

Formalisation de l'amazighe standard avec NooJ

NEJME Fatima Zahra^{1,1} BOULAKNADEL Siham^{1,2}

(1) LRIT, Faculté des Sciences, Université Mohammed V-Agdal, Rabat, Maroc

(2) IRCAM, Avenue Allal El Fassi, Madinat Al Irfane, Rabat-Instituts, Maroc
fatimazahra.nejme@gmail.com, Boulaknadel@ircam.ma

RÉSUMÉ

Depuis l'antiquité, le patrimoine amazighe est en expansion de génération en génération. Cependant, l'accès au domaine des nouvelles technologies de l'information et de la communication (NTIC) s'avère primordial pour sauvegarder et exploiter ce patrimoine et éviter qu'il soit menacé de disparition.

Dans cette perspective, et dans le but de développer des outils et des ressources linguistiques, nous avons entrepris de construire un module NooJ pour la langue amazighe standard (Ameur et al., 2004). Le présent article propose une formalisation de la catégorie nom permettant de générer à partir d'une entrée lexicale son genre (masculin, féminin), son nombre (singulier, pluriel), et son état (libre, annexion).

ABSTRACT

Formalization of the standard Amazigh with NooJ

Since antiquity, the Amazigh patrimony is expanding from generation to generation. However, the access to the domain of new Information and Communication Technologies (NICT) proves to be primordial to safeguard and exploit this patrimony and to prevent that it will be threatened of disappearance.

In this perspective, and in the context of developing tools and linguistic resources, we undertook to build a module NooJ for the standard Amazigh language. This paper proposes a formalization of the category name allowing to generate from a lexical entrance its gender (male, female), its number (singular, plural), and its status (free, annexation).

MOTS-CLÉS : La langue amazighe, NooJ, Morphologie flexionnelle.

Keywords : Amazigh language, NooJ, Inflectional morphology.

1 Introduction

La langue amazighe du Maroc est considéré comme un constituant éminent de la culture marocaine et ce par sa richesse et son originalité. Cependant il a été longtemps écarté sinon négligé en tant que source d'enrichissement culturel. Mais grâce à la création de l'Institut Royal de la Culture Amazighe (IRCAM), cette langue a pu être aménagée et son introduction assurée dans le domaine public notamment dans l'enseignement, l'administration et les médias. Elle a pu avoir une graphie officielle, un codage propre dans le standard Unicode, des normes appropriées pour la disposition d'un clavier amazighe et des structures linguistiques qui sont en phase d'élaboration avec une démarche progressive. La première phase de cette démarche été initiée par la

construction des lexiques (Kamel, 2006; Ameur et al., 2009), l'homogénéisation de l'orthographe et la mise en place des règles de segmentation de la chaîne parlée (Ameur et al., 2006), et par l'élaboration des règles de grammaire (Boukhris et al., 2008). De ce fait elle a eu sa chance de se positionner dans la société globale de l'information.

Cependant, l'amazighe reste encore une parmi les langues peu dotées informatiquement (les langues- π (Berment, 2004)) à cause de la limite des outils informatiques liés à son traitement automatique, ce qui rend difficile son adhésion à ses consœurs dans le domaine des nouvelles technologies de l'information et de la communication (NTIC). Par conséquent, un ensemble de recherches scientifiques et linguistiques sont lancées pour remédier à cette situation. L'un des volets prioritaire de ces recherches, est de concevoir et réaliser des applications capables de traiter d'une façon automatique des données linguistiques.

C'est dans ce contexte, que se situe notre contribution qui s'ajoute aux efforts de la communauté scientifique pour la construction d'outils et de ressources linguistiques en langue amazighe standard du Maroc. L'un de nos objectifs est la formalisation du vocabulaire amazighe : nom, verbe et particules. Dans cet article nous nous sommes restreint dans un premier lieu à la formalisation de la catégorie nom. Pour ce faire, nous avons opté pour l'utilisation de la plateforme linguistique de développement NooJ, compte tenu de ses avantages, pour la construction d'un module pour l'amazighe, dont l'objectif est de l'utiliser dans l'enseignement au Maroc.

