

*République Algérienne Démocratique et Populaire*

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

Université Mouloud Mammeri de Tizi Ouzou

*Faculté de Génie Electrique et d'Informatique*

*Département d'Informatique*



## **Mémoire de fin d'études**

En vue de l'obtention du diplôme de

Master en Informatique

Spécialité : Ingénierie des systèmes d'information

**Thème :**

**Méthode hybride pour la classification de mots dans  
un dictionnaire de langue amazighe**

Réalisé par :

**BENGHOUBA Youcef**

**Devant le jury composé de :**

<b>Président</b>	M <sup>me</sup> BOURKACHE Ghenima	U.M.M. Tizi-Ouzou.
<b>Examinatrice</b>	M <sup>elle</sup> YESLI Yasmine	U.M.M. Tizi-Ouzou.
<b>Promoteur</b>	M <sup>f</sup> BIBI Mohand Ouamer	U.A/Mira Béjaia.
<b>Co-Promotrice</b>	M <sup>me</sup> AOUGHLIS Farida	U.M.M. Tizi-Ouzou.

---

*Tizi Ouzou 2019/ 2020*

---



## Dédicaces

*Je dédie ce modeste travail à mes chers parents  
qui n'ont jamais cessé de m'encourager,  
qui ont toujours cru en moi.*

*À mes frères et ma sœur pour leurs soutiens, à toute ma  
famille et à mes chers amis qui sont restés auprès  
de moi tout au long de mon travail et qui m'ont donné  
le courage pour le réussir.*

# Remerciements

*Je tiens à exprimer toute ma reconnaissance à mon directeur de mémoire, Monsieur BIBI Mohand Ouamer, et aussi Madame Aoughlis Farida, Je les remercie de m'avoir encadré, orienté, aidé et conseillé.*

*Mes remerciements s'adressent également aux membres de jury qui ont accepté d'évaluer mon travail.*

*Je tiens à exprimer mes sincères remerciements à tout le corps professoral et administratif de l'université Mouloud Mammeri de Tizi-Ouzou, et ceux de L'université Abderrahmane Mira de Bejaia. Enfin, je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.*

## Table Des Matières :

<b>INTRODUCTION GENERALE</b> .....	8
<b>Chapitre 1 : La recherche d'information et le traitement automatique de langues</b>	
Introduction.....	10
1. La recherche d'information (RI) .....	10
1.1 Définitions .....	10
1.2 Processus de la recherche d'information .....	11
1.2.1 L'indexation .....	12
1.2.1.1 Les types d'indexation .....	12
Indexation manuelle.....	12
Indexation automatique .....	12
Indexation supervisé .....	13
1.2.1.2 Le processus d'indexation .....	13
1.2.2 L'appariement requête-document ou l'interrogation .....	13
1.2.3 La reformulation de la requête .....	14
2.Le Traitement Automatique des Langues (TAL) .....	14
2.1 Définition .....	14
2.2 Les niveaux d'analyse en TAL .....	14
2.2.a L'analyse morphologique .....	15
2.2.b L'analyse syntaxique .....	15
2.2.c L'analyse sémantique .....	15
2.2.d L'analyse pragmatique .....	15
3.Les applications du TAL .....	15
4. Les problèmes majeurs du TAL .....	16
Conclusion .....	16

## **CHAPITRE 2 : La langue amazighe**

1. La langue amazighe .....	17
2. Le lexique de la langue amazighe .....	18
3. La morphologie amazighe .....	20
3.1 Morphologie flexionnelle.....	20
4. La formation de mots amazighs .....	22
4.1 Formation de mots par dérivation .....	23
4.2 Formation de mots par composition .....	24
5. La classification des mots amazighs .....	25
Avantage et inconvénients .....	26
Conclusion .....	27

## **CHAPITRE 3 : Construction et formalisation d'un dictionnaire**

Introduction.....	28
1. Racinisation (stemming) .....	28
2. Représentation formelle du dictionnaire .....	29
3. Le principe de l'algorithme .....	30
* Algorithme de racinisation .....	31
* Racinisation des mots amazighs .....	32
4. Le corpus .....	33
5.L'application .....	34
6.Le processus de fonctionnement de l'application .....	35
7. LA classification en utilisant la méthode hybride .....	37
Conclusion .....	38
<b>CONCLUSION GENERALE .....</b>	<b>39</b>
Perspectives .....	39
Bibliographie .....	40

## Liste des figures et tableaux

Figure 1.1: Processus de recherche d'information.....	11
Figure 1.2 : Processus d'indexation .....	12
Figure 1.3: Les différents niveaux d'analyse d'un texte.....	14
Figure 2.1 : famille chamito-sémitique.....	17
Figure 2.2 : Les groupes amazighophones .....	18
Figure 2.3 : Les classes grammaticales en amazigh. ....	19
Figure 2.4: Formation du féminin à partir d'un nom masculin .....	21
Figure 2.5: Formation de mots par racine .....	24
Figure 3.1 : Processus de racinisation .....	28
Figure 3.2 : Arbre d'exécution de l'algorithme .....	32
Figure 3.3 : La répartition des mots dans notre corpus .....	33
Figure 3.4: La fenêtre principale .....	34
Figure 3.5 : Créer une table dans la base de données « mot.db » .....	35
Figure 3.6 : Ajouter des données dans la table. ....	35
Figure 3.7 : Supprimer des données dans la table. ....	35
Figure 3.8 : Afficher les données (racine et mot). ....	36
Figure 3.9 : Extrait de l'algorithme de racinisation .....	36
Figure 3.10 : Résultat d'exécution de l'algorithme. ....	37
Figure 3.11 : Recherche d'un mot dans le dictionnaire.....	37
Figure 3.12 : La répartition des mots en utilisant la méthode hybride.....	38
<hr/>	
Tableau 2.1 : Avantages et Inconvénients de la classification par ordre alphabétique .....	25
Tableau 2.2 : Avantages et Inconvénients de la classification par racine.....	26
Tableau 3.1 : Les classes grammaticales .....	29

---

# *INTRODUCTION GENERALE*

---

Le travail réalisé dans ce mémoire s'inscrit dans le cadre général du Traitement Automatique des Langues Naturelles (TALN), qui est une discipline qui se situe à la limite de la linguistique, l'informatique et de l'Intelligence Artificielle (IA); elle vise le développement d'applications informatiques capables de traiter de façon automatique des données exprimées dans une langue. Les applications du TALN ont connu un grand essor notamment pour les langues comme le français et l'anglais. Néanmoins, il reste encore des efforts à fournir pour un nombre non négligeable de langues peu ou pas du tout étudiées dans la communauté du TAL. Ainsi, ce travail est réalisé dans le cadre de la modernisation et de la promotion de la langue amazighe.

Notre travail de recherche porte sur l'étude de la langue amazighe qui, malgré les efforts alloués pour son aménagement, présente encore des difficultés dans la façon dont les mots seront classés dans un dictionnaire. Dans cette perspective, le développement d'un Algorithme hybride pour l'extraction de la racine et la classification de mots s'avère nécessaire dans le but d'appliquer les outils du traitement automatique pour à la langue amazighe.

Ce mémoire est organisé selon le plan suivant :

- **le chapitre 1** introduit les concepts de base de la recherche d'information (RI). Nous commençons par quelques définitions, ensuite nous allons présenter le processus RI en passant par l'indexation, ses types et son processus. Puis, nous allons expliquer le concept du Traitement Automatique des Langues (TAL). Enfin, nous concluons par les applications et les problèmes majeurs du TAL.

- **Le chapitre 2** présente la langue amazighe. Nous décrivons le lexique et la morphologie amazighe. Ensuite, nous illustrons la formation des mots amazighs. Enfin, nous présentons une méthode de classification de ces mots.

- **Le chapitre 3** montre notre contribution qui se présente par le développement d'une application qui utilise notre algorithme, et pour ensuite l'exécuter sur un corpus de mots. A la fin nous allons conclure par une comparaison entre le résultat obtenu et la méthode classique de classification.

---

# *Chapitre 1 :*

## *La recherche d'information et le traitement automatique des langues*

---

### Introduction

Le traitement automatique des langues (TAL) et la recherche d'information (RI) sont deux disciplines dont l'interaction s'est renforcée ces dernières années. Le traitement automatique des langues s'intéresse aux traitements informatisés qui mettent en jeu des techniques linguistiques: analyse de texte, traduction automatique, classification des mots, etc. L'objectif est la représentation des données textuelles à différents niveaux de compréhension (morphologique, syntaxique), ou l'arrangement d'un ensemble de mots dans des catégories (dictionnaire, index). Ce chapitre a pour but de présenter les notions de la RI et du TAL. Nous allons commencer par présenter les concepts de base de la RI en décrivant son processus. Après, nous passerons en vue Le TAL, tout en expliquant ses différents niveaux d'application.

## 1. La recherche d'information (RI)

### 1.1 Définitions

#### ❖ **La recherche d'information :**

La RI est un domaine qui fournit des outils et techniques qui permettent de représenter, stocker, organiser, rechercher et retrouver, dans une masse documentaire existante, le ou les documents contenant l'information qui répond au besoin exprimé par l'utilisateur. Gerard Salton (Salton G., 1968) l'a défini comme suit:

*"Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information."*

### Un système de recherche d'information :

(HARTER,1986) : « *Un Système de Recherche d'Information (SRI) est un dispositif que s'interpose entre les usagers potentiels et la collection d'information* ».

Un SRI est un ensemble de programmes informatiques, qui a pour but d'établir une correspondance pertinente entre l'information recherchée par l'utilisateur, représentée généralement par le biais d'une requête, et l'ensemble des documents disponibles.

## 1.2 Processus de la recherche d'information

Le processus de la RI est composé de trois fonctions principales : l'indexation des documents et des requêtes, l'appariement requête-document (interrogation) et la reformulation de la requête.

