

Développement de ressources pour la langue amazighe : Le Lexique Morphologique El-AmaLex

NEJME Fatima Zahra¹ BOULAKNADEL Siham^{1,2} ABOUTAJDINE Driss¹

(1) LRIT, Unité Associée au CNRST (URAC 29), Faculté des Sciences, Mohammed V-Agdal, Rabat, Maroc.

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

RESUME

Notre étude s'inscrit dans le cadre de doter la langue amazighe d'outils et de ressources linguistiques nécessaires pour son traitement automatique. Ce traitement ne peut être réalisé sans un lexique électronique qui doit être aussi complet que possible. Dans ce contexte nous avons entrepris de développer un lexique morphologique "El-AmaLex" pour l'amazighe à large couverture et qui regroupe un ensemble d'informations morphologiques liées à chaque entrée.

ABSTRACT

Amazighe Language Development Resources: The Morphological lexicon El-AmaLex.

Our study is part of the development of Amazighe tools and language resources for its automatic processing. This treatment can not be achieved without an electronic lexicon which should be as complete as possible. In this context we have undertaken to develop a morphological lexicon "El-AmaLex" for Amazighe which includes a set of morphological information related to each entry.

MOTS-CLES : La langue amazighe, TALN, NooJ, Lexique morphologique, Morphologie flexionnelle, Morphologie dérivationnelle.

KEYWORDS: Amazigh language, NLP, NooJ, Morphological lexicon, Inflectional morphology, derivational morphology.

1 Introduction

Durant ces dernières décennies, entraînées par l'évolution permanente de l'informatique et la constante informatisation de nos sociétés, les technologies du langage ont connu des progrès majeurs. Le déploiement de ces technologies pour des langues peu-dotées représente un challenge important. Cependant, les technologies du langage sont toujours pleinement confrontées au manque de ressources lexicales et représentent ainsi peu la diversité des langues. Comme beaucoup d'autres, la langue Amazighe est fortement touchée par cette lacune.

Au Maroc, la langue amazighe a toujours disposé d'un statut restreint, bien qu'elle soit considérée comme la plus ancienne langue d'Afrique du Nord ; elle est fortement employée sur une aire très vaste par environ 50% de la population marocaine (Boukous, 1995). Grâce aux efforts de revitalisation et standardisation de la langue, la langue amazighe est maintenant une langue qui en

possède tous ces attributs: dotée d'une graphie officielle, un codage propre dans le standard Unicode, une grammaire, une orthographe ainsi qu'un vocabulaire très riche et une littérature orale riche. Certes ces étapes de standardisation sont élémentaires et essentielles mais ne sont pas suffisantes pour qu'une langue peu dotée informatiquement telle que l'amazighe puisse franchir le seuil de la mondialisation informatique et de rejoindre ses consœurs dans ce domaine.

Dans cette perspective s'inscrit de nombreuses recherches scientifiques, principalement celles se focalisant sur la correction orthographique (Es Saady et al., 2009), la reconnaissance optique des caractères (Fakir et al., 2009), et celles s'occupant de la conception et la réalisation des ressources et outils linguistiques (Talha et al., 2015; Outahajala et al., 2015; Nejme et al., 2012a ; Nejme et al., 2012b ; Nejme et al., 2012c ; Nejme et al., 2013a). Or, pour mieux mener ce chantier de construction de ressources qui s'est ouvert à la langue amazighe, il s'avère primordial de doter la langue amazighe d'un lexique morphologique essentiel à son traitement automatique. En ce sens, nous présentons dans cet article "El-AmaLex", un lexique morphologique de la langue amazighe à large couverture, implémenté avec la plateforme linguistique NooJ. Ce travail présente une continuation de nos travaux déjà réalisés (Nejme et al., 2012b ; Nejme et al., 2013a ; Nejme et al., 2013b ; Nejme et al., 2015).

