du paysage linguistique, celle de la côte, teintée de particularités dialectales swahili de la région et celle de l'intérieur plus influencée par :

a) l'anglais :

<i>Eproni</i> , tablier	<i>bishopu</i> , évêque
<i>setla</i> , settler	<i>sabchifu</i> , sous chef
<i>bachela</i> , célibataire	<i>weita</i> , mhudumu
<i>ambaseda</i> , ambassade	<i>urtineja</i> , adolescent
<i>chifu</i> , chef	<i>buchari</i> , boucherie
<i>shule ya upili</i> , école secondaire	

b) certaines langues locales :

<i>bewa</i> , campus
<i>barambee</i> , cri de ralliement, motion de solidarité.

Mutahi (1986 : 107) cite l'exemple des livrets de vulgarisation de l'Inades où on trouve *moori*, génisse, *nduma*, racine d'une flèche. Ces deux termes, totalement inconnus en kiswahili, sont empruntés au kikuyu.

Au delà de ces variétés, des divergences subsistent non seulement d'une université à l'autre mais parfois aussi d'un campus à l'autre. À cela s'ajoutent les créations individuelles, surtout issues de la côte où se dessine une tendance à réhabiliter des termes archaïques (poésie classique, folklore...) ayant déjà des équivalents acceptés à Nairobi ou Dar es Salaam. Marshad (1993) et Mdee (1995) offrent des exemples parlants :

	Particularisme	Terme standard
<i>runinga</i>	<i>televisheni</i>	télévision
<i>ulughbai</i>	<i>isimu</i>	linguistique
<i>ulughbai wa kijumuia</i>	<i>isimujamiis</i>	ociolinguistique
<i>mdiplomasi</i>	<i>mwanadiplomasia</i>	diplomate
<i>itakidi</i>	<i>itikadi</i>	idéologie
<i>fizia</i>	<i>fizikia</i>	physique
<i>tabaa</i>	<i>kero</i>	ennui, gêne
<i>tengeo</i>	<i>kizaazaa</i>	chaos, désordre
<i>turbaa</i>	<i>maarifa</i>	connaissances
<i>usburuzi</i>	<i>saikolojia</i>	psychologie
<i>peketa</i>	<i>danganya</i>	escroquer, mentir
<i>tamsa</i>	<i>futa</i>	effacer
<i>tarakilishi</i>	<i>kompyuta</i>	ordinateur

Il arrive qu'une notion soit traduite par trois équivalents différents selon les régions. *Pression sanguine* est dit *mtutu wa damu* en Tanzanie, *bubari* à Mombasa et *presha* partout ailleurs au Kenya (Mutahi 1986: 113).

La collaboration avec Dar es Salaam a souvent été handicapée par les choix politiques des deux pays. Le kiswahili tanzanien était principalement destiné à la diffusion de l'*Ujamaa*. Les Kenyans s'en sont méfiés car ils craignaient qu'un impérialisme linguistique n'en cache un idéologique²⁴. Les définitions de certains termes tels que donnés par le dictionnaire monolingue swahili²⁵ n'étaient pas du genre à rassurer les voisins du nord fortement imprégnés d'une tradition libérale:

Bepari (ma): 1. *Mfanyibiashara mkubwa; tajiri.*
2. *mwenye kumiliki rasilimali na njia kuu za uchumi katika kuzalisha mali kwa faida yake mwenyewe.* 3. *Mmojawapo katika tabaka la wanyonyaji.* 1. Grand commerçant, riche.
2. Celui qui possède beaucoup d'investissements et de grandes possibilités économiques dans la production de richesses à des fins personnelles. 3. L'un des membres de la classe des exploiters. (Tuki 1981: 18)

Kabaila (ma) 1. *Mtu apendaye kumfanya mtu mwengine amtumikie na yeye akastarebe.* 2. *Mtu aliye na majumba na ardhi na kupangisha kwa faida yake binafsi au za watu wachache.* 3. *Mnyonyaji mwenye majumba mengi ya kupangisha na ardhi kubwa ya kupangisha au inayolimwa na watu wengine wengi kwa ajili ya faida yake.* 1. Quelqu'un qui aime qu'une autre personne travaille pour lui alors qu'il se repose.
2. Quelqu'un qui possède beaucoup d'immeubles et des terres en rente pour son seul bénéficiaire ou celui d'une poignée

24 Dans une boutade, Alexandre (1989: 23) souligne cette appropriation linguistique du kiswahili: *Sasa si Kiswahili tena, imekuwa (sic) kitanzania*, maintenant ce n'est plus du kiswahili, c'est du kitanzania. Le préfixe *ki-* désigne la plupart des langues bantu comme le kikuyu, kigogo, kinyarwanda, kirundi...

25 Édité pour la première fois en 1981, il a connu une réédition annuelle (et parfois plusieurs!) depuis 1983 sans que son contenu ne varie d'un iota.

26 *Chama cha Kiswahili cha Taifa*, Association nationale de Kiswahili.

27 Actuelle République démocratique du Congo.

28 Paradoxalement le mot authenticité ne sera rendu dans aucune langue africaine du Zaïre.

de personnes. 3. Un exploitateur qui possède beaucoup d'immeubles à louer et de grands terrains en rente ou cultivés par d'autres personnes pour son seul bénéficiaire. (Tuki 1981: 93-94).

C'est ainsi que ces deux termes, à l'origine signifiant respectivement boutiquier d'origine indienne (Abdulaziz 1984: 197) et aristocrate, féodal ont été retenus pour traduire « capitaliste » et « exploitateur » dans le contexte de l'*Ujamaa*.

L'histoire a évolué et cette vision marxisante n'apparaît plus dans le dernier dictionnaire kiswahili-anglais (Tuki 2001), plus politiquement correct.

Non contents du *statu quo* linguistique en vigueur dans leur pays et d'une dépendance de fait par rapport à la politique linguistique tanzanienne, quelques intellectuels kenyans se mobilisent pour insuffler plus de vie au CHAKITA²⁶. Ils revendiquent un rôle plus clair pour la langue swahili éclipsée par l'anglais et veulent faire entendre le point de vue national dans le concert d'autres pays swahiliphones.

