

Le projet Base de données lexicographiques panfrancophone du réseau Étude du français en francophonie

Évariste Ntakirutimana

Département des Langues modernes
Université nationale du Rwanda, Butare
entakirutimana@nur.ac.rw



Synergies Afrique des Grands Lacs n° 1 - 2012
pp. 67-74

Résumé : L'établissement d'une base de données lexicographiques panfrancophone (BDLP) exige beaucoup d'attention, de patience, de jugement et de minutie aussi bien dans la sélection des particularismes que dans leur présentation pour une mise en ligne. Cet article se propose de faire le point sur la structure, la présentation et la consultation d'une BDLP.

Mots-clés : base de données ; recherche simple ; recherche transversale ; BDLP ; particularité lexicale ; vedette ; fiche.

Abstract: The development of a panfrancophone lexicographical data base (BDLP) requires important attention, patience, judgment and meticulousness in the selection of lexical particularisms as well as in their presentation for their online reference. This paper describes the structure and the design of a BDLP and indicates how to refer to.

Keywords: data base ; simple reference ; transversal reference ; BDLP ; lexical particularism ; headword ; lexical slip.

Introduction

La Base de données lexicographiques panfrancophone (BDLP) est un projet de couverture internationale qui relève de la lexicographie différentielle. Son objectif général est de constituer et de regrouper les bases représentatives des particularités lexicales du français de chacun des pays et/ou de chacune des régions de la Francophonie. Les différentes bases disponibles sont conçues de manière à pouvoir être consultées séparément (recherche simple) ou comme un unique corpus (recherche transversale) et à servir de complément au *Trésor de la langue française informatisé*, implanté au centre Analyses et traitements informatiques du lexique français (ATILF) de Nancy (www.bdlp.org).

Le projet de constitution de bases de données regroupant les particularités lexicales de chaque pays, voire de chaque région, de la Francophonie est une entreprise du *Trésor des vocabulaires français*, lancée par le professeur Bernard Quemada dans les années 80. Il est patronné par l'Agence universitaire de la Francophonie qui l'appuie via son réseau *Étude du français en francophonie*.

La BDLP internationale a été mise en ligne le 18 mars 2004, dans le cadre de la Semaine de la Francophonie. Cette mise en ligne a suscité l'intérêt des chercheurs qui, par la suite, se sont mis à l'œuvre pour augmenter le nombre de bases. Depuis 2004, les bases ont été annuellement complétées et mises à jour en mars, toujours dans le cadre de la Semaine de la Francophonie. Actuellement, on comptabilise 20 bases de données en ligne et 23 en cours de réalisation. L'établissement d'une BDLP demande beaucoup d'attention, de patience, de jugement et de minutie. Mais une fois complétée et mise en ligne, sa consultation devient captivante surtout quand on effectue une recherche transversale.

1. Aide pour l'établissement d'une BDLP

Les équipes travaillant sur les bases respectives bénéficient ordinairement de l'aide technique nécessaire. Elle consiste en une formation, de trois à quatre semaines, aux techniques d'établissement de la BDLP à l'aide du logiciel *FileMaker Pro*. Tout le monde doit recevoir les mêmes consignes de manière à travailler uniformément et, partant, aboutir à une présentation uniforme des résultats. La formation aux différentes conventions de codage se déroule, jusqu'à nouvel ordre, à l'université Laval (Québec, Canada) qui abrite le *Trésor de la langue française au Québec* (TLFQ).

Une aide financière est généralement octroyée aux chercheurs du Sud dont les moyens sont plus limités. Cette aide consiste en la mise à disposition de fonds modestes permettant d'acheter les ouvrages de référence utiles pour l'exemplification (romans, nouvelles, manuels scolaires, journaux, etc.) et le petit matériel. Elle est en principe répartie sur deux ans de recherche.