Dans la suite de cet article, nous présentons dans la section 2 un descriptif des particularités de la langue amazighe du Maroc. Puis, nous détaillons dans la section 3 les étapes de construction de notre lexique morphologique. Alors que nous consacrons la section 4 à la conclusion.

2 Particularités de la langue amazighe

2.1 Historique

La langue amazighe connue aussi sous le nom du berbère ou Tamazight (ⵜⴰⴳⴷⵓⴷⴰⵢⵜ), est une branche de la famille de langue afro-asiatique (chamito-sémitique) (Greenberg, 1966; Ouakrim, 1995) séparée en deux : langues berbères du Nord et du Sud. Elle présente la langue d'une population appelée "Imazighen" qui se présente à l'heure actuelle dans une dizaine de pays allant depuis le Maroc, avec 50% de la population globale (Boukous, 1995), jusqu'à l'Égypte, en passant par l'Algérie avec 25%, la Tunisie, la Mauritanie, la Libye le Niger et le Mali (Chaker, 2003).

Au Maroc, sur le plan linguistique, l'amazighe est caractérisée par la prolifération des dialectes en raison de facteurs historiques, géographiques et sociolinguistiques. Elle se répartit en trois grandes régions dialectales qui couvrent l'ensemble des régions montagneuses : au nord-est, le Rif avec le dialecte Tarifite ; au centre, le Moyen-Atlas et une partie du Haut-Atlas avec le dialecte Tamazighte ; au sud et sud-ouest, le Haut-Atlas, l'Anti-Atlas et Sous, le domaine chleuh avec le dialecte Tachelhite.

Jusqu'à 1994 l'amazighe a été exclusivement réservée au domaine familial (Boukous, 1995). Cependant, suite au discours royal en 2001, l'amazighe est devenue une langue institutionnelle par la création de l'IRCAM. Et grâce à la constitution de 2011, l'amazighe a joui auprès de sa consœur l'arabe d'un statut d'une langue officielle.

2.2 Caractéristiques de l'amazighe du Maroc

2.2.1. Système d'écriture

Depuis l'antiquité la langue amazighe possède sa propre écriture, Tifinaghe, qui est toujours utilisée chez les Amazighes des zones sahariennes (Touaregs). Néanmoins, le long de l'histoire, particulièrement depuis la fin des années 1960, cette écriture a assimilé des changements et des développements dont l'objectif est de fournir à la langue Amazighe un système alphabétique standard plus adéquat et utilisable pour tous les parlars amazighes actuels. Ainsi en 2003, partant d'un héritage aussi bien ancien que moderne et contemporain, l'IRCAM a développé un système d'alphabet sous le nom de Tifinaghe-IRCAM. 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., 2004). Cet alphabet comporte :

- 27 consonnes dont : les labiales (ⵀ, ⵀ, ⵀ), les dentales (ⵏ, ⵏ, ⵏ, ⵏ, ⵏ, ⵏ, ⵏ), les alvéolaires (ⵉ, ⵉ, ⵉ, ⵉ), les palatales (ⵇ, ⵇ), les vélares (ⵔ, ⵔ), les labiovélares (ⵔ, ⵔ), les uvulaires (ⵛ, ⵛ), les pharyngales (ⵏ, ⵏ) et la laryngale (ⵏ);
- 2 semi-consonnes : ⵏ et ⵏ ;
- 4 voyelles : trois voyelles pleines ⵏ, ⵏ, ⵏ et la voyelle neutre (ou schwa) ⵏ qui a un statut assez particulier en phonologie amazighe.

