

Analyse Automatique de la Morphologie Nominale Amazighe

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

RÉSUMÉ

Dans le but de préserver le patrimoine amazighe et éviter qu'il soit menacé de disparition, il semble opportun de doter cette langue de moyens nécessaires pour faire face aux enjeux de l'accès au domaine de l'Information et de la Communication (TIC). Dans ce contexte, et dans la perspective de construire des outils et des ressources linguistiques pour le traitement automatique de cette langue, nous avons entrepris de construire un système d'analyse morphologique pour l'amazighe standard du Maroc. Ce système profite des apports des modèles à états finis au sein de l'environnement linguistique de développement NooJ en faisant appel à des règles grammaticales à large couverture.

ABSTRACT

Morphological analysis of the standard Amazigh language using NooJ platform

In the aim of safeguarding the Amazigh heritage from being threatned of disappearance, it seems opportune to equip this language of necessary means to confront the stakes of access to the domain of New Information and Communication Technologies (ICT). In this context, and in the perspective to build tools and linguistic resources for the automatic processing of Amazigh language, we have undertaken to develop a system of a morphological description for standard Amazigh of Morocco. This system uses finite state technology, within the linguistic developmental environment NooJ by using a large-coverage of morphological grammars covering all grammatical rules.

MOTS-CLÉS : La langue amazighe, TALN, NooJ, analyse morphologique, morphologie flexionnelle, morphologie dérivationnelle.

KEYWORDS : Amazigh language, NLP, NooJ, morphological analysis, inflectional morphology, derivational morphology.

1 Introduction

La langue amazighe du Maroc est considérée 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 malgré son usage important (environ 50% de la population). Toutefois, au cours des dernières années, la société marocaine a connu beaucoup de débat sur la langue et la culture amazighe. Ainsi, la création d'une nouvelle institution gouvernementale, à savoir l'Institut Royal de la Culture Amazighe (IRCAM), a permis à cette langue ainsi qu'à sa culture de retrouver leur place légitime dans de nombreux domaines. Par conséquent, cette langue a pu être aménagée et son introduction assurée dans le domaine public notamment dans l'enseignement, l'administration et les

médias. Cette création lui a permis d'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. La démarche d'élaboration a 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).

Toutefois, en traitement automatique du langage naturel (NLP), l'amazighe, comme la plupart des langues non européennes¹, souffre encore de la rareté des outils de traitement automatique du langage, ce qu'elle ne permet pas à cette langue de rejoindre ses consœurs dans le domaine des nouvelles technologies de l'information et de la communication (NTIC). En ce sens, étant donné que toute analyse linguistique doit passer par une première étape d'analyse morpho-lexicale, qui consiste à tester l'appartenance de chaque mot du texte au lexique de la langue, nous avons entrepris de construire un système d'analyse morphologique pour l'amazighe standard du Maroc. Ce système profite des apports des modèles à états finis au sein de l'environnement linguistique de développement NooJ en faisant appel à des règles grammaticales à large couverture.

Le présent article se structure autour de trois volets: le premier présente un descriptif de la langue amazighe, le deuxième expose le module d'analyse de la langue amazighe standard en utilisant la plateforme linguistique NooJ, le troisième expose l'expérimentation et l'évaluation de la réalisation alors que le dernier volet est consacré à la conclusion et aux perspectives.

2 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, l'amazighe se répartit selon trois grandes zones régionales : le Tarifit au Nord, le Tamazight au Maroc central et au Sud-Est et le Tachelhit au Sud-Ouest et dans le Haut-Atlas. Chacun de ces dialectes comprend des sous-dialectes ou dialectes locaux constituant le deuxième type.

En adoptant l'amazighe comme langue officielle du Maroc, l'IRCAM s'est engagée à réaliser un processus de standardisation² de la langue amazighe (Ameur et al., 2004a), qui a pour vocation d'uniformiser les structures et d'atténuer les divergences, en éliminant les occurrences non distinctives qui entraînent souvent des problèmes d'intercompréhension. Ce processus de standardisation consiste en plusieurs étapes à savoir : adapter une graphie

¹ Langues peu dotées informatiquement (les langues-π (Berment, 2004)).

