

## Essai de dictionnaire *tasaḥlit*<sup>1</sup> (parler kabyle d'Aokas)-français : conception lexicographique et modélisation informatique

Abdelaziz BERKAI<sup>1</sup> & Paul ANDERSON<sup>2</sup>

1, Département de Langue et Culture Amazighes

1, Université A. Mira de Bejaïa

2, Ingénieur en informatique et chercheur indépendant

### Agzul

Amagrad-a, akken i t-id-yemmal uzwel-is yakan, yebda yef sin yeḥricen yemgaraden. Amezwaru d asegi n tizri d tesnarrayt i yettwadefren deg usali n yiwen usegzawal amsinwaltasaḥlit (*tameslayt n Weqqas*)-tafransist i yellan d tazrawt d Duçtura n tesnilest n tmaziyt. Aḥric wis sin d tazrawt tansenselkimt yef wudem alikeṭrunan n usegzawal-a ara yessifessen anadi n yisallen deg-s d tmerna d ubeddel n wiyad i usimyer d uwenneε-ines.

### Abstract

The study, as its title suggests, is twofold. The first considers the theoretical and methodological approaches having been adopted in the elaboration essay of a *bilingual dictionary Tasaḥlit (Aokas language)-French* elaborated as a part of doctoral thesis in Berber linguistics. The second examines the computer design of an electronic version of this dictionary which would allow both an easy interrogation of the work and the equally easy possibility of intervening to improve or enrich it.

**Keywords:** dictionary, electronic version, bilingual lexicography, Kabyle (Aokas language), French.

La présente contribution se compose de deux parties complètement distinctes<sup>2</sup> :

- I. Une brève présentation théorique d'un *Essai de dictionnaire* fait dans le cadre d'un doctorat de linguistique amazighe ayant pour intitulé : *Essai d'élaboration d'un dictionnaire tasaḥlit (parler d'Aokas)-français* (disponible en ligne sur le site de l'Université de Tizi-Ouzou) ;

---

<sup>1</sup>*Tasaḥlit* (de l'arabe *saḥel* « littoral ») désigne les parlers kabyles du littoral extrême oriental (Aokas, Souk Letnin, Melbou...). Cette désignation qu'utilisent les habitants de la région ne recoupe pas exactement ce que Naït-Zerrad appelle le groupe linguistique extrême oriental, un des quatre grands groupes qu'il distingue en Kabylie, qui comprend par exemple le parler des « AtSmail » qu'il cite dans son article et qui n'est pas un parler *tasaḥlit*. Il connaît par exemple le phénomène de labiovélarisation qui n'est pas attesté dans les parlers du littoral à l'est de Béjaïa ni même d'ailleurs dans cette dernière ville et ses environs. AytSmailon dit par exemple *egg*<sup>w</sup> « être cuit ».

<sup>2</sup>La première est rédigée en français, la seconde en anglais.

II. Une présentation par Paul Anderson d'une modélisation informatique de cet *Essai de dictionnaire* pour en faire un outil électronique facile à consulter et à modifier.

## I. Théorie et méthodologie

Notre contribution porte sur une expérience d'élaboration d'un dictionnaire bilingue parler kabyle d'Aokas<sup>3</sup>-français dans le cadre d'une thèse de doctorat soutenue en 2014 (v. bibliographie). Nous voudrions aborder ici rapidement trois aspects de notre recherche lexicographique :

1. La méthode utilisée pour le recueil et la constitution du corpus ;
2. Les spécificités lexicales du parler d'Aokas au niveau kabyle et pan-berbère ;
3. Les choix macro- et microstructurels faits pour le traitement lexicographique des données.

### 1. Le corpus

Le lexicographe dispose à travers le corpus, pour peu qu'il soit correctement constitué, d'un matériau qui lui permet d'étudier les unités lexicales dans leur milieu naturel et de connaître précisément leur fonctionnement sémantique, morphologique et syntagmatique et même sociolinguistique en sachant le contexte, la situation d'usage et le statut du locuteur (âge, origine géographique, profession, etc.).

Un corpus lexicographique « normalement » constitué doit satisfaire, dans une certaine mesure, à deux principes :

- a. **Le principe d'extensivité** : le corpus doit inclure le maximum de domaines et de registres possibles impliqués par les objectifs et la nature du dictionnaire ;
- b. **Le principe d'exhaustivité** : il doit représenter de façon exhaustive le lexique d'un domaine particulier, concerné bien entendu par la nomenclature (v. Lafage, 1997 : 88).

Notre essai de dictionnaire est élaboré *in extenso* à partir d'un corpus. Nous avons conçu deux types de sous-corpus complémentaires : le « sémasiologique » qui repose sur un discours « disponible » (à recueillir) et l'« onomasiologique » qui résulte d'enquêtes ciblées qui partent de signifiés, notions ou concepts, pour rechercher les dénominations qui leur correspondent. Trouver un seul signe linguistique et ses diverses acceptions peut être le résultat de la combinaison des deux démarches à la fois. Trouver, par une approche onomasiologique le signifiant en partant de l'un de ses signifiés, puis, par une démarche sémasiologique, les autres signifiés en partant du signifiant en question. Ce sont évidemment les deux approches combinées qui nous ont permis de constituer le nôtre<sup>4</sup>. Un corpus varié constitué de divers types de textes : des poèmes, des proverbes, des

---

<sup>3</sup>Aokas (*Weqqas*, pour ses habitants), chef-lieu de Daïra (sous-préfecture), est une petite ville côtière située sur la route nationale n° 9, à 25 km à l'Est de Béjaïa sur le littoral méditerranéen.

<sup>4</sup> Nous avons développé cette idée dans un article intitulé : « L'intérêt du corpus et une idée de sa constitution en lexicographie amazighe », dans Tiziri N. (dir.), *Corpus : exploitation, théories et méthodes, Iles D Imeslin*° 5, 2013, p. 281-293.

devinettes, des contes, des textes en prose, des énoncés disponibles, sollicités ou produits ad hoc et recoupés, ainsi que de simples mots. Un corpus dont nous avons dégagé environ 3500 racines discriminées par le sens et ayant fourni quelque 5500 entrées et des milliers d'équivalents dans la langue cible.

Beaucoup de ces racines sont spécifiques au parler d'Aokas non seulement à l'échelle du kabyle, mais bien au niveau panberbère.

Notre dictionnaire est élaboré *in extenso* à partir d'un corpus dont les données sont systématiquement traitées : du premier jusqu'au dernier mot, quel que soit son appartenance catégorielle.

Nous avons conçu deux types de sous-corpus complémentaires : le

« sémasiologique » qui repose sur un discours « disponible » (à recueillir) et l'« onomasiologique » qui résulte d'enquêtes ciblées qui partent de signifiés, notions ou concepts, pour rechercher les dénominations qui leur correspondent.