1.2.2.3. Le Congo²⁷

Comme nous l'avons vu plus haut, le Congo a toujours fait figure de marginal dans le mouvement swahili. Jusque dans les années 1950, le comité avait tenté en vain de rallier les territoires belges (Rwanda-Urundi et Congo) au mouvement de standardisation sans parvenir à les convaincre (Mbaabu 1994: 32-33). Cet échec trouve son explication d'une part par l'éloignement de la région du centre névralgique du kiswahili. D'autre part, dans un contexte de concurrence et de méfiance coloniales, il aurait été aberrant qu'une puissance encourage l'emploi d'une langue dont les leviers de commande conditionnant son évolution ne se trouvent en ses mains. C'est ainsi que le kiswahili standard sera fortement marqué « anglophone » tandis que le kiswahili du Congo, lui, évoluera dans un environnement francophone.

Le Congo indépendant effacera toutes les traces de la politique coloniale qui privilégiait l'enseignement en langues nationales (ciluba, kikongo, lingala, kiswahili). Dans le souci de rattraper le temps perdu, il instituera l'enseignement du et en français. Plus tard, au milieu des années 70, une volte-face sera opérée et la politique de l'authenticité²⁸ prônera le retour aux valeurs africaines.

Les quatre langues nationales seront plus ou moins réhabilitées, le lingala bénéficiant d'un appui implicite du président Mobutu.

Le besoin en terminologie va se créer, car les anciens manuels élaborés par les missionnaires utilisaient le *kiswahili bora*, meilleur kiswahili de la côte qui n'avait rien à voir avec celui pratiqué quotidiennement au Congo-Zaïre. S'aligner sur l'expérience tanzanienne aurait élargi ce fossé et les linguistes congolais se mirent à inventer de nouveaux termes. Il est évident que l'opération ne connaîtra pas le même succès qu'en Tanzanie où le kiswahili n'avait pas de concurrent et où l'adhésion du pouvoir était totale. Il en est de même pour l'accueil réservé aux nouveaux termes qui n'intéressent que certaines catégories de personnes (musulmans, journalistes, instituteurs) déjà familières avec le *kiswahili bora*, bon, meilleur kiswahili, variété haute proche du kiswahili standard.

Comme en Tanzanie et au Kenya, les chercheurs congolais recourent aux procédés classiques comme le décrivent Paluku (1990) et Ntita (1993).

En voici quelques exemples :

a) Dérivation :

conducteur, *mwongozi* (<*ongoza*, conduire)
 priorité, *pendeleo* (<*pendelea*, privilégier)
 abattoir, *kichinjio* (<*-chinja*, égorger)
 croissance, *ukuzi* (<*kuu*, croître)

b) Composition

Séismomètre, *kipimatetemeko* (-*pima tetemeko*, mesurer tremblement)
 Fusion, *myeyungano* (-*yeyusha -ungana*, diluer joindre)

29 Il faut toutefois noter que de façon ponctuelle, une coopération entre Tuki de Dar-es-Salam et le Celta (Centre d'études linguistiques théoriques et appliquées) de Kinshasa a souvent existé. Récemment une équipe de Tanzanie a effectué un séjour scientifique en République démocratique du Congo pour s'informer de ce qui est fait sur le kiswahili. Par ailleurs, la plupart des speakers de la télévision et de la radio

s'alignent plus ou moins sur la forme swahili standard. Mais il est évident que l'usage de l'anglais par les uns et du français par les autres limite cette coopération même si la volonté reste réelle.

c) Emprunt

Langues locales:
 volcan, *kirunga*

Français

écologie, *ekolojianation*, *nasioni*
 carbone, *karboni* capitaine, *kapiteni*
 carotte, *karoticamp*, *kampi*
 acide phosphorique, *asidi fosforasi* téléphone, *telefoni*
 gaz carbonique, *gazi ya karbonimilliard*, *miliardi*
 ballet, *balecinéaste*, *simeasti*
 choux, *shu azote*, *azoti*

Anglais via kiswahili standard

avance, *advansi*

d) Calque, transposition et extension sémantique

microphone, *kibuyu*, gourde
 aliment énergétique, *chakula cha kuleta nguvu*, aliment pour apporter la force
 comptabilité, *hesabu ya biasbara*, décompte des affaires
 inflation, *kupungua samani ya feza*, diminution de la valeur financière
 tribunal, *shariani*, *sharia + ni*, littéralement lieu de la loi
 éolienne, *bomba inayotembeza na upepo*, instrument qui fonctionne au vent
 chlorophylle, *umbijani*, *umbi+jani*, forme feuille

La plupart de ces exemples viennent du Letac (1983), où les chercheurs, tout en se rapprochant de leurs collègues de Tanzanie²⁹, n'en sont pas moins restés fidèles au dialecte local. De façon générale, le kiswahili officiel de la République démocratique du Congo est resté uniforme car les publications relatives à son enseignement provenaient pour la plupart de Bukavu. Cependant, pour des raisons de brassages urbains ou de substrats linguistiques, il s'est développé ici et là des formes populaires de cette langue.

Tout comme les autres secteurs, l'activité terminologique au Congo est paralysée par une situation politique et socio-économique des plus précaires et peu favorable à tout épanouissement intellectuel.

1.2.3 Mondialisation et terminologie swahili

Les mutations politiques et économiques qui se sont opérées dans les pays swahiliphones ont un grand impact sur la langue swahili qui, comme nous l'avons vu, a dû s'adapter à ces nouvelles réalités. Ce processus s'accéléra avec le phénomène de la mondialisation et la diffusion des Nouvelles technologies de l'information et de la communication³⁰. Vu la rapidité et l'instantanéité de la diffusion des informations, il est difficile de ne pas réagir rapidement à celles-ci et de les retransmettre directement. Si d'emblée les médias semblent les premiers concernés, les acteurs de tous ordres (politiques, économiques) ou tout autre swahiliphone qui accède à ces informations se retrouvent également impliqués.

Pour ce qui concerne le sort de la langue swahili, cette situation a pour corollaire deux faits :

La démission des terminologues attirés

Les organismes habituellement chargés de la création et du contrôle terminologique, la plupart du temps parastataux, semblent relégués au second plan. Désormais, leur place a été prise par les différents médias (TV, radio, journaux...), majoritairement aux mains des groupes privés disposant de plus de moyens que ces institutions officielles. Les médias puisent leur information des dépêches émanant des grandes agences occidentales (Reuters, AP, AFP) ou de grandes chaînes de radiodiffusion mondiale. Quels que soient les sujets traités (science, politique) ils essayent de rendre l'information le plus rapidement possible au public swahiliphone.

