

Pour une indexation raisonnée des corpus littéraires kabyles

Pour une indexation raisonnée des corpus littéraires kabyles

Patrice Pognan & Mohand Akli Salhi
INALCO, STIH-Lalic. Sorbonne & DLCA. Tizi- Ouzou

Ce texte présente un projet en cours portant sur l'indexation des corpus littéraires et leurs traitements informatiques. Trois parties le constituent. La première propose un état de ces corpus en tentant de les situer dans le champ littéraire. La seconde explicite l'opération d'indexation (objet, méthodologie et objectif). La troisième est réservée à la présentation de la base de données conçue pour cette opération d'indexation.

Etat des corpus littéraires

Panorama général

D'essentiellement orale, la littérature kabyle passe progressivement de l'oralité pure et mitigée à l'écriture et à l'oralité médiatisée. Le processus de transcription est ancien de près de deux siècles. La création par le biais de l'écriture est de plus en plus importante ; elle touche pratiquement tous les genres et elle tend à être menée aussi bien par les hommes que par les femmes et occupe, de manière de plus en plus significative, de nouveaux espaces (livre, presse, internet, etc.).

Plusieurs types d'agents ont participé à la collecte, la transcription et la traduction de la littérature orale kabyle. Depuis le début de la transcription, on dispose d'une masse textuelle considérable grâce aux voyageurs européens, aux administrateurs et militaires français, aux missionnaires religieux, aux instituteurs et écrivains kabyles d'expression française, aux universitaires (étrangers et autochtones), aux écrivains d'expression kabyle, aux militants culturalistes et aux étudiants suivant des études en langue et culture amazighes (dans les universités algériennes et étrangères).

Depuis quelques décennies, un mouvement de renouveau littéraire, plus généralement culturel, est observé aussi bien en Kabylie

qu'en diaspora. Ce mouvement apparaît, au plan poétique, sous forme :

a. de nouveaux genres (romans, nouvelle, théâtre) et de textes à identité générique hybride,

b. de nouvelles tendances textuelles dans la poésie écrite et dans la poésie chantée, et, au plan institutionnel, sous forme :

a. d'un nouveau personnel créateur (homme et femme, jeunes et moins jeunes), hautement instruit pour une bonne partie,

b. de l'émergence d'un embryon d'édition en kabyle. Certaines maisons d'édition semblent faire de l'amazighité leur créneau principal. C'est le cas notamment de Tira Editions et des Editions Achab. Cette émergence de l'édition privée s'ajoute à celle de l'édition étatique (ENAG, ANEP et notamment HCA),

c. l'apparition timide certes mais non moins réelle de la critique.

Cette dernière prend forme dans la presse écrite (mensuel *Izuran*, quotidiens comme *La Dépêche de Kabylie*, *Liberté*, *L'Expression*), dans la presse audio (certaines émissions sont animées par des écrivains) et sur les sites internet.

En somme, la longue opération de transcription, la dynamique de renouvellement et l'intégration de plus en plus importante de la littérature kabyle dans la Toile ont engendré l'établissement par l'écriture d'une masse considérable de corpus littéraires. De nombreux inventaires bibliographiques, dont certains sont critiques, leur sont, en partie, consacrés (Lacoste-Dujardin, Galand, Chaker, Chaker et Bounfour, Boughchine, Brenier-Estrine, etc.)

Visibilité des corpus

Toutefois, si on reconnaît l'existence et la diversité de ces corpus, il reste que leur localisation pose problème. Leurs signalements bibliographiques ne suffisent pas dans la mesure où ils sont géographiquement et temporellement éparpillés. A titre d'exemple, les corpus transcrits durant la période coloniale sont difficilement accessibles ; ceux édités en Europe, essentiellement en France, ne sont pas commercialisés en Algérie. Ces nombreux corpus demeurent par conséquent inconnus du large public.

De la même manière, la variation caractérisant la littérature orale traditionnelle aussi bien dans les pratiques sociologiques (agents producteurs, rites et conditions de performance par exemple) que textuelles et génériques (dénomination des types poétiques par exemple) constitue un obstacle dans l'explicitation, au niveau analytique, du champ littéraire. Ce qui empêche à un certain degré la visibilité de certains corpus.

