

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université A. Mira de Béjaïa  
Faculté des Sciences Exactes  
Département d'informatique



## **Mémoire de fin d'étude**

En vue de l'obtention du diplôme de Master en Informatique  
professionnel

Option : Génie Logiciel (GL)

## *Thème*

---

*Conception et réalisation d'un traducteur  
Tamazight –Français sous Android.*

---

**Encadrés par :**

Mme. Yaici Malika.  
Mme. Aloui Soraya.

**Réalisé par :**

M<sup>elle</sup> Bessaoudi Sonia.  
M<sup>elle</sup> Chilla Habiba.

**Devant le jury composé de :**

**Présidente :** Mme. El Bouhissi Houda.  
**Examineur :** Mr. Akilal Karim.

# *Dédicace*

*La page de la dédicace est généralement celle qui est écrite en dernier et lue en premier, aussi je doute que ce mémoire puisse faire exception à la règle.*

*Nous dédions ce mémoire avec un énorme plaisir, un cœur ouvert et une immense joie :*

***A nos chers parents,***

*Que nulle dédicace ne puisse exprimer ce que nous leurs devons, pour leur bienveillance, leur affection et leur soutien que ça soit au niveau moral ou financier... Trésors de bonté, de générosité et de tendresse, en témoignage de notre profond amour et notre grande reconnaissance « Que Dieu vous garde et vous protèges ».*

***A nos chères sœurs et nos frères,***

*En témoignage de notre sincère reconnaissance pour les efforts qu'ils ont consenti pour l'accomplissement de ce projet. On leur dédie ce modeste travail en témoignage de notre grand amour et notre gratitude infinie.*

***A tous nos amis,***

*Pour leur aide et leur soutien moral leurs encouragements durant l'élaboration du travail de fin d'étude en leur souhaitant un brillant avenir.*

***A toutes nos Familles***

***A tous ceux dont l'oubli du nom n'est guère celui du cœur ...***

# *Remerciement*

*On dit souvent que le trajet est aussi important que la destination. Les cinq ans de master nous ont permis de bien comprendre la signification de cette phrase toute simple. Ce parcours, en effet, ne s'est pas réalisé sans défis et sans soulever de nombreuses questions pour lesquelles les réponses nécessitent de longues heures de travail.*

*Nous tenons à la fin de ce travail à remercier ALLAH le tout puissant de nous avoir donné la foi et de nous avoir permis d'en arriver là.*

*En premier lieu nous tenons à remercier profondément notre promoteur Mme **Yaici Malika** d