Le rôle de la presse en général et écrite en particulier a toujours été déterminant dans la diffusion et la promotion des nouveaux termes swahili. Ainsi dans les années 80, les termes suivants ont été forgés par les journalistes et sont entrés directement dans la langue swahili :

30 En kiswahili *TEKNOHAMA: TEKNOlojia ya HAbari na MAwasiliano*, Technologie de l'information et de la communication.

31 <http://www.ippmedia.com/guardian/2003/05/16/guardian3.asp>

<i>kuboresha</i>	<i>bora</i>	améliorer
<i>UKIMWI</i>	<i>Ukosefu wa KInga MWilini</i>	sida
<i>siasa kali</i>	politique tranchante	politique extrémiste
<i>watafutaji</i>	<i>-tafuta</i> , chercher	prospecteurs
<i>polisireli</i>	police rail	police des chemins de fer
<i>mtumba</i>		voiture d'occasion
Aujourd'hui nous avons :		
<i>mtandao</i>	< <i>kutanda</i> , étendre	réseau, Internet
<i>utandawazi</i>	réseau ouvert	globalisation, mondialisation
<i>ujumbe wa</i>	message écrit	
<i>maandishi</i>	au téléphone	SMS, texto
<i>kwa simu,</i>		

Cependant, même si les journalistes appartenant aux différents médias s'acquittent tant mal que bien de leur devoir d'informer, ils sont souvent confrontés à de multiples problèmes. N'ayant pas de formation en traduction et encore moins en terminologie, ils ne peuvent pas prétendre cerner avec exactitude la notion ou le sens des termes et du message qu'ils veulent rendre. *Traduttore, traditore*, ceci aboutit souvent à des articles en kiswahili qui ne sont pas nécessairement fidèles au message d'origine. Le public, très friand de ce genre de nouveautés et fier de les entendre en kiswahili, s'en contente et s'en fait vite l'écho. Rares sont les termes contestés ou contredits, séquelles d'une culture où ce qui est dit et surtout écrit par des intellectuels est vérité. Il existe une sorte de pacte entre les différents protagonistes pour s'entraider.

L'autre problème lié à cet exercice de traduction est l'enracinement d'une forme de kiswahili dont la syntaxe est très fortement influencée par l'anglais. Les critiques formulées à l'encontre du comité resurgissent plus de 70 ans après, la seule différence étant que peu de personnes prennent leur plume pour contester. Une certaine résignation s'installe. Les swahiliphones, au lieu de penser en kiswahili ce qui a été formulé en anglais, transposent directement en kiswahili le message tel quel en anglais comme le montre cet exemple³¹ tiré de deux journaux tanzaniens *The Guardian* et *Nipashe* :

Hamad urges CUF supporters to play rough if...

*Friday, May 16, 2003.
By Mashaka Mgeta, Pemba*

The Secretary General of the opposition Civic United Front (CUF), Seif Shariiff Hamad, has urged party supporters to prepare for a 'head on collision' with state organs if there will be a breach of peace during the parliamentary by-elections on May 18.

Addressing an election campaign rally for Chake Chake, Chonga and Ziwani constituencies at Chanjawiri village, he said CUF had unearthed a plot to contravene electoral laws during the by-elections in 17 constituencies. « We would want to inform President Amani Abeid Karume that we (CUF) promised to protect the CUF/CCM accord, but he (Karume) is bothering us. I swear that we shall fight against the state, » he said.

He accused CCM in Zanzibar of transporting a group of youths from Unguja Island to sabotage voting in Mkenyageni.

Hamad said the move was aimed at depriving residents of Mkenyageni of their right to elect a member of parliament and a representative of their own choice.

Ataka wafuasi wa CUF wajiandaa

*Friday, May 16, 2003.
Na Mashaka Mgeta, Pemba*

Katibu Mkuu wa CUF, Maalim Seif Shariiff Hamad, amewataka wafuasi wa chama bicho kujiweka tayari kwa mapambano dbidi ya vyombo vya dola, endapo kutatokea ukiukaji wa sheria siku ya Uchaguzi mdogo, ambayo ni kesho kutwa.

Alisema hayo jana katika hotuba yake ya kuwanadi wagombea Ubunge na Uwakilishi wa CUF katika majimbo ya Chake Chake, Chonga na Ziwani. Katika mkutano huo kwenye kijiji cha Chanjawiri, Malim alisema CUF imegundua kuwepo njama, zenye lengo la kukiuka sheria na kanuni za uchaguzi huo.

« Tunamuambia Rais Karume kuwa CUF tuliabidi kuulinda Muafaka, lakini yeye anatuletea fyokofyoko, walabi ladbim, naelekeza macho yangu mbinguni, tutapambana na dola na serikali yake, » aliahidi.

Maalim Seif alidai kuwa CCM Zanzibar chini ya Rais Amani Abeid Karume, imeleta vijana kutoka Unguja huku Pemba kufanya bujuma ili wapiga kura wa jimbo la Mkenyageni wasitumie baki ya kumchagua Mbunge na Mwakilishi wao.

Vers une certaine uniformisation

32 Voix de l'Allemagne, BBC,
Voix de l'Amérique...

35 <http://www.google.com/intl/sw/>

33 <http://www.yale.edu/swahili/>

36 Wanjiku Ng'ang'a, *Automatic Word Sense Disambiguation of Kiswahili words*, Thèse de doctorat cours, Université de Helsinki.

34 <http://www.unizh.ch/spw/aflring/aliswahili/indexE.html> et <http://www.africa.uga.edu/Kiswahili/doe/kikonini.html>

37 Une équipe de l'Inalco l'utilise actuellement dans la confection d'un dictionnaire kiswahili-français.

La mondialisation a cet avantage de briser d'une certaine manière les barrières politiques, culturelles et sociales. Cette circulation rapide de l'information et des idées a un effet sur l'uniformisation du kiswahili entre le Kenya et la Tanzanie d'une part et l'Afrique orientale et le reste du monde (Afrique centrale, institutions impliquées dans la diffusion du swahili disséminées sur différents continents) d'autre part. Les interventions des correspondants ou des interviewés congolais ou ougandais diffusées par les radios internationales³² reflètent ce souci de s'aligner sur le kiswahili de l'Est auxquels ils sont de plus en plus exposés.