² 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).

standard normalisée sur une base phonologique; adapter un lexique de base commun; appliquer les mêmes règles orthographiques, les mêmes consignes pédagogiques, et les mêmes formes néologiques; et enfin exploiter la variation dialectale afin de sauvegarder la richesse de la langue. La suite de cet article est focalisée sur l'amazighe standard du Maroc.

2.2 Alphabet amazighe

En se basant sur le système original, l'IRCAM a développée un système d'alphabet sous le nom de Tifinaghe-IRCAM. 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.

2.3 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.

3 Morphologie Nominale de l'amazighe standard du Maroc

La langue amazighe présente une morphologie riche et complexe. Les mots peuvent être classés en trois catégories morphosyntaxiques: Nom, Verbe et Particules (Boukhris et al., 2008). Dans cet article, nous nous intéressons à la catégorie nom.

1. Nom

En amazighe, le nom est une unité lexicale formée d'une base et d'un ou plusieurs affixes. Cette base résulte de la combinaison d'une racine et d'un schème (Boukhris et al., 2008). Le nom possède deux caractéristiques. La première est qu'il peut prendre différentes formes à savoir: une forme simple (ⵓⵔⵗⵓⵣⵓ [argaz] "homme"), une forme composée (ⵉⵎⵓⵔⵉⵙⵏⵉⵢⵓⵏⵉ [buhyyuf] "la famine") ou bien une forme dérivée (ⵓⵎⵓⵙⵏⵉⵢⵓⵏⵉ [amsawaɗ] "la communication"). Dans cet article, nous nous intéressons aux noms simples et aux noms dérivés.

- Les noms simples

L'amazighe distingue deux grandes sous-classes de noms simples : les noms propres et les noms communs.

• Les noms propres

Les noms propres désignent soit des personnes (ⵏⵏⵓ [hnu]) ou des lieux (noms de villes ou villages), aussi dits « toponymes », comme ⵓⵣⵓⵣⵓ [azru]. Dans ce travail, nous nous intéresserons à la formalisation des noms de personnes. Ce type de nom n'est pas sujet à la flexion.

- Les noms communs

Les noms communs peuvent être soit abstraits soit concrets. Ces derniers peuvent, à leur tour, être soit animés (+∘⊂+⊗E+ [tamtudt] “femme”) ou inanimés (+⊗⊂⊂⊂⊂ [tigmmi] “maison”).

- Les noms dérivés

A partir d’une racine verbale, les noms dérivés sont formés par une préfixation ou suffixation d’un morphème de dérivation plus des variations intra-radicales. Le nombre et la nature de ces formes varient selon le statut du verbe auquel ils se rattachent. Ainsi, le nom d’action, le nom d’agent, le nom d’instrument et le nom de qualité sont constitués.

- Le nom d’action

Le nom d’action se forme avec des préfixes associés à des modifications intra-radicales. Les principaux procédés de dérivation sont : préfixation de ∘ [a], préfixation de ∘ [u], préfixation de ⊗ [i], préfixation et affixation du morphème du féminin +...+ [t...t], préfixation de ¶ aux noms empruntés et intégrés : ⊂⊂⊂ [ḥmu] “ê. chaud”-> ¶⊂⊂ [ḷḥmu] “chaleur”.

- Le nom d’agent

Le nom d’agent dérive d’un verbe d’action par la préfixation de l’un des éléments suivants : ∘ [a], ∘⊂ [am]/∠ [an], ⊗⊂ [im] et ⊗ [i] (∘⊂⊂ [akr] “diffamer”-> ∘⊂∘⊂⊂ [amakr] “diffamateur”).