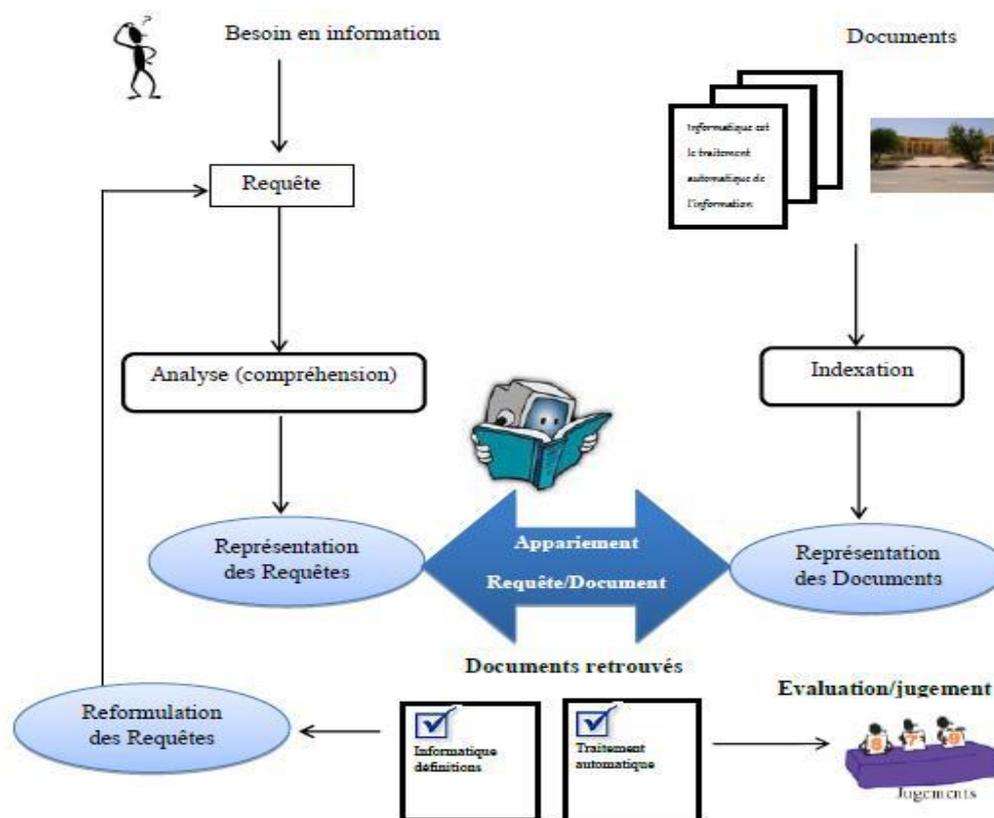


Figure 1.1: Processus de recherche d'information.

## 1.2.1 L'indexation

L'indexation est une étape primordiale dans le processus de recherche d'information. Pour chaque document, on procède à l'élimination des images et l'extraction du texte afin de le décomposer. Son résultat constitue un descripteur du document ou de requête. Ce descripteur est souvent sous forme d'une liste de tokens (termes) ou groupe de termes significatifs pour le document.

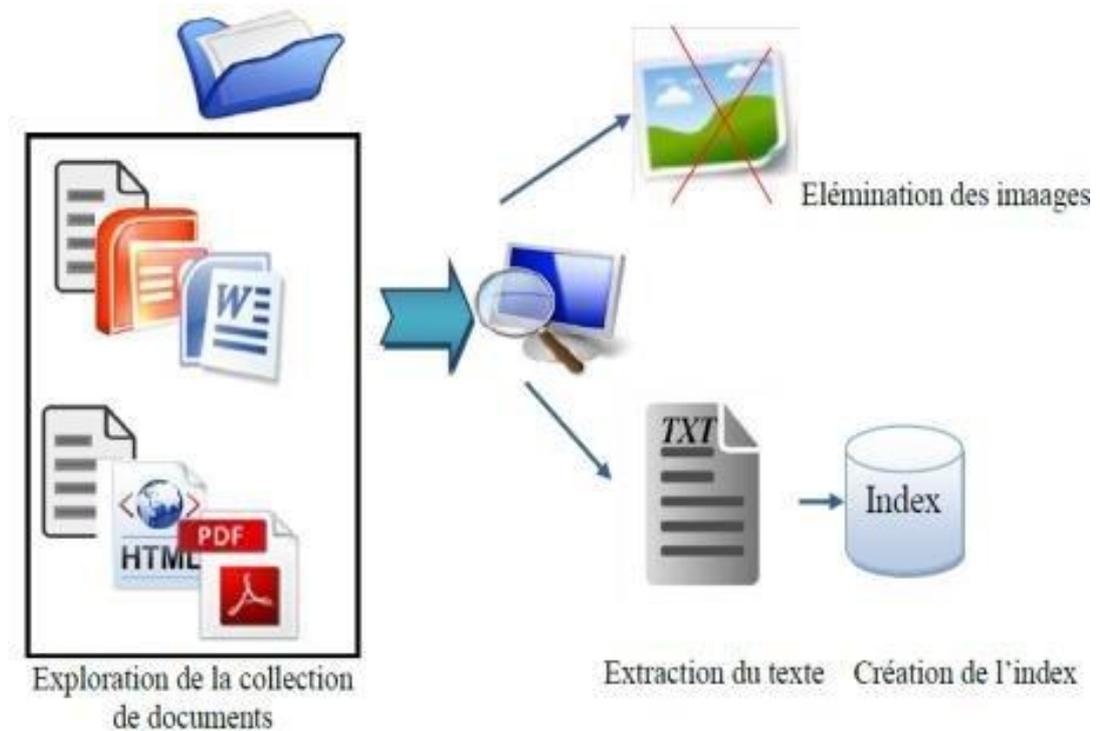


Figure 1.2 : Processus d'indexation

### 1.2.1.1 Les types d'indexation

- ❖ **Indexation manuelle** : Les documents sont analysés par des experts dans le domaine. Cette méthode d'indexation présente l'avantage d'être plus performante en termes de précision des documents retrouvés, en réponse à une requête utilisateur donnée, par contre elle est coûteuse en termes de temps et d'effort nécessaire.
- ❖ **Indexation Automatique** : C'est un processus complètement automatisé qui se charge d'extraire les termes caractéristiques du document, particulièrement adaptée aux corpus volumineux.

- ❖ **Indexation Supervisé** : L'indexation est faite automatiquement, alors que le choix final reste au spécialiste qui intervient souvent pour choisir les termes significatifs.

### 1.2.1.2 Le processus d'indexation

Le processus général d'indexation des documents est appliqué sur chacun des documents de la collection et comprend plusieurs étapes :

- a- Tokénisation (Segmentation):** Cette phase représente la segmentation des documents en unités, cette segmentation est basée en général sur la ponctuation et sur une liste de séparateurs, le résultat de cette étape est un ensemble de mots. La segmentation consiste à analyser chacun des documents afin d'extraire les termes représentatifs du contenu, un texte peut être découpé en plusieurs unités différentes chacune d'elles nécessitant un traitement bien défini.
- b- Elimination des mots vides:** C'est l'ensemble de prépositions, pronoms, conjonctions, déterminants, etc. qui permettent de lier les mots d'une phrase pour la structure. La liste des mots vides est prédéfinie pour chaque langue.
- c- Choix des termes:** Cette étape transforme les textes en une liste de termes ; elle se fait selon plusieurs possibilités, par exemple :
  - c.1 La normalisation linguistique** : Elle est basée sur l'application des techniques de racinisation et de lemmatisation.
    - c.1.1 La lemmatisation:** C'est l'obtention de la forme canonique (le lemme) d'un mot.
    - c.1.2 La racinisation (Stemming):** consiste à ramener un terme à sa forme la plus simple (racine).

### 1.2.2 L'appariement requête-document ou l'interrogation

La comparaison entre le document et la requête revient à calculer un score. Cette étape permet de comparer la représentation d'un document à celle de la requête. Ce score prend en considération les descripteurs, ainsi que leurs pondérations (poids) dans la représentation de la requête et la représentation du document.

### 1.2.3 La reformulation de la requête

Le principe de ce processus est de modifier la requête initiale de l'utilisateur en rajoutant des termes significatifs ou/et la ré-estimation de leurs poids.

## 2. Le Traitement Automatique des Langues (TAL)

### 2.1 Définition

(Delafosse, 1999) : « *Le Traitement Automatique des Langues a pour objectif de traiter des données linguistiques (textes) exprimées dans une langue dite "naturelle" ».*

Le traitement automatique des langues (TAL) ou bien le traitement automatique des langues naturelles (TALN) est un domaine de recherche interdisciplinaire. Il repose sur le fait de vouloir créer des programmes informatiques capables de traiter automatiquement des langues naturelles en se basant sur la linguistique, les formalismes et l'informatique afin d'extraire des informations sur un document.

### 2.2 Les niveaux d'analyse en TAL

Le processus d'automatisation d'une application nécessite en général une étape d'analyse du texte ou document-source (corpus), qui se fait à son tour par la décomposition de cette analyse en sous-tâches à savoir :

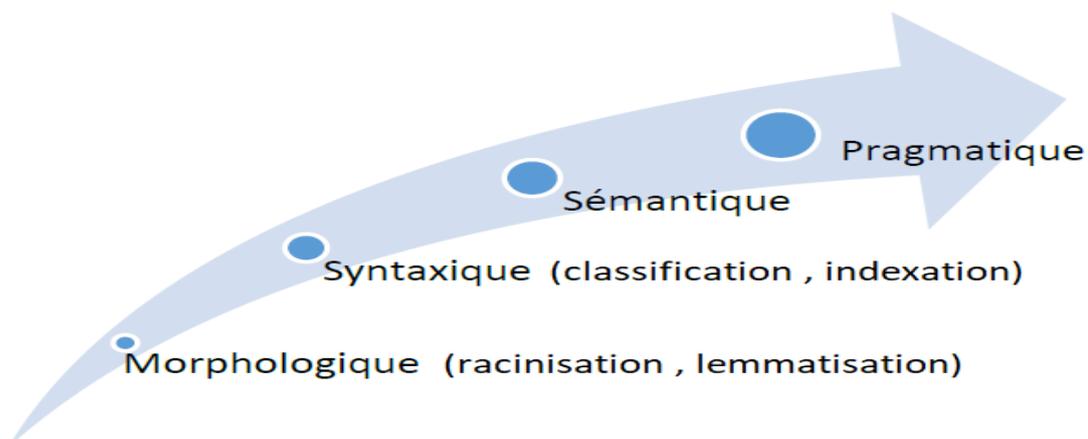


Figure 1.3: Les différents niveaux d'analyse d'un texte

**2.2.a L'analyse morphologique** présente la première étape d'un traitement linguistique de données textuelles, permettant de reconnaître les mots sous les différentes structures (conjugaison, déclinaison,), et de mettre en correspondance ces mots une fois reconnus avec les informations disponibles. Dans ce niveau, deux termes sont à étudier, la flexion et la dérivation :

**- La flexion**

La flexion est un processus qui consiste à modifier le radical d'un mot pour lui adjoindre certains types d'éléments

**- La dérivation**

C'est un procédé de formation de mots nouveaux par addition, suppression ou remplacement d'un élément grammatical d'un mot simple.