1.2.4 Applications des NTIC au kiswahili

La modernisation du kiswahili ne se limite pas seulement à l'enrichissement de son lexique. Elle passe également par les applications d'outils existants sur cette langue et à la création de nouveaux outils qui lui sont complètement dédiés. Faute de les signaler tous, nous ne mentionnerons que les plus importants.

On peut trouver sur Internet: outre deux grands dictionnaires (kiswahili-anglais et anglais-kiswahili)³³, plus d'une dizaine de méthodes d'enseignement du kiswahili, de profondeur inégale, y sont répertoriées. Les plus consistantes sont *Ali Swabili* et *Kiko*³⁴. *Google* a par ailleurs une page de recherche entièrement en langue swahili³⁵.

Des chercheurs de l'université d'Helsinki ont mis en place *Salama (Swabili Language Manager)*. Il s'agit d'une série d'outils informatiques permettant le traitement et l'analyse de textes swahili. Il comporte entre autres un correcteur orthographique swahili pour *Word 97* et ultérieur, un analyseur morphologique appelé *Swatwol* (Hurskainen 1999). Un programme de traduction automatique anglais-kiswahili est en préparation³⁶ et une thèse sur la terminologie assistée de l'ordinateur y a déjà été soutenue (Sewangi 2001).

Une version cédérom des deux dictionnaires bilingues (Tuki 1996 et 2001) est déjà disponible et un outil lexicographique comme le logiciel *Shoebbox*, très pratique, se prête très bien au kiswahili³⁷.

Conclusion

Par rapport aux autres langues africaines, le kiswahili est une langue favorisée. Entré très tôt en contact avec le monde extérieur, symbole de modernité, il s'enrichira des apports culturels et linguistiques aussi bien de l'orient que de l'occident. D'autres circonstances positives contribueront à son affermissement comme véritable moyen de communication et d'information : la puissance économique de Zanzibar, un système colonial favorable à son développement et plus tard la foi du peuple tanzanien en ses capacités.

La mondialisation et les nouvelles technologies le trouvent déjà outillé et apte à servir dans tous les domaines du savoir, même les plus actuels. Cependant il convient de signaler que malgré tout, ces résultats positifs dans la course vers la modernisation sont obtenus en ordre dispersé. Le manque d'harmonisation entre les différentes institutions, aussi bien publiques que privées, nationales et internationales déroutent et parfois découragent. Ce serait le prochain défi à relever en ce début de siècle, au moment où une énième résolution de l'Union africaine³⁸ a confirmé le kiswahili comme première langue africaine de travail.

Jean de Dieu Karangwa,
 Institut national des langues et civilisations orientales, Paris,
 France.
 jeandedieu.karangwa@inalco.fr

38 Anciennement Organisation de l'unité africaine (OUA).

Bibliographie

- Abdulaziz (M.H.), 1971 : « Tanzania's National Language Policy and the Rise of Swahili Political Culture » in Whiteley (W.H.), *Language Use and Social Change: Problems of Multilingualism with Special Reference to Eastern Africa*, OUP: London, p. 160-178.
- Abdulaziz (M.H.), 1984 : « Aspects of lexical and semantic elaboration in the process of modernization of Kiswahili », in MAW (J.) et Parkin (D.) (eds), 1984 : p. 195-213.
- ACCT (éd.), 1983 : *Lexique thématique de l'Afrique centrale, Zaïre: kiswahili, activités économiques et sociales 1*, Cerdotola.
- Akida (H.), 1958 : « Correspondence », in *Journal of East African Swahili Committee*, p. 85-89.
- Akida (H.), 1971-1976 : « Msamiati wa Muda wa Elimuviumbe – Biology Terminology », *Kiswahili*, 1971, vol. 41/2, p. 95-101 ; 1972, vol. 43/1, p. 97-101 ; 1973, vol. 43/2, p. 116-120 ; 1975, vol. 45/1, p. 57-63 ; 1976, vol. 46/1, p. 76-81.
- Alexandre (P.), 1989 : « La nébuleuse swahili », dans Alexandre, (P.) et Rombi, (M.-F.), 1989, p. 21-24.
- Alexandre (P.) et Rombi (M.-F.), 1989 : *Le swahili et ses limites : ambiguïté des notions reçues*, Paris : Edicef.
- Bakhressa (S.K.), 1992 : *Kamusi ya Maana na Matumizi*, OUP: Nairobi.
- Bakita (éd), 1992 : *Kamusi ya Istilabi za Sayansi na Teknolojia*, Bakita and Educational Publishers & Distributors.
- Bakita, 1993 : *Mwongozo wa Usanifishaji Istilabi*, (ronéo)
- Besha (R.M.), 1995 : « Nafasi ya Msamiati wa Magazeti katika Kamusi ya Kiswahili Sanifu », dans Kiango (J.G.) et Mdee (J.S.) (éd.), 1995, p. 61-72.
- Crozon (A.), 1998 : « Dire pour séduire : Langages et politique en Tanzanie », dans Martin (D.C.), *Nouveaux langages de politique en Afrique orientale*, Karthala-Itra : Paris, p. 115-185.
- Gurnah, (A.M.), 1974 : « Agricultural Terms in Kiswahili. Part I. Crop Science – Msamiati wa Elimu ya Ukulima », dans *Kiswahili*, vol. 44/2, p. 32-44.
- Hurskainen (A.), 1999 : « SALAMA Swahili Language Manager », dans *Nordic Journal of African Studies*, vol. 8/2, p. 139-157.
- Irira (S.D.), 1995 : *Kamusi Awali ya Sayansi na Teknolojia*, Dar es Salaam : Ben & Company Ltd.
- Issa (O.), 2000 : « Le développement des médias : l'exemple de la Tanzanie », dans *Afrique Contemporaine*, décembre.
- Karangwa (J. de D.), *Le Kiswahili dans l'Afrique des Grands Lacs : Contribution sociolinguistique*, Thèse de doctorat, Paris : Inalco.
- Khalid (A.), 1977 : *The Liberation of Swahili from European appropriation*, Nairobi : EALB.
- Kiango (J.G.) et Mdee (J.S.) (éds.), 1995 : *Utafiti na Utungaji wa Kamusi*, Dar es Salaam : Tuki.