- Le nom d’instrument

Ce type de nom est formé sur la base des schèmes ∘ [a]/∠ [as] associé à des modifications vocaliques ou consonantiques : ∠⊂¶ [rgl] “fermer”-> ∠∠∠⊂¶ [asrgl] “couvercle”.

- Le nom de qualité

Le nom de qualité est généralement dérivé des verbes dits de qualité ou d’état. Les procédés de formation de ce type de nom sont: (1) préfixation de ∘ [a] et alternance vocalique, (2) préfixation de ∘⊂ [am]/∠ [an] suivie parfois d’une variation intra ou post-radical, (3) préfixation de ⊗ [i] et variation intra-radical et (4) préfixation de ∘ [u] accompagné parfois de l’ infixation de ⊗ [i] : ⊂⊂∠ [qmr] “ê. étroit”-> ∘⊂⊂⊂∠ [uqmir] “étroit”.

La deuxième caractéristique d’un nom amazighe correspond à la flexion : il varie en genre (féminin, masculin), en nombre (singulier, pluriel) et en état (libre, annexion).

- Le genre

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

- ✓ Le nom masculin: il commence généralement par une des voyelles initiales: ∘ [a], ⊗ [i] ou bien ∘ [u]: ∘⊂⊂ [udm] “visage”. Cependant, il existe certains noms qui font exception : ⊗⊂⊂∠ [imma] “(ma) mère”.
- ✓ Le nom féminin: celui-ci est généralement de la forme +...+ [t...t], à l’exception de certains noms qui ne portent que le + [t] initial ou le + [t] final du morphème du féminin : +∠⊂∠ [tadla] “gerbe”.

Dans le cas général, le féminin est formé à partir du radical d’un nom masculin par

l'ajout du morphème discontinu +...+ [t...t]: $\xi\Theta\xi$ [isli] "marié" -> $+\xi\Theta\eta\xi+$ [tislit] "mariée".

- Le nombre

Le nom amazighe, qu'il soit masculin ou féminin, possède un singulier et un pluriel. Ce dernier est obtenu selon quatre types: le pluriel externe, interne, mixte et le pluriel en $\xi\Lambda$ [id].

- ✓ Le pluriel externe est obtenu par une alternance vocalique accompagnée par une suffixation de l [n] ou l'une de ses variantes (ξ [in], \circ [an], $\circ\gamma$ [ayn], Ll [wn], $\circ\text{Ll}$ [awn], Ll [wan], $\text{L}\xi$ [win], H [tn], $\gamma\xi$ [yin]). $\circ\text{X}\text{X}\circ\text{L}$ [axxam] -> $\xi\text{X}\text{X}\circ\text{L}$ [ixxamn] "maisons", $+\circ\Theta\circ+$ [tarbat] "fille" -> $+\xi\Theta\Theta\circ+\xi$ [tirbatin] "filles".
- ✓ Le pluriel interne (ou brisé) est obtenu par une alternance vocalique plus un changement de voyelles internes. $\circ\Lambda\circ\circ$ [adrar] -> $\xi\Lambda\circ\circ\circ$ [idurar] "montagnes".
- ✓ Le pluriel mixte est formé par une alternance d'une voyelle interne et/ou d'une consonne plus une suffixation par l [n]. $\xi\eta\xi$ [ili] "part" -> $\xi\eta\circ$ [ilan] "parts", ou bien par une alternance vocalique initiale accompagnée d'un changement vocalique final \circ [a] plus une alternance interne $\circ\text{L}\text{X}\text{X}\circ\circ$ [amggaru] "dernier" -> $\xi\text{L}\text{X}\text{X}\circ\circ$ [imggura] "derniers".
- ✓ Le pluriel en $\xi\Lambda$ [id]: ce type de pluriel est obtenu par une préfixation $\xi\Lambda$ [id] du nom au singulier. Il est appliqué à un ensemble de cas de noms à savoir les noms à initiale consonantique, des noms propres, des noms de parenté, des noms composés, des numéraux, ainsi que pour les noms empruntés et intégrés $\text{X}\circ\eta\xi$ [xali] " (mon) oncle" -> $\xi\Lambda \text{X}\circ\eta\xi$ [id xali].