Par ailleurs, l'évolution historique n'est pas la même pour toutes les régions kabyles. Au plan de la littérature orale, certaines sont plus conservatrices que d'autres. On observe même des évolutions particulières dans certaines régions dues, entre autres, aux contacts avec l'autre comme cela semble être le cas avec les Pères Blancs et l'émigration. A cause de son aspect archaïque et / ou de son aspect novateur et émergent, un corpus (ou une pratique littéraire) peut ne pas (ou plus) être visible y compris pour la communauté dans laquelle il se pratique.

L'opération d'indexation

Les premières indexations

Des bibliographies, des tables, des dictionnaires et des index abordant globalement la langue et la culture amazighes sont à la disposition des utilisateurs (chercheurs, étudiants et lecteurs particuliers). Les corpus littéraires y sont signalés, présentés et parfois même critiqués. Ces documents peuvent être considérés comme des outils d'indexation. A titre d'illustration :

a. le Dictionnaire de la Culture Kabyle de Lacoste-Dujardin présente pas moins d'une centaine de mots expliquant et parfois illustrant soit un motif récurrent dans les contes, soit la biographie d'un auteur, soit les conditions de performance d'un type poétique, etc.

b. le Dictionnaire des auteurs algériens d'Achour Chourfi signale près d'une centaine d'auteurs d'expression kabyle avec des données biographiques et bibliographiques.

c. Les chroniques bibliographiques berbères présentées dans l'Annuaire de L'Afrique du Nord, tenues respectivement par Lionel

Galand, Salem Chaker, puis, par la suite, Claude Brenier-Estrine, présentent les documents inventoriés, dont ceux relatifs à la littérature, avec des annotations, des commentaires et, pour certains d'entre eux, des critiques.

Généralement, ces documents prévoient des index permettant des recherches faciles et rapides. Par les commentaires et les annotations des textes littéraires qu'ils proposent, ces documents sont des outils de référence dans l'opération d'indexation proposée ici, notamment quand le corpus à indexer n'est pas accessible. Le tableau suivant propose une typologie de ces documents en donnant comme exemple les plus connus :

Bibliographies		Index, tables et revues		Dictionnaires	
Bibliographies signalétiques	Bibliographies commentées	Index et tables	Revue	Dictionnaires biographiques	Dictionnaires de culture
Lacoste-Dujardin Boughchi-che	Galand Chaker Brenier-Estrine	Lanfry (FDB) Lacoste-Dujardin	BEB Awal Tisuraf EDB Tifin Ayamun.com	Cheurfi Chaker	Lacoste-Dujardin

Définition et méthodologie de l'indexation

Afin de présenter les corpus littéraires aussi bien pour les besoins de la recherche universitaire que pour la consommation culturelle, il y a lieu de procéder à leur indexation. L'opération d'indexation désigne dans le cas qui nous concerne ici le processus destiné à représenter, au moyen de termes descripteurs, la forme, le contenu, le mode de performance et de diffusion, le type d'agent producteur et / ou le métalangage de ces corpus. Autrement dit, il s'agira de traiter des corpus littéraires sous forme de mots clés en établissant entre ces derniers des relations. Le traitement préconisé dans notre démarche consiste à catégoriser, à caractériser et, éventuellement, à illustrer ces corpus. Les étapes de catégorisation et de caractérisation s'attacheront à présenter ces corpus à partir de :

1. leurs auteurs (créateurs et autres) et éventuellement de leurs récepteurs
2. leurs types et genres
3. leurs textes et des modalités de leurs textualités
4. leurs conditions de création, de diffusion et de réception (situations de performance)
5. discours tenus sur eux

Plus concrètement, il sera question de proposer des indexations relatives aux données caractérisant une partie ou la totalité d'un corpus. Ces données peuvent être d'ordre lexical, textuel, générique, institutionnel et métalittéraire. Ces données se formaliseront en mots clés (ou mots descripteur) qui prennent en charge l'explicitation d'une réalité du corpus (formelle, thématique, sémiotique, stylistique, institutionnelle ou autre). En l'état actuel du travail, ces données se déclinent comme suit :