Le présent article se structure autour de trois volets: le premier présente un descriptif des particularités de la langue amazighe du Maroc, et le deuxième expose le module NooJ, ainsi qu'un exemple de notre dictionnaire, et de grammaires flexionnelles, alors que le dernier volet est consacré à la conclusion et aux perspectives.

2 Particularités de la langue amazighe

2.1 Historique

L'amazighe connu aussi sous le nom du berbère ou Tamazight (tamazɣt), est une famille de langues séparée en deux branches : langues berbères du Nord et du Sud. Elle présente la langue d'une population appelée « Imazighen » qui s'est installée depuis l'antiquité sur un espace géographique allant depuis le Maroc, avec 50% de la population globale (Boukous, 1995), jusqu'à l'Égypte avec environ 27%, en passant par l'Algérie avec 25%, la Tunisie avec 5% à 10%, le Niger et le Mali (Chaker, 2003).

Au Maroc, l'amazighe se répartit selon deux types de dialectes: les dialectes régionaux et les dialectes locaux. Pour le premier type, nous avons trois grandes variétés régionales : le Tarifit au Nord, le Tamazight au Maroc central et au Sud-Est et le Tashelhit au Sud-Ouest et dans le Haut-Atlas. Chacun de ces dialectes comprend des sous-dialectes ou dialectes locaux constituant le deuxième type. A titre d'exemple, le dialecte régional Tamazight contient un ensemble de sous-dialectes, dont nous citerons: le Tamazight de Béni-Mellal, le Tamazight d'Errachidia, le Tamazight de Ait Sadden, etc.

La langue amazighe connaît une grande richesse au niveau de son vocabulaire. Ainsi, un seul sens est rendu de plusieurs façons dans chaque dialecte ou sous-dialecte. Par

exemple : tête = « ixf, aqrru, ukhsas, azllif, axshash, ajdjif ».

2.2 Caractéristiques de la langue amazighe standard

Dans cet article, nous allons restreindre notre étude sur l'amazighe standard du Maroc. Depuis quelques années, le Maroc s'est engagé pour réaliser un processus de standardisation¹ de la langue amazighe (Ameur et al., 2004a), qui a pour vocation d'uniformiser les structures et à atténuer les divergences, en éliminant les occurrences non distinctives qui entraînent souvent des problèmes d'intercompréhension. Ce processus de standardisation consiste à :

- adopter une graphie standard normalisée sur une base phonologique ;
- adopter un lexique de base commun ;
- appliquer: les mêmes règles orthographiques, les mêmes consignes pédagogiques, et les mêmes formes néologiques ;
- exploiter la variation dialectale afin de sauvegarder la richesse de la langue.

2.2.1 Système d'écriture

En se basant sur le système original, l'IRCAM a développé un système d'alphabet sous le nom de Tifinaghe-IRCAM (voir annexe 1). Il s'écrit de gauche à droite. Cet alphabet standardisé est basé sur un système graphique à tendance phonologique. Cependant, il ne retient pas toutes les réalisations phonétiques produites, mais uniquement celles qui sont fonctionnelles (Ameur et al., 2004b). Il est composé de 27 consonnes, 2 semi-consonnes, 3 voyelles pleines et une voyelle neutre.

A partir de ces propriétés morphologiques, l'amazighe peut être considéré comme une langue complexe dont les mots peuvent être classés en trois catégories morpho-syntaxiques : nom, verbe et particules (Boukhris et al., 2008).