- L'état

En amazighe, le nom est concerné par la variation d'état. Ainsi, nous distinguons deux: l'état libre et l'état d'annexion.

- ✓ L'état libre: la voyelle initiale du nom ne subit aucune modification. 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 [d] "c'est".
- ✓ L'état d'annexion est fondé sur une variation formelle qui affecte la première syllabe des noms en cause dans des contextes syntaxiques déterminés. Il prend l'une des formes suivantes: alternance vocalique \circ [a]/ \circ [u] ou bien maintien de la voyelle initiale et ajout d'un L [w] au cas des noms masculins à initiale \circ [a] ($\circ\text{O}\text{X}\circ\text{X}$ [argaz] "homme" -> $\circ\text{O}\text{X}\circ\text{X}$ [urgaz]), addition d'un L [w] pour ceux à initial \circ [u] et d'un γ [y] aux noms à voyelle ξ [i] ($\xi\eta\Theta$ [ils] "langue" -> $\gamma\xi\eta\Theta$ [yils]). Pour les noms féminin, cet état est défini soit par la chute ou le maintien de la voyelle initiale ($+\circ\text{L}\text{Y}\circ\text{O}+$ [tamYart] "femme" -> $+\text{L}\text{Y}\circ\text{O}+$ [tmYart]).

4 Analyse automatique du nom amazighe

Durant ces dernières années, le traitement du langage naturel a montré plus d'intérêt pour la langue amazighe. La construction des ressources appropriées pour cette langue devient une nécessité vitale pour effectuer des analyses efficaces. Dans cette contribution, nous nous intéresserons à la formalisation de la catégorie nom en utilisant la plateforme linguistique NooJ.

4.1 La plateforme NooJ

NooJ, publié en 2002 par Max Silberztein (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 (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, permet d'effectuer des recherches et des traitements dans les textes à partir d'expressions régulières et associent leurs reconnaissances à la liste d'annotations correspondante (étiquette grammaticale, des informations sémantiques, la traduction, etc.). Pour ce faire, il se base sur l'utilisation d'un ensemble d'opérateurs de transformations (eg. <L> : déplacement vers la gauche, <RW> : aller à la fin du mot, etc), prédéfinis dans NooJ, à l'intérieur des graphes décrivant les 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.

4.2 Formalisation des noms communs et dérivés

Toute analyse linguistique doit passer par une première étape d'analyse lexicale, qui consiste à tester l'appartenance de chaque mot du texte au lexique de la langue. Pour mener à bien cette étude, nous avons commencé notre travail par l'étape de formalisation du vocabulaire de la langue amazighe et nous nous intéresserons, dans cet article, à la formalisation (flexionnelle et dérivationnelle) de la catégorie nom.

Cette formalisation est basée sur l'utilisation de certaines commandes génériques prédéfinies telles que:

- <L> : déplacement vers la gauche,
- <R> : déplacement vers la droite,
- : suppression du dernier caractère,

- <S> : suppression de caractère courant,
- <RW> : aller à la fin du mot.
- Etc.

4.2.1 Morphologie flexionnelle

En amazighe, le nom peut présenter trois flexions différentes, en fonction de la variation du genre (masculin, féminin), du nombre (singulier, pluriel) et d'état (état libre, état d'annexion). Afin de formaliser ces flexions, nous avons dû créer par le biais de graphes incorporé au logiciel NooJ, un ensemble de graphes décrivant les modèles de flexions en amazighe standard du Maroc. Cette étude présente l'implémentation des règles de flexion associée à chaque entrée lexicale nom et permettant de générer toutes les formes fléchies : genre, nombre et état.