2.2.2. Encodage Unicode

Depuis l'adaptation de Tifinaghe comme graphie officielle au Maroc pour la langue amazighe, l'encodage Tifinaghe est devenu nécessaire. Pour cette raison, des efforts considérables ont été investis par le centre des études et systèmes d'information et de communication de l'IRCAM. Ces efforts ont abouti à un codage Unicode constitué de quatre sous-ensembles de caractères Tifinaghe à savoir : l'ensemble de base de l'IRCAM, l'ensemble étendu de l'IRCAM, et d'autres lettres néo-Tifinaghe ainsi que des lettres Touareg moderne. Les deux premiers sous-ensembles constituent les ensembles de caractères choisis par l'IRCAM.

2.2.3. Aperçu de la morphologie de l'amazighe

La langue amazighe présente une morphologie riche et complexe. Les mots peuvent être classés en quatre catégories morphosyntaxiques: nom, verbe, pronoms et mots outils (Boukhris et al., 2008).

- *Nom* : En amazighe, le nom est une unité lexicale formée d'une racine et d'un schème, qui indique un substantif ou un adjectif. Il peut être sous une forme simple (ⵏⵏⵏ [argaz] "homme"), composée (ⵏⵏⵏⵏⵏ [buhyyuf] "la famine") ou dérivée (ⵏⵏⵏⵏⵏ [amsawaf] "la communication"). Il varie en genre, en nombre et en état. Le nom existe en deux genres : le masculin et le féminin. Le masculin commence en général par une des voyelles initiales ⵏ [a], ⵏ [i] ou ⵏ [u]. Quant au féminin, il est généralement formé à base du masculin par l'ajout du morphème discontinu ⵏ...ⵏ [t...t]. Cependant, il possède un singulier et un pluriel qui se réalise selon trois types. Le pluriel externe est formé par une alternance vocalique et par la suffixation de l'indice ⵏ [n] ou l'une de ses variantes (ⵏⵏⵏⵏⵏ [axxam] "maison" → ⵏⵏⵏⵏⵏ [ixxamn] "maisons"). Le pluriel interne est constitué d'une alternance vocalique accompagnée d'un changement de voyelles internes (ⵏⵏⵏⵏⵏ [tiymst] "dent" → ⵏⵏⵏⵏⵏ)

[tiymas] “dents”). Le pluriel mixte est formé par l’alternance d’une voyelle interne et/ou finale accompagnée par une suffixation de l [n] (ⵏⵉⵍ [ili] “part” → ⵏⵉⵏⵏ [ilan] “parts”) ; ou une alternance vocalique initiale accompagnée d’un changement vocalique final ◦ [a] et d’une alternance interne (ⵏⵉⵏⵏⵓⵔ [amggaru] “dernier” → ⵏⵉⵏⵏⵓⵔⵓ [imggura] “derniers”). En outre, le nom a deux états : l’état libre ou il ne subit aucune modification et l’état d’annexion qui se manifeste par une variation des noms à initiale vocalique (ⵓⵔⵗⵓⵎ [argaz] → ⵓⵔⵗⵓⵎ [urgaz] “homme”). L’état d’annexion est réalisé quand le nom a la fonction de sujet lexical postposé au verbe et est précédé d’un coordonnant, de termes d’attribution, de termes d’appartenance et de filiation, de nom de nombre ou d’une préposition à l’exception de ⵎ/ⵓ [al/ar] “jusqu’à” et ⵓⵎ [bla] “sans”.