2. Structure d'une BDLP

Tout le travail se fait en vue de la mise en ligne des données qui facilite la consultation et l'échange. Pour ce faire, un modèle de fiche a été conçu à l'aide du logiciel *FileMaker Pro*. La fiche contient les informations suivantes :

2.1. En-tête de la fiche : 14 renseignements

- L'index BDLP web : il sert à ordonner les différentes vedettes relatives à un même terme, à créer la liste des entrées pour la recherche simple et, enfin, à créer les liens hypertextes.
- La vedette : elle correspond à la forme à décrire.
- La catégorie grammaticale : elle donne les renseignements grammaticaux en rapport avec la vedette.
- L'état des données : on indique dans ce champ à quelle étape se trouve la recherche (sommaire, en cours, avancée).
- L'inventaire : il réfère au pays ou à la région concernée.
- La zone : elle est en quelque sorte liée à l'inventaire. Au bout du compte, les pays francophones sont conventionnellement répartis en huit zones, à savoir l'Afrique subsaharienne, l'Amérique du Nord, les Caraïbes, l'Europe, l'Extrême-Orient, le Maghreb, le Moyen-Orient et l'océan Indien.
- Le sens : dans le cas précis de la BDLP, le sens sert à structurer les différentes vedettes d'un même mot.

- La prononciation : on indique comment prononcer la vedette.
- La définition : elle doit absolument rendre compte de l'emploi de la vedette.
- La syntagmatique : on y propose un exemple d'emploi de la vedette.
- Les remarques : dans cette rubrique, on fournit d'éventuelles informations additionnelles en rapport avec la vedette, tant sur le plan de sa graphie que de sa catégorie grammaticale et de sa prononciation.
- Les variantes graphiques, morphologiques et polymorphiques : on donne respectivement les variantes graphiques, morphologiques et polymorphiques de la vedette, quand elles existent.

2.2. Données dictionnairiques : 7 renseignements

- Le registre : il renseigne sur le niveau de langue qui caractérise l'emploi concerné (familier, littéraire ou soigné, populaire, vulgaire).
- La marque diachronique : elle renseigne sur la vitalité de l'emploi décrit (disparu, historique, néologisme, vieilli et vieux).
- La marque stylistique : ce champ donne des précisions relatives à l'emploi décrit dans la fiche.
- La dérivation, la composition et les mots de même famille : on y note toutes les formes issues des procédés de dérivation, de composition et les formes relevant de la même famille que la vedette.
- Les commentaires linguistiques : on y consigne toutes les informations qui ne concernent directement ni la forme de la vedette, ni sa catégorie grammaticale, ni sa prononciation.
- Les avis et recommandations : ils font état d'avis ou de recommandations émanant d'organismes officiels, quand il y en a.
- La clé d'indexation : elle complète l'index BDLP web en permettant, dans le cas des vedettes complexes contenant plus d'un particularisme ou n'en contenant aucun, de rattacher la vedette à un autre mot qui apparaîtra, au même titre que celui que comprend l'index BDLP web, dans la liste des entrées de la recherche simple dans la version web.

2.3. Données encyclopédiques et historiques : 8 renseignements

- L'origine de l'emploi : la fiche propose, d'une part, une liste de valeurs cliquables indiquant le phénomène qui est à l'origine de la vedette. D'autre part, elle propose une liste de langues pouvant être la source directe de l'emploi en vigueur.
- La première attestation : on mentionne l'année au cours de laquelle la vedette a été utilisée pour la première fois. Mais comme on doit s'en douter, cette information n'est pas toujours facile à préciser.
- L'étymon proche : on y indique l'étymon de la vedette, ce qui permet d'effectuer des recherches par étymon dans la version en ligne.
- L'étymon du FEW : cette rubrique indique l'étymon sous lequel on repérera la vedette dans le *Französisches etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes* (Wartburg (von), 1922).
- L'étymologie/historique : on y fournit les renseignements d'ordre étymologique et historique de la vedette, dans la mesure du possible.
- Encyclopédie : on y consigne toute information de nature encyclopédique relevant de l'emploi particulier donné dans la fiche.
- Bilan métalinguistique : on y fait le bilan des commentaires d'ordre métalinguistique relatifs à l'emploi donné dans la fiche.
- Renvois à d'autres langues : on mentionne les équivalents connus dans une ou plusieurs autres variétés que le français de référence.

2.4. Champ sémantique : 4 renseignements

- Classement conceptuel : il s'agit d'une liste de valeurs cliquables permettant de bien situer le signifié dans trois champs sémantiques présélectionnés qui distinguent /hommes/, /animaux/ et /choses/.
- Synonymes : synonymes ou quasi-synonymes de la vedette dans la variété décrite.
- Antonymes : comme dans le cas des synonymes, on note les contraires de la vedette dans la variété décrite.
- Renvois onomasiologiques : on indique les termes relevant du même champ sémantique que la vedette.