1. Agents littéraires
2. Groupes et institutions littéraires
3. Lexiques littéraires
4. Motifs et thèmes
5. Performance littéraire
6. Personnages littéraires
7. Types et genres littéraires

Cette répartition forme les catégories de l'index général. Chaque catégorie est composée de l'ensemble des mots clés qui couvrent son volet. Une fois établis, les mots clés seront mis en relation entre eux afin de permettre de situer les réalités et faits indexés dans leurs contextes élargis. Les relations entre les mots clés sont de différentes natures ; les plus importantes sont, d'après nos premières constatations :

1. La relation d'opposition
2. La relation de contiguïté
3. La relation de variation (morphologique et / ou sémantique)
4. La relation d'évolution (historique)
5. La relation de catégorisation

La mise en relation des mots clés peut prendre deux formes :

1. relation entre les mots pris isolément. Des renvois ponctuels seront prévus entre les mots clés traitant de la même réalité du corpus.

2. relation entre les groupes de mots (participant à la caractérisation d'un volet du corpus, par exp.: forme, performance, contenu). C'est ce type de relation qui définira les catégories de l'index général.

Dans le but de ne pas imposer aux réalités du corpus des dénominations, les mots descripteurs doivent prioritairement être repris des pratiques littéraires et formulés dans la langue même de ce corpus. Ce qui justifie l'option ethnolinguistique et ethnopoétique dans l'opération d'indexation. Ceci ne doit pas cependant empêcher de proposer des mots clés en français quand la réalité du corpus n'est exprimée dans la langue kabyle.

Les objectifs et les retombées documentaires

Les objectifs recherchés à travers cette opération d'indexation des corpus littéraires sont de trois ordres. Le premier ordre concerne le volet de la recherche : cette indexation ambitionne de mettre à la disposition des chercheurs berbérissants et autres des données littéraires kabyles circonstanciées (lexicales, textuelles, génériques et autres) à des fins comparatives. Le second ordre touche à la visibilité du champ littéraire kabyle en présentant, sous forme encyclopédique et synthétique, la masse des corpus et les pratiques littéraires. Le troisième et dernier ordre concerne le volet didactique : cette opération participera à la préparation des données littéraires (textuelles, génériques, institutionnelles et historiques) pour l'enseignement.

Conséquemment à ces types d'objectifs et en termes de retombées documentaires, ce projet ambitionne de proposer des documents de type :

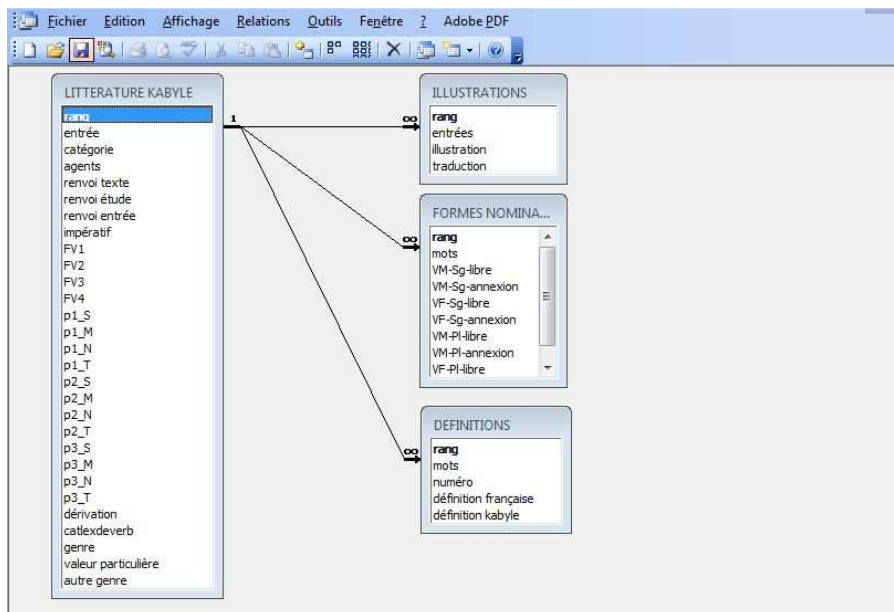
1. Dictionnaire de littérature kabyle (général et / ou partiel)
2. Dictionnaire biobibliographique des auteurs en kabyle
3. Petite (s) anthologie (s) de textes littéraires
4. Bibliographie (s) des textes littéraires (générale et / ou partielle)
5. Bibliographie (s) des études littéraires (générale ou suivant les genres littéraires)