**2.2.b L'analyse syntaxique:** Etudier les contraintes entre les catégories morphosyntaxiques qui doivent être prises en compte pour la description des séquences de mots "acceptables" dans une langue donnée.

**2.2.c L'analyse sémantique:** prend comme unité d'analyse la phrase, et conduit à représenter sa partie significative.

L'analyseur sémantique décrit le sens des mots de la phrase ; ces mots sont identifiés par l'analyse morphologique et regroupés en structures par l'analyse syntaxique.

**2.2.d L'analyse pragmatique:** prend en compte le contexte d'énonciation. Ce qui fait que cette analyse est la plus difficile car certains énoncés ne se comprennent que dans un contexte géographique, historique ou culturel donné.

### 3. Les applications du TAL

Dans la vie de tous les jours, le TAL trouve de nombreuses applications telles que:

- La reconnaissance de l'écriture manuelle.
- La correction orthographique
- Le résumé d'un texte.
- La recherche d'information.
- La traduction automatique
- Les systèmes de dialogue homme-machine.
- la classification de mots ou leur arrangement suivant des catégories.

## 4. Les problèmes majeurs du TAL

### ❖ L'ambiguïté :

Il est souvent nécessaire à un système de traitement des langues de pouvoir déterminer la structure d'une phrase donnée ou le sens d'un mot en contexte, Cette ambiguïté se manifeste par la multitude d'interprétations possibles pour chacune des entités linguistiques pertinentes pour un niveau de traitement, comme par exemple :

- Ambiguïté des graphèmes (lettres) dans le processus d'encodage orthographique.
- Ambiguïté des terminaisons dans les processus de conjugaison et d'inflection.
- Ambiguïté dans les propriétés grammaticales et sémantiques.
- Ambiguïté de la fonction grammaticale des groupes de mots.
- Ambiguïté sur l'interprétation du contexte.
- l'ambiguïté sur la racine réel d'un mot.

### ❖ L'implicite :

La machine ne dispose pas de connaissance d'arrière-plan, ce qui rend la compréhension complète des énoncés difficile, voire impossible, si l'on ne dispose pas de bases de connaissances additionnelles, donnant accès à la fois à un savoir sur le domaine en général et sur le contexte de l'énonciation.

## Conclusion

Les techniques du traitement automatique des langues permettent d'extraire des informations plus riches que de simples unités lexicales. Ces niveaux d'analyse que ce soit morphologique, syntaxique, sémantique ou pragmatique, Ont été partiellement utilisées en RI pour améliorer les méthodes d'appariement, les représentations des contenus des documents et le processus de recherche. Le TAL est aussi d'une importance capitale dans la classification des mots et leur arrangement selon des catégories.

---

# CHAPITRE 2 :

## La langue amazighe

---

### 1. La langue amazighe

La langue amazighe, connue aussi sous le nom de berbère ou Tamazight (ⵜⴰⴳⴷⵓⴷⴰⵢⵜ [tamaziɣt]), est une langue afro-asiatique classée parmi les langues chamitiques au sein de la famille chamito-sémitique (Greenberg, 1966; Ouakrim, 1995). Cette dernière comprend, outre le berbère, le sémitique, le couchitique, l'égyptien (ancien) et le tchadique.

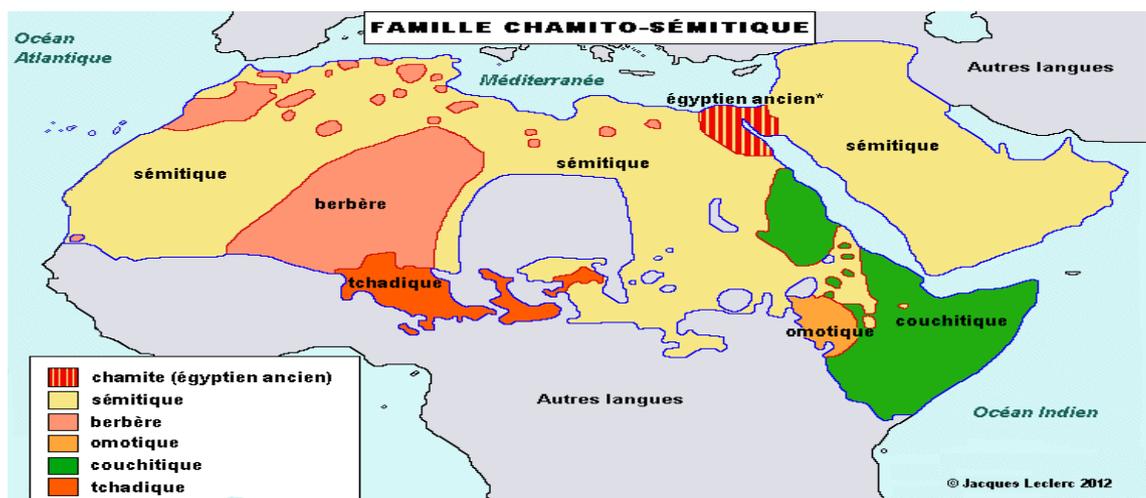


Figure 2.1 : Famille chamito-sémitique.

La langue berbère est considérée comme la langue autochtone de l'Afrique du Nord, elle regroupe un ensemble de parlers répartis de façon discontinue dans toute l'Afrique du Nord, allant de l'est de l'Égypte au littoral marocain en passant par la Libye, la Tunisie, le Niger, l'Algérie, le Mali et la Mauritanie.

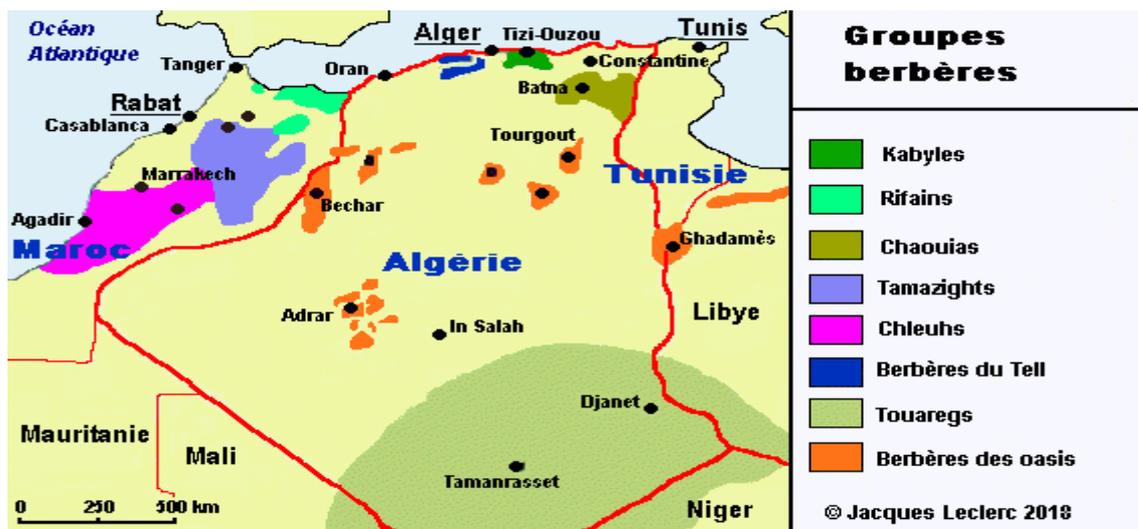


Figure 2.2 : Les groupes amazighophones

## 2. Le lexique de la langue 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).

Pratiquement parlant, la classe verbale et la classe nominale sont les classes qui jouent un rôle prépondérant dans toute la morphologie amazighe. Chacune de ces deux classes possède des caractéristiques formelles et fonctionnelles qui la distinguent.

### ❖ Le nom :

En amazigh, 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 verbe :

Les verbes en amazigh sont des mots simples qui présentent la base de la morphologie amazighe, étant donné qu'ils représentent une large classe morphologique qui est remarquablement riche et que d'autres verbes peuvent être dérivés d'eux. Le verbe peut donc prendre deux formes : simple et dérivé. La classification des verbes simples se fait suivant le nombre de consonnes qu'ils comprennent; ainsi on distingue les monolitères ([g] "être"), formés d'une seule consonne radicale, les bilitères (2 consonnes) ([ns]"passer la nuit"), les trilitères (3 consonnes) [lmd] "apprendre", les quadrilitères (4 consonnes) ([dryl] "être aveugle"), etc.

❖ **Les pronoms :**

Le pronom désigne tout élément susceptible de se substituer à un nom ou à un groupe nominal. (Boukhris et al, 2008 ; Sadiqi, 1997).

❖ **Les 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 subordonnants et les conjonctions (Boukhris et al., 2008). Généralement, les mots outils sont invariables.