La sémasiologie est l'approche adoptée généralement en lexicographie et qui a l'avantage de recueillir les mots dans leurs contextes discursifs en fonction de leur fréquence dans le discours : plus ils sont fréquents, plus leur occurrence est grande dans le corpus. Mais il se trouve qu'une partie non négligeable de « mots courants » ne se retrouve pas dans ce type de recueil. Ce sont des mots usuels « qui se dérobent à la statistique » (J. Picoche). Des mots très connus mais pas nécessairement sollicités. Ce sont les mots qu'on appelle « disponibles » et qui complètent les « fréquents » dans l'ensemble des mots usuels. Une autre approche est nécessaire pour les « débusquer ». C'est précisément celle qui suit le chemin inverse en partant des signifiés, relevant d'un domaine particulier, pour rechercher les signifiants qui leur correspondent. C'est la deuxième approche évoquée ci-dessus, à savoir l'onomasiologie. Elle nous permet, à titre d'exemple, de recueillir tous ou l'essentiel des vocables relatifs au corps humain. Ce qui n'est pas possible de faire avec la première approche. Nous avons donc deux types de corpus que nous avons appelé par convention « corpus sémasiologique » et « corpus onomasiologique ».

Trouver un seul signe linguistique et ses diverses acceptions peut être le résultat de la combinaison des deux démarches à la fois. Trouver, par une approche onomasiologique, le signifiant en partant de l'un de ses signifiés, puis, par une démarche sémasiologique, les autres signifiés en partant du signifiant en question.

### **1.1. Le sous-corpus sémasiologique**

C'est un corpus constitué à partir de deux sources différentes : des textes écrits dans le parler d'Aokas, notamment les travaux de Slimane Rahmani, Louisa Yaker née Rahmani, Allaoua Rabhi et Ourida Aïssou, et d'autres recueillis par nous-mêmes auprès de nos informateurs. Un corpus varié, constitué de différents types de textes : contes, proverbes, devinettes, textes en prose, poèmes, etc.

Concernant les textes écrits dans le parler d'Aokas et traduits, mis à part les cas de confusion dans la transcription ou dans la traduction, notre

transcription du discours rapporté en kabyle et traduit en français est quasiment la même que celle de l'auteur. Son adaptation systématique ne nous aurait pas permis, en cas d'erreur, de nous corriger par un retour au corpus. Les adaptations sont donc faites plutôt au niveau de la partie dictionnaire. S'agissant des textes que nous avons nous-mêmes recueillis leur traduction et transcription (restitution du sens et de la forme sonore) s'est faite avec l'aide précieuse de nos informateurs qui sont souvent les auteurs des textes en question. Les informations sont souvent recoupées et complétées avec d'autres informateurs.

## **1.2. Le sous-corpus onomasiologique**

Nous sommes partis ici de notions relevant de certains champs lexicaux contenant des vocables qui sont, pour une bonne part, d'usage courant dans la langue : le vocabulaire de la cuisine, de la literie, du corps humain, des plantes, des arbres, des animaux, etc. Une liste de notions, généralement définies (en kabyle), de chaque champ lexical est établie et nous demandions à nos informateurs pour chacune d'entre elles de nous donner son « nom », et ses différentes variantes morphologiques (genre, nombre et état d'annexion), dans leur parler. Cela concerne surtout les substantifs concrets qui ne sont pas fréquents dans le discours, alors que les mots grammaticaux et les verbes y sont relativement bien représentés.

Cette partie est donc conçue pour compléter la première concernant surtout certains substantifs relevant des mots disponibles qui risqueraient d'échapper aux mailles trop larges d'un modeste corpus textuels. Mais par souci d'exhaustivité, nous avons essayé de recueillir l'essentiel du vocabulaire de chaque champ, de sorte que beaucoup de mots recueillis ne sont pas catégorisables dans les mots « disponibles ».

A cela s'ajoute un important **corpus complémentaire** constitué de 1731 énoncés produits ou sollicités auprès de nos informateurs. Un mélange de séma- et d'onomasiologie. Certains énoncés sont produits par nos informateurs au cours de notre enquête, nous les avons transcrits et traduits avec leur concours. Ils relèvent donc de la sémasiologie. D'autres par contre sont produits par nous-mêmes, avec leur traduction précise, et nous avons cherché auprès de nos informateurs à avoir la forme et le sens corrects dans le parler d'Aokas. Ils relèveraient donc de l'onomasiologie. On retrouve ici évidemment beaucoup de verbes qui complètent donc les listes nominales évoquées ci-dessus.

## **1.3. Les enquêtes de terrain**

Nos enquêtes se déroulaient d'abord dans les sièges de certaines associations locales, à la maison de la culture ou à la maison de jeunes, mais par la suite l'essentiel des entretiens que nous avons eus avec nos informateurs se faisaient dans notre véhicule. Le contact est plus direct, la discussion plus ouverte et libre. Il n'y a personne à interférer pour couper, détourner ou carrément arrêter les discussions comme ce fut souvent le cas dans les locaux évoqués ci-dessus. Nous garions notre véhicule dans des endroits calmes et beaux dont la région regorge, notamment sur des chemins

forestiers ou au bord de la mer, et nous interrogeons tranquillement nos informateurs, souvent accompagnés d'une boisson, en prenant note sur un cahier. La bière que la plupart de mes informateurs apprécient favorise la détente et la décontraction et délient les langues. Les entretiens se déroulent généralement le week-end et durent en moyenne deux à trois heures.

Nous recueillions souvent au terme d'un entretien des mots et des énoncés nouveaux à rajouter au corpus, ainsi que des corrections portant sur le sémantisme ou la morphologie d'autres mots déjà recueillis. Au cours de notre travail de complétion du corpus et du traitement lexicographique des unités concernées par l'ajout ou la correction, surgissent toujours de nouveaux questionnements concernant une acception, une désinence ou carrément l'existence de tel ou tel autre mot. Ce qui entraîne un va-et-vient permanent entre le terrain et le bureau. Et le terrain appelle toujours le terrain. Le travail sur le lexique n'est jamais fini.

La plupart de nos principaux informateurs sont de la gent masculine. Des gens qui s'intéressent à leur langue et qui sont plus ou moins disponibles, les uns plus que les autres. Certains d'entre eux sont des poètes dont une partie plus ou moins importante des textes est recueillie dans notre corpus. Ils se sont montrés disponibles non seulement pour les enquêtes concernant leurs textes, mais pour l'ensemble du corpus.

En plus de ces informateurs réguliers, nous avons eu beaucoup d'autres de façon occasionnelle, notamment des étudiants et étudiantes de la région auxquels nous demandions des renseignements précis à recueillir auprès de vieilles personnes de leur entourage. Une foule d'informations est ainsi recueillie auprès de cette catégorie de la population. L'information est soit transcrite ou enregistrée avec un dictaphone ou un portable. Elle est ensuite discutée et recoupée avec d'autres informateurs.

#### **1.4. La transcription**

Nous avons adopté une transcription à tendance phonétique pour mieux montrer l'usage de la langue et faciliter son analyse. Les assimilations sont par exemple rendues au niveau de l'écrit, de même que les dissimilations.