2.2.2 Nom

En amazighe, le nom est une unité lexicale formée d'une racine et d'un schème. Il possède deux caractéristiques, la première est qu'il peut prendre différentes formes à savoir: une forme simple (argaz "homme"), forme composée (ⵜⴰⴳⴷⴰⵏⵜ "la famine") ou bien forme dérivée (ⵜⴰⴳⴷⴰⵏⵜ "la communication"). La deuxième caractéristique correspond à la variation, il varie en genre (féminin, masculin), en nombre (singulier, pluriel) et en état (libre, annexion).

1. Le genre : le nom amazighe connaît deux genres, le masculin et le féminin.

Le nom masculin: il commence généralement par une des voyelles initiales: ⵏ 'a', ⵢ 'i' ou bien ⵓ 'u', à titre d'exemple: ⵙⴰⵏⵓⵏ "visage", ⵙⴰⵏⵓⵏ "tête". Cependant, il existe certains nom qui font l'exception: ⵙⴰⵏⵓⵏ " (ma) mère", ⵙⴰⵏⵓⵏ " (ma) fille", ⵙⴰⵏⵓⵏ " (ma) sœur".

¹ La standardisation de l'amazighe s'impose d'autant plus avec son introduction dans le système éducatif, et avec le rôle que cette langue est appelée à jouer « dans l'espace social, culturel et médiatique, national, régional et local » (cf. article 2 du Dahir portant création de l'IRCAM).

Le nom féminin : celui-ci est généralement de la forme +...+ 't...t', à l'exception de certains noms qui ne portent que le + initial ou le + final du morphème du féminin: +oΛηο "gerbe", QQEε̄ɣ+ "fatigue". Dans le cas général, le féminin est formé à partir du radical d'un nom masculin par l'ajout du morphème discontinue +...+ 't...t': ε̄Oηε "marié" -> +ε̄Oηε+ "mariée". Dans le cas des noms composés, le féminin est formé par une préfixation du morphème à valeur attributive (Oε "celui à / ayant"), à valeur d'appartenance ou d'affiliation (ε̄, oɣ+ "celui / ceux appartenant à, relevant de"): Oε ε̄Ληηηο "menteur" -> E ε̄Ληηηο "menteuse".

2. Le nombre : le nom amazighe, qu'il soit masculin ou féminin, possède un singulier et un pluriel. Ce dernier est obtenu selon trois types: le pluriel externe, pluriel interne et le pluriel mixte.

Le pluriel externe : le nom ne subit aucune modification interne, et le pluriel est obtenue par une alternance vocalique accompagné par une suffixation de 'l' ou une de ses variantes (ε̄l, o, oɣ, ll, oll, ll, ll, ll, ll, ll, ll, ll): oXXoE -> ε̄XXoE "maisons", +oOo+ -> +ε̄Oo+ε̄l "filles".

Le pluriel interne (ou brisé): le pluriel brisé est obtenue par une alternance vocalique plus un changement de voyelle internes (oΛoO -> ε̄ΛoO "montagnes").

Le pluriel mixte: ce pluriel est formé par une alternance d'une voyelle interne et/ou d'une consonne plus une suffixation par 'l' (ε̄ηε "part"-> ε̄ηηο "parts"); ou bien par une alternance vocalique initiale accompagné d'un changement vocalique final a 'a' plus une alternance interne (oEεεo "dernier" -> ε̄Eεεo "derniers").

Le pluriel en ε̄Λ : ce type de pluriel est obtenu par une préfixation de ε̄Λ du nom au singulier. Il est appliqué à un ensemble de cas de noms à savoir : des noms à initiale consonantique, des noms propres, des noms de parenté, des noms composés, des numéraux, ainsi que pour les noms empruntés et intégrés (Xoηε "mon) oncle"-> ε̄Λ Xoηε).

3. L'état : nous distinguons deux états pour les noms amazighs, l'état libre (EL.) et l'état d'annexion (EA.).

L'état libre : dans cet état, la voyelle initiale du nom ne subit aucune modification: oOxo* "homme", +oEoO+ "terre, pays". Le nom est en état libre lorsqu'il s'agit : d'un mot isolé de tout contexte syntaxique, d'un complément d'objet direct, ou bien d'un complément de la particule prédictive Λ "c'est".