- *Verbe* : En amazighe, le verbe peut prendre deux formes : simple ou dérivée. Le verbe simple est composé d’une racine et d’un radical. Alors que la forme dérivée est obtenue à partir des verbes simples par la préfixation de l’un des morphèmes suivants: ⵓ [s]/ⵓⵓ [ss], ⵜⵜ [tt] et ⵏ [m]/ⵏⵏ [mm]. La première forme correspond à la forme factitive, la deuxième marque la forme passive et la troisième désigne la forme réciproque. Le verbe, qu’il soit simple ou dérivé, se conjugue selon quatre aspects : l’aoriste, l’accompli, l’inaccompli et l’accompli négatif et possède trois modes : l’indicatif, le participe et l’impératif (l’impératif simple et intensif).
- *Pronom* : Le paradigme des pronoms comprend : les pronoms personnels (ⵏⵏⵏ [nkk] “moi”), les pronoms possessifs (ⵏⵉⵎ [winu] “le mien”), les pronoms démonstratifs (ⵏⵓⵏ [wad] “celui-ci”), les pronoms interrogatifs (ⵏⵓ [ma] “que”) et les pronoms indéfinis (ⵏⵓⵓ [kra] “quelque chose”). Généralement, certains pronoms sont flexionnelles telles que les pronoms possessifs (ⵏⵉⵎⵏ [winns] “le sien” → ⵏⵉⵎⵏⵏ [winnsn] “le leur”).
- *Mots outils* : Les mots outils 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 adverbes de lieu, de temps, de quantité et de manière; les prépositions; les subordinants et les conjonctions (Boukhris et al., 2008). Généralement, les mots outils sont invariables.

3 El-AmaLex : Lexique Morphologique

L’élaboration de ressources linguistique électronique est devenue une nécessité pour les langues véhiculaires, compte tenu de l’impact que représente la disponibilité de telles ressources à plus qu’un niveau. Les lexiques morphologiques représentent l’une des ressources les plus sollicitées. C’est dans cette optique que se situe notre travail dont le but est d’élaborer un lexique morphologique pour la langue amazighe en utilisant la plateforme linguistique NooJ.

3.1 NooJ : Plateforme de développement linguistique

NooJ¹, publié en 2002 par Max Silberztein (Silberztein et Tutin, 2004 ; Silberztein, 2007), est un environnement de développement linguistique qui permet de construire et de gérer des dictionnaires électroniques et de grammaires formelles à large couverture afin de formaliser les différents phénomènes linguistiques qui sont : l’orthographe, la morphologie (flexionnelle et dérivationnelle), le lexique (de mots simples, mots composés et expressions figées), la syntaxe

¹ <http://www.nooj4nlp.net/>

(locale, structurelle et transformationnelle), la désambiguïsation, la sémantique et les ontologies. Ces descriptions, formalisées à l'aide des machines à états finis tels que les automates à états finis, les transducteurs à états finis et les réseaux de transition récursifs ; peuvent ensuite être appliquées pour traiter des textes et corpus de taille importante afin de localiser les modèles morphologiques, lexicologiques et syntaxiques, lever les ambiguïtés et étiqueter les mots simples et composés.

Le module morphologique de NooJ, utilisé tout au long de cet article, se base sur des opérateurs de transformations (eg. <L> : déplacement vers la gauche, <RW> : aller à la fin du mot, etc) à l'intérieur des formes et des graphes morphologiques décrivant des règles grammaticales à large couverture. Ces transformations fonctionnent sur une pile et nécessitent un temps de transformation en $O(n)$. Ainsi, elles garantissent une correspondance en un temps linéaire entre le lemme et sa forme flexionnelle ou dérivationnelle. Quant aux grammaires morphologiques, elles sont construites en utilisant l'éditeur de graphes de NooJ et représentées sous forme de transducteurs à états finis (FST). Elles représentent des séquences de lettres et associent leurs reconnaissances à la production des informations lexicales correspondantes (étiquette grammaticale, un ensemble d'informations sémantiques, etc.).

3.2 Construction du lexique El-AmaLex

Etant donné que l'amazighe fait partie des langues peu dotées informatiquement et vu que jusqu'à ce jour nous n'avons pas de dictionnaire sous format électronique, nos sources de repérage ont été des dictionnaires en papier. Ainsi, nous nous sommes basés sur un ensemble de ressources à savoir: la nouvelle grammaire de la langue amazighe (Boukhris et al., 2008), le dictionnaire de Taifi (Taifi, 1988), le vocabulaire de la langue amazighe (Ameur et al., 2006), le vocabulaire des médias (Ameur et al., 2009a), le vocabulaire grammaticale (Ameur et al., 2009b), le lexique scolaire (Aagnaou et al., 2011) et le manuel de conjugaison amazighe (Elabdellaoui et al., 2012). Les entrées lexicales ont été collectées manuellement en suivant deux étapes : (1) construction d'entrées lexicales à partir des ressources ci-dessus et (2) vérification et nettoyage manuelle des entrées obtenues. Une description du lexique obtenu est fournie dans le tableau suivant (cf. Table 1).