La consonne *p* est retenue dans notre système de transcription et dispose même d'une rubrique spéciale eu égard aux nombreux emprunts aux français qui sont totalement intégrés dans la langue : *apulis* (*ipulisen*) « policier », *apayaş* (*ipayaşen*) « matelas ; paille », *apuţu* (*ipuţuyen*) « poteau (haut supportant les fils téléphoniques ou électriques) », *apansyu* (*ipansyuyen*) « pension (allocation périodique) », *plastik* « le plastique (matière) », *apurtabl* « portable (téléphone) », etc. Cette consonne possède même une fonction expressive dans cette langue, à travers des exemples comme *pappa* (langage enfantin) « pain ». La fonction expressive de la consonne *p* est évidemment attestée dans tous les parlers kabyles. En Grande Kabylie elle est même utilisée comme marqueur du langage féminin dans des mots où elle remplace chez cette catégorie de locuteurs la bilabiale occlusive sonore tendue labiovélarisée ou non [B/B<sup>w</sup>]. Des mots comme *metabburt* « porte », *ebb<sup>w</sup>* « cuire », *abbu* « fumée », etc. sont articulés *tappurt*, *epp<sup>w</sup>*, *appuchez* la gent féminine. Dans certains villages comme Azrou dans la région de Chemini (à

60 km à l'ouest de Béjaïa), la bilabiale occlusive tendue [B]est systématiquement articulée comme une sourde [P] par tous les locuteurs.

## 2. Quelques spécificités lexicales du parler *tasahlit* d'Aokas

### 2.1. Spécificités sémantiques et/ou formelles

Beaucoup de mots (forme et/ou sens) du parler d'Aokasne sont pas attestés non seulement dans les parlers de la Soummam et de Grande Kabylie (désormais S et GK) plus connus et dont une partie importante du lexique est recueillie dans le dictionnaire kabyle-français de J.-M. Dallet, mais même *a priori* au niveau panberbère si l'on s'en tient évidemment aux seuls matériaux lexicographiques existants. En voici quelques-uns :

- *tajla* (*tajliwin*) n. f. (arch.) : « don héréditaire [d'une femme dans le traitement de certaines maladies] » ;
- *abelliṭ(u-ibelliṭen)* n. m. : « absurdité, invraisemblance » ;
- *berber* [*bərbəʁ*] (*iberber*, *tberbir*, *aberber*) v. tr. : « être pleinement satisfait de qqch, en avoir fait plus qu'il n'en faut » ;
- *jallaj*, n. m. sing. : « cholestérol » ;
- *mend(imend, tmenday, amend)* v. tr. : « être dans l'attente d'être fait » ;
- *nnettwa(nnettwa)* n. f. : « parenté proche, proches parents » ;
- *rzem(irzem, ulirzim, rezzem, arzam/rrezza)* (plus usité) v. tr. : « prendre une quantité (d'eau, en partic.) sans remplir son récipient, remplir à moitié un récipient » ;
- *slulli(islulli, slullay, aslulli)* v. intr. : « se marier ; se faire beau/belle (comme au jour du mariage) » ;
- *ssugger(issugger, ssuggur, asugger)* v. tr. : « s'en sortir, se tirer (d'affaire), se débrouiller » ;
- *tisrit(te-/tisriyin(te-))* n. f. : « peau épaisse des animaux en général, qu'on emploie parfois péjorativement pour l'homme » ;
- *zmet(izmet, ulizmit, izemmet, azmat)* v. mixte : « être fort, impétueux ; faire qqch de façon forte, impétueuse » ;
- *zzinzer(izzanzer, ttinzir, azinzer)* v. intr. : « aller et venir continuellement, faire la navette » ;
- *amezrub(u-/imezruben)* n. m. : « petit poème de neuf vers ; poème court et concentré » ;
- *suq(isuq, ttsuq, asuq/ssuqan/tasuqit)* v. intr. : « devenir rare (fruit en fin de saison) » (opposé à *ssemṭi* « commencer à mûrir » > *ssuqan/isuqan*, n. m. : « début de la fin de qqch ; vers la fin (saison, récolte...) »), etc.

### 2.2. Des mots et des « désignations »<sup>5</sup> culturels

---

<sup>5</sup> Voir notre article intitulé : « Quelques problèmes lexicographiques que posent l'établissement d'équivalences et leur organisation dans un dictionnaire kabyle-français », dans Van Campenhoutd, M., Lino, T. et Costa, R. (dir.), *Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité*, Actes des

*Essai de dictionnaire tassahtit<sup>l</sup> (parler kabyle d'Aokas)-français : conception lexicographique et modélisation informatique*

Des mots désignant des réalités spécifiques à une langue-culture et ne possédant donc pas d'équivalents dans la langue cible ou des réalités qui existent et qui ne sont pas désignées dans l'autre langue. En voici des exemples dans la langue d'Aokas pris dans la partie dictionnaire :

- *Tiqiğğet*(*tq-*) n. f. : « jeu collectif, sorte de golf, qui se joue avec une plaque de liège qu'on essaie de pousser à l'aide d'un bâton pour la loger dans un trou » ;
- *tabsist*(*te-*) n. f. (arch.) : « après tamisage de caroubes et de l'orge grillés dans l' « afan » (grand plat généralement en terre), les deux farines obtenues sont mélangées dans la proportion d'une mesure d'orge pour deux mesures de caroubes. C'est le résultat de ce mélange qu'on appelle *tabsist* » ;
- *timerdal* (*tm-*) n. f. pl. : « préservatrices : certains produits alimentaires (les premiers d'une série) qui ne doivent être mangés que par les membres de la famille pour « préserver » leur productivité » (v. équiv. *tinehjab*) ;
- *tizenzeyt*(*tz-/tizenzay* (*tz-*) n. f. : « cordelette faite avec des poils de la queue de vache » ;
- *uṭuṭ*(*w-*) n. m. (arch.) : « mélange de farine d'orge ou de blé et beaucoup de sel et de l'eau froide, remué à la main jusqu'à ce qu'il n'y ait plus de grumeaux. La peau de mouton ou de bouc y séjourne pendant un mois pour servir d'outre à grains *aylu* [le sens renvoyant à la mauvaise odeur n'est pas archaïque] » ;
- *bayṣur*, n. m. (arch.) : « mélange de beurre fondu et d'huile d'olive à parts égales (qui permettait autrefois d'économiser le beurre réservé aux invités) ; mélange de beurre et du miel liquide » ;
- *taceyyact* (*t-/taceyyacin* (*t-*) n. f. : « celle qui fait vivre (tatouage (*ticret*) en forme de croix ou de point fait au front d'un enfant au troisième jour de sa naissance, lequel tatouage est censé lui assurer longue vie) ; pièce d'argent offerte à la jeune mariée qui la trouera et la fera porter plus tard à son enfant dans la coiffure en guise de porte-bonheur » ;
- *redded*[*rəddəd*] (*iredde*, *tredde*, *aredde*) v. intr. : « laisser des résidus de nourriture salis par ses pattes (en parlant d'une bête) » ;
- *acembir*<sup>6</sup>(*u-/icembiren*) n. m. (arch.) : « bande d'étoffe noire de sept coudées au moins avec laquelle les femmes kabyles ornaient leurs coiffures. L'usage de l'*acembir*tend à disparaître à Aokas [déjà en 1936] » ;
- *bixbix*[*bixbix*], n. m. sing. : « matière brûlée complètement, très légère et réduite à l'état de cendre, qui se dégage d'un feu et flotte dans l'air », etc.

### 2.3. Des formes lexicales intéressantes au plan étymologique

Des formes qui justifieraient en l'occurrence la thèse de la racine trilitère qui se serait désagrégée avec le temps en bi et même en monolittère. C'est ce que montrent les racines trilitères des mots suivants dont il ne subsiste que deux ou même une seule consonne dans les parlers de la Soummam et de Grande Kabylie :

---

jours d'étude de Lisbonne, 15-17 octobre 2009, Paris, Editions des archives contemporaines, 2011, p. 540.