La base de données

Une base de données est un outil informatique qui permet de stocker des informations en les structurant. Les structures de notre base de données sont strictement celles présentées ci-dessus, utiles à l'analyse littéraire et à ses applications. En aucun cas, le chercheur ne doit être contrarié ou même limité par l'outil «base de données». Au contraire, ce dernier doit démultiplier les applications possibles à partir d'un seul ensemble de données. C'est pourquoi nous pouvons prévoir les 5 types de documents énoncés ci-dessus.

La base de données est actuellement en développement. Un des avantages de ce type d'outil, en plus d'être un lieu unique de conservation pérenne de données, est sa facilité de modification, en ajouts, en retranchements, en précisions diverses. Dans le cas présent, nous utilisons une base de données relationnelle dont la conception fondamentale est tabulaire. Cela veut dire que nous répartissons les données dans différentes tables qui sont reliées entre elles suivant certains types de relations (souvent de «un à plusieurs», parfois de «un à un»). De là vient le terme «relationnel».

Nous présentons ci-dessous la structure de la base de données dans son état actuel :



On y reconnaît quatre tables différentes :

- la table dénommée «LITTERATURE KABYLE». C'est la table première, la table mère.
- la table «DEFINITIONS» comportant deux définitions en français et en kabyle.
- la table «ILLUSTRATIONS»
- une table «FORMES NOMINALES».

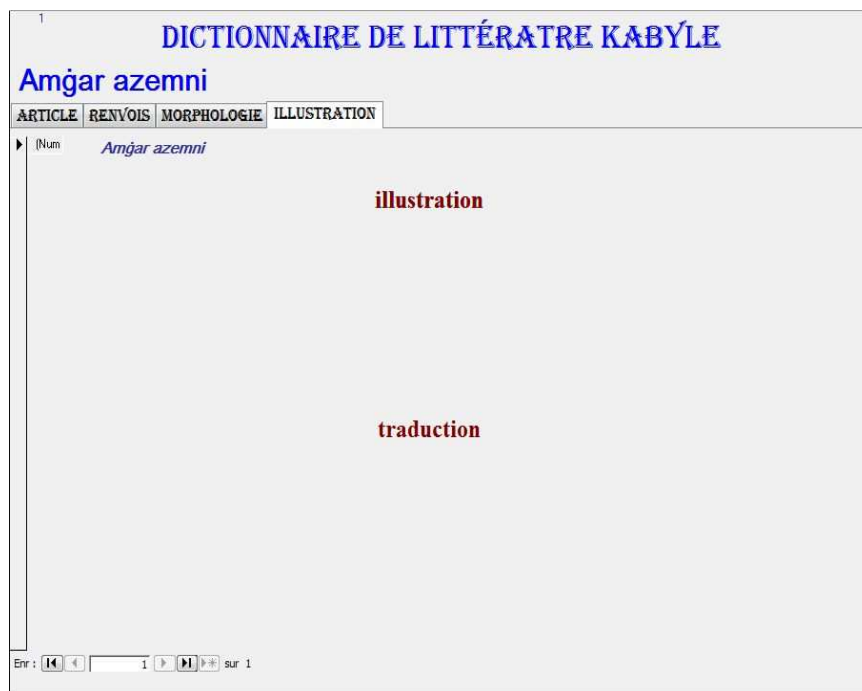
Cette dernière table n'est qu'une partie d'une description grammaticale des termes destinée à recevoir les résultats de la flexion automatique des substantifs. Elle est subordonnée à la table mère où se trouve l'essentiel des informations grammaticales.