L'état d'annexion : cet état est fondé sur une modification de l'initiale du nom dans des contextes syntaxiques déterminés. Il prend l'une des formes suivantes: alternance vocalique a/u au cas des noms masculins (oOxo* "homme" -> oOxo*), chute de la voyelle initiale au cas des noms féminins (+oEyoO+ "femme" -> +EyoO+), addition d'un ll ou ε̄ aux noms à voyelle o ou ε̄ (ε̄ηO "langue" -> ε̄ε̄ηO), maintien de la voyelle initiale a avec apparition de la semi-consonne ll seulement au cas du masculin; le féminin ne subit aucune modification (oOo "jour" -> ll oOo [masc.], +oΛΛoO+ "maison" -> +oΛΛoO+ [femin.]). L'état d'annexion est réalisé dans les contextes syntaxiques suivants : lorsque le sujet lexical suit le verbe, après une préposition, et après un coordonnant.

2.2.3 Verbe

En amazighe, le verbe peut prendre deux formes : simple ou dérivée. Le verbe simple est composé d'une racine et d'un radical. Par contre le verbe dérivé est obtenu à partir des verbes simples par une préfixation de l'un des morphèmes suivants : ⵜ/ ⵜⵜ, ++ et ⵏ/ ⵏⵏ. La première forme (ⵜ/ ⵜⵜ) correspond à la forme factitive ou la forme en ⵜ, la deuxième marque la forme passive ou la forme en ++, et la troisième désigne la forme réciproque ou en ⵏ. Le verbe, qu'il soit simple ou dérivé, se conjugue selon quatre thèmes : l'aoriste, l'inaccompli, l'accompli positif et l'accompli négatif.

2.2.4 Particule

Les particules sont un ensemble de mots amazighs qui ne sont ni des noms, ni des verbes, et jouent un rôle d'indicateurs grammaticaux au sein d'une phrase. Cet ensemble est constitué de plusieurs éléments à savoir :

- Les particules d'aspect, d'orientation et de négation;
- Les pronoms indéfinis, démonstratifs, possessifs et interrogatifs;
- Les pronoms personnels autonomes, affixes sujet, affixes d'objet direct et indirect, compléments du nom ordinaire et de parenté, compléments de prépositions;
- Les adverbes de lieu, de temps, de quantité et de manière;
- Les prépositions;
- Les subordonnants et les conjonctions.

3 Module NooJ pour L'Amazighe

« NooJ (Silberztein, 2007) est une plateforme de développement linguistique qui offre un ensemble d'outils et méthodologies permettant de formaliser des langues tout en construisant, gérant et accumulant un grand nombre d'application de traitement automatique des langues (TAL), et les appliquant à des corpus de taille importante». Il permet de formaliser différents niveaux et composantes des langues naturelles, à savoir: l'orthographe, la morphologie (flexionnelle et dérivationnelle), le lexique (de mots simples, mots composés et expressions figées), la syntaxe locale et désambiguïsation, la syntaxe, la sémantique et les ontologies. Pour chacun de ces niveaux, NooJ propose une méthodologie, un ou plusieurs formalismes adaptés, des outils-logiciels de développement et un ou plusieurs analyseurs automatiques de textes.

Actuellement, les utilisateurs de NooJ forment un public très varié en extension, ce qui a permis de développer des ressources linguistiques à large couverture dans une vingtaine de langues (arabe, arménien, bulgare, catalan, chinois, anglais, français, hébreu, hongrois, italien, polonais, portugais, espagnol, vietnamien et biélorusse).