<sup>6</sup> Ce serait un emprunt au turc attesté même en arabe algérien : *cenbir* « voile rouge couvrant le visage de la mariée, conduite de la chambre où elle dîne à la chambre nuptiale où l'attend son mari » du turc *techenbér* « voile avec lequel les femmes couvrent leur tête ou serrent leur front » (Ben Cheneb, Mohamed, 2012 : 39).

- *wureg(iwureg, ttwurug, tiwargin)* v. tr. : « rêver, faire un rêve » >*tawuregt(tw-/tiwurigin (tw-)* n. f. : « rêve » (v. *rrjuj* « cauchemar(s) ») (*argu/targit* dans les parlers S et GK) ;
- *kfil(ikfil, ttekfil, akfil)* v. intr. : « cailler, se cailler » (v. *ikkil/iqil* « lait caillé ») (*ikil*, dans les parlers S et GK) ;
- *waqqr(iwaqqr, ttwaqqr, awaqqr)* v. intr. : « goûter, s'égoutter » (*iqi*, dans les parlers S et GK) ;
- *tadgalt(te-/tidgalin (te-)* n. f. : « veuve ; célibataire » / *adgal(u-, we-/idgalen)* n. m. : « célibataire, célibataire endurci (péj.) » > *tibudgelt(tb-)* n. f. : « célibat, viduité, veuvage » (*ağğal/tagğalt/tuğğla* dans les parlers S et GK) ;
- *iyejd(i-, yi-/iyejden)* n. m. : « chevreau » (var. *iyid*, Ait Bouaïssi, v. *agermic*) > *iyejdeṭ (ty-/tiyejdin (ty-)* n. f. : « chevrette, biquette » (*iyid*, dans les parlers S et GK), etc.

## 2.4. D'autres racines ayant des formes manquantes dans les autres parlers

- *ayyaṭ (w-/iyeṭṭen)* (arch.) n. m. : « bouc » (seule la forme du féminin *tayaṭ* « chèvre » est attestée dans S et GK) ;
- *asekkur(u-/isekkuren)* (attesté à Ait Bouaïssi) n. m. et coll. : « mâle de perdrix ; perdrix (coll.) » (équival. *aḥaqul*) (*ihiqel*, dans les parlers S et GK) ;
- *klu (kliy, ikla, kellu, tikli/aklaw)* v. intr. (usité surtout à Ait Bouaïssi, mais aussi à Tizi n Berber) : « aller ; marcher » (seule la forme nominale *tikli* « marche » est attestée dans les parlers S et GK) ;
- *mmuli(immula, mmulay, amuli)* v. intr. : « s'ombrager, être ombragé, être couvert (temps) » > *amalu* « ombre » ;
- *anji/aneğği(w-/injiyen ; tanjit (tinjiyin (te-)* adj. et n. : « d'en haut, supérieur » ;
- *anedda(u-/ineddayen ; tanneddayt (tn-/tineddayin (tn-)* adj. et n. : « qui se trouve au fond, au bas, bas, inférieur... » [on retrouve ici l'adjectivateur *an* comme préfixe] ;
- *wejjeṭ(iwejjeṭ, ttwejjit, awejjeṭ)* v. intr. : « s'évaser, prendre la forme d'un sac, d'une poche » (seule la forme nominale est attestée dans les parlers S et GK) ;
- *cceqder(icceqder, cceqdir, aceqder)* v. tr. : « briser, ébrécher ; couper en morceaux » (seule la forme nominale est attestée dans les parlers S et GK), etc.

## 2.5. Des faux-amis

Des mots ayant des signifiants et une morphologie identique et fort probablement la même étymologie, mais des signifiés différents. Nous utilisons ici la notion de « faux-amis » à l'intérieur d'un même système linguistique, alors qu'elle est généralement utilisée pour des mots relevant de deux langues différentes. Nous la distinguons néanmoins de celle de l'homonymie qui ne peut fonctionner qu'à l'intérieur d'un système « homogène » où les mots objets de la comparaison sont utilisés par les mêmes locuteurs dans leurs différentes acceptions, ce qui n'est pas le cas des exemples donnés ici<sup>7</sup>.

---

<sup>7</sup>Koessler l'inventeur de la notion considère l'homonymie comme un des trois types de faux-amis à l'intérieur d'un système linguistique comme le français avec ses différents dialectes (Kiss M., 2002 : 42).

*Essai de dictionnaire tasahlit<sup>l</sup> (parler kabyle d'Aokas)-français : conception lexicographique et modélisation informatique*

- *nnerni*(*innerni*, *ttnerni*, *annerni/tinnernit*) v. intr. : « naître » (« grandir » dans les parlers S et GK) ;
- *aqecwal*(*u-/iqecwalen*) n. m. : « brindille ; petite chose insignifiante, négligeable » (« grande corbeille » dans S et GK) ;
- *taqadumt*(*tq-/tiqadumin*, *tiqadam* (*tq-*) n. f. : « hachette pour couper la viande, hansart, couperet » (hom. *taqadumt* « visage, gueule (péj.) ») (« visage, gueule » dans S et GK) ;
- *aqamum*(*u-/iqumam/iqamumen* ; *taqamumt*(*tq-/tiqamumin* (*tq-*) ou *tiqumam* (*tq-*) fém. dim.) n. m. : « bouche » (syn. *imi*, *aqemmuç*) (« bec » dans S et GK) ;
- *nanna*, n. f. sing. de parenté : « grand-mère, grand-maman » (terme de respect qui désigne la grande sœur ou une femme plus âgée de la famille paternelle ailleurs) ;
- *rceq*(*irceq*, *ulirciq*, *recceq*, *arcaq*) v. intr. : « être embarrassé, s'embourber, s'empêtrer ; être coincé » (v. *uhel*, moins usité) (« donner de l'argent aux musiciens, danseurs... dans S et GK) ;
- *asyar*(*u-*, *we-/isyaren*) n. m. : « 1. charrue 2. (arch.) bois (syn. *akeccuç*, seul usité dans la langue parlée) ; planche ; cercueil (au pl.) » (« bois », ailleurs), etc.

## 2.6. A titre symbolique enfin

Nous avons trouvé le sens d'un mot très couramment usité dans le parler d'Aokas, *tiwwa* « dos », que M. Mammeri note dans son *Poèmes kabyles anciens* comme « non connu » (*uryettwassen ara unamekbbwawal-a*<sup>8</sup>). *Tibbwas* dans le livre écrit d'un seul tenant<sup>9</sup>.

*Lukanattebbreyyibbwas*

*Annegtiyersi*

*Annaweçlâad'artilas*

*Anneyzanreçsi*

*ççbeh ad ay-d-yaf f tibbwas*<sup>10</sup> (traduit : Au matin il nous y trouverait).

Deux autres mots *agulizet uyyiz* sont également notés dans ce classique de la littérature kabyle comme étant des mots dont le sens n'est pas connu par l'auteur : *urnessin ara anamek n sin wawalen-a* (note 111, p. 150). Ces mots sont aussi attestés dans le parler d'Aokas. Le premier avec même une autre variante *agaluz* « reste, résidu », et le second serait un adjectif dérivé du verbe *Ïeyyez* « égrainer, dépiquer ».