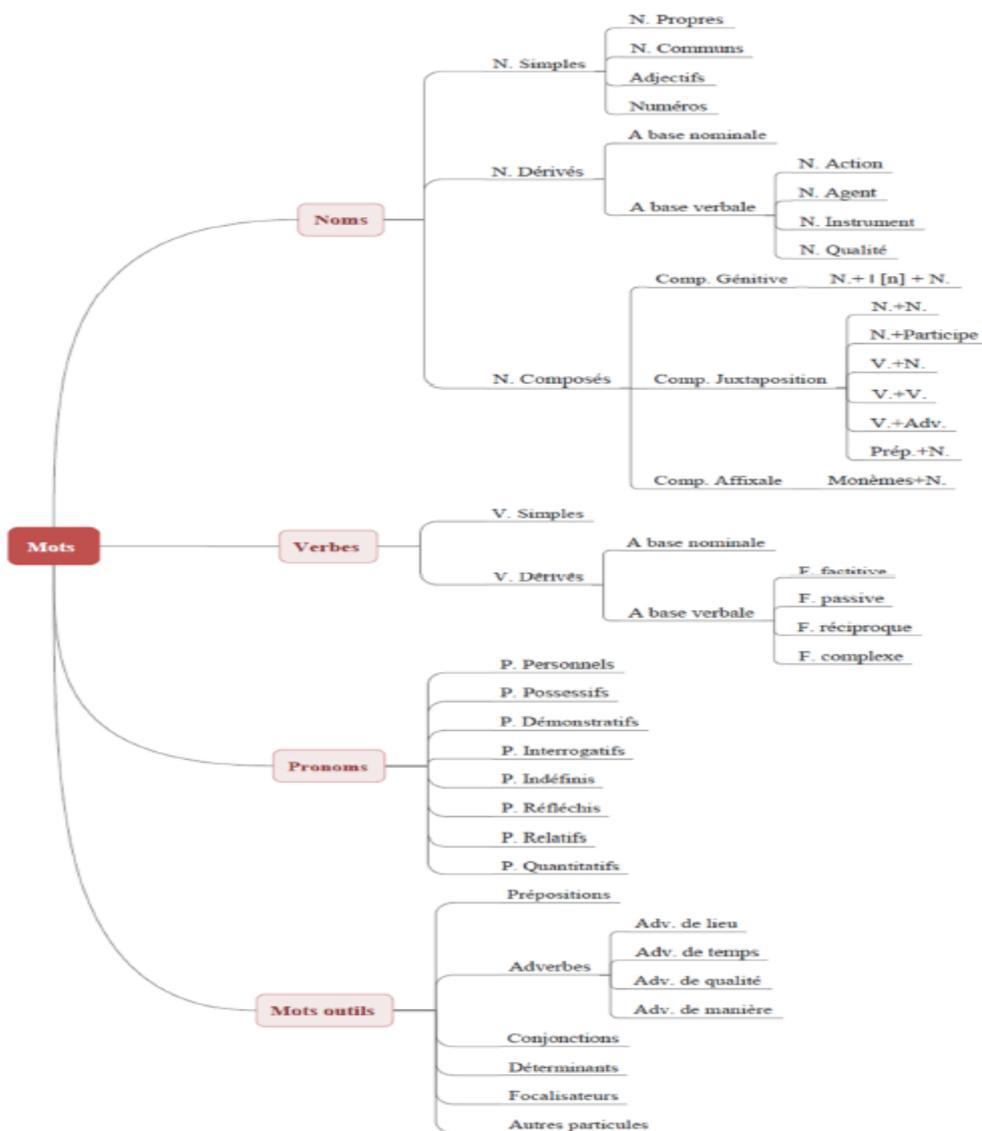


Figure 2.3 : Les classes grammaticales en amazigh.

### 3. La morphologie amazighe :

La langue amazighe est riche en termes de flexion, de dérivation et de complexité qu'elle produit. Comme la plupart des langues chamito-sémitiques, la morphologie de l'amazigh a une structure multiniveau. Ces mots ont une structure morphologique complexe. Cette complexité est due au fait que le processus de la formation de ces mots se fait d'abord à partir d'une racine, qui est ensuite encadrée dans un schème auquel peut s'ajouter un nombre variable d'affixes.

#### 3.1 Morphologie flexionnelle

La morphologie flexionnelle est une méthode de création lexicale qui consiste à créer un nouveau terme en associant les désinences du genre, du nombre et d'état à une base lexicale déjà existante.

La langue amazighe est une langue fortement flexionnelle. Elle emploie, pour la conjugaison du verbe et la flexion du nom, des indices d'aspect, de mode, de temps, de personne, de genre, de nombre et d'état, qui sont en général des suffixes et préfixes à antéposer ou à postposer (Boukhris et al., 2008).

#### ❖ Flexion des noms

Les morphèmes flexionnels sont les éléments morphologiques qui donnent à la catégorie nominale ses valeurs grammaticales. En amazigh, le nom peut présenter trois flexions différentes, en fonction de la variation du genre, du nombre et d'état.

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

**1.a Le nom masculin:** commence généralement par une des voyelles initiales: [a] [i] ou [u]. Cependant, il existe certains noms qui font exception tels que [fad] "soif" ou [laz] "faim".

**1.b Le nom féminin:** Dans le cas général, le féminin est formé sur le masculin par la préfixation et la suffixation du morphème discontinu [t...t], il y a aussi des noms féminins singulier qui commence par [t] et se termine par une voyelle [a] ou [i] tels que [tala] "fontaine", [tili] "ombre", et aussi des noms féminins qui n'ont pas de masculin [tafat] "lumière". Il y a aussi des exceptions de noms féminin qui commence par une voyelle : [imma] "(ma) mère", [ulli] "chèvres". Nous pouvons former le féminin à partir d'un thème masculin.

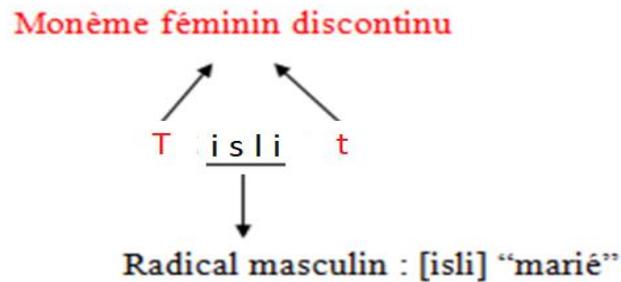


Figure 2.4: Formation du féminin à partir d'un nom masculin

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

**2.a Le pluriel régulier:** qu'on appelle également « externe », étant donné que le nom ne subit pas de modification interne, est obtenu par une alternance vocalique de la première voyelle accompagnée par une suffixation de [n] ou l'une de ses variantes ([in], [an], [ayn], [wn], [awn], [wan], [win], [tn], [yin]) [axxam] → [ixxamn] “maisons”.

**2.b Le pluriel interne (ou brisé) :** est obtenu par une alternance vocalique plus un changement de voyelles internes (et plus rarement d'une consonne), sans suffixe ([adrar] → [idurar] “montagnes”).

**2.c Le pluriel mixte :** est formé par une alternance d'une voyelle interne et/ou finale plus une suffixation par [n], ou bien par une alternance vocalique initiale accompagnée d'un changement vocalique final [a] plus une alternance interne ([anggaru] “dernier” → [inggura] “derniers”).

### 3. L'état :

L'opposition flexionnelle entre l'état d'annexion (dit aussi état construit) et l'état libre (ou neutre) est un phénomène morphosyntaxique qui concerne une grande partie des noms en amazigh. Il s'agit d'une variation morphologique qui affecte la première syllabe du nom lorsque celui-ci remplit les conditions nécessaires à la formation de l'état d'annexion.

### 3.a L'état libre :

Le nom en état libre est un mot isolé de tout contexte syntaxique, l'état libre se manifeste sous forme d'un complément d'objet direct, ou bien d'un complément de la particule prédictive [argaz], [wagi d argaz], [zrigh argaz].

### 3.b L'état d'annexion :

Elle est fondée 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 [a] [u] [argaz] "homme" [urgaz]) ou bien maintien de la voyelle initiale et ajout d'un [w] au cas des noms masculins à initiale [a] ; préfixation d'un [w] pour les noms à initiale [u] ; préfixation d'un [y] aux noms à voyelle initiale [i] ([ils] "langue" [yils]).

### ❖ Flexion des verbes :

La conjugaison des verbes décrit la variation de leurs formes en fonction des circonstances. En amazigh, les principales valeurs grammaticales que les morphèmes flexionnels verbaux expriment sont : l'aspect, le mode et la désinence.

**1. La valeur aspectuelle :** l'aspect est un trait grammatical. Il dénote la façon dont l'état exprimé par le verbe est envisagé du point de vue de son développement indépendamment du moment où l'on parle ;

**2. La valeur modale :** montre la manière dont l'action exprimée par le verbe est conçue et présentée. L'action peut être mise en doute, affirmée comme réelle ou éventuelle. Le verbe amazigh dénote trois modes, à savoir : indicatif, impératif et participial.

**3. La valeur désinentielle :** la désinence est un suffixe grammatical qui désigne des traits grammaticaux. En amazighe, on distingue trois types de désinences verbales ou indices de personne : les désinences de la forme indicative, de la forme impérative et de la forme participiale.

## 4. La formation de mots amazighs :

En berbère, les mots sont formés selon les deux procédés bien connus, la dérivation et la composition. La dérivation lexicale se réalise par l'ajout d'un schème

vocalique à une racine consonantique. La composition consiste en la juxtaposition ou l'agglutination de mots censés avoir un comportement autonome dans la langue.

#### 4.1 Formation de mots par dérivation :

*<< Tamazight, appartenant à la famille chamito-sémitique, est une langue à « racines » ; cela signifie que chaque mot, quel que soit sa complexité, peut se réduire à une suite de phonèmes (consonnes) qui constitue sa base lexicale et à cette base se rattachent des mots, constituant ce qu'on appelle une série morphologique. Si la racine forme l'ossature du mot, c'est le schème, également ensemble de phonèmes, ici voyelles et consonnes, qui lui permet de se réaliser en tant que mot et de s'intégrer dans une catégorie grammaticale : nom, verbe, adjectif...>> (D. Cohen, 1968).*

La racine kabyle qui exprime l'idée générale de « manger » illustre cette organisation du vocabulaire :

ečč « manger » avec différents schèmes : 'i \_\_ a' exprime le prétérit ;

ssečč / ccečč : « faire manger » : s- « préfixe actif-passif »

mmečč « être mangé » ; mm- : « préfixe passif »

myečč « se manger, se dévorer réciproquement » : my- « préfixe «réciproque » ;

učči : « fait de manger, nourriture » : u – i « schème du nom d'action verbale et du nom concret » ;

ameččay « gros mangeur » : am- « préfixe de nom d'agent ». Il faut ajouter pour les verbes, les formes nominales associées aux noms d'action verbale :

ečč : učči, tuččit

ccečč : acečči

mmečč : amechči

myečč : amyečči

mcečči : amcečči

Chaker (1991 : 124 - 128) ajoute que l'amalgame de ces deux éléments, racine et schème, forme le thème sur lequel se greffent les modalités obligatoires inhérentes à la catégorie syntaxique du mot (le genre et le nombre pour les noms et l'aspect pour les verbes). Ce procédé général se schématise :

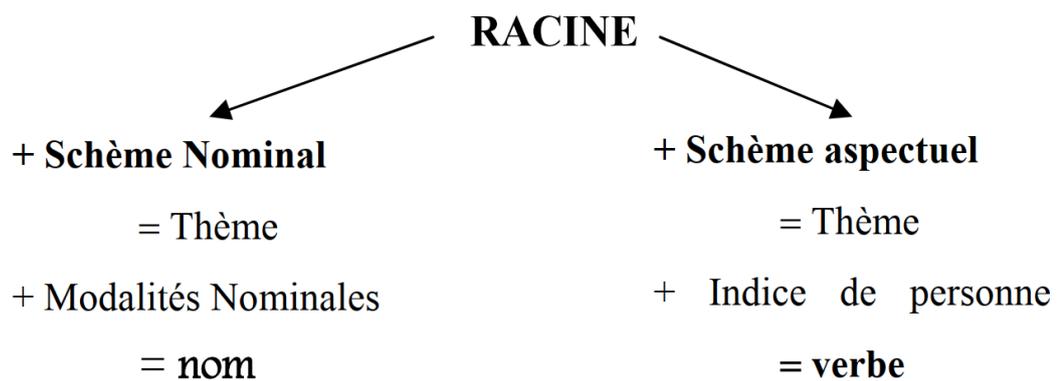


Figure 2.5: Formation de mots par racine

La racine lmd « apprendre », par exemple, est commune à toutes les catégories d'unités d'une même base. Pour qu'un terme passe d'une catégorie grammaticale à l'autre, il suffit de procéder à la substitution de son schème initial par celui de la catégorie désirée. Par exemple, le verbe lmed (c1 c2 c3) devient anelmad « étudiant » par l'utilisation du schème nominal an\_\_a\_\_ de nom d'agent.