1. Le genre

Étant donné que le nom féminin est généré à partir du nom masculin en appliquant une préfixation et une suffixation par le morphème discontinu + [t], nous avons construit la règle suivante :

La règle dans NooJ	Explication	Exemple
<LW>+<RW>+	La règle ajoute un + [t] au début et à la fin du nom	ⵉⵎⵓⵔⵉ [isli] "marié" -> +ⵉⵎⵓⵔⵉⵓ+ [tislit] "mariée".

TABLE 1 – Règle permettant de passer d'un nom masculin à un nom féminin

2. Le nombre

En amazighe, quatre type de pluriel sont distingués. Le processus de construction et de reconnaissance de ces pluriels ne sont pas homogènes et n'obéissent pas à des règles précises (i.e., ajout d'une ou plusieurs lettres à la forme de singulier, suppression d'une ou plusieurs lettres) et ils sont généralement imprévisibles. A cet effet, nous avons fixé comme premier objectif, le recensement du maximum des règles afin de reconnaître et gérer toutes les formes fléchies possibles de chaque pluriel. Ainsi, pour chaque type, nous avons créé, en se basant sur la nouvelle grammaire de l'amazighe (Boukhris et al., 2008) et sur le dictionnaire de Taifi (Taifi, 1988), un ensemble de règles. Ces règles sont basées sur les schèmes, par exemple pour le pluriel externe on a défini les règles suivant le début et la fin du nom: si le nom féminin est de la forme '+ⵉ...V+' [ta...Vt] la règle de construction du pluriel sera une suffixation de ⵉ [in] par contre si le nom féminin est de la forme '+...C+' [ta...Ct] la règle sera une alternance vocalique de la première voyelle accompagnée par une suffixation de ⵉ [in] et suppression du dernier + [t]. Afin d'illustrer l'implémentation de ces règles, nous allons montrer par la suite un exemple pour chaque type de pluriel.

- Le pluriel externe

Pour le pluriel externe, nous avons créé, en se basant sur les schèmes, 26 règles (14 règles pour les noms masculins et 12 pour les noms féminins).

La règle dans NooJ	Explication	Exemple
<LW>ξ<S><RW>+l	La règle concerne les noms de la forme ‘o...o’ [a...a] et consiste à une alternance vocalique initiale accompagnée par une suffixation de +l [tn].	oɔɪɥξ [amɲɥi] “dispute” -> ξɔɪɥξ+l [imɲɥitn] “disputes”.
<LW><R>ξ<S><RW>ξl	Cette règle concerne les noms de la forme ‘+o...Ct’ [ta...Ct] et consiste à une alternance vocalique de la première voyelle accompagnée par une suffixation de ξl [in] et suppression du dernier + [t].	+oʌʒʌξO+ [taħudirt] “courbette”-> +ξʌʒʌξOξl [tiħudirin] “courbettes”.

TABLE 2 – Exemple de règles pour le pluriel externe

- Le pluriel interne

Pour ce type de pluriel, nous avons pu recenser, jusqu'à présent, 29 règles (10 pour les noms masculins et 19 pour les noms féminins).

La règle dans NooJ	Explication	Exemple
<LW><R>ξ<S><RW><B2>o	Cette règle concerne les noms de la forme ‘TVcNvT’ et consiste à une alternance vocalique première voyelle accompagnée par un changement de la dernière voyelle plus suppression du dernier t.	+oʒξʒɪɪξ+ [tazizwit] “abeille” -> +ξʒξʒɪɪo [tizizwa] “abeilles”
<LW>ξ<S><RW><L3>ξ <R2>o	Cette règle concerne les noms de la forme ‘VCnVCVC’ et consiste à une alternance vocalique initiale accompagnée par un changement des	oθoɥ%θ [abaɥus] “singe” -> ξθ%ɥoθ [ibuɥas] “singes”

	voyelles internes.	
--	--------------------	--

TABLE 3 – Exemple de règles pour le pluriel interne

- Le pluriel mixte

Le pluriel mixte regroupe une suffixation et une alternance interne. Pour cela, nous avons créé 39 règles (30 pour les masculins et 9 pour les féminins).