Nous avons, par ailleurs, concernant la « théorie du dictionnaire » abordé en deux parties différentes les deux « structures » complémentaires de celui-ci, à savoir la macro et la microstructure.

<sup>8</sup> Mammeri M., 2009, *Poèmes kabyles anciens*, Editions Mehdi, Tizi-Ouzou, note 115, p. 154.

<sup>9</sup> Exemples repris dans nos articles sur l'archaïsme parus dans les revues *Timsal n Tmazightn*<sup>o</sup> 4 et *Awaln*<sup>o</sup> 43-44 (v. bibliographie).

<sup>10</sup> Si un jour j'avais le pouvoir

Je nouerais l'affaire

Nous irions jusqu'aux frontières de l'ennemi

Y installer des tranchées

Au matin il nous y trouverait (p. : 155)

Traduction très habile qui ne prend pas en charge le segment *f tibbwa-s* « sur son dos » : « au matin il nous trouverait sur son dos ».

### 3. Au niveau lexico-sémantique et lexicographique

#### 3.1. La macrostructure

Certains problèmes qui se sont posés à nous, et qui se posent plus généralement à la lexicographie amazighe, ont trouvé ici une proposition de solution. C'est ainsi que nous avons proposé d'introduire des néologismes et pérégrinismes dans le dictionnaire pour rendre compte de l'usage actuel de la langue. C'est la fonction même d'un dictionnaire d'attester un usage connu (*bbipi* « biper », *lanṭirnat* « internet », *apurṭabl* « téléphone portable », *tilelli* « liberté », *Taferka* « Afrique »<sup>11</sup>...). Pour les emprunts non adaptés morphophonologiquement à la langue d'accueil, nous avons proposé de les adapter graphiquement en en donnant la transcription phonétique, à l'instar du mot *aklimatizur* [aklimatizœr] « climatiseur ». Pour le traitement des homonymes, nous avons proposé la solution polysémique et le regroupement des signifiés quand ils sont apparentés et leur intersection sémique n'est pas un ensemble vide. Ils sont dégroupés en homonymes lorsque l'un de ces deux critères n'est pas satisfait, a fortiori si la forme à laquelle ils sont rattachés a des flexions spécifiques à chaque signifié (par ex. *azal* « valeur »/*azal* « journée » ; *bnu*(*lebni/bnu* + COD) « construire »/*bnu*(>*bennu/bnuyef* « escompter »). Pour le classement alphabétique des entrées par racines consonantiques, pratique la plus répandue en lexicographie amazighe, nous avons proposé de tenir compte, en plus, de la constance de la tension et du redoublement consonantique pour diminuer le nombre de racines homonymes. La tension étant notée, comme en phonétique à la suite de L. Galand, en majuscule. Nous avons proposé aussi de tenir compte de l'article défini arabe *el* ou français *l(e)/l(a)/l(es)* lorsque le radical de la langue « prêteuse » n'est pas connu dans la langue emprunteuse. En voici des exemples pris dans la partie dictionnaire concernant la prise en compte de la tension, du redoublement consonantique ainsi que de l'article défini *l* :

- nDm**     ▪ *nnuddem* (*innuddem*, *tnuddum*, *nuddam*) v. intr. : « dormir » (syn. *ṭtes*, plus usité et *gen*)  
*Segṭelliurinnuddem*, *aritturar d medden* « il ne s'est pas couché depuis hier ; il ne fait que jouer avec les gens »  
 ▪ *naddam/nuddam*, n. m. : « sommeil » (syn. *ṭtes*, plus usité)  
*Lukan ay ṭṣleḥnaddamadilaqirbeḥumcic* « si le sommeil était utile le chat serait fortuné » ; *ay aberbar n nuddam ! Arwaḥyer mmi ad innuddem* « Ô berbar (berceuse) du sommeil ! Viens chez mon fils pour qu'il s'endorme »
- qjaj**     ▪ *qejqej* (*iqejqej*, *ttqejqij*, *aqejqej*) v. intr. : « produire un bruit du ventre

<sup>11</sup>*Taferka* qui signifie « propriété terrienne » en kabyle est un emprunt-calembour qui imite la forme *Afrique/Africa* qui serait elle-même d'origine berbère et dériverait de l'étymon *afri/ifri* « grotte » en berbère auquel les Romains ont ajouté le suffixe latin *-ca* pour donner *Africa* (*Nova*) qui désigna d'abord le Nord de la Tunisie où habitaient à l'époque les *Afri* (les habitants de la grotte pour les Romains), avant de désigner l'Afrique du Nord et parla suite l'ensemble du continent. Voir à ce propos F. Decret et M. Fantar, 1981, *L'Afrique du Nord dans l'antiquité. Des origines au Ve siècle*, Paris, Payot.

*Essai de dictionnaire tasahlit<sup>1</sup> (parler kabyle d'Aokas)-français : conception lexicographique et modélisation informatique*

(ou d'estomac), un borborygme »

*Ittqejqijuebbuṭ-iw* « j'ai le ventre qui gargouille »

▪ *aqejqej (u-)* n. m. : « borborygme, gargouillement »

**qrre** ▪ *taqraruṭ(te-/tiqrarucin (te-))* n. f. (Ait B.) : « petite courge au début de son développement, courgette » (var. *taqlalujt*)

**lml** ▪ *lmalu* [o], n. m. : « tornade ; cyclone »

*Ikkerlmalu* « il s'est déclenché une tornade »

**Lss** ▪ *llsas (llsisan)* n. m. : « fondation, base, fondement »

*Llsas n uxxam* « les fondations de la maison » ; *tameṭṭut d llsas, argaz d ajyualemmas* « la femme est la fondation, l'homme la poutre faitière » ;

*ad ṣelliḡfell-ak a nmbi \* d kečč ay d llsas n dḍin* « ma prière soit sur toi ô prophète \* tu es l'assise de notre religion » ; *tennezlaztmurt f llsas* « le pays est secoué dans ses fondements » ; *ass-a i uzekka d'att-nebnu f llsas*

« aujourd'hui pour demain nous la construirons à la base »

### 3.2. La microstructure

Nous avons adopté un programme de microstructure qui prend en charge quasiment l'ensemble des éléments d'analyse de ce niveau pour un dictionnaire synchronique de langue qui ne s'intéresse pas à l'étymologie qui pourrait avoir toute sa place dans un dictionnaire encyclopédique ou historique. En voici les éléments :

- Transcrire phonétiquement de façon partielle et non systématique des mots-entrées dont la graphie pourrait présenter une difficulté de prononciation, comme *tacercurt*[r] « petite cascade » (vs *tacercurt* « boucle d'oreille »). Il n'est pas intéressant de transcrire phonétiquement des mots comme *ifelfel* « piment ; poivron », *ayilas* « tigre ; lion »... ;
- Donner systématiquement la catégorie grammaticale du mot-entrée en le situant dans l'une des neuf parties du discours connues en en donnant les différentes flexions quand il s'agit d'un mot variable ;
- Donner systématiquement la valence des verbes pour permettre à des usagers actifs du dictionnaire de les utiliser correctement ;
- Donner quand c'est possible la marque d'usage du mot traité, qui permet à l'utilisateur passif du dictionnaire de mieux le situer dans le discours et à l'utilisateur actif de mieux l'utiliser ou de l'utiliser correctement ;
- Donner des exemples d'usage pour les différents sens et acceptions d'un mot polysémique qui permettent de voir ces mots en discours et de mieux saisir leur sémantisme qui pourrait compléter, clarifier et même éventuellement corriger celui donné en équivalence. Ils peuvent même tenir lieu de définition/équivalence. Ils permettent aussi de montrer une construction particulière (proverbe, maxime...) ou fréquente (locution, collocation...) qui éclairent l'utilisateur sur certaines occurrences et cooccurrences, sur la combinatoire du mot-entrée en discours. Ils peuvent montrer aussi des aspects intéressants de la vie sociale et culturelle de la communauté linguistique concernée.