#### 4.2 Formation de mots par composition :

Selon Jean Dubois (1994 : 106) « *Par composition, on désigne la formation d'une unité sémantique à partir d'éléments lexicaux susceptibles d'avoir par eux-mêmes une autonomie dans la langue.* »

Les unités lexicales 'lexèmes' formant les mots composés, qu'ils soient agglutinés ou séparés par des blancs, sont des lexèmes libres ayant, chacun à part, une autonomie syntaxique dans d'autres contextes du discours. Par exemple, les deux mots composant 'amagraman' « aulne », par agglutination, sont le verbe 'mager' et le nom 'aman' ; les deux unités ont un fonctionnement autonome dans la langue.

Chaker (1991 : 179) fait remarquer que la composition « *n'est en berbère qu'un phénomène sporadique, peu productif...* ». Il parle de composition par agglutination ou collage de mots simples comme l'illustre l'exemple suivant : asyarsif « arbuste poussant au bord des rivières », de : 'asyar' « bois » et 'asif' « rivière » (Nom + Nom).

## 5. La classification des mots amazighs :

[J. Rey-Debove (1971 : 22)] : « *Pour qu'un ouvrage puisse être consulté, il faut un ordre formel ; et s'il ne peut être consulté, ce n'est plus un dictionnaire* ».

### ❖ Le classement par ordre alphabétique des mots :

Ce type de rangement du lexique est le plus usité. Malgré son caractère banal, il reste le plus pragmatique grâce à sa commodité qui rend facile la tâche du lexicographe ainsi que celle du consulteur. Les mots sont pris tels qu'ils sont transcrits puis classés selon l'ordre alphabétique canonique. Exemple : akal ; argaz ; izi ; tamurt ; etc.

#### • Avantages et Inconvénients:

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• ne nécessite aucune connaissance préalable de la langue ;</li> <li>• très pratique pour chercher un mot, connaissant son orthographe (adopté par toutes les langues à alphabet, il est devenu universel) ;</li> <li>• utile en vue de la normalisation et la diffusion de la langue.</li> </ul>	<ul style="list-style-type: none"> <li>• ne tient pas compte des rapports entre les mots ;</li> <li>• problème de surcharge des mots en amazigh à voyelle initiale a/i/t/u (les féminins ayant un masculin doivent aussi figurer dans la liste du lexique car ils ont un sens différent du masculin).</li> </ul>

Tableau 2.1 : Avantages et inconvénients de la classification par ordre alphabétique

### ❖ Le classement par ordre alphabétique des racines :

Ces dernières décennies, cette méthode de classement domine la lexicographie amazighe.

« *A notre avis, le classement par racine s'impose, cela est dû à la nature de la langue amazighe dont le mot simple est composé d'une racine et d'un schème* » (Serhoual, 2002 : VIII).

« *Cette méthode [classement par racines], il faut le reconnaître, nous offre donc, en principe, l'avantage de présenter les mots en les intégrant à des familles lexicales ayant en commun le sens minimum porté par la racine. Ceci ne serait qu'une*

*justice rendue à la nature éminemment dérivationnelle du berbère et permettrait au consultant, en même temps que de consulter le dictionnaire et de s'initier à la langue et à sa logique propre* ». Écrivait Rabdi (2004 : 200).

La racine est la base consonantique d'un mot qui reste après que l'on a retranché tout élément à valeur grammaticale ou dérivationnelle, par exemple : le mot 'tixxamin', chambres, on enlève les marques du féminin pluriel 'ti—in' et la voyelle 'a' pour trouver la racine 'XXM'.

- **Pourquoi la classification par racine :**

Comme toutes les langues chamito-sémitiques, les mots de la langue amazighe sont formés à partir d'un noyau consonantique appelé racine. Par exemple, la racine RBH se retrouve dans 'rbeḥ', gagner ; 'ṭwarbeḥ', être gagner ; 'mserbaḥ', se faire gagner ; 'urbiḥ', porte-chance. Une classification par ordre alphabétique des mots obligerait à classer ces mots sous des lettres différentes ce qui détruirait l'unité de cette famille de mots. Ainsi, en adoptant cette méthode on trouvera que la majorité des noms masculins se verrait traitée sous la lettre A et la quasi-totalité des noms féminins sous la lettre T. Ce qui nous pousse à utiliser la classification par racines.

- **Avantages et Inconvénients:**

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>• les locuteurs « sentent » ou « ont conscience » de la racine ;</li> <li>• tient compte des rapports entre les mots (groupement par famille) ;</li> <li>• compatible avec le comparatisme inter-dialectal et utile pour la reconstruction diachronique.</li> </ul>	<ul style="list-style-type: none"> <li>• nécessite la connaissance des racines des mots ou la technique d'extraction de la racine ;</li> <li>• difficilement accessible aux profanes (qui ne maîtrisait pas la grammaire amazighe).</li> <li>• Absence d'harmonisation ; la méthode change d'un auteur à un autre ;</li> <li>• Une racine donnée peut correspondre à plusieurs sens très éloignés les uns des autres.</li> </ul>

Tableau 2.2 : Avantages et inconvénients de la classification par racine

## Conclusion :

Dans ce chapitre, nous avons donné un aperçu général sur la langue amazighe. Nous avons présenté son lexique et sa morphologie qui est très complexe, ce qui représente un défi que les chercheurs en traitement automatique de langue amazighe doivent relever. Nous avons aussi donné un aperçu sur la formation et la classification de mots et montré la nécessité d'un classement de mots selon la racine. Mais pour les profanes qui ne maîtrisent pas la langue amazighe, il leur sera difficile de consulter un tel dictionnaire. Pour cela, afin de faciliter la tâche, la racine réelle ou formelle sera dégagée à partir du Système Grammatical Amazigh (SGA).

Dans la suite de ce travail, nous allons présenter une méthode hybride basée à la fois sur la racine et l'ordre alphabétique à l'intérieur de celle-ci, pour la classification de mots dans un dictionnaire amazigh.

---

## CHAPITRE 3 :

### Construction et formalisation d'un dictionnaire

---

#### Introduction

Dans ce chapitre, nous allons construire un dictionnaire de langue amazighe en se basant à la fois sur l'ordre alphabétique des mots et celui de leurs racines.

On va commencer par trouver les racines des mots puis on va les classer par ordre alphabétique des racines, et à l'intérieur de chaque racine les mots sont classés par ordre alphabétique, et ce en développant un algorithme approprié.

#### 1. Racinisation (stemming)

La racinisation (stemming en anglais) est le nom donné au procédé qui vise à donner le radical ou stamme d'un mot. Elle cherche à rassembler les différentes variantes flexionnelles et dérivationnelles d'un mot autour d'un stamme. La racine d'un mot correspond à la partie du mot restante après la suppression des différentes variantes qui existe au sein du mot (suffixe, préfixe, voyelles.etc). Contrairement au lemme qui correspond à un mot réel de la langue, la racine ne correspond généralement pas à un mot réel.

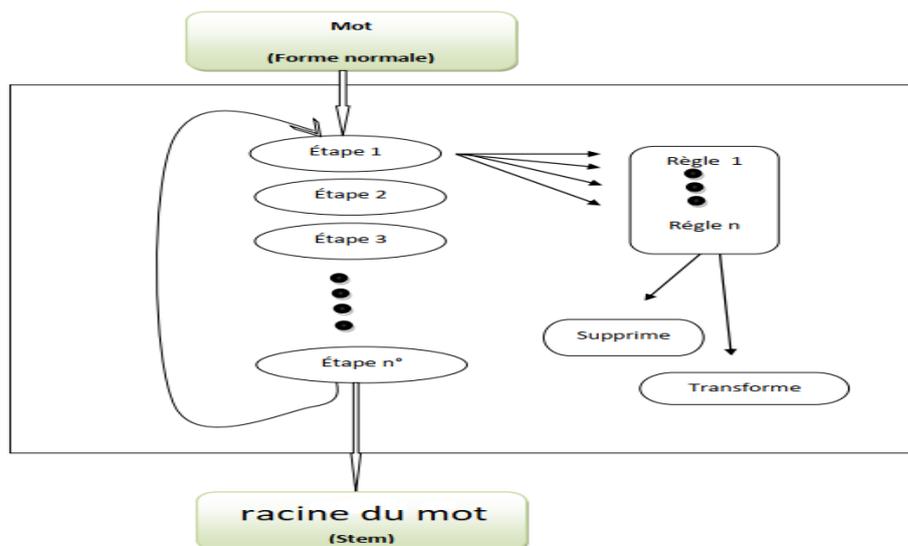


Figure 3.1 : Processus de racinisation

## 2. Représentation formelle du dictionnaire

Dans ce dictionnaire on va présenter les champs suivants :

### ➤ Le mot

- Les mots sont insérés comme suit :
- Les noms masculins ou féminins singulier.
- Les verbes conjugués à la deuxième personne du singulier du mode impératif.
- Les conjonctions, particules, pronoms, adverbes et prépositions.

### ➤ La catégorie grammaticale

La première information que nous avons attribuée aux entrées de notre dictionnaire correspond à la catégorie grammaticale. Ainsi, chaque entrée est associée, de façon non ambiguë, à une catégorie grammaticale désignée par un code. Le tableau suivant présente les abréviations utilisées pour chaque partie:

La classe grammaticale	Code Catégorie	Exemple
Nom	N	Argaz « homme » Tala « fontaine »
Verbe	V	Ettu « oublies »
Adjectif	Adj	ameqqran « grand »
Conjonction	Conj	Ar « jusqu'à »
Particule	Part	Aked « et, avec »
Pronom	Pron	Nekk « moi »
Adverbe	Adv	Assa « aujourd'hui » Melmi « quand » Axxir « mieux »

Tableau 3.1 : Les classes grammaticales

### 3. Le principe de l'algorithme

Dans cet algorithme, le système de notation qu'on va utiliser est celui des berbérissants scientifiques :

a, b, c, č, d, đ, e, f, g, ğ, h, ħ, i, j, k, l, m, n, ɣ, q, r, ř, s, š, t/ṭ, ṭ, u, w, x, y, z, ž, ε.