Catégories grammaticales		Nombre d'entrées
Noms	Simple	17622
	Composés	1577
	Dérivés	462
Verbes		3695
Pronoms		235
Mots outils		443
Totale		24034

TABLE 1 – Les entrées lexicales

Nous présentons par la suite la structure des entrées de notre lexique ainsi que les diverses informations qui peuvent être associées à chaque entrée.

3.2.1. Structures des entrées

Chaque entrée dans le lexique présente généralement les items suivants :

- Le lemme,
- La catégorie lexicale,
- Les traits syntactico-sémantiques,
- L’ensemble des informations morphologiques : genre, nombre et état,
- Et la traduction s’elle existe.

La première information que nous avons attribuée aux entrées de notre lexique correspond à la catégorie grammaticale désignée par un code écrit en majuscules. Le tableau suivant présente quelques abréviations utilisées pour chaque partie du discours.

Catégories	Code	Exemples
Nom	N	◦Oᄆ◦ᄆ [argaz] “homme”
Verbe	V	◦ᄆᄆ [aws] “aide”
Adjective	ADJ	ᄆᄆᄆᄆ [uffir] “clandestin”
Adverbe	ADV	ᄆ◦ [da] “ici”
Préposition	PREP	ᄆ [i] “à”
Pronoms	PRON	ᄆᄆᄆ [nkk] “moi”

TABLE 2 – Liste des catégories grammaticales

La deuxième information introduit des champs d’informations syntactico-sémantiques aux entrées du lexique. Quelques informations sont citées dans le tableau ci-dessus.

Trait syntactico-sémantiques	Code	Exemples
Abstrait	Abs	+◦ᄆᄆᄆ [tayri] “amour”
Animal	Anl	ᄆᄆᄆᄆ [uccn] “chacal”
Concrète	Conc	+◦ᄆᄆᄆᄆ [tadlist] “petit livre”
Alimentation	Alim	+◦ᄆᄆᄆᄆ [taxsayt] “courgette”
Date/ heure	Date	ᄆᄆᄆᄆ [yunyu] “Juin”
Organisation	Org	+◦ᄆᄆᄆᄆᄆ [tamssntit] “entreprise”

TABLE 3 – Liste des traits syntactico-sémantiques

Et finalement nous avons proposé un équivalent de traduction. En outre ces informations linguistiques, un ensemble de paradigmes flexionnels sont associées, à savoir: le genre, le nombre et l’état. Ces informations sont ajoutées après une formalisation du lexique.

3.2.2. Formalisation des entrées

Compte tenu de la perspective purement informatique de notre lexique, chaque entrée doit être rattachée à un ensemble d'informations morphologiques nécessaires pour que le système puisse générer et/ou reconnaître toutes les formes d'un mot.

Après une étude que nous avons fait en se basant sur nos entrées collectés et en nous inspirons des travaux (Boukhris, 2008) (Moujahid, 1981) (sadiqi, 1997) (Quitout, 1997) (Elabdellaoui et al., 2012) (Oulhaj, 2000) (Berkai, 2007), nous avons pu relever :

- (1) Pour la formalisation flexionnelle : 224 paradigmes pour les noms (une description détaillée est fournie dans les travaux (Nejme et al., 2013a) et (Nejme et al., 2016)), 484 paradigmes pour les verbes (une description détaillée est fournie dans le travail (Nejme et al., 2015)) et 8 paradigmes pour les pronoms (une description détaillée est fournie dans le travail (Nejme et al., 2013b)).
- (2) Pour la formalisation dérivationnelle : 125 paradigmes pour les noms (une description détaillée est fournie dans les travaux (Nejme et al., 2013a) et (Nejme et al., 2016)) et 50 paradigmes pour les verbes (une description détaillée est fournie dans le travail (Nejme et al., 2016)).