Compte tenu de ces avantages, nous avons entrepris de construire un module NooJ pour la langue amazighe. Notre but est la formalisation du vocabulaire de cette langue. Cependant, dans cette contribution nous visons une formalisation de la catégorie nom permettant ainsi de générer à partir d'une entrée lexicale son genre (masculin, féminin),

son nombre (singulier, pluriel), et son état (libre, annexion). A cet effet, nous avons construit un exemple de dictionnaire contenant un ensemble de noms de test. Chaque nom est associé à un ensemble d'informations linguistiques, tels que la catégorie grammaticale, le paradigme flexionnel. Ce paradigme est décrit et stocké dans des grammaires flexionnelles, et permet de reconnaître toutes les formes flexionnelles correspondantes.

3.1 Formalisation des règles morphologiques

Cette étude présente l'implémentation des règles de flexion permettant de générer à partir d'un nom ses informations flexionnelles : genre, nombre et état.

Ainsi, nous avons formalisé ces paradigmes flexionnels à l'aide d'une collection de graphes et de sous graphes présentant des grammaires flexionnelles qui décrivent les modèles de flexion en amazighe (genre, nombre et état), et qui sont stockées dans le fichier des flexions « Flexion.nof » qui se présente comme suit :

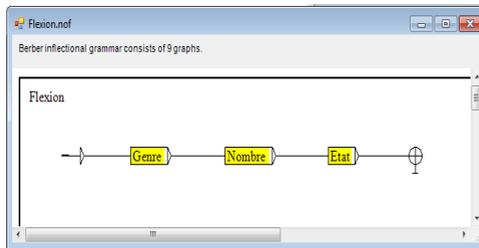


FIGURE 1 – Paradigme flexionnel

Ce graphe contient trois sous graphes: le premier « Genre » présente le genre, le deuxième « Nombre » concerne le suffixe du nombre, et le troisième « Etat » qui présente l'état d'annexion. Chaque formalisation de sous graphe peut contenir un ou bien plusieurs sous graphes. Ces formalisations sont basées sur l'utilisation de certaines commandes génériques prédéfinies: <LW> déplacement au début du lemme, <RW> déplacement à la fin du lemme, <R> déplacement vers la droite, <S> suppression du caractère courant.

3.1.1 Genre

Afin de formaliser le genre, nous avons construit ce graphe qui permet de générer à partir d'un nom masculin son correspondant féminin. La règle consiste à ajouter le morphème discontinu + 't' au début et à la fin du nom.

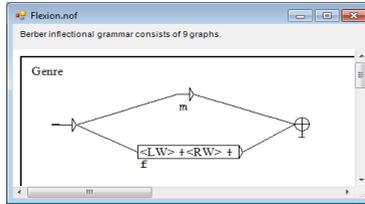


FIGURE 2 – Sous graphe Genre

3.1.2 Nombre

En amazighe, le pluriel prend trois formes variées : le pluriel externe, pluriel interne et le pluriel mixte. Pour chacun de ces types, Les formes du pluriel sont nombreuses et généralement imprévisibles, ce qui rend cette langue assez complexe au niveau morphologique. Dans cet article, nous avons choisi de consacrer plus d'importance au pluriel externe (ou régulier). Le pluriel externe est formé, généralement, par une alternance vocalique accompagné par une suffixation de 'l' ou une de ses variantes (ⵍ, ⵎ, ⵏⵍ, ⵎⵍ, ⵍⵍ, ⵎⵍ, ⵍⵍ, ⵎⵍ, ⵍⵍ, ⵎⵍ, ⵍⵍ).

Notre approche tient compte de formaliser une quantité suffisante de descriptions de suffixations. Ainsi, Nous nous sommes basés pour l'implémentation des règles flexionnelles sur les travaux de (Boukhris et al., 2008) et ceux de (Oulhaj, 2000). Nous allons spécifier deux catégories :

Noms masculins :

Étant donné que les noms masculins commencent généralement par l'une des voyelles : 'o', 'ɛ', 'ɔ', l'alternance vocalique ne concerne dans ce cas que les noms commençant par un 'o' qui va se transformer en 'ɛ'. Or, pour la suffixation, nous avons pu définir quatre règles générales, que nous avons adoptés afin d'établir les formes fléchies :