- King'ei (K.), 1999: «Swahili Technical Terminology: Problems of Development and Usage in Kenya», in *Afrikanistische Arbeitspapiere*, vol. 60, p. 147-160.
- Knappert (I.), 1989: «Les mots swahili empruntés au grec, aux langues romanes et américaines», dans Alexandre, (P.) et Rombi, (M.-F.), 1989: p. 41-57.
- Madalla (A.), 1972: «Mechanical Engineering Terminologies», in *Kiswahili*, vol. 42/1, p.75-86.
- Marshad (H.A.), 1993: *Kiswahili au Kiingereza? (Nchini Kenya)*, Nairobi: Jomo Kenyatta Foundation.
- Martin (D.C.), 1988: *Tanzanie, l'invention d'une culture politique*, Paris: Presses de la Fondation des sciences politiques, Karthala.
- Maw (J.) et Parkin (D.), 1984: *Swahili Language and Society*, Wien: Beitrage zur Afrikanistik, Band 23.
- Mbaabu (I.), 1991: *Historia ya Usanifishaji wa Kiswahili*, Longman Kenya.
- Mdee (J.S.), 1995: «Kiswahili cha Kenya na Tanzania katika Kamusi ya Kiswahili», Kiango (J.G.) et Mdee (J.S.) (éds.), 1995, p. 42-60.
- Mdee (J.S.), 1980: *The Degree of Acceptability of New Swahili Terms: Speakers Response Analysis*, MA thesis, University of Dar es Salaam.
- Mdee (J.S.), 1986: «Matatizo ya kuunda Istilahi kama yanavyojitokeza katika Kiswahili», in *Kiswahili*, vol. 53/1-2, p. 115-127.
- Mhina (G.A.), 1976: *La planification linguistique en Tanzanie*, Éditions du base, Kisangani.
- Mkude (D.), 1984: «The Fate of Standard Swahili», Maw (J.) et Parkin (D.) (eds.), p. 25-34.
- Mlacha (S.A.K.) (éd.), 1995: *Kiswahili na vyombo vya habari*, Dar es Salaam: Tuki.
- Mlacha, (S.A.K.), *Kamusi ya sheria Kiingereza Kiswahili*, Dar es Salaam: Tuki.
- Musole (M.), 1989 et 1990: *Tujifunze Kiswahili II et III*, Manuels pour l'élève de 4^e et 5^e Primaire, Éditions Loyola.
- Mutahi (K.), 1986: «Swahili Lexical Expansion: Prospects and Problems», in *Kiswahili*, vol. 53/1-2, p. 104-114.
- Mwansoko (H.J.M.) et Tumbo-Masabo (Z.N.Z.), 1996: *Matumizi ya Kiswahili Bungeni*, Dar es Salaam: Tuki.
- Mwansoko (H.J.M.), 1991: «The Aftermath of Post Independence liberalism in Swahili Standardization», in *Kiswahili*, vol. 58, p. 1-11.
- Mwansoko (H.J.M.), 1995: «Dhima ya Vyombo vya Habari katika Ukuzaji, Usanifishaji na Uenezaji wa Istilahi», Mlacha (S.A.K.) (éd.), 1995, p. 1-13.
- Mzee (H.), 1995: «Uhuru na Mzalendo katika kuendeleza na Kukuza Kiswahili», dans Mlacha (S.A.K.), (éd.), 1995, p. 79-83.
- Ntita (N.), 1993: «Terminologie et développement linguistique au Zaïre», dans *Terminologies nouvelles*, n° 9, p. 80-83.
- Ohly (R.), 1982: *Swahili-the diagram of crises*, Beiträge zur Afrikanistik, Band 15, Wien-Dar es Salaam.
- Paluku (Mb.), 1990: «Méthodes de travail en terminologie au Zaïre», dans *Terminologies nouvelles*, n° 3, p. 37-41.
- Philippson (G.), 1970: «Étude de quelques concepts politiques swahili dans les œuvres de J.K. Nyerere», dans *Cahiers d'études africaines*, vol. 40/1, p. 530-545.
- Polome (E.C.), 1983: «Standardization of Swahili and the Modernization of the Swahili Vocabulary», Hagège, (Cl.), dans *La réforme des langues III*, Hamburg: Buske Verlag, p. 53-77.
- Sewangi (S.S.), 2001: *Computer-Assisted Extraction of Terms in Specific Domains: The case of Swahili*, PhD Thesis, University of Helsinki.
- Sewangi (S.S.), 2002: «Isimu Kompyuta na Kiswahili: Matatizo na Matarajio», communication présentée au colloque *Kiswahili: Lugha na Mikondo ya Fikera*, Nairobi, 3-6 octobre 2002.
- Temu (C.W.), 1984: «Kiswahili Terminology: Principles adopted for the Enrichment of the Kiswahili Language», in *Kiswahili*, vol. 51/1-2, p. 112-126.
- Tuki (éd.), 1996: *English Swahili Dictionary Kamusi ya Kiingereza Kiswahili*, Dar es Salaam: Tuki.
- Tuki (éd.), 1981: *Kamusi ya Kiswahili sanifu*, Nairobi: OUP.
- Tuki (éd.), 1990a: *Kamusi Sanifu ya Isimu na Lugha*, Dar es Salaam: Tuki/Unesco/Sida.
- Tuki (éd.), 1990b: *Kamusi Sanifu ya Biolojia, Fizikia na Kemia*, Dar es Salaam: Tuki/Unesco/Sida.
- Tuki (éd.), 2001: *Kamusi ya Kiswahili-Kiingereza Swahili-English Dictionary*, Dar es Salaam: Tuki.
- Tumbo (Z.), 1982: «Towards a Systematic Terminology Development», in *Kiswahili*, vol. 49/1, p. 87-98.
- Tumbo-Masabo (Z.N.Z.) et Mwansoko (H.J.M.), 1992: *Kiongozi cha Uundaji wa Istilahi za Kiswahili*, Dar es Salaam: Tuki.
- Whiteley (W. H.), 1969: *Swahili: The Rise of a National Language*, London: Methuen and Co, Ltd.
- Whiteley (W.H.), 1974: «Msamiati wa Muda wa Saikolojia», in *Kiswahili*, vol. 44/2, p. 75-81.