- Donner enfin, quand c'est possible et de façon non systématique, l'information analogique utile à un usage actif du dictionnaire (un synonyme, parasynonyme, hyponyme, hyperonyme, variante, etc.).

### 3.3. L'équivalence interlinguistique

Nous avons essayé, en particulier concernant les locutions, de faire en sorte que l'équivalence donné ménage sens, style et conditions d'usage et soit accompagnée d'une traduction littérale qui permet à l'utilisateur qui en a besoin de mieux voir le fonctionnement des différents mots composant la locution et son degré de figement. Leur classement s'est fait généralement sous le verbe, le nom, l'adjectif ou l'adverbe qu'elles contiennent. Dans la version numérique on peut la traiter sous tous ses composants grâce à la disponibilité de l'espace.

L'existence de *mots culturels* dans la langue source qui n'ont pas d'équivalents dans la langue cible oblige à recourir aux définitions explicatives caractéristiques de la lexicographie monolingue. Cela montre que la définition, par-delà l'hétéronymie dominante en lexicographie bilingue, est loin d'être l'apanage de la lexicographie monolingue. On y trouve, par exemple, beaucoup de définitions hyperonymiques, inclusives ou encore logiques/aristotéliennes qui est le type le plus utilisé en lexicographie monolingue. En voici des exemples :

**mw**     ▪ *tamawayt(tm-/timawayin (tm-))* n. f. : « pieu ayant la forme de Y, placé au milieu de la pièce, il supporte la poutre centrale dans la maison kabyle traditionnelle »

*Qqarenimiyaren : tameṭṭuttettrebbiargaz-is ; d nettat ay t-tamawayt n wexxam*« les vieux disent : la femme éduque son mari ; c'est elle qui constitue la poutre centrale de la maison »

**ck**     ▪ *cuk(icuk, tucuk, acuk)* v. tr. indir. : « psalmodier des versets coraniques dans l'oreille de certains malades (généralement psychiques) afin de les guérir en faisant sortir, c'est ce que l'on croit, par la puissance du coran le mal qui les habitent. Cette pratique qui est le fait de marabouts ou de talebs est plus connue aujourd'hui (et ailleurs) sous le nom arabe de *rruqya* »

*Cuk-as g-mezzuy \* ulama'yab d amezduy*« psalmodie-lui (du coran) dans l'oreille \* même absent il sera parmi nous »

Voici un exemple de traitement d'un verbe polysémique où les acceptions proches sont séparées par des points-virgules ou par des virgules lorsqu'elles sont très proches et par des chiffres (1, 2...) des acceptions divergentes. Les différents sens sont illustrés par des exemples comportant entre parenthèses leur référence précise dans le corpus. Les locutions comportant le mot-entrée sont traitées à part à la fin de l'article.

▪ *ddu (ddiy, idda, teddu/tteddu, tikli/tiddi)* v. intr. : « 1. marcher ; avancer, s'avancer, venir (avec usage de la particule de direction) ; aller ; fouler (le sol) ; accompagner, fréquenter ; être (r)apporté, (r)amené (avec partic. de dir.) ; tendre à, viser 2. subir 3. se comporter 4. voler, enlever, subtiliser 5. se réaliser, s'accomplir 6. se marier (pour une femme) (ellipse de *ddu t-tisli*« litt. aller comme nouvelle mariée ») »

1. *Tedduy, tedduy, g lqæaufiyamezzuy*« je marche et je marche, par terre je trouve une oreille » (*d agersal*« c'est le champignon »); *aka-yaleejeb r-*

*Essai de dictionnaire tasahlit<sup>l</sup> (parler kabyle d'Aokas)-français : conception lexicographique et modélisation informatique*

- Rebbiiteddu-dd* « voilà que le monstre (de Dieu) s'avanceit » ; *ittedduiyzeriyzer al ittcernun* « il va le long d'un ravin en hurlant » (*d aman* « c'est l'eau ») ; *d'ik-iččlqaea-nn (te)tteddut* « il te mangerait et la terre que tu foules » ; *aql-ay g lqernxmesřac \* ultuřit d w'alateddut* « nous sommes au siècle quizième \* tu ne trouves pas qui fréquenter » ; *a w'iddan d yetbiren* « heureux qui accompagne les colombes » ; *d'itlebayekfi n tyilastw'i dd-iddu g tejliř n mmi-s* « il exigera du lait de lionne ramené dans la peau de son fils » ; *i uqerruyumeksa ay tetteđdu* « c'est le meneur qui est visé (litt. c'est vers la tête du berger qu'elle se dirige) »
2. *Fell-iteddunt \* almiqqley d anedda* « je les subis (étym. elles passent sur moi) \* jusqu'à ce que je devienne le dernier des derniers »
  3. *I nřill i lejnasamekteđdun* « qu'on regarde les (autres) peuples comment ils se comportent »
  4. *Ma tedheriyewelřařatedda, tenza g yiwen n ssuq* « lorsqu'on s'aperçoit qu'une chose a été enlevée et vendue dans un marché » ; *setta t-teffahin-ennak-ddan* « les six pommes qu'on t'a volées » ; *d'ak-ddunidrimen-ik* « il te subtiliseront ton argent »
  5. *Ayenicreř g-řulan-isidda-y-edd* « ce qu'il a exigé de ses gendres fut réalisé » ; *ddan-ak-edd !* « on te l'a apportée (eau)/la chose s'est réalisé (le fait de te les rapporter) ? »
  6. *Ma d'imma-s mi d'idduteqcict d'iččxamsaduru* « Quant à la mère, elle reçoit cinq douros (25 fr.) le jour du mariage de sa fille »
- *Ddugaryiccer d lřifer*(loc.) : « se trouver dans une situation difficile, être empêtré »

Voici un autre exemple concernant le nom et les différents types d'informations (morphologiques, pragmatiques et analogiques) qu'on peut y trouver :

- mLq** ▪ *amelluq(u-/imelluřen)* n. m. (très fam. ou vulg.) : « couille, testicule » (dim. *tamelluqt*; syn. *agluř* ; v. *timexmuxin, dduzan, lmateryal*)  
*Amdan s yiwenumelluq \* inřeyyar f ayenictuq \* uřdes berk ay itkařar* « l'homme à une seule couille \* embarrassé par ce dont il est frustré \* c'est tout seul qu'il se démerde »  
 ▪ *tamelluqt(tm-/ti-in (tm-))* n. f. : « petitecouille (d'un enfant) »