Ces tables sont fabriquées par le concepteur de la base (définies, structurées et dotées d'opérations de calcul, ici sur la langue kabyle, par exemple flexion automatique des substantifs). Elles ne sont vues ni par les personnes qui «renseignent» la base de données (si elles sont différentes du concepteur), ni par les utilisateurs, qui disposent d'une interface graphique leur permettant de charger la base ou de la

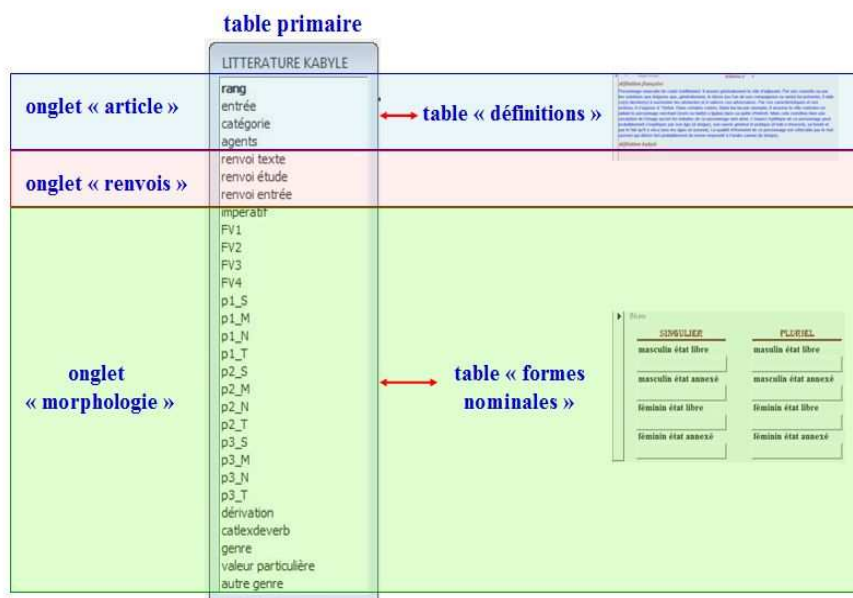
consulter. Nous donnons ci-dessous une première vue de cette interface avec le premier item de la base :

The screenshot displays a web interface for a Kabyle literature dictionary. At the top, the title 'DICTIONNAIRE DE LITTÉRATURE KABYLE' is centered. Below it, the entry 'Amğar azemni' is shown. A navigation bar contains four tabs: 'ARTICLE', 'RENOIS', 'MORPHOLOGIE', and 'ILLUSTRATION'. The 'ARTICLE' tab is active. The entry details include the word 'Amğar azemni', its category 'personnage littéraire', and a dropdown menu for 'agents'. The main content area features two definitions: 'définition française' and 'définition kabyle'. The French definition describes the character as a traditional male hero who helps overcome obstacles. The Kabyle definition is currently blank. At the bottom, there is a search bar with the text 'Enr : 1 sur 1' and navigation icons.

Un élément important du confort de consultation sont les onglets (article, renvois, morphologie, illustration). Au nombre de 4 comme les tables, ils ne correspondent pas nécessairement à celles-ci. Seul l'onglet «illustration» est identique à la table portant le même nom. Ils comportent deux champs principaux, l'un dénommé «illustration» est destiné à recevoir un texte ou un extrait de texte en kabyle permettant d'illustrer l'entrée portée sur l'interface et dupliquée sur l'onglet «article», l'autre champ intitulé «traduction» propose la traduction en français du premier champ:



En fait, les onglets proposent un regroupement thématique d'éléments pouvant appartenir à des tables différentes. Ainsi, la table principale est-elle répartie entre «article», «renvois» et «morphologie». La sous-table «définitions est présentée dans l'onglet «article» et la sous-table «formes nominales» dans l'onglet «morphologie».