1. La première: si le nom est monosyllabique, il y a une suffixation de l'indice 'ⵍⵍ' : ⵍⵍⵍ (tête) -> ⵍⵍⵍⵍⵍⵍ.

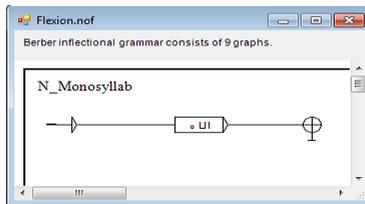


FIGURE 3 – Pluriel des noms monosyllabiques

2. La deuxième: si le nom commence et se termine par 'ɛ' nous ajoutons une suffixation de 'ⵍ': ⵍⵍⵍⵍ (marié)-> ⵍⵍⵍⵍⵍ.

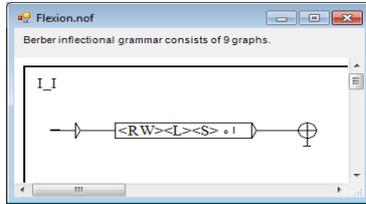


FIGURE 4 – Pluriel des noms en ‘...ɛ’

3. La troisième: si le nom commence et se termine par ‘o’, la voyelle initiale se transforme en ‘ɛ’, et une suffixation de l’indice ‘+’ est appliqué: $\circ\text{O}\text{O}$ (bureau)-> $\varepsilon\text{O}\varepsilon\text{O}\circ+\text{I}$.

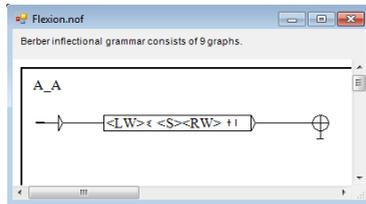


FIGURE 5 – Pluriel des noms en ‘...o’

4. La quatrième: si le nom commence par ‘o’ et se termine par une consonne, la voyelle initial se transforme en ‘ɛ’ et une suffixation de ‘+’ est appliqué: $\circ\Lambda\text{H}\varepsilon\text{O}$ (livre)-> $\varepsilon\Lambda\text{H}\varepsilon\text{O}+\text{I}$.

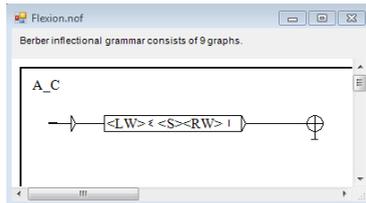


FIGURE 6 – Pluriel des noms en ‘...c’ (c : consonne)

Noms féminins :

Le nom féminin commence et se termine par un ‘+’. Ainsi, nous avons défini deux critères à la base de deux lettres, la lettre qui suit le premier ‘+’ et l’autre qui précède le dernier ‘+’.

1. Le premier : si le nom est de la forme ‘+...v+’ (v : voyelle), la voyelle ‘o’ est transformé en ‘+’ et une suffixation de ‘ɛ+’ est appliquée.

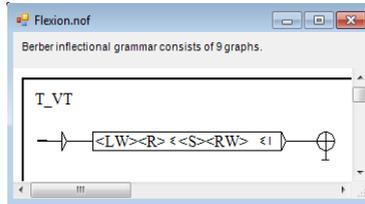


FIGURE 7 – Pluriel des noms féminins ‘+o...v+’ (v : voyelle)

2. Le deuxième : si le nom est de la forme ‘+o...c+’ (c : consonne), la voyelle ‘o’ est transformé en ‘ɛ’, le dernier ‘+’ est supprimé, et une suffixation de l’indice ‘€l’ est appliquée.