## Conclusion de la partie lexicographique

Lorsque nous avons résolu de mettre un point final à notre *Essai de dictionnaire* du parler d'Aokas dans le cadre de notre doctorat, nous avons encore dans notre « répertoire investigatif » des mots à traiter et pour lesquels il nous a fallu encore d'autres enquêtes de terrain. Et chaque enquête appelle irrémédiablement une autre pour clarifier ou résoudre des problèmes posés par certaines données recueillies. Cela peut concerner le sémantisme de certains mots, leur morphologie, leur emploi, leur prononciation ou toute autre information nécessaire à leur traitement lexicographique. Le travail sur le lexique n'est jamais fini, ni dépourvu d'erreurs et de lacunes. Le nôtre n'est qu'une ébauche, intitulé d'ailleurs « Essai », pour l'élaboration d'un dictionnaire du parler d'Aokas qui aura besoin, pour un travail

plus complet, d'un corpus plus important et plus varié. Nous continuerons donc nos enquêtes pour essayer d'y parvenir dans les prochaines années. Nous essaierons de le faire avec le souci de compléter, comme nous avons commencé à le faire ici, les informations grammaticales et sémantiques par les informations pragmatiques et analogiques nécessaires.

## **II. Computer modeling of dictionary “Modélisation informatique du dictionnaire”**

### **Aokas dialect Tasaḥlit/French dictionary: IT design for an electronic implementation**

Berber dictionaries have tended to be written in Microsoft Word. Once they reach a few thousand entries, they become difficult to maintain, especially if the task is to be shared with a small team, or if the format of the entries is to be changed. It is difficult to find formatting mistakes and keep the structure consistent. Entries cannot be browsed on the web. And the data cannot easily be reused for other projects like spellcheckers.

The solution is to create an electronic dictionary.

The current scope of the project is to enable editing and simple consultation of Berkai's Aokas dialect Tasaḥlit/French dictionary by a small group of users, and formatting of the dictionary for printing.

This article will outline the design that the project arrived at, which is an evolution of the author's electronic version of Bouamara's monolingual ISSIN dictionary (Bouamara, 2010). First, however, it will explain the theoretical context and technical aim, as well as the approach taken and an examination of some of the existing tools. The conclusion will give an overview and evaluation, and suggest further work to extend the design and exploit the data.

### **Theoretical context**

A dictionary can be thought of as a set of abstract data representing its author's judgement of linguistic reality, presented as a concrete view that is clear for a human reader.

That printed human view with its macro- and microstructure is usually highly compressed to save space. For example, grammatical homonyms are often listed together in one entry.

The order of text items helps the reader to navigate – often the most common or important information comes first.

For finding a word in a printed dictionary, the view is indexed by headword or a higher level organising key like a root.

Indexes are less important in an electronic dictionary, because it is possible to jump directly to a word or any of its listed forms, so the view can be different. It is navigated and read differently.

With that picture in mind, the project aim was a design that supported Dr. Berkai's lexicographic decisions, and technically, kept two processing levels

*Essai de dictionnaire tashlit<sup>1</sup> (parler kabyle d'Aokas)-français : conception  
lexicographique et modélisation informatique*

separate - a lexical data level for machines, and a print view formatting level for humans.

That structure makes it easier to export whichever type of data is needed for other projects.

Some examples of machine use of lexical data are morphologies for lemmatizers and rule-based machine translation, and spellcheckers. Prof. Bouamara's ISSIN database already supplies Kabyle data for an open source spellchecker for Firefox and OpenOffice/LibreOffice\*.

## **Approach**

Research in Natural Language Processing and machine readable dictionaries has given the world shared conceptual models, data models and formats that can be reused and extended.

The project uses these frameworks and formats to develop the lexical data structures and printed views. That means that outside researchers or new members of the team can easily understand and use project data, as there is a clear description.

Another aspect of the project approach is that the user interface is available in Berber.

The project also chose to use open source platforms and components. The intention is to make the software parts that can be used for other dictionaries freely available.

## **Background**

The project needed a flexible tool built from freely available components. For the data to be easy to maintain in a choice of database engines, the database needed to follow a standard.

The existing electronic dictionary tools tended to fall into one of the following groups:

- Closed specialized private commercial platforms, not freely available or free to use and with custom data structures
- Tools created for research projects but not available for reuse.
- Wiki-style tools where the entries are stored as flat text, not data structures.
- Hosted platforms – free to use but not download, with limited control over the text formatting, and where the data structures are not necessarily built from standard parts.
- Desktop applications (complicating collaboration and maintenance)

The path taken was to extend the ISSIN electronic dictionary (see fig. 1). Some additional features required for Dr. Berkai's dictionary were, for example, a root index, and usage labels on definitions.

Two resources were used in the design:

- One is the Lexical Markup Framework (LMF), an ISO standard for NLP and machine readable dictionary lexicons. It provides a standard skeleton for

Abdelaziz BERKAI & Paul ANDRESON

the conceptual model of the dictionary's underlying lexical data. The project extends it for Berber, and several research projects have information online about their own customizations.

\* <https://addons.mozilla.org/fr/firefox/addon/tira-n-teqbaylit/>  
<http://extensions.libreoffice.org/extension-center/tira-n-teqbaylit>  
<https://sourceforge.net/projects/tiranteqbaylit/>

- The other is the Text Encoding Initiative (TEI). It provides an extensible document structure for the dictionary view and is used when generating the printed dictionary. The ISSIN formatter is also TEI-based.

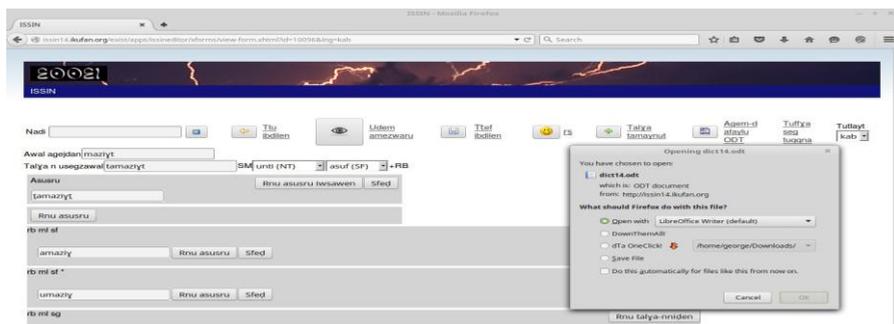


Fig. 1: ISSIN electronic dictionary – downloading formatted .odt document

## Design

- The design is a web application, hosted on a central server for easy maintenance and administration.
- The dictionary is formatted from detailed standards-based data structures, not simply stored as text like in a wiki. The previous, ISSIN, database structure was custom-built, inspired by a mix of standards.
- The lexical data and the human view are clearly separated as explained above, much more so than in the ISSIN tool. The lexical data can be used without being familiar with a particular dictionary view. The view layer groups, positions, and labels the lexical data, formatting it into readable compressed text. Data at both layers are easy to understand and to export from the system, or even to import into it.
- As in the ISSIN tool, the output is ODT open document format text files, but Microsoft Word files, pdf's or TeX typesetting are also feasible.
- For technical simplicity, as with the ISSIN, the current version is based on XML throughout, and runs on one integrated platform, the eXist database. That may change to a JSON database in future.