On va classer les mots par racines. A l'intérieur de chaque racine, les mots suivent l'ordre alphabétique. La lettre qui correspond à chaque mot est forcément la première consonne rencontrée, différente des consonnes qui participent au système grammatical amazigh :

« m, n, s, t /ṭ, w, y ».

Une fois la lettre choisie, la racine est alors constituée de cette lettre et des consonnes situées après elle dans le mot.

Si un mot comporte exclusivement des consonnes grammaticales, la recherche de la lettre correspondante au mot est la consonne rencontrée après la première voyelle, quelle que soit par ailleurs celle-ci : a, e, i, u ; et s'il n'y a pas de consonne après la voyelle, on prend la première consonne tout court.

Notons qu'une racine peut comporter des consonnes grammaticales si elles sont justifiées.

#### Exemples :

Asemblal : aller à la lettre L et la racine LL.

Tutlayt : aller à la lettre L et la racine LY.

Times : aller à la lettre M et la racine MS.

Tasa : aller à la lettre S et la racine S.

Tatut : aller à la lettre T et la racine T.

Wa : aller à la lettre W et la racine W

Notons que la racine d'un mot peut être réel et/ou formelle Ainsi :

\* Une racine d'un mot est dite formelle, si elle comporte à la fois des consonnes non grammaticales et grammaticales justifiées ou non. En particulier une racine peut comporter uniquement des consonnes grammaticales justifiées ou non.

\* Une racine d'un mot est dite réelle, si les lettres qui forment cette racine sont des consonnes non grammaticales avec la possibilité d'avoir des consonnes grammaticales si elles sont justifiées. En particulier une racine réelle peut comporter

uniquement des consonnes grammaticales justifiées à condition qu'elles soient justifiées.

Exemple :

La racine du mot « amḍan » est « ḍn » la lettre n est justifier, « ḍn » est une racine réelle.

La racine du mot « acebḥan » est « chḥ » la lettre n renvoi à la qualité d'un adjectif.

### Algorithme de racinisation

Entrée : Mot

Sortie : la racine du mot

Début

Mettre toutes les lettres en minuscule ;

parcourir le mot :

Si on trouve une consonne non grammaticale :

```
{
  Supprimer la lettre t au début et à la fin du mot
  Supprimer toutes les lettres avant la consonne non grammaticale
}
```

Sinon

```
{
  Prendre la première consonne trouvée après une des voyelles (a, e, i, u)
  Supprimer la lettre t au début et à la fin du mot // signe du féminin //
  Supprimer les toutes lettres avant la consonne
}
```

Supprimer les voyelles

Afficher la racine

Fin

### 3.1 Racinisation des Mots amazighs

En amazigh, la racine d'un mot est la partie restante après avoir supprimé les différentes variantes qui le composent (suffixes, préfixes, voyelles). Nous avons élaboré un arbre illustrant la démarche à suivre pour extraire la racine d'un mot.

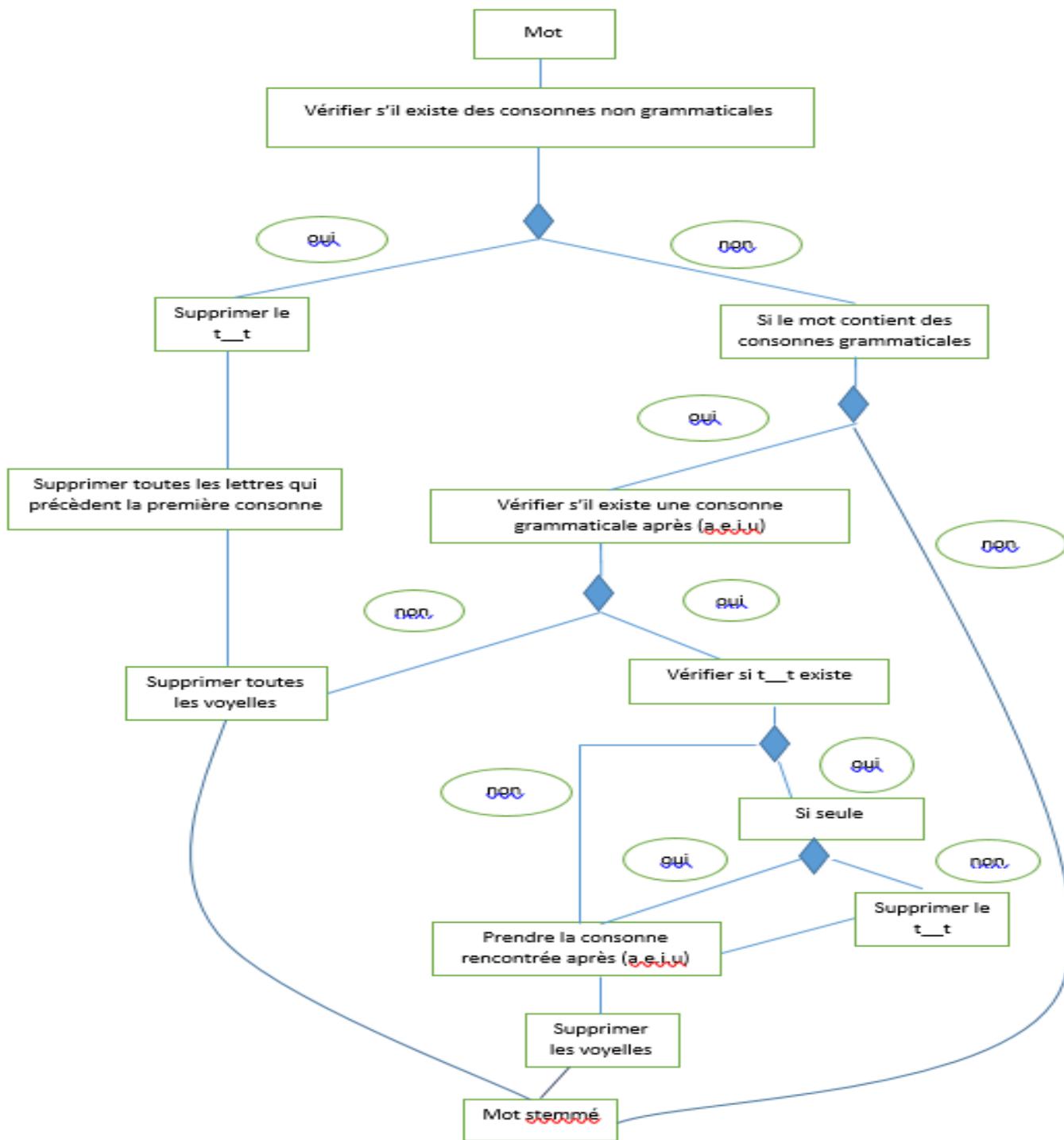


Figure 3.2 : Arbre d'exécution de l'algorithme

## 4. Le corpus

Nous allons exécuter notre algorithme sur un ensemble de 259 mots (noms, verbes, conjonctions, etc.). Après la classification de ces mots par ordre alphabétique, on a construit un diagramme circulaire pour montrer la répartition des mots selon la première lettre de chacun d'eux.

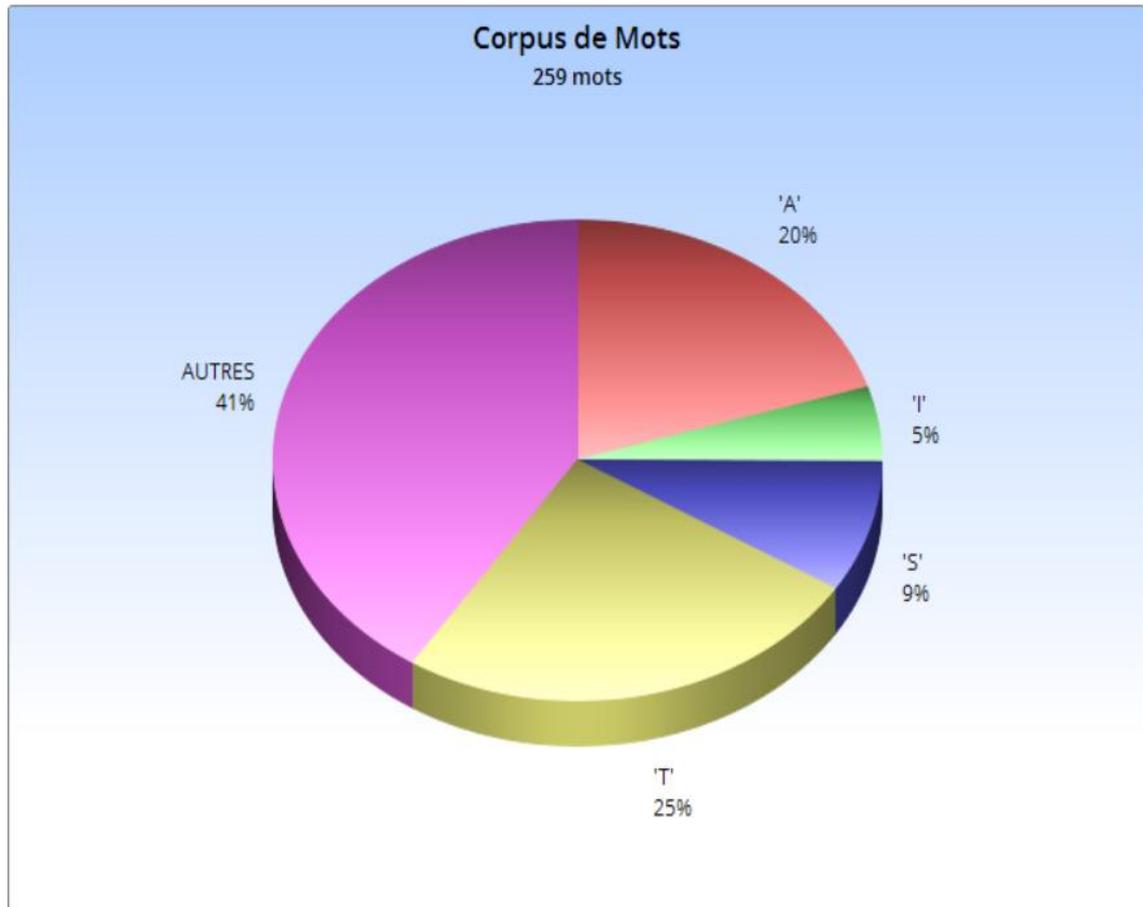


Figure 3.3 : La répartition des mots dans notre corpus

## 5. L'application

Dans ce qui suit, nous allons donner un aperçu sur l'application qu'on a développée; notons qu'on a utilisé « Python » comme langage de programmation