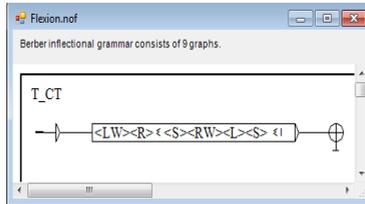


FIGURE 8 – Pluriel des noms féminins ‘+o...c +’ (c : consonne)

3.1.3 État

Afin de formaliser l’état d’annexion, nous avons distingué deux cas: le cas masculin et le cas féminin.

- Le nom masculin : l’état d’annexion est défini par modification de l’initiale du nom dans des contextes syntaxiques déterminés. Nous citerons le cas de l’initiale ‘ɣ’, la règle consiste à ajouter un ‘ɣ’ au début du nom, il devient ‘ɣɣ’: ɣɣɣ (mouche)-> ɣɣɣɣ.

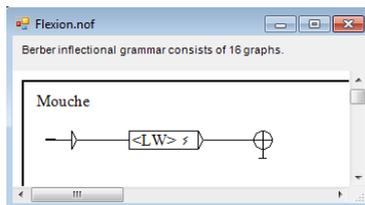


FIGURE 9 – Graphe du paradigme flexionnel « Mouche »

- Le cas féminin est défini par la chute de la voyelle initiale : +o€O+ (pays)-> +€O+.

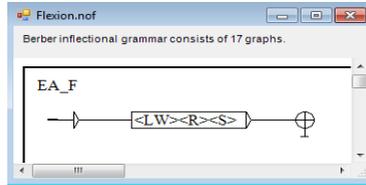


FIGURE 10 – État d'annexion des noms féminins

Conclusion et perspectives

Cet article décrit une formalisation de la catégorie nom en langue amazighe standard, au sein de l'environnement linguistique de développement NooJ. Cette démarche permet de générer à partir d'une entrée lexicale son genre, son nombre et son état. Ainsi, nous avons construit un dictionnaire contenant un ensemble de noms, accompagnés d'un ensemble de grammaires présentant le paradigme flexionnel et permettant de reconnaître toutes les formes fléchies correspondantes.

Certes, le développement de cet outil ne présente qu'une étape préliminaire pour notre but qui est la formalisation du vocabulaire amazighe.

Références

- AMEUR M., BOUMALK A. (DIR) (2004a). Standardisation de l'amazighe, Actes du séminaire organisé par le Centre de l'Aménagement Linguistique à Rabat, 8-9 décembre 2003, Publication de l'Institut Royal de la Culture Amazighe, Série : Colloques et séminaires.
- AMEUR M., BOUHJAR A., BOUKHRIS F., BOUKOUSS A., BOUMALK A., ELMEDLAOUI M., IAZZI E., SOUIFI H. (2004b). Initiation à la langue amazighe. Rabat, Maroc: IRCAM.
- AMEUR M., BOUHJAR A., BOUKHRIS F., BOUKOUSS A., BOUMALK A., ELMEDLAOUI M., IAZZI E. (2006). Graphie et orthographe de l'amazighe. Rabat, Maroc : IRCAM.
- AMEUR M., BOUHJAR A., BOUMALK A., EL AZRAK N., LAABDELAOUI R. (2009). Vocabulaire de la langue amazighe (amazighe-arabe). Rabat, Maroc: IRCAM.
- BERMENT V. (2004). Méthodes pour informatiser des langues et des groupes de langues peu dotées, Thèse de doctorat de l'Université J. Fourier - Grenoble I, France.
- BOUKHRIS F., BOUMALK A., ELMOUJAHID E., SOUIFI H. (2008). La nouvelle grammaire de l'amazighe. Rabat, Maroc: IRCAM.
- BOUKOUS A. (1995), Société, langues et cultures au Maroc: Enjeux symboliques, Casablanca, Najah El Jadida.
- CHAKER S. (2003), Le berbère, Actes des langues de France, 215-227.
- GREENBERG J. (1966). The Languages of Africa. Mouton, USA: The Hague.
- KAMEL S. (2006). Lexique Amazighe de géologie. Rabat, Maroc: IRCAM.