*Essai de dictionnaire tashlit<sup>1</sup> (parler kabyle d'Aokas)-français : conception  
lexicographique et modélisation informatique*

- A configurable tool is now used to break up the original text into data fields for the database.

## **Analysis**

Storing entries as data structures, not flat text, has some advantages. The presentations can be changed easily. Specific data can be extracted if needed for other projects. And data can be merged between dictionaries.

One disadvantage is that adding new features such as verb transitivity to a dictionary means updating the software. Fortunately, instead of editing entries, adding data for the new features can be automated. Features can also be added by hand to the generated dictionary text.

For the moment, the design is suitable for monolingual dictionaries, or for the half of bilingual dictionaries that defines unknown words, but not yet for the half that gives equivalents of known words. It is possible to adapt it to other dictionary data by adding lexical features, creating views, and configuring the data loader.

The LMF conceptual model can be realised in data models in different ways. That means the exported data could be used easily in different forms by diverse projects according to their needs and experience – for example NLP research (RDF format) or web application development (JSON-LD format, TEI or JSON).

## **Conclusion of computer modeling part**

The design forms a web application for consulting and editing the Aokas dictionary data and for formatting it into a document. Final adjustments can be made with a text editor before printing.

The design is built from open source components and is flexible for adaptation to other dictionaries.

The database type is well known in a university and publishing context, making maintenance and administration easy.

To make further use of the dictionary, both as data for NLP use and as text for human readers, the design lets you export clearly described, standards-based data formats.

Some examples of possible future work:

### *For the public*

- Web and smartphone applications and browser plugins that would allow the public to consult the dictionary, do advanced searches and store the results. These could make use of cloud computing for performance and easy maintenance.

- Adapting the design for other dictionaries such as Tamashek and Mozabite.

*For research and development.*

- Creating API's to expose dictionary information on the semantic web, to link ontologies to Berber, and to allow application developers to use the dictionary text.
- Building an electronic morphology for lemmatizers and rule-based machine translation. That project has already started.
- Enriching the data with further semantic relationships, for applications like machine translation, word nets and thesauri.
- Extending the design for two-way bilingual dictionaries for humans and for machine translation.

#### *For dictionary compilers*

- Issue tracker integration to link reported problems with corrections to the dictionary.
- Importing field data collected with e.g. FieldWorks.
- Allowing editing and generation of dictionaries with related data (pocket size, web, concise)
- Investigating techniques for linking dictionaries to a corpus, both for research and to help in dictionary compilation.

## **Bibliographie**

Ben Cheneb, Mohamed, 2012 : *Les mots turcs et persans conservés dans le parler algérien*, Publication du cinquantenaire de l'Université d'Alger 1962-2012.

Berkaï, Abdelaziz, 2011 : « Quel programme microstructurel en lexicographie berbère ? », dans Naït-Zerrad, K. (éd.), *La standardisation du berbère à la lumière des évolutions récentes en Europe et dans le Nord de l'Afrique*, Actes du colloque organisé à l'INALCO (Paris) 6-7 octobre 2008, *Revue des Etudes Berbère*, Vol. 5, pp. 25-45.

Berkaï, Abdelaziz, 2011 : « Quelques problèmes lexicographiques que posent l'établissement d'équivalences et leur organisation dans un dictionnaire kabyle-français », dans Van Campenhoudt, M., Lino, T. et Costa, R. (dir.), *Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité*, Actes des journées d'étude de Lisbonne, 15-17 octobre 2009, Paris, Editions des archives contemporaines, pp. 537-547.

Berkaï, Abdelaziz, 2013 : « Quelques problèmes macrostructurels en lexicographie berbère », in A. Zavaglia et Ph. Humblé (coord.), *La lexicographie d'aujourd'hui en contextes francophone et brésilien*, *Synergies Brésil*, n° 11, pp. 49-65.

Berkaï, Abdelaziz, 2013 : « L'intérêt du corpus et une idée de sa constitution en lexicographie amazighe », in Figziri N. (dir.), *Corpus : exploitation, théories et méthodes*, *Iles D Imesli*, n° 5, 2013, pp. 281-293.

Berkaï, Abdelaziz, 2014 : *Essai d'élaboration d'un dictionnaire tasahlit (parler d'Aokas)-français*, thèse de doctorat sous la direction du Professeur Md Akli Haddadou, 3 volumes, Université de Tizi-Ouzou.

*Essai de dictionnaire tashlit<sup>l</sup> (parler kabyle d'Aokas)-français : conception lexicographique et modélisation informatique*

- Berkaï, Abdelaziz, 2014 : « L'archaïsme en tamazight (kabyle) : analyse et identification », dans Dourari A. (dir.), *La néologie, les corpus informatisés et les processus d'élaboration des langues de moindre diffusion*, Timsal n Tmazight<sup>n</sup> 4, revue du Centre National Pédagogique et Linguistique pour l'Enseignement de Tamazight (CNPLET), pp. 78-91.
- Berkaï, Abdelaziz, 2015 : « L'archaïsme linguistique : un essai d'identification en berbère (kabyle) », in Yacine T. (dir.), *Langue et politique en Afrique du Nord*, Awal, n°43-44, 2015, pp. 133-144.
- Bouamara, Kamal, 2010 : *ISSIN - Asegzawal n teqbaylit s teqbaylit*. TiziOuzou : L'Odyssee.
- lexicalmarkupframework.org. *Lexical Markup Framework (LMF)*.  
<http://www.lexicalmarkupframework.org/>
- Dallet, Jean-Marie, 1982 : *Dictionnaire kabyle-français (parler des Ait Menguellat, Algérie)*, Paris, SELAF, 1052 p. + XXXIX.
- Decret, François et Fantar, Mhamed, 1981 : *L'Afrique du Nord dans l'antiquité. Des origines au Ve siècle*, Paris, Payot.
- Galand, Lionel, 2002 : *Etudes de linguistique berbère*, Leuven-Paris : Peeters.
- Kiss, Monika, 2002 : « Les pièges du vocabulaire bilingue : les faux amis », *Revue d'Etudes Française* n° 7, pp. 41-55.
- Lafage, Susanne, 1997 : « De quelques principes apparemment contradictoires dans la constitution d'un corpus lexicographique différentiel », in C. Frey et D. Latin, *Le corpus lexicographique : Méthodes de constitution et de gestion*, Actes des troisièmes journées scientifiques du réseau thématique de recherche « Etude du français en francophonie », Paris, Editions Duculot, pp. 87-100.
- Mammeri, Mouloud, 2009 : *Poèmes kabyles anciens*, Tizi-Ouzou, Editions Mehdi.
- Naït-Zerrad, Kamal, 2004 : « Kabylie : Dialectologie », in 26 | Judaïsme – Kabylie, Aix-en-Provence, Edisud (« Volumes », n° 26). En ligne à : <http://encyclopedieberbere.revues.org/1433>
- Picoche, Jacqueline, 1992 : *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire*, Paris, Nathan.
- Romary, Laurent, 2015 : *TEI and LMF crosswalks*. JLCL - Journal for Language Technology and Computational Linguistics, 30 (1).  
<https://hal.inria.fr/hal-00762664v4/document>
- Text Encoding Initiative. *P5: Guidelines for Electronic Text Encoding and Interchange. Chapter9: Dictionaries*. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>