2.5. Francophonie : 3 renseignements

- Les boutons : il est question de deux renseignements permettant de situer la vedette par rapport au français standard. Le premier demande si la réalité appartient à la variété décrite ou pas. Le second demande s'il s'agit d'un mot particulier par son statut ou pas.
- Les commentaires « intrazones » et renvois à d'autres zones : Dans cette rubrique, on mentionne les autres zones dans lesquelles la vedette est utilisée. Cela permet de créer des liens entre les différentes bases et d'effectuer la recherche transversale dans la version en ligne.
- Le français de référence : cette rubrique réfère à l'unité lexicale du français standard qui équivaut sémantiquement à la vedette.

2.6. Les citations : 9 renseignements

- Citation : on y consigne les extraits des sources écrites pour illustrer l'attestation du terme mis en cause.
- Référence : on mentionne la référence bibliographique de la citation.
- Médium : ce champ fait référence au support duquel la citation a été tirée (source orale ou écrite).
- Type de source : on précise la nature du document d'où la citation a été extraite (archives, littérature, textes administratifs, presse, sites internet, etc.).
- Année : année de publication de la source.
- Auteur : nom de l'auteur de la source.
- Métalinguistique : on a affaire ici à de valeurs cliquables permettant d'indiquer s'il s'agit de sources métalinguistiques ou de citations métalinguistiques.
- Renseignements bibliographiques : renseignements en rapport avec l'auteur ou la source.
- Fichier BDLPcitat : fichier permettant de modifier le contenu des fiches Citations en cas de besoin. On y accède par le menu Fenêtre dans la fiche.

2.7. Répartition géolinguistique : une liste de régions prédéterminées est fournie. On a juste à sélectionner celle qui correspond à la base impliquée.

2.8. Fichiers multimédias : il est possible d'insérer dans les fiches des sons, images, musiques et vidéos. Les fichiers « son » permettent au lecteur d'entendre la prononciation des termes ou des syntagmes. Les fichiers « image », « musique » et « vidéo » donnent les illustrations visuelles ou sonores de la vedette en question.

2.9 Fichiers BDLPderiv : fichiers qui contiennent deux types d'informations : d'une part, la mention de tous les dérivés de la vedette, quand ils sont disponibles dans la variété décrite, et, d'autre part, les renvois onomasiologiques. On accède au fichier BDLPderiv par le menu Fenêtre dans la fiche.

3. Consultation de la BDLP

Une fois les fiches apprêtées et bien codées, elles sont mises en ligne, après vérification par l'équipe du TLFQ. Chaque base est repérable par le nom et le drapeau du pays concerné. Deux possibilités de recherche du mot sont offertes : l'utilisateur peut effectuer une recherche simple ou une recherche transversale.

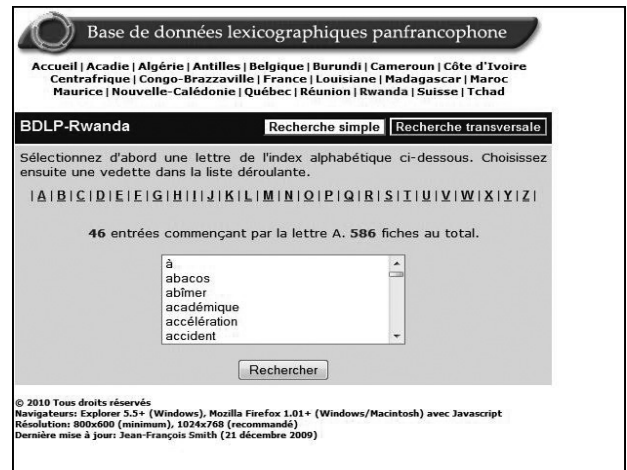
a) Recherche simple

La recherche simple consiste à rechercher un terme dans une base déterminée. Dans ce cas précis, on clique sur le drapeau ou le nom de la base souhaitée. Un clic sur le drapeau du Rwanda (<http://www.BDLP.org/rwanda.asp>) donne l'icône suivante :