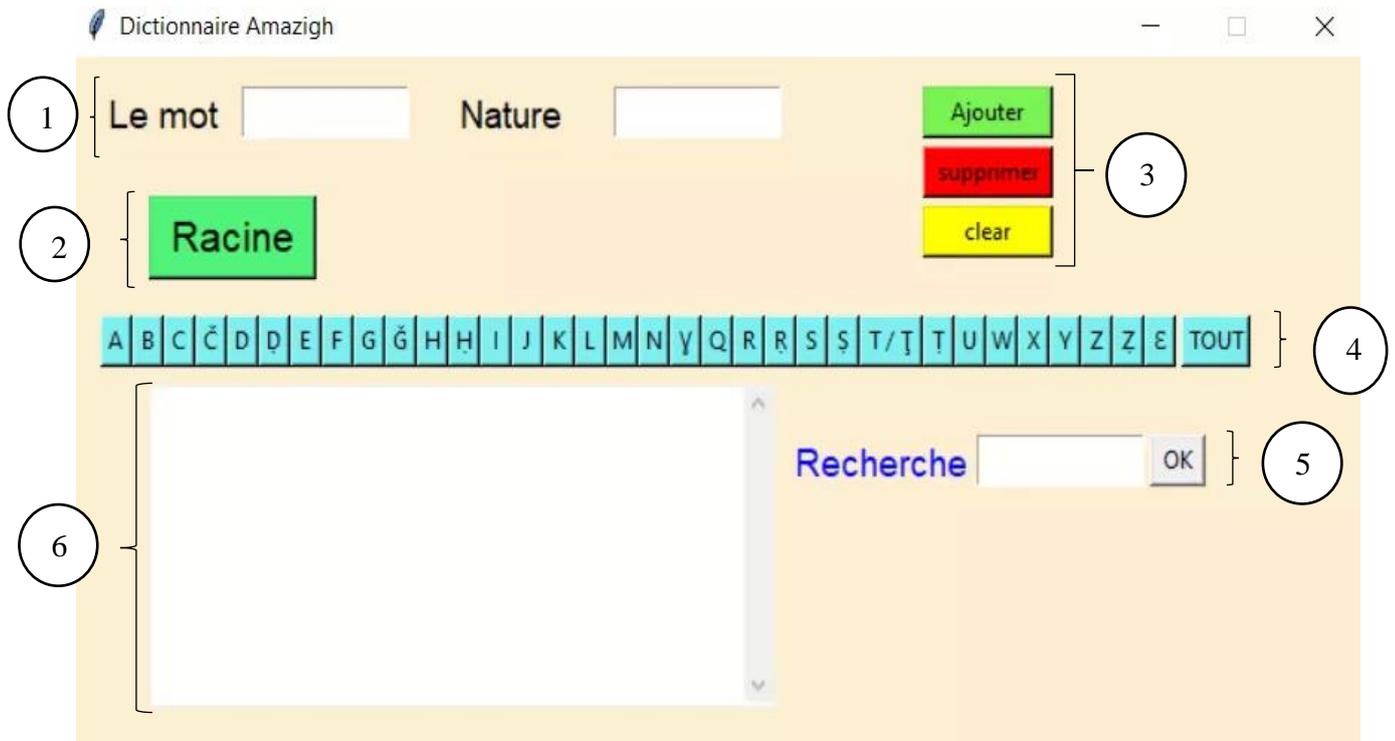


Figure 3.4: La fenêtre principale.

### Légendes

1. Les champs pour saisir les mots et leur nature.
2. Le bouton « Racine » pour extraire et afficher la racine du mot saisi.
3. Boutons pour manipuler les mots saisis :
  - Ajouter : sauvegarder le mot dans la base de données.
  - Supprimer : pour supprimer une entrée dans la base de données.
  - Clear : effacer les entrées dans les champs de saisie mot et nature.
4. L'alphabet amazigh, chaque caractère est représenté sous forme d'un bouton.
5. Une zone de recherche d'un mot dans le dictionnaire.
6. Une zone pour afficher des données de la base de données.

## 6. Le processus de fonctionnement de l'application :

On a commencé par créer une base de données qu'on a nommé « mot.db », à l'intérieur de cette base, on a créé une table « dictionnaire » qui contient 4 éléments (colonnes) : id, mot, nature et racine . Et ce en utilisant la requête SQL « CREATE TABLE », ensuite on a défini les fonctions nécessaires pour la manipulation de notre base de données (insert, remove, fetch).

```
def __init__(self, db):
    self.conn = sqlite3.connect(db)
    self.cur = self.conn.cursor()
    self.cur.execute(
        "CREATE TABLE IF NOT EXISTS dictionnaire (id INTEGER PRIMARY KEY, mot,nature,racine)")
    self.conn.commit()
```

Figure 3.5 : Créer une table dans la base de données « mot.db ».

On utilise la requête « INSERT INTO » pour insérer les données (mot, nature, racine), tout en respectant l'ordre des colonnes.

```
def insert(self, mot,nature,racine):
    self.cur.execute("INSERT INTO dictionnaire VALUES (NULL, ?, ?, ?) ",
                    (mot,nature,racine))
    self.conn.commit()
```

Figure 3.6 : Ajouter des données dans la table.

Si on clique sur le bouton supprimer, le programme va exécuter la requête « DELETE FROM » afin de supprimer les données sélectionnées.

```
def remove(self, id):
    self.cur.execute("DELETE FROM dictionnaire WHERE mot=?", (id,))
    self.conn.commit()
```

Figure 3.7 : Supprimer des données de la table.

Les requêtes « SELECT » et « ORDER BY » permettent d'afficher les données, tout en respectant un ordre précis

```
def fetch(self):
    self.cur.execute("SELECT racine,mot FROM dictionnaire ORDER BY racine ,mot")
    rows = self.cur.fetchall()
    return rows
```

Figure 3.8 : Afficher les données (racine et mot).

Après la création de la base de données, on a procédé à la création de la fonction 'racinisation' qu'on a mis au point avec le langage de programmation Python. Pour extraire la racine d'un mot, il suffit donc d'appuyer sur le bouton « Racine », après avoir donné le mot et sa nature. Alors l'algorithme peut s'exécuter et nous afficher la racine du mot introduit, par la suite, on peut ajouter le mot et sa racine à notre base de données en appuyant sur le bouton ajouter.

```
for lettre in a:
    if lettre not in cg and lettre not in v:
        if a[0] == 't' and a[-1] == 't':
            a = a[1:-1]

        for letter in a:
            if letter not in v:
                racine += letter
        for lettre in racine:
            if lettre not in cg:
                r = racine[racine.find(lettre):]
                break
        break

for lettre in a:

    if len(a) < 8:
        e = a + ' '
        if e[0] not in c and e[1] not in c and e[2] not in c and e[3]
            for i in range(len(a)):
                if e[i + 1] in cg and e[i] in v:
                    racine = a[i:]
                    if a[0] == 't' and a[-1] == 't' and racine != 't':
                        racine = racine[0:-1]
                    for letter in racine:
                        if letter not in v:
                            r += letter
                    break
```

Figure 3.9 : Extrait de l'algorithme de racinisation .

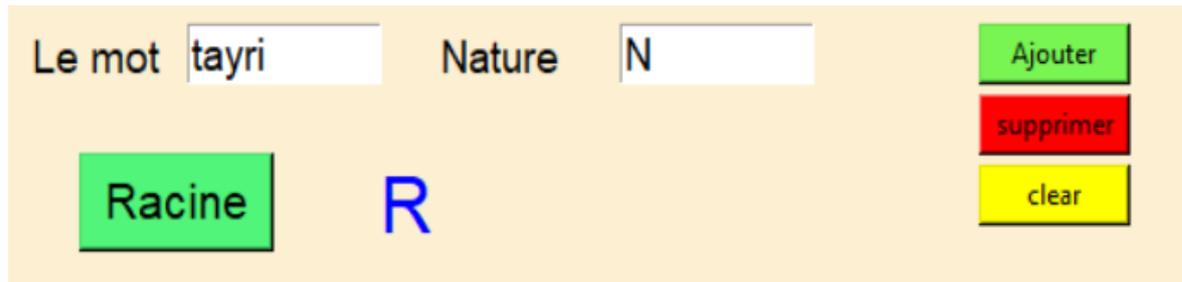


Figure 3.10 : Résultat d'exécution de l'algorithme.

On peut aussi savoir comment rechercher dans un dictionnaire un mot dont on ne connaît pas la racine, Pour cela, il suffit d'écrire le mot dans la zone de recherche et appuyer sur OK.

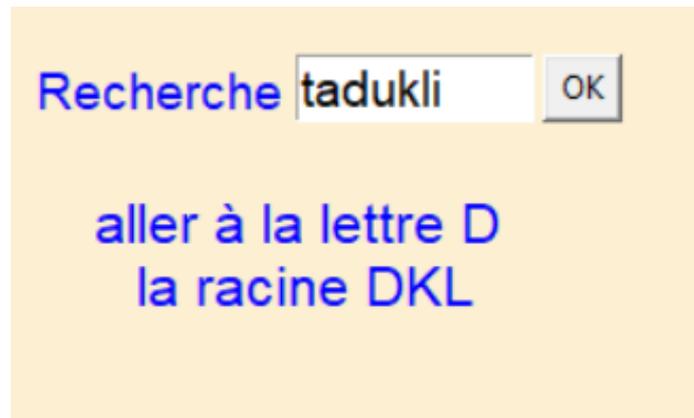


Figure 3.11 : Recherche d'un mot dans le dictionnaire.

## 7. La classification en utilisant la méthode hybride

En prenant notre corpus, on a trouvé, par l'ordre strictement alphabétique, que la majorité des mots se trouvent dans les lettre (A20%) et (T,25%). (Voir figure 3.3).

