



HAL
open science

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

Chantal Enguehard, Chérif Mbodj

► To cite this version:

Chantal Enguehard, Chérif Mbodj. Flore : un site coopératif pour recueillir et diffuser les noms des plantes dans les langues africaines. Les cahiers du RIFAL, 2004, Le traitement informatique des langues africaines, 23, pp.47-54. hal-01094951

HAL Id: hal-01094951

<https://hal.science/hal-01094951>

Submitted on 17 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

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

RÉSUMÉ

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 visibles par 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 : XML.

Mots 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 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 : chapitre2)) 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 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). A 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

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)

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.

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'il leur interdit l'accès aux nouvelles technologies, et en particulier à Internet, dans leur langue. 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 »⁴

2 CHOIX DU MEDIUM

La diffusion des connaissances a été longtemps monopolisée par les ouvrages imprimés (revue 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.

³ Séminaire tenu à Mons en Belgique, du 16 au 22 mars 2002. Ce séminaire a été 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.

⁴ Voir B. Harmegnies et R. Renard, 2002, *Rapport d'évaluation du Séminaire sur le respect de la diversité linguistique et l'accès au cyberspace*, p.2.

- Le partage de grandes quantités de données électroniques est facilité par l'utilisation de formats comme 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, 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, 1997 : 92) pour le *Moringa oleifera* Lam : 'nevöday' et 'nöböday' alors que dans (Von Maydell, 1983 : 447) est énoncé 'benaile'. Ces trois dénominations sont grossièrement erronées : les caractères 'ö' et 'v' n'existent pas dans la langue wolof, tandis que 'benaile' 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, 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 croisé : ɓ; ou le n palatal : ŋ), 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 jòola, le peul⁵, 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.

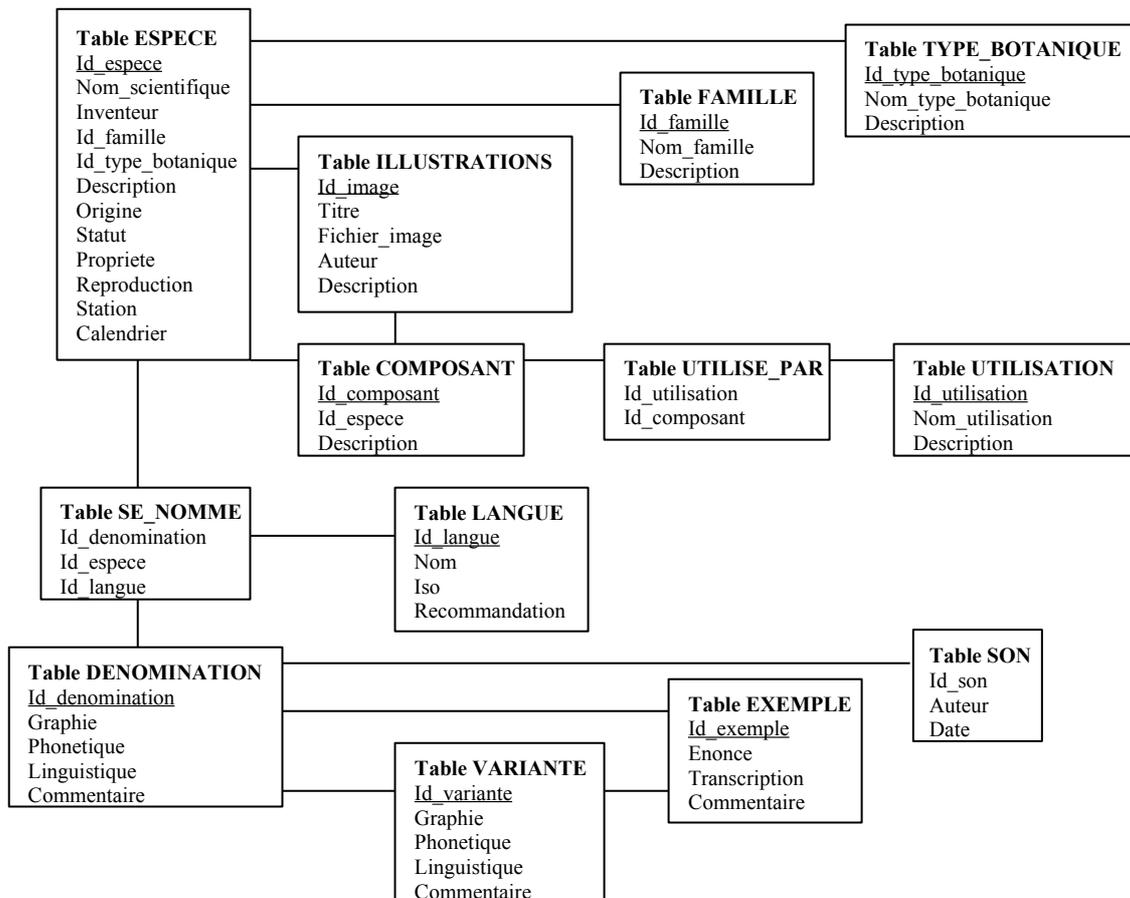
Chaque nom est assorti de sa graphie, de sa phonétique, d'informations linguistiques et d'un commentaire libre.

Les informations linguistiques diffèrent selon les 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.

⁵ Appelé également pula(a)r, ful ou fulfulde.

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 (appelée 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'espèce⁶ 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 utilisation*

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, 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.

⁶ bourgeon, branche, cendre, écorce, feuille, fleur, fruit, gomme, épine, graine, inflorescence, racine, tronc.

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, personnels 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

Enoncer 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 dans cette langue préalablement saisies ; 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 Etats-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 croisé : °; 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 ŋ), 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 Etat 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 certains composants 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 on veut que l'Afrique, par le virage culturel que permet de réaliser Internet, puisse s'approprier les sciences et les techniques.

*Chantal Enguehard,
Institut de Recherche en Informatique de Nantes,
Nantes,*

France,

chantal.enguehard@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) ». Agence de coopération culturelle et technique, ISBN 92.90.28.208.8, Paris.

Chanard (C.), Popescu-Belis (A.), 2001. « Encodage informatique multilingue : application au contexte du Niger ». Cahiers du Rifal, 22, p.33-45, décembre.

Durand (C.-X.), 2002. « La nouvelle guerre contre l'intelligence II. La manipulation mentale par la destruction des langues ». O.E.I.L., F.X. deGuibert, Paris.

Fortin (D.), Lô (M.), Maynard (G.), 1997. « Plantes médicinales du Sahel ». CECI 1990, ENDA 1990. ISBN 2-920114-0-2.

von Maydell (H.-J.), 1983. « Arbres et arbustes du Sahel, leurs caractéristiques et leurs utilisations ». GTZ.

Mbodj (Ch.), 1994, « L'activité terminologique au Sénégal », dans *Terminologies nouvelles*, n° 11, pp.3-9

Oumarou (I. A.), 2000. « Zarma ciine kaamuusu kayna », Niamey, ACCT ; Editions 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 "mérissa", 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 Sahel" CECI 1997, ENDA 1997. ISBN 2-920114-0-2, 1990.
- von Maydell H.-J., "Arbres et arbustes du Sahel, 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 :

- Enoncé : Hano kaŋ koo ga wura hay, talka si duu goobu kaŋna a catu.