Dans l'onglet «article», celui qui est vu à l'ouverture de la base de données et destiné à décrire l'entrée, le premier champ marque la catégorie à laquelle appartient le mot clé. Quand la catégorie est «Agent littéraire», une fenêtre «Agent» est prévue pour permettre de préciser le type d'agent littéraire :

DICTIONNAIRE DE LITTÉRATURE KABYLE

Ameddaḥ

ARTICLE | **RENOIS** | MORPHOLOGIE | ILLUSTRATION

Ameddaḥ catégorie agent littéraire agents poète

3 Ameddaḥ définition n° 1

définition française

Type de poète traditionnel déclamant ses poèmes dans les marchés hebdomadaires ou en se déplaçant de village en village. Les poèmes qu'il récite peuvent être de sa propre composition ou appartenir à d'autres poètes (tous types confondus). Il peut également reprendre ceux consacrés par la tradition comme les fameuses tiqsidin*. Etymologiquement, le terme ameddaḥ est un emprunt à la langue arabe et signifie « celui qui fait (ou dit) des louanges (notamment à l'adresse du prophète Mohamed) ». Parmi les imeddaḥen (pl. de ameddaḥ), que la collecte poétique a répertoriés, figurent eli U eemruš (Mammeri, 1988, pp.). Actuellement, la communauté des imeddaḥen a disparu ; seul Muḥend n Yidir (de son vrai nom ?????, ancien animateur de la chaîne II, né en ?????) continue à déclamer des poèmes (dont une bonne partie de Si Mohand) dans les marchés.

définition kabyle

Enr : 1 sur 1

Cet onglet intègre ensuite la sous-table «définitions» destinée à compléter les informations de base afférentes à l'entrée. Dans le corps des définitions, il est prévu de marquer les mots (clés) qui constituent eux-mêmes des entrées d'article. Ce marquage par un signe (* par exemple) permet de mettre en relation l'objet d'explication avec les autres mots clés.

L'onglet «renvois» assure, comme son nom l'indique, les liens internes à la base.

The screenshot shows a web interface for a Kabyle literature dictionary. At the top, the title 'DICTIONNAIRE DE LITTÉRATURE KABYLE' is displayed in blue. Below it, the entry 'Amğar azemni' is shown. A navigation bar contains four tabs: 'ARTICLE', 'RENVOIS', 'MORPHOLOGIE', and 'ILLUSTRATION'. The 'RENVOIS' tab is currently selected. Under this tab, there are two sections: 'enregistrement lié' and 'renvois bibliographiques'. In the 'enregistrement lié' section, the text 'entrée liée' is followed by a green link 'Settut'. In the 'renvois bibliographiques' section, there are two sub-sections: 'texte littéraire' with a white input field, and 'étude' with a white input field.

Cet onglet est consacré aux renvois bibliographiques du mot descripteur (entrée) expliqué dans l'article. Deux types de renvois sont prévus. Le premier donnera les références du corpus (texte littéraire) que le mot décrit ; il est nommé dans la base de données «Texte». Le deuxième présente les données bibliographiques des études traitant totalement ou partiellement de l'objet de l'article. Il est désigné par le champ «Etude».

Dans l'exemple présenté, le personnage «amğar azemni» (l'entrée de l'enregistrement) est lié au personnage féminin «Settut». Par un double clic sur le champ (les champs contenant un lien sont en vert), nous sommes reconduits sur la consultation de l'enregistrement «Settut» :

DICTIONNAIRE DE LITTÉRATURE KABYLE

Settut

ARTICLE RENVois MORPHOLOGIE ILLUSTRATION

Settut catégorie **personnage littéraire** agents

2 Settut définition n° 1

définition française

Personnage féminin maléfique dans le conte traditionnel. Ce personnage est présenté comme une femme méchante, malveillante et laide mais qui, en se déguisant, peut prendre une image d'une femme, généralement vieille, de bonne apparence. Ce personnage est, dans plusieurs contes, à l'origine du départ du héros de sa maison familiale ou à l'origine des malheurs de ce dernier (ou l'un de ses compagnons, amis ou membres de sa famille). Settut joue généralement le rôle d'opposant ; soit c'est elle qui manipule un roi (ou un personnage puissant) soit c'est elle qui aide l'anti-héros dans ses entreprises contre le gentil et jeune héros. Par le rôle actantiel qu'elle assure, en tant que protagoniste dans le conte traditionnel, elle s'oppose à amgar azemni*. Quand il est observable, le destin de ce personnage est le châtimeur et parfois la mort.

définition kabyle

Enr : 1 sur 1

Enfin, l'onglet «*Morphologie*» concerne la description morphosyntaxique du lexique constituant les mots clés retenus et définis. Cet ensemble de données est important notamment pour la catégorie «Lexique littéraire» qui contient le lexique des performances littéraire, les dénominations des agents littéraires et le métalangage (traditionnel et moderne).