La règle dans NooJ	Explication	Exemple
<LW>ξ<S><R4>∅<RW>∅	La règle concerne les noms de la forme 'aCCCaC _(n) u' et consiste à une alternance vocalique initiale accompagnée par un changement vocalique interne et finale.	∅CXX∅∅ [amggaru] "dernier"-> ξCXX∅∅ [imggura] "derniers"
<RW>ξLξl	Cette règle concerne les noms de la forme 't∅...∅' [ta...a] et consiste à un changement de voyelle finale accompagné par une suffixation par Lξl [win].	+∅LI∅ [tawja] "famille"-> +∅LIξLξl [tawjiwin] "familles"

TABLE 4 – Exemple de règles pour le pluriel mixte

- Le pluriel en ξΛ [id]

Etant donné que ce type de pluriel est obtenu par une préfixation ξΛ [id] du nom au singulier, nous avons construit une règle qui permet d'ajouter le ξΛ au début de chaque nom à initiale consonantique, noms propres, noms de parenté, noms composés, les numéraux, ainsi que pour les noms empruntés et intégrés.

La règle dans NooJ	Explication	Exemple
<LW>ξΛ "	La règle concerne les noms à initiale consonantique et consiste à une préfixation de ξΛ [id].	X∅Mξ [xali] "(mon) oncle" -> ξΛ X∅Mξ [id xali]

TABLE 5 – Exemple de règles pour le pluriel en ξΛ [id]

3. L'état

En amazighe, nous distinguons deux état (cf. 3-1): état libre et d'annexion. Pour ce dernier, nous avons construit 9 règles (4 pour les masculins et 5 pour les féminins).

La règle dans NooJ	Explication	Exemple
<LW><S>§	La règle concerne les noms masculins à initiale ◦ [a] non constant et consiste à une alternance vocalique ◦ [a]/§[u].	◦OX◦ⵎ [argaz] "homme" -> §OX◦ⵎ [urgaz]
<LW><R2>	La règle concerne les noms féminins formés à partir des noms masculins à voyelle non constant et consiste à la chute de la voyelle initiale du nom féminin.	+◦ⵏⵓⵔ◦+ [tamɣart] "femme" -> +ⵏⵓⵔ◦+ [tmɣart].

TABLE 6 – Exemple de règles pour l'état d'annexion

4.2.2 Morphologie dérivationnelle

Outre la morphologie flexionnelle, Il existe des procédés morphologiques et lexicaux par lesquels les noms prennent un ensemble de formes autre que simple. L'une de ces principales formes est la forme dérivée.

La morphologie dérivationnelle s'occupe de la création des mots appartenants à des catégories souvent différentes de celle de base. En amazighe, nous pouvons distinguer deux types de dérivation: la dérivation grammaticale et la dérivation expressive, auxquelles s'ajoutent la dérivation affixale moderne, flexionnelle et la dérivation par analogie (Berkai, 2007). Dans notre cas, nous nous intéresserons au premier type qui est la dérivation grammaticale et plus précisément à la dérivation nominale sur une base verbale. Dans ce cas, le processus de dérivation est basé sur des procédés de préfixation ou de suffixation d'un morphème rattaché à la base lexicale en maintenant le noyau sémantique de l'action ou de l'état exprimé par le verbe d'origine. Ainsi ils sont constitués le nom d'action, le nom d'agent, le nom d'instrument et le nom de qualité.

Afin de formaliser ces dérivations, nous avons dû créer 63 règles : 28 règles pour les noms d'action, 8 pour les noms d'agents, 8 pour les noms d'instruments et 19 règles pour les noms de qualité. Ces règles décrivent un ensemble de modèles de dérivations en amazighe standard du Maroc. Cependant, elles ne sont pas exhaustives et ne couvrent pas l'ensemble des phénomènes dérivationnelles de l'amazighe.