En cliquant sur *Basculer vers la recherche simple* (http://www.BDLP.org/simple.asp?base=BDLP_rwanda), on obtient les informations suivantes :

Le nombre de fiches disponibles dans la base s'affiche automatiquement. On peut alors naviguer dans la base en recherchant les termes par ordre alphabétique. En se positionnant sur l'initiale du terme à rechercher, tous les termes commençant par cette initiale s'affichent. Il ne reste plus qu'à choisir le terme voulu et à cliquer sur *Rechercher*.



b) Recherche transversale

La recherche transversale consiste à rechercher un terme dans toutes les bases disponibles. Cette option permet de voir rapidement l'expansion du terme. On y accède en se positionnant sur base n'importe quelle base de la page d'accueil (BDLP-Rwanda ici), et en cliquant sur *Recherche transversale* (www.BDLP.org/transv.asp?base=BDLP_rwanda), on obtient les données suivantes :

The screenshot shows the 'Base de données lexicographiques panfrancophone' interface. At the top, there are navigation links for various regions: Accueil | Acadie | Algérie | Antilles | Belgique | Burundi | Cameroun | Côte d'Ivoire | Centrafrique | Congo-Brazzaville | France | Louisiane | Madagascar | Maroc | Maurice | Nouvelle-Calédonie | Québec | Réunion | Rwanda | Suisse | Tchad. Below this, the 'BDLP-Rwanda' header is visible, with buttons for 'Recherche simple' and 'Recherche transversale'. A menu bar includes 'Recherche de mots', 'Citations', 'Marques', 'Grammaire', 'Champ sémantique', 'Avis et critiques', 'Géolinguistique', 'Étymologie/Historique', and 'Statut de l'emploi'. A 'Francophonie' button is highlighted. The main text explains that the search covers all rubrics and allows filtering by content. A 'Lancer la recherche' button is at the bottom. Footer information includes copyright for 2010, browser requirements (Explorer 5.5+, Mozilla Firefox 1.01+), resolution (800x600), and the last update date (21 décembre 2009).

Un clic sur l'option *Francophonie* (www.BDLP.org/transv.asp?base=BDLP_rwanda), donne les informations suivantes:

On coche l'option *Toutes les bases*, puis on clique sur l'option *Recherche de mots*. On obtient *ipso facto* une case dans laquelle il faut inscrire le terme à rechercher. Pour l'exemple, si l'on inscrit le terme *maquis* que l'on retrouve dans la BDLP-Rwanda, puis que l'on clique sur l'option *Rechercher*, on obtient la vedette en question, la définition et la liste des pays dans lesquels le terme est attesté (www.BDLP.org/resultats.asp?base=BDLP_rwanda&mode=trans).

This screenshot shows the 'Sélection par base' and 'Sélection par zone' sections of the search interface. Under 'Sélection par base', there are checkboxes for various regions: Acadie, Antilles, Burundi, Centrafrique, Côte d'Ivoire, Maurice, Madagascar, Nouvelle-Calédonie, Réunion, Suisse, Algérie, Belgique, Cameroun, Congo-Brazzaville, France, Louisiane, Maroc, Québec, Rwanda, Tchad, and 'Toutes les bases'. Under 'Sélection par zone', there are checkboxes for: Afrique subsaharienne, Amérique du Nord, Caraïbe, Europe, Extrême-Orient, Maghreb, Moyen-Orient, and Océan Indien.