En bref – Nouvelles du Rifal

Le Rifal et la normalisation internationale en matière de langue

L'Agence intergouvernementale de la Francophonie a, depuis quelques années, élargi sa coopération avec le monde de la normalisation, notamment en participant activement au Comité technique 37 de l'Iso (Terminologie et autres ressources linguistiques). L'Agence est associée à ce comité technique de l'Iso par l'intermédiaire du Rifal (Réseau international francophone d'aménagement linguistique), accrédité par l'Iso comme organisme de liaison.

L'intérêt du Rifal et de l'Agence pour la normalisation internationale tient au fait que le caractère référentiel des documents normatifs s'étend autant aux données techniques qu'ils contiennent qu'à la langue qu'ils véhiculent. De plus, les normes techniques ont souvent force de loi dans la mesure où de nombreuses lois et les textes réglementaires nationaux font une référence explicite et obligatoire à telle ou telle norme technique indépendamment de la langue dans laquelle cette norme est rédigée.

Dans le contexte de la mondialisation et de l'uniformisation internationale des normes, l'harmonisation des normes nationales ou régionales se fera le plus souvent par l'adoption des normes internationales, lesquelles sont le plus souvent élaborées en anglais et, dans un certain nombre de cas, traduites en français, comme c'est le cas pour l'Iso et la Cei, pour ne prendre que ces deux exemples.

Il s'ensuit donc une pression linguistique sur les pays dont la ou les langues ne font pas partie des langues officielles de la normalisation internationale ou, s'agissant d'une langue officielle, celle-ci n'est qu'une langue de traduction.

Sur le plan linguistique, outre le problème de la langue des normes, deux types de normes peuvent influencer les langues elles-mêmes et leur usage :

1. les normes qui sont susceptibles de toucher le système des langues, notamment les normes pour l'informatisation des langues et plus particulièrement les normes utilisées dans les domaines suivants :

- l'informatique documentaire ;
- la création de systèmes experts ;
- les outils d'analyse et de génération de textes ;

- les outils pour la traduction assistée par ordinateur, la lexicomatique, et la terminotique ;
- les outils d'aide à la rédaction et à l'édition de textes ;
- les outils de saisie automatique de l'écrit ;
- les outils de traitement automatique de la parole ;
- la localisation ;
- les outils inforoutiers (navigateurs, moteurs de recherche, etc.).

2. Les normes terminologiques, qu'il s'agisse des normes sur les principes et méthodes terminologiques ou des normes terminologiques elles-mêmes (vocabulaires), ces dernières exerçant une pression terminologique constante susceptible de favoriser l'emprunt massif de termes aux langues dominantes dans lesquelles les normes sont rédigées par les autres langues.

En participant au Comité technique 37, le Rifal souhaite apporter le point de vue des pays de la Francophonie, qui ne sont pas tous membres de l'Iso. Cette participation s'appuie sur les quelques principes et objectifs suivants :

1. Promouvoir la rédaction ou la traduction des normes internationales dans le plus grand nombre des grandes langues véhiculaires transnationales (par exemple, outre le français : l'espagnol, le portugais, le chinois, l'allemand, l'arabe, etc. de même que dans un certain nombre de langues transnationales du Sud).

2. S'assurer de l'applicabilité des normes internationales à toutes les langues.

3. S'assurer que les normes internationales rendent possible l'informatisation de toutes les langues, notamment en ce qui a trait aux systèmes d'écriture (polices de caractères, accentuation des caractères, création de claviers, saisie et lecture de gauche à droite et inversement, etc.).

Ce dernier principe doit soutenir la possibilité de l'informatisation du plus grand nombre de langues possible, car, au XIX^e siècle, l'informatisation d'une langue est devenue un critère incontournable de survie de cette langue, tout comme ce fut le cas jadis de l'écriture puis de l'existence d'imprimés dans cette langue.

Pour illustrer par un exemple l'effet réel de la participation du Rifal, le Sous-comité 1 du CT 37 (Terminologie et autres ressources linguistiques – Principes et méthodes) vient

de créer un nouveau groupe de travail sur la socioterminologie, qui prendra en compte les travaux en cette matière menés notamment par le Rint (ancien réseau de l'Agence) et du Rifal. Ce groupe de travail sera notamment chargé d'élaborer des directives pour l'interprétation et l'usage des autres documents normatifs du Comité technique 37 dans la perspective de la diversité culturelle et linguistique. L'introduction de ce nouveau groupe de travail pourrait conduire à une nouvelle conception de la normalisation et de l'harmonisation de la terminologie, au moment où les exigences du marché linguistique oscillent entre le pôle de la localisation et celui de l'intercompréhension.

Louis-Jean Rousseau,

Président du SC1 du CT 37 de l'Iso,

Agence intergouvernementale de la Francophonie.

Le Rifal, et l'informatisation des langues et l'ingénierie linguistique

C'est sur le thème de l'informatisation des langues que le Réseau international francophone d'aménagement linguistique (Rifal) a tenu la réunion de son Assemblée générale à Beyrouth du 6 au 8 novembre 2002. Cette réunion était accompagnée d'une journée d'étude sur l'informatisation des langues africaines, dans la perspective de la création et de l'implantation prochaines de banques de terminologie en réseau sur Internet dans certains pays du Sud. Il s'agit ici de la conception et de l'implantation de banques de textes et de banques de terminologie multilingues en réseau pour les pays du Sud. Le projet doit servir à rassembler les terminologies déjà élaborées, à élaborer de nouvelles terminologies en réseau entre pays partageant les mêmes langues, à diffuser la terminologie (français-langues partenaires¹) sur Internet et à éditer des lexiques et vocabulaires pour la diffusion

1 On entend ici par « langue partenaire » toute langue qui coexiste avec la langue française et avec laquelle sont aménagées des relations de complémentarité et de coopération fonctionnelles dans le respect des politiques linguistiques nationales.

d'imprimés. Les résultats escomptés sont un transfert technologique et un transfert de connaissances conduisant à l'élaboration et à la diffusion de la terminologie sur Internet et, donc, à l'informatisation des langues du Sud.