La règle dans NooJ	Explication	Exemple
$\langle LW \rangle \circ \langle RW \rangle \langle L \rangle \varepsilon$	La règle concerne la dérivation d'un nom d'action à partir d'une entrée verbale et qui consiste à une préfixation par \circ [a] accompagnée par une insertion pré-finale de la voyelle ε [u].	$\ast\ast\ast\ast$ [zznz] "vendre" -> $\circ\ast\ast\ast\ast$ [azznuz] "vente" (nom d'action).
$\langle LW \rangle \circ \Theta \langle RW \rangle \langle B \rangle \varepsilon$	La règle concerne la dérivation d'un nom d'instrument à partir d'une entrée verbale et consiste à une préfixation par $\circ\Theta$ [as] accompagnée par un changement de la voyelle finale ε [u]/ ε [i].	$\ast\ast$ [gnu] "coudre" -> $\circ\Theta\ast\ast$ [asgni] "grosse aiguille" (nom d'instrument)

TABLE 7 – Exemple de règles de dérivations

4.3 Formalisation des noms propres

Afin de formaliser la classe de noms propres nous avons commencé par l'élaboration d'un dictionnaire électronique « EDicAMPN » (Electronic Dictionary for Amazigh Proper Noun) de prénoms, reconnaissable par le biais de l'étiquette $\langle N + \text{Prénom} \rangle$, qui contient, à l'heure actuelle, 424 entrées de prénoms simples et proprement amazighes ($\Theta\ast\ast$ [hnnu]) et environ 500 entrées de prénoms étrangers, sous format simple et composé, transcrits ($\circ\ast\ast\circ\ast$ [ahmad]). Cependant, ce dictionnaire ne contient ni les compositions de prénoms telles que $\ast\ast \circ \varepsilon \ast \Lambda \circ$ [Ali Riða] ou bien $\ast \circ \varepsilon \ast \Lambda \circ \ast \ast \circ$ [Marie-Laure], ni les prénoms islamiques tels que $\varepsilon \Theta \ast$ $\ast \ast \ast \circ \ast \ast$ [Ibn Tofayl], etc. Afin de reconnaître ce genre de prénoms, nous avons construit une grammaire locale présentée ci-dessous :

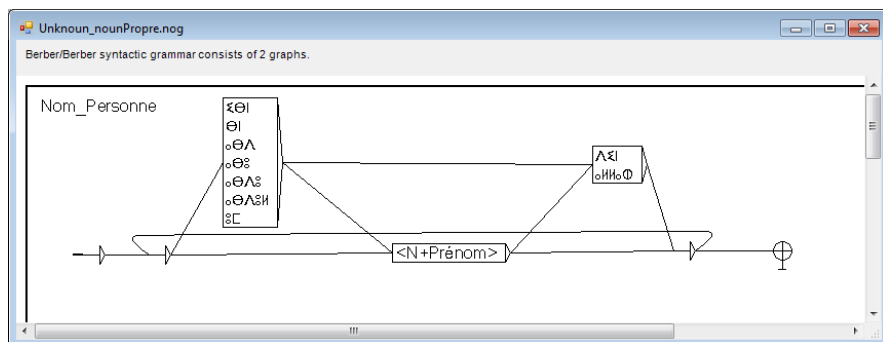


FIGURE 1 – Grammaire morphologique pour la reconnaissance des prénoms composés.

A l'aide de cette grammaire, le système peut reconnaître les prénoms composés à partir de ceux appartenant à notre dictionnaire ainsi que les noms islamiques composés des éléments lexicaux tel que ⵎⵓ [um] "la mère de" etc.

5 Test et évaluation

Afin d'évaluer nos ressources basées sur les grammaires flexionnelles ainsi que dérivationnelles, nous avons construit un corpus contenant :

- Une liste, basée sur le dictionnaire de Taifi (Taifi, 1988) et sur le vocabulaire de la langue amazighe (Ameur et al., 2009), contenant 4524 entrées de noms communs (sous formes simple et fléchies).
- Une liste, basée sur la nouvelle grammaire de la langue amazighe (Boukhris et al., 2008), contenant à l'heure actuelle, 113 entrées de noms dérivés.
- Une liste de 200 entrées de prénoms étrangers, sous format simple et composé, transcrits.