Base de données lexicographiques panfrancophone		
Accueil Acadie Algérie Antilles Belgique Burundi Cameroun Côte d'Ivoire Centrafrique Congo-Brazzaville France Louisiane Madagascar Maroc Maurice Nouvelle-Calédonie Québec Réunion Rwanda Suisse Tchad		
Rechercher vers la recherche simple		
Nouvelle recherche transversale		
Vedette(s)	Définition(s)	Francophonie
1. maquis 01. (n. m.)	(Arg. étudiant). Chambre prêtée à un étudiant qui ne sait pas où se loger, ou partagée avec lui.	
2. maquis 01. (n. m.)	Retraite, fait de se retenir, de vivre à l'écart, en vue de se préparer pour une épreuve (examen pour des élèves, match important pour des sportifs, concert pour des groupes musicaux).	
3. maquis 01. (n. m.)	Retraite, fait de se retenir, de vivre à l'écart, en vue de se préparer pour une épreuve (examen pour des élèves, match important pour des sportifs, concert pour des groupes musicaux).	
4. maquis 01. (n. m.)	Logement qu'on partage avec quelqu'un sans en être le propriétaire ni l'ayant droit.	
5. maquis 02. (n. m.)	(Arg. étudiant). Situation de l'étudiant qui n'est pas logé et qui partage la chambre d'un autre étudiant.	
6. maquis 02. (n. m.)	Lieu où l'on se rebrte afin de préparer une épreuve importante. Loc.Aller dans le maquis, être au maquis. ("faire retraite").	
7. maquis 02. (n. m.)	Lieu où l'on se rebrte afin de préparer une épreuve importante.	
8. maquis 03. (n. m.)	Débit de boisson, restaurant plus ou moins clandestin.	
9. maquis (donner le...) 03. (loc. verb.)	(Arg. étudiant). Partager sa chambre (avec un étudiant qui n'est pas logé).	

Conclusion

L'établissement d'une base de données est un travail de longue haleine, mais passionnant. Il exige rigueur et minutie extrêmes et sollicite beaucoup d'informations d'ordre linguistique et paralinguistique. Chaque information est codée de manière précise. Toutes ces exigences visent à faciliter la mise en ligne et, donc, la consultation des données. Le fichier en ligne est interactif à plus d'un titre. Nous aimerions inviter les linguistes, à participer activement à cette entreprise d'envergure internationale et d'importance historique pour le français. Que chacun reste convaincu que sa contribution aura une part déterminante pour l'avenir du trésor de la langue française.

Pour des raisons de commodité, nous avons fait ici l'économie d'un certain nombre de difficultés liées à l'établissement d'une BDLP, au nombre desquelles la sélection des particularismes, l'emprunt aux langues locales et les techniques de codage. Il n'est, en effet, pas toujours facile de savoir si telle unité lexicale de son propre parler est un particularisme ou pas, surtout quand elle est utilisée dans tout le pays ou toute la région, comme c'est le cas pour un certain nombre d'unités lexicales du français du Rwanda. Un regard externe est souvent requis pour mieux discerner. Le français du Rwanda, à titre d'exemple retient beaucoup de termes du français de Belgique en raison des circonstances historiques qui lient la Belgique aux pays des Grands Lacs africains.

Certaines unités lexicales locales, voire régionales, sont francisées pour des raisons variées allant de la méconnaissance du français (incompétence ou moindre effort) au manque d'équivalents voire à l'impossibilité de traduction. La tendance générale est de prendre tous les termes du terroir et de leur donner une apparence française. Cela trompe le chercheur et gonfle inutilement le nombre de particularismes, comme c'est le cas chez Jouannet (1984).

Les techniques de codage dans la BDLP sont nombreuses et font l'objet d'une formation d'au moins trois semaines. Elles ont été élaborées dans le but ultime d'harmoniser la recherche et, partant, de faciliter la mise en ligne et la consultation des résultats. Le nombre important de conventions et de champs à compléter rend le travail très ardu. D'après notre expérience, les données à inscrire dans les champs *Encyclopédie* et *Champ sémantique* ne sont pas toujours évidentes. Comme d'ailleurs le choix entre *Réalité propre à la variété décrite* et *Mot particulier par son statut* que l'on retrouve dans le champ *Francophonie*.

Bibliographie

Bacon, N., C. Poirier et J-F Smith, 2004. *Base de données lexicographiques panfrancophone (BDLP) : guide de saisie des données*. Québec : TLFQ.

Jouannet, F., 1984. *Le français au Rwanda : enquête lexicale*. Paris : SELAF.

Massoumou, O., et A.-J.-M. Queffélec, 2007. *Le français en République du Congo sous l'ère pluripartiste (1991-2006)*. Paris : Éditions des Archives contemporaines, AUF.

Ntakirutimana, E., 2010. « Le français au Rwanda ». *Revue des Observatoires du français contemporain en Afrique*, n°25, pp. 19-31.

Wartburg (von), W., 1922. *Französisches etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, t. 1. Bonn : Kurt Schroeder.