Après l'exécution de notre algorithme sur le même corpus, on a constaté que les mots sont répartis sur plusieurs lettres et les mots qui sont de la même famille se trouvent souvent dans la même racine (la racine de : axxam , taxxamt est XXM ).

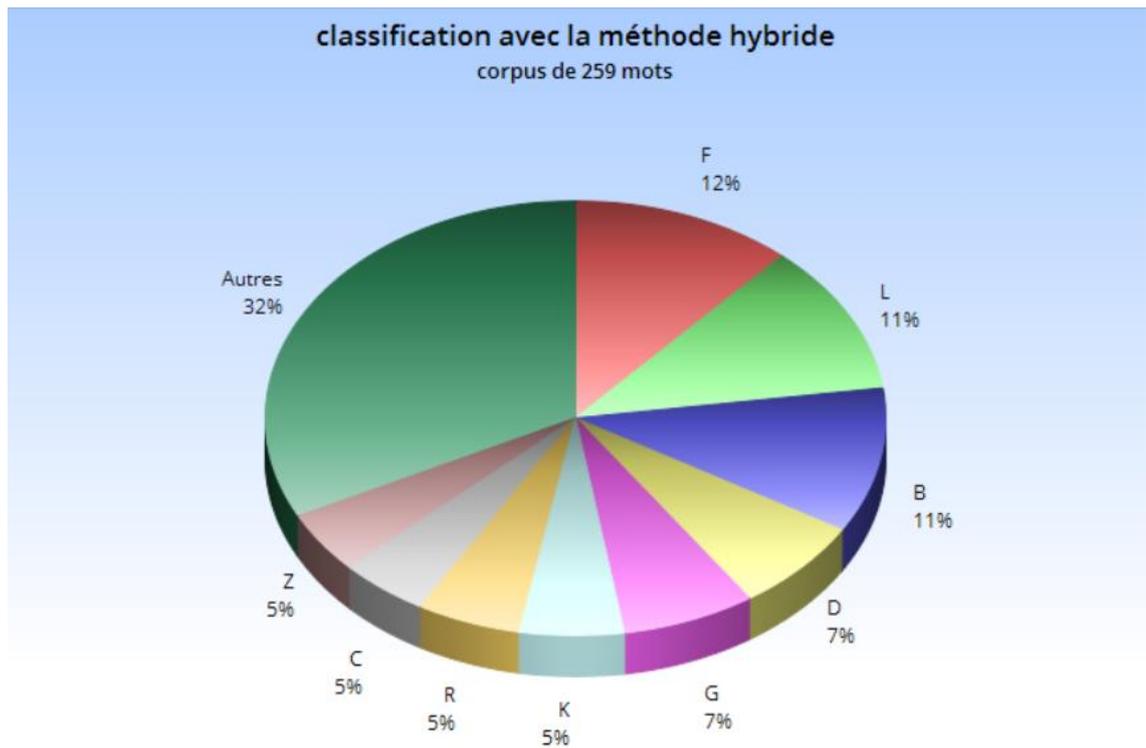


Figure 3.12 : La répartition des mots en utilisant la méthode hybride.

## Conclusion

Nous avons développé un algorithme qui exécute une méthode hybride pour la classification de mots dans un dictionnaire amazigh. Cette méthode se base sur le fait d'extraire la racine formelle d'un mot en fonction des lettres qui le composent.

On a aussi développé une interface graphique, en utilisant le Python comme langage de programmation, pour bien illustrer le travail effectué.

L'application de cette méthode sur un corpus nous a montré la possibilité d'avoir un dictionnaire de langue amazighe, dont les mots seront répartis sur un grand nombre de lettre pas sur deux ou trois lettres comme on a toujours vu dans les autres dictionnaires avec l'ordre strictement alphabétique. De plus, le classement des mots à la fois par racine et par ordre alphabétique permet de retrouver facilement un mot dans le dictionnaire, en plus du fait que les mots de la même racine sont regroupés souvent ensemble.

---

## *CONCLUSION GENERALE :*

---

Le but principal convoité par ce mémoire était de concevoir un algorithme qui a pour objectif l'extraction des racines des mots de la langue amazighe et de les classer, en se basant à la fois sur l'ordre alphabétique des racines et celui des mots qui existent dans chaque racine. Ceci peut servir à confectionner un dictionnaire facile à consulter. Cet objectif est atteint grâce aux divers aspects que nous avons intégrés.

### *Perspectives :*

Les perspectives de notre travail sont multiples. Elles concernent l'amélioration de l'application en proposant une deuxième version caractérisée par :

- Une interface graphique plus agréable pour l'utilisateur.
- Une forme d'affichage plus sophistiqué.
- Ajouter des définitions aux mots.

D'autre part, on peut apporter des modifications sur l'algorithme développé en ajoutant :

- un tokeniseur afin qu'il puisse traiter des textes entiers.
- une fonction pour extraire des racines réelles.

## Bibliographie

- [1] Aoughlis ,F. (2014). Noms et adjectifs dérivés en Tamazight, 6ème Edition de la Conférence Internationale sur Les Technologies d'Information et de Communication pour l'Amazigh, TICAM'14, IRCAM, Rabat, 24 et 25 Novembre 2014. (Actes à paraître aux éditions de l'IRCAM).
- [2] Ameer, M., Ansar, K., Bouhjar A., Boumalk A., El Azrak N., Laabdelaoui R. (2011). Vocabulaire grammatical de l'amazigh : Application phraséologique. Série:Lexiques N°7, IRCAM, Rabat, Maroc
- [3] Ataa Allah, F. et Boulaknadel, S. (2010a). Pseudo-racinisation de la langue Amazighe. In Proceeding of Traitement Automatique des Langues Naturelles. Montréal, Canada.
- [4] Ataa Allah, F. et Jaa, H. (2009). Etiquetage morphosyntaxique: Outil d'assistance dédié à la langue amazighe. Actes du 1er Symposium international sur le traitement automatique de la culture Amazighe. Morocco.
- [5] Boukhris, F. et al. (2008), La nouvelle grammaire de l'amazighe, Rabat, IRCAM
- [6] Chaker, S. (1973). Le système dérivationnel verbal berbère (dialecte kabyle). Paris, EPHE, Université René Descartes (thèse doctorat de 3ème cycle).
- [7] Chaker, S. (1983). Un parler berbère d'Algérie (Kabylie) : syntaxe. Aix-en Provence, Publications de l'Université de Provence.
- [8] Chaker, S. (1991). Manuel de linguistique berbère I. Alger, Editions Bouchène,291p.
- [9] Chaker, S. (1996).Manuel de linguistique berbère II, Syntaxe et diachronie,Alger.
- [10] COHEN D.(1968) Les langues chamito-sémitiques, Le langage, Paris, NRF Gallimard (« La Pléiade »).
- [11] DALLET, J-M. (1985), Dictionnaire Français-Kabyle,(parler des Ait Mengellat, Algérie), Paris, SELAF.
- [12] Fatima Sadiqi (1997). Grammaire du berbère, Paris : Harmattan,
- [13] Gerard Salton et M. J. McGill (1983), Introduction to Modern Information Retrieval, McGraw-Hill, Université du Michigan.

- [14] Greenberg, J. (1966). *The Languages of Africa*. Indiana University, Université de Californie.
- [15] J. Rey-Debove (1971), *Étude linguistique et sémiotique des dictionnaires français Contemporains*, De Gruyter Mouton Edition.
- [16] Lionel Delafosse (1999) *Glossaire de Linguistique Computationnelle*
- [17] M. Serhoual (2002), *Dictionnaire tarifit-français*
- [18] Nait-Zerrad, K. (1995). *Grammaire du berbère contemporain (kabyle)I, Morphologie*, Alger ; Enag Edition.
- [19] Nait-Zerrad, K. (2001). *Grammaire moderne du kabyle, tajerrumt tatrart n teqbaylit*, Editions Karthala, Paris.
- [20] RABDI, L.(2004),*Etudes linguistiques en vue de confection d'un dictionnaire de langue berbère (kabyle)*, Thèse de Doctorat de sciences du langage et traductologie, Université Paris III-Sorbonne Nouvelle
- [21] René Basset, Ernest Leroux.(1893). *Notes de lexicographie berbère*, paris.
- [22] Stephen P. HARTER.(1986). *Online Information Retrieval: Concepts, Principles and Techniques*. Orlando: Academic Press. Inc.
- [23] TAIFI (Miloud). (1990). *AWAL - Cahiers d'Etudes Berbères; L'altération des racines berbères : la diachronie dans la synchronie*.

## Résumé

Durant les dernières décennies, le Traitement Automatique des Langues (TAL) a connu des progrès considérables sur le plan de la qualité des résultats obtenus, ainsi que sur la diversité des outils utilisés. Cependant, la plupart des recherches entreprises ont été restreintes à un nombre de langues majoritairement occidentales. Toutefois, l'avènement de l'Internet a déclenché un début d'évolution de cette situation en mettant en avant la problématique du multilinguisme. Ainsi, ces vingt dernières années ont vu l'accomplissement d'un nombre assez important de travaux de recherche sur des langues tels le japonais, le chinois et l'arabe. Néanmoins, il reste encore des efforts à fournir pour un nombre non négligeable de langues peu ou pas du tout étudiées dans la communauté du TAL, en particulier le cas de l'amazigh.

Dans cette optique, l'objectif de ce mémoire est d'apporter une contribution au sujet de l'analyse morphologique de l'amazigh en proposant une application qui utilise un algorithme pour extraire la racine réelle ou formelle des mots amazighs, pour ensuite les classer en vue de la confection d'un dictionnaire.

---

Mots clés : Traitement automatique de la langue amazighe ; classification de mots amazighs ; Dictionnaire électronique ; Morphologie flexionnelle et dérivationnelle ; Analyseur morphologique.

## Abstract

Over the last few decades, language processing (LP) has seen considerable progress in terms of quality of the results obtained and the diversity of available tools. However, most studies has been restricted to a number of Western languages, and the advent of the World Wide Web has triggered an early evolution of this situation by highlighting the issue of multilingualism. Thus, the last twenty years have seen a list of research accomplishments of a large number of languages such as Japanese, Chinese and Arabic. Nevertheless, for less resourced languages, many efforts need to be made. This is particularly the case of the amazighe language

In this perspective, the objective of this thesis is to contribute to the subject of the morphological analysis of amazigh by proposing an application that uses an algorithm to extract the root of amazigh words and then to classify its words in a dictionary.

---

Key words: Amazighe language processing; classification of Amazigh words; Electronic dictionary; Inflectional and derivational morphology; Morphological analyzer.