		Nombre de formes reconnues		Nombre de formes non reconnues	
		Nombre	%	Nombre	%
Résultats	Noms communs	4205	93%	319	7,05%
	Noms propres	594	95,19%	28	4,5%
	Noms dérivés	103	91,15%	10	8,84%

TABLE 8 – Evaluation de nos ressources linguistiques

D'après les résultats obtenus, l'analyse lexicale de ce corpus, montre une bonne couverture à l'aide de nos ressources morphologiques.

Les formes non reconnues sont classées en 3 sous-ensembles:

- Un total de 319 occurrences de noms communs non reconnues dues principalement aux problèmes au niveau de flexion. Par exemple pour le nom ⵎⵓⵏⵓⵢⵓ [aḍḍra], il est associé à la règle flexionnelle qui concerne les noms sous formes '...o' [a...a] et qui consiste généralement à une alternance vocalique initiale accompagnée par une suffixation de ⵏ [tn]. Cependant, sa forme fléchie de base est ⵎⵓⵏⵓⵢⵓⵏⵓ [aḍḍratn]. Ce qui présente des cas particuliers pour la langue amazighe.
- Un total de 28 occurrences de prénoms. Ceci est dû aux problèmes de transcriptions. Chaque nom peut être transcrit de différentes manières par exemple pour le nom عبد الله, il peut être transcrit ⵎⵓⵏⵓⵢⵓⵏⵓ, ⵎⵓⵏⵓⵢⵓⵏⵓⵏⵓ ou bien ⵎⵓⵏⵓⵢⵓⵏⵓⵏⵓ. Pour notre grammaire nous n'avons défini que la forme la plus utilisée.
- Un total de 10 occurrences de formes dérivées générées à partir des verbes et qui

n'existent pas.

Conclusion et perspectives

L'objectif principal visé par ce travail était de formaliser la morphologie flexionnelle ainsi que dérivationnelle pour la catégorie morpho-syntaxique nom (nom simple et dérivé). Pour ce faire, nous avons construit un système qui regroupe un ensemble de grammaire flexionnelle et dérivationnelle ainsi qu'une grammaire pour la reconnaissance des prénom. La construction de ce système morphologique, en utilisant la technologie à états finis incorporé dans la plateforme de développement linguistique NooJ, nous a permis d'annoter les noms présentés soit sous une forme fléchie ou bien dérivée ainsi que les prénom.

D'après les résultats obtenus, le système affiche un taux de reconnaissance très encourageant. Par ailleurs, nous envisageons de :

- Ajouter d'autre règles de flexion afin d'inclure tous les cas exceptionnels pour la langue amazighe.
- Formaliser les autres catégories morphosyntaxiques : verbes et particules.
- Développer un corpus de textes amazighs pour l'évaluation.

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. (2006a). « Graphie et orthographe de l'amazighe ». Rabat, Maroc : IRCAM.
- AMEUR M., BOUHJAR A., BOUKHRIS F., ELMEDLAOUI M., IAZZI E. (2006b). « 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. (2009). « Vocabulaire des médias (Français-Amazighe-Anglais-Arabe) ». Série : Lexiques N°3, IRCAM, Rabat, Maroc.
- BERKAI A. (2007). « Lexique de la linguistique Français-Anglais-Berbère : précédé d'un essai de typologie des procédés néologiques ».
- 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 ». The Hague.

KAMEL S. (2006). « Lexique Amazighe de géologie ». Rabat, Maroc: IRCAM.

OUAKRIM O. (1995). « Fonética y fonología del Bereber », *Survey at the University of Autònoma de Barcelona*.

REVUZ.D. (1991). « Dictionnaires et lexiques : méthodes et algorithmes ». *Thèse de doctorat Institut Blaise Pascal*, Paris, France.

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) ».