Suite à la phase de formalisation des paradigmes flexionnels et dérivationnels, nous avons procédé à l'application de ceux-ci au lexique des lemmes "El-AmaLex" afin de générer la liste de toutes formes fléchies potentielles. En effet, à l'issue de la phase de compilation de notre lexique, nous obtenons à partir de chaque entrée lexicale l'ensemble des formes qui s'y rattachent. Par exemple, lors de la phase de compilation, nous partons de l'entrée lexicale simple telle que :

$\mathbb{R}\mathbb{R}\mathbb{O}, V+\text{Simple}+\text{Bilitère}+\text{Reg}+\text{FLX}^2=\text{C}'\text{C}'\text{C_bbj}+\text{DRV}^3=\text{Forme2_}\mathbb{C}\mathbb{F}\text{-8:CC_vc+FR}=\text{enlever}+\text{ôter}+\text{supprimer}+\text{AR}=\text{نزع}$

nous obtenons, en moyenne, 197 entrées secondaires incluant :

- 98 formes verbales fléchies telle que :

$+\mathbb{R}\mathbb{R}\mathbb{O}\mathbb{A}, \mathbb{R}\mathbb{R}\mathbb{O}, V+\text{Simple}+\text{Bilitère}+\text{Reg}+\text{FLX}=\text{C}'\text{C}'\text{C_bbj}+\text{DRV}=\text{Forme2_}\mathbb{C}\mathbb{F}\text{-8:CC_vc+FR}=\text{enlever}+\text{ôter}+\text{supprimer}+\text{AR}=\text{نزع}+\text{Acc}+2+\text{f}+\text{s}$

$\mathbb{R}\mathbb{R}\mathbb{O}\mathbb{H}, \mathbb{R}\mathbb{R}\mathbb{O}, V+\text{Simple}+\text{Bilitère}+\text{Reg}+\text{FLX}=\text{C}'\text{C}'\text{C_bbj}+\text{DRV}=\text{Forme2_}\mathbb{C}\mathbb{F}\text{-8:CC_vc+FR}=\text{enlever}+\text{ôter}+\text{supprimer}+\text{AR}=\text{نزع}+\text{Acc}+1+\text{f}+\text{s}$

- 99 formes qui regroupent la forme dérivée avec ses flexions:

$\mathbb{I}\mathbb{C}\mathbb{F}\mathbb{R}\mathbb{R}\mathbb{O}\mathbb{O}, \mathbb{R}\mathbb{R}\mathbb{O}, V+\text{Dérivé}+\text{Réciproque}+\text{Préf } \mathbb{C}\mathbb{F}+\text{FLX}=\text{C}'\text{C}'\text{C_bbj} +\text{DRV}=\text{Forme2_}\mathbb{C}\mathbb{F}\text{-8:CC_vc}+\text{FR}=\text{enlever}+\text{ôter}+\text{supprimer}+\text{AR}=\text{نزع}+\text{Acc}+1+\text{f}+\text{p}$

Les entrées ainsi produites sont structurées de façon différente de celle du lexique de lemmes. En lisant de gauche à droite nous trouvons les éléments suivants :

- le mot dans sa forme fléchie ;
- un premier séparateur de champ, c'est-à-dire la virgule “,” ;
- le mot dans sa forme non fléchie, c'est-à-dire le lemme ;

² Le code “+FLX” (Paradigme flexionnel) permet d'indiquer le nom des règles de flexions du mot,

³ Le code “+DRV” (Paradigme dérivationnel) permet d'associer chaque mot avec les variantes dérivationnelles correspondantes.