Rappelons que le Rifal, qui regroupe une vingtaine d'États et gouvernements, a été créé par l'Agence intergouvernementale de la Francophonie en 2000 pour la réalisation de sa programmation en matière de langues. Le réseau est notamment chargé des questions de terminologie et d'ingénierie linguistique.

La tenue de cette table ronde avait pour objectif la mise en relief de l'importance de travailler à lever les obstacles linguistiques et techniques à l'informatisation des langues africaines et créoles, dans la mesure où l'on peut affirmer aujourd'hui que l'informatisation d'une langue est devenue un facteur déterminant de son usage et de sa survie, comme ce fut jadis le cas de l'imprimerie. Il existe en effet des obstacles linguistiques, telle l'absence de standardisation de l'orthographe de certaines langues, mais aussi et surtout, des obstacles techniques comme, par exemple, la présence de caractères spéciaux pour lesquels il n'existe pas de polices de caractères dans les normes techniques existantes. Les exposés présentés lors de la table ronde de Beyrouth et publiés dans le présent numéro des *Cahiers du Rifal* démontrent l'existence au sein du réseau d'une expertise solide en matière d'informatisation des langues. À la tenue de cette table ronde, le Rifal vient d'ajouter une *Vitrine-forum de l'ingénierie linguistique* qui s'est tenue à Montréal, le 13 juin 2003. Cette importante manifestation, dont les travaux seront publiés dans la prochaine livraison de cette revue, a été organisée en étroite collaboration avec l'Agence universitaire de la Francophonie et plus particulièrement avec le Réseau Lexicographie, terminologie et traduction. Cette vitrine-forum est destinée à favoriser l'émergence et l'évolution des outils d'ingénierie linguistique et à assurer la promotion et la diffusion des produits technologiques francophones.

À ces nouvelles activités du Rifal s'ajoute notamment la création d'un inventaire des produits d'ingénierie linguistiques francophones. Il s'agira de rassembler dans le site Internet du réseau toute l'information sur les logiciels ou outils informatiques francophones traitant les

langues, comme les aides à la rédaction, les outils de traduction, de terminologie et de lexicographie assistées par ordinateur, les outils de traitement de la parole, les systèmes experts, les outils de traitements de texte, les outils d'informatique documentaire, les outils d'analyse et de génération de texte, les outils inforoutiers (navigateurs, moteurs de recherche, etc.), les outils d'aide à la localisation, etc.

Le projet de banques de données évoqué ci-dessus est associé au programme de formation du RIFAL en direction des pays du Sud qui porte notamment sur la mise dans l'Internet de contenus linguistiques et terminologiques en français et en langues africaines. Le programme de formation a pris la forme de stages tenus sur place par des formateurs bénévoles.

Des séances de formation ont eu lieu en 2000 et en 2001 à Haïti, Dakar, Niamey, Kinshasa, Bangui et Antananarivo. En 2001, Le Rifal a organisé en novembre et décembre une série de stages de formation intitulée « Banques de données multilingues français-langues partenaires : modélisation et partage via Internet ». Dans sa programmation de 2002, le Rifal a organisé des sessions de formation dans neuf pays du Sud (Centrafrique, Congo démocratique, Haïti, Madagascar, Mali, Maroc, Mauritanie, Niger, Sénégal) sur la constitution de banques de textes en français et dans les langues partenaires au format XCES, un standard d'encodage de corpus. En 2003 et 2004, d'autres sessions seront organisées dans les mêmes conditions pour une initiation au BTML et l'installation des premières banques de données décrites ci-dessus.

Louis-Jean Rousseau,

Agence intergouvernementale de la Francophonie.

Les Cahiers du Rifal

Les *Cahiers du Rifal*, revue scientifique du Réseau international francophone d'aménagement linguistique (Rifal), paraissent chaque année à la suite d'un appel à contributions. La thématique abordée est fixée par le Réseau et la direction scientifique du numéro est confiée à une ou plusieurs personnalités actives dans le domaine abordé. La revue publie également des actes de colloque et de séminaires.

Les articles destinés à paraître dans les *Cahiers du Rifal* doivent être soumis au(x) directeurs scientifiques du numéro à paraître.

Direction scientifique de ce numéro:
Unité mixte de recherche *Langage, langues et cultures d'Afrique noire* (Llacan, Paris).

Comité scientifique:

Daniel Blampain (Institut supérieur de traducteurs et interprètes, Bruxelles),
Marcel Diki-Kidiri (Centre national de la recherche scientifique, Paris),
Abdelkader Fassi-Fehri (Institut d'études et de recherches pour l'arabisation, Rabat),
John Humbley (Université de Paris VII),
Chérif Mboj (Centre de linguistique appliquée de Dakar),
Marie-Claude L'Homme (Université de Montréal),
Silvia Pavel (Bureau de la traduction, Hull),
Eric Wehrli (Université de Genève).

Comité de rédaction:

Daniel Blampain (Institut supérieur de traducteurs et interprètes),
Martine Garsou (Communauté française de Belgique),
Louis-Jean-Rousseau (Agence intergouvernementale de la Francophonie),
Marc Van Campenhoudt (Institut supérieur de traducteurs et interprètes).

Secrétaire de rédaction:

Marc Van Campenhoudt,
Institut supérieur de traducteurs et interprètes, 34 rue Joseph Hazard,
B-1180 Bruxelles, Belgique,
marc.van.campenhoudt@euronet.be.

Éditeur responsable:

Martine Garsou, Communauté française de Belgique, 44 bd Léopold II,
B-1080 Bruxelles, Belgique.

Conception:

Patrice Junius, *Alternatives Tbéâtrales*.

Conseillers techniques:

Serge Paulus et Christian Chanard.

Photocomposition et impression:

Édition et imprimerie.

Présentation des tapuscrits:

Les normes de présentation des articles ainsi que la feuille de style associée peuvent être téléchargées sur le site internet du Rifal: www.rifal.org.

Descriptif bibliographique:

UMR Llacan, dir., *Le traitement informatique des langues africaines*, dans *Cahiers du Rifal*, n° 23, novembre 2003, Bruxelles, Agence intergouvernementale de la Francophonie et Communauté française de Belgique, ISSN: 1015-5716.