La description prévue peut s'appliquer tant aux formes nominales qu'éventuellement, si cela se révélait nécessaire, à des formes verbales. L'onglet tient compte, pour le moment, de cette éventualité.

DICTIONNAIRE DE LITTÉRATURE KABYLE

Amɣar azemni

ARTICLE
RENOIS
MORPHOLOGIE
ILLUSTRATION

formes verbales

FV1 aoriste <input style="width: 80%;" type="text"/> FV2 inaccompli aoriste <input style="width: 80%;" type="text"/> intensif <input style="width: 80%;" type="text"/> impératif <input style="width: 80%;" type="text"/>	FV3 accompli prétérit <input style="width: 80%;" type="text"/> FV4 accompli négatif prétérit <input style="width: 80%;" type="text"/> négatif <input style="width: 80%;" type="text"/>
--	--

combinatoire préfixale

	pos. 3	pos. 2	pos. 1
factitif	<input type="checkbox"/> p3_S	<input type="checkbox"/> p2_S	<input type="checkbox"/> p1_S
réciproque	<input type="checkbox"/> p3_M	<input type="checkbox"/> p2_M	<input type="checkbox"/> p1_M
réfléchi, expressif	<input type="checkbox"/> p3_N	<input type="checkbox"/> p2_N	<input type="checkbox"/> p1_N
passif	<input type="checkbox"/> p3_T	<input type="checkbox"/> p2_T	<input type="checkbox"/> p1_T

formes nominales

dérivation

catégorie

Amɣar azemni

SINGULIER	PLURIEL
masculin état libre <input style="width: 90%;" type="text"/>	masculin état libre <input style="width: 90%;" type="text"/>
masculin état annexé <input style="width: 90%;" type="text"/>	masculin état annexé <input style="width: 90%;" type="text"/>
féminin état libre <input style="width: 90%;" type="text"/>	féminin état libre <input style="width: 90%;" type="text"/>
féminin état annexé <input style="width: 90%;" type="text"/>	féminin état annexé <input style="width: 90%;" type="text"/>

Enr : 1 sur 1

genre autre genre

valeur particulière

La base de données ainsi préparée est en attente de développements ultérieurs qui seront marqués par des modifications simples : ajouts éventuels, mais aussi peut-être des suppressions, changement de formats, ajout de liens, programmation de procédures de calcul en morphologie flexionnelle ou pour la gestion de l'interface.

Références bibliographiques

Boughiche L., 1997, *Langues et littératures berbères des origines à nos jours. Bibliographie internationale*, Awal/Ibis Press, Paris.

Chaker S., 1992, *Une décennie d'études berbères, 1980-1990 : bibliographie critique*, Bouchène, Alger.

Chaker S. et Bounfour A., 1994, *Langues et littératures berbères : chronique des études*, Inalco/l'Harmattan, Paris.

Citti P. et Detrie M. (dir.), 1992, *Le champ littéraire*, Vrin, Paris.

Dubois Jacques, 1983, *L'institution littéraire : Introduction à une sociologie*, Nathan/Ed. Labor, Brussels.

Fonkoua R. et Halen P. 2001 (Textes réunis par), *Les champs littéraires africains*, Karthala.

Galand L., 1979, *Langue et littérature berbères, vingt-cinq ans d'études*, Cnrs, Paris.

Galand-Pernet Paulette, 1998, *Littératures berbères. Des voix. Des lettres*, PUF, Paris.

Lacoste-Dujardin C., 2005, *Dictionnaire de la culture berbère de Kabylie*, La Découverte.

Lanfry J., 1974, *Table chronologique et index des articles parus au «fichier» de 1946 à 1972*, Le Fichier Périodique, n° 124, 1974 (IV).