- un second séparateur, c'est-à-dire la deuxième virgule “,” ;
- l'étiquette grammaticale (V) ;
- un troisième séparateur, c'est-à-dire le signe “+” ;
- l'ensemble de toutes les informations corrélées à la forme fléchie séparées par des signes “+” :
 - ✓ les informations syntactico-sémantiques as: +Simple,
 - ✓ la ou les informations flexionnelles correspondant à la forme as: +Acc+1+f+p.

Les entrées ainsi produites sont décrits dans le tableau suivant (cf. Table 4).

Catégories grammaticales	Nombre d'entrées	Nombres de formes générées
Noms	19661	141006
Verbes	3695	481472
Pronoms	235	741
Mots outils	443	443
Totale	24034	623662

TABLE 4 – Les entrées lexicales après compilation du lexique El-AmaLex

4 Conclusion

L'objectif principal visé par ce travail était d'élaborer notre lexique électronique pour la langue amazighe en utilisant la technologie à états finis incorporé dans la plateforme linguistique de développement NooJ. Pour ce faire, nous avons collecté notre lexique et ensuite construit un système qui regroupe un ensemble de grammaires flexionnelles et dérivationnelles. Ainsi, le lexique “El-AmaLex” construit associe chaque entrée lexicale, qu'elle soit nominale, verbale ou autre, à l'ensemble des informations correspondantes ainsi que le ou les paradigmes flexionnels et dérivationnels qui permettent la génération automatique de toutes les formes fléchies et dérivés. La liste des formes fléchies potentielles générée inclut des informations linguistiques utiles pour les tâches d'analyses ultérieures telles que : le lemme, la catégorie grammaticale, les informations flexionnelles (genre, nombre et l'état), etc. Ce lexique pourra être utilisé, par la suite, par différentes applications en traitement automatique des langues en particuliers l'analyse des textes.

Références

- AGNAOU F., BOUZANDAG A., EL BAGHDADI M., EL GHOLB H., KHALAFI A., OUQUA K., SGHIR M. (2011). « Lexique scolaire ». IRCAM, Rabat, Maroc.
- AMEUR M., BOUJAR A., BOUKHRIS F., BOUKOUSS A., BOUMALK A., ELMEDLAOUI M., IAZZI E., SOUFI H. (2004). « Initiation à la langue amazighe ». Rabat, Maroc: IRCAM.
- AMEUR M., BOUJAR A., BOUKHRIS F., ELMEDLAOUI M., IAZZI E. (2006). « Vocabulaire de la langue amazighe (Français-Amazighe) ». Série : Lexiques N°1, IRCAM, Rabat, Maroc.