Les éditeurs

Agence intergouvernementale de la Francophonie

Direction des langues et de l'écrit
13, quai André Citroën
75015 Paris
France
<http://agence.francophonie.org>

Communauté française de Belgique

Service de la langue française
44 bd Léopold II
B-1080 Bruxelles
<http://www.cfwb.be/franca>
et
Commissariat général
aux relations internationales
2, Place Saintelette
B-1080 Bruxelles
Belgique
<http://www.wbri.be>

Liste des numéros parus sous le titre *Terminologies nouvelles*

- Consultables à partir du n° 14
à l'adresse www.rifal.org
- 1, mai 1989: *Le Rint: objectifs et perspectives*
 - 2, décembre 1989: *La formation en terminologie*
 - 3, juin 1990: *Harmonisation des méthodes en terminologie* (actes des séminaires de Talence et de Hull)
 - 4, décembre 1990: *Numéro général*
 - 5, juin 1991: *Terminologie et informatique*
 - 6, décembre 1991: *Terminologie et développement I* (actes du séminaire de Rabat)
 - 7, juin 1992: *Numéro général*
 - 8, décembre 1992: *Terminologie et environnement*
 - 9, juin 1993: *Terminologie et développement II* (actes du séminaire de Cotonou)
 - 10, décembre 1993: *Phraséologie* (actes du séminaire de Hull)
 - 11, juin 1994: *Numéro général*
 - 12, décembre 1994: *Implantation des termes officiels* (actes du séminaire de Rouen)
 - 13, juin 1995: *Terminologie et entreprise*
 - 14, décembre 1995: *Numéro général*
 - 15, décembre 1996: *Banques de terminologie* (actes de la table ronde de Québec)
 - 16, juin 1997: *Enquêtes terminologiques*
 - 17, décembre 1997: *Terminologie et formation*
 - 18, juin 1998: *Terminotique et documentation*
 - 19, décembre 1998 - juin 1999: *Terminologie et intelligence artificielle* (actes du colloque de Nantes)
 - 20, décembre 1999: *De nouveaux outils pour la néologie*
 - 21, décembre 2000: *Terminologie et diversité culturelle*

Liste des numéros parus sous le titre *Cahiers du Rifal*

- Consultables à l'adresse
www.rifal.org
- 22, décembre 2001: *Développement linguistique: enjeux et perspectives*
 - 23, novembre 2003: *Le traitement informatique des langues africaines*

Membres du Rifal

Afrique de l'Ouest (Bénin, Guinée, Mali, Mauritanie, Niger, Sénégal)

Coordination:

Centre de linguistique appliquée
Université Cheikh Anta Diop
Dakar-Fann, Sénégal
Téléphone: 221.825.01.26
Télécopie: 221.824.71.47
Courriel: chembodj@ucad.refer.sn

Canada

Bureau de la traduction
Travaux publics et Services
gouvernementaux Canada
Immeuble Richelieu
975, bd Saint-Joseph, 5^e étage
Gatineau (Québec) K1A 0S5, Canada
Téléphone: 1.819.997.33.00
Télécopie: 1.819.997.19.93
Courriel: gabriel.huard@pwgsc.gc.ca

Communauté française de Belgique

Service de la langue française
Ministère de la Culture
et des Affaires sociales
44, boulevard Léopold II
1080 Bruxelles, Belgique
Téléphone: 32 2.413.22.95
Télécopie: 32 2.413.28.94
Courriel: martine.garsou@cfwb.be

France

Délégation générale à la langue française
6, rue des Pyramides
75001 Paris, France
Téléphone: 33.1.40.15.36.70
Télécopie: 33.1.40.15.36.76
Courriel: bernard.cerquiglini@culture.fr

Haïti

Faculté de linguistique appliquée
Université d'État d'Haïti
B.P. 668
Rue Dufort n° 38
Port-au-Prince, Haïti
Téléphone: 509 245.12.33
Télécopie: 509 245.91.53
Courriel: Ssylvestre.fla@ht.refer.org

Madagascar

Centre des langues de l'Académie
Académie malgache
B.P. 6217
Antananarivo 101, Madagascar
Téléphone: 261.20.22.624.39
Télécopie: 261.20.22.313.61
Courriel: tenymalagasy@dts.mg

Maroc

Institut d'études et de recherches
pour l'arabisation
B.P. 6216, Instituts
Rabat, Maroc
Téléphone: 212.77.730.09
Télécopie: 212.77.720.65
Courriel: fassi@atlasnet.net.ma

Québec

Office québécois de la langue française
125, rue Sherbrooke Ouest
Montréal (Québec) H2X 1X4, Canada
Téléphone: 1.514.873.07.97
Télécopie: 1.514.873.34.88
Courriel: nicole.rene@oqlf.gouv.qc.ca

République centrafricaine

Institut de linguistique appliquée
Université de Bangui
B.P. 1450
Bangui, République centrafricaine
Téléphone: 236.61.67.59
Télécopie: 236.61.78.90
Courriel: ila@intnet.cf

République démocratique du Congo

Centre de linguistique théorique
et appliquée
B.P. 4956
Kinshasa/Gombé,
République démocratique du Congo
Téléphone: 243.12.62.910
Télécopie: 243.12.46.197
Courriel: anyembwe@yahoo.fr

Roumanie

Institutul de lingvistica Iorgu Iordan
Casa Academiei
Calea 13 Septembrie nr. 13
76100 Bucuresti, Roumanie
Téléphone: 40.1.410.34.09
Télécopie: 40.1.410.34.10
Courriel: ioanar@fx.ro

Suisse

Chancellerie de la Confédération suisse
Services linguistiques centraux
Section française
CH-3003 Berne, Suisse
Téléphone: 41 31. 324. 11. 49
Télécopie: 41 31. 324. 11. 04
Courriel: anne.marie.gendron@bk.admin.ch

Union latine

Union latine
131, rue du Bac
75340 Paris Cédex 07, France
Téléphone: 33.1.45.49.60.60
Télécopie: 33.1.45.44.45.97
Courriel: d.prado@unilat.org

Bulletin d'abonnement

Je soussigné souhaite recevoir gratuitement les *Cahiers du Rifal*.

Nom: _____

Entreprise, organisme: _____

Fonction: _____

Ce bulletin d'abonnement est à adresser à l'organisme représentant votre pays ou, à défaut, au secrétariat général du Rifal.

Adresse: _____



Notes

A series of horizontal dotted lines for writing notes.