- AMEUR M., BOUHJAR A., BOUMALK A., EL AZRAK N., LAABDELAOUI R. (2009a). « Vocabulaire des médias (Français-Amazighe-Anglais-Arabe) ». Série : Lexiques N°3, IRCAM, Rabat, Maroc.
- AMEUR M., BOUHJAR A., BOUMALK A., EL AZRAK N., LAABDELAOUI R. (2009b). « Vocabulaire grammatical ». Lexiques N°5, IRCAM, Rabat, Maroc.
- BERKAI A. (2007). « Lexique de la linguistique Français-Anglais-Berbère : précédé d'un essai de typologie des procédés néologiques ».
- 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.
- ES SAADY Y., AIT OUGUENGAY Y., RACHIDI A., EL YASSA M., MAMMASS D. (2009). « Adaptation d'un correcteur orthographique existant à la langue Amazighe : cas du correcteur Hunspell ». Actes du 1er symposium international sur le traitement automatique de la culture amazighe. pp. 149-158.
- FAKIR M., BOUIKHALENE B., MORO K. (2009). «Skeletonization methods evaluation for the recognition of printed tifinaghe characters». In Proceedings of the 1er Symposium International sur le Traitement Automatique de la Culture Amazighe, pp. 33—47. Agadir, Morocco.
- GREENBERG J. (1966). « The Languages of Africa ». The Hague.
- LAABDELAOUI R., BOUMALK A., IAZZI M., SOUIFI H. and ANSAR K. (2012). «Manuel de conjugaison de l'Amazighe ». IRCAM, Rabat, Morocco.
- MOUJAHID M. (1981). « La classe du Nom dans un parler de la langue tamazighete, le tachelhiyt d'Ighrem (Souss- Maroc) ». Thèse de 3ème cycle, Paris V, Université René Descartes.
- NEJME F., BOULAKNADEL S., ABOUTAJDINE D. (2013a). « Analyse Automatique de la Morphologie Nominale Amazighe ». Actes de la conférence du Traitement Automatique du Langage Naturel (TALN). Les Sables d'Olonne, France.
- NEJME F., BOULAKNADEL S., ABOUTAJDINE D. (2013b). « Finite State Morphology for Amazighe Language». In Proceeding of International Conference on Intelligent Text Processing and Computational linguistics (CICLing). Samos, Greece.
- NEJME F., BOULAKNADEL S., ABOUTAJDINE D. (2012a). « Toward an amazigh language processing ». Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP), COLING, December, pages 173 - 180. Mumbai.
- NEJME F., BOULAKNADEL S. (2012b). « Formalisation de l'Amazighe standard avec NooJ ». Actes de la conférence JEP-TALN-RECITAL. Grenoble, France.
- NEJME F., BOULAKNADEL S., ABOUTAJDINE D. (2012c). «Vers un dictionnaire électronique de l'Amazighe». Actes de la Conférence Internationale sur les Technologies d'Information et de Communication pour l'AMazighe (TICAM). Rabat, Maroc.
- NEJME F., BOULAKNADEL S., ABOUTAJDINE D. (2015). «Automatic Processing of Amazighe Verbal Morphology: a New Approach for Analysis and Formalization». In International Review on Computers and Software (IRECOS). doi.:<http://dx.doi.org/10.15866/irecos.v10i5.5755>.

- NEJME F., BOULAKNADEL S., ABOUTAJDINE D. (2016). « AmAMorph: Finite State Morphological Analyzer for Amazighe ». CIT. Journal of Computing and Information Technology, vol. 24, no. 1, pp. 91-110.
- OULHAJ L. (2000). GRAMMAIRE DU TAMAZIGHT. IMPRIMERIE NAJAH ELJADIDA
- OUTAHAJALA M., BENAJIBA Y., ROSSO P., ZENKOUAR L. (2015). « Using Confidence And Informativeness Criteria To Improve POS Tagging In Amazigh ». In Journal of Intelligence and Fuzzy Systems 28, pp. 1319—1330. doi: 10.3233/IFS-1 41417.
- OUAKRIM O. (1995). « Fonética y fonología del Bereber », *Survey at the University of Autònoma de Barcelona*.
- QUITOUT M. (1997). « Grammaire berbère (rifain, tamazight, chleuh, kabyle) ». Paris, L'Harmattan.
- SADIQI F. (1997). « Grammaire du Berbère ». Paris. L'Harmattan.
- SILBERZTEIN M. et TUTIN A. (2004). « NooJ : Un outil TAL de corpus pour l'enseignement des langues et de la linguistique ». *Journée ATALA TAL et apprentissage des langues*.
- SILBERZTEIN M. (2007). « An Alternative Approach to Tagging ». NLDB 2007: 1-1.
- TAIFI M. (1988). « Le lexique berbère (parlers du Maroc central) ».
- TALHA M., BOULAKNADEL S., ABOUTAJDINE D. (2015). « L'apport d'une approche symbolique pour le repérage des entités nommées en langue amazighe ». EGC 2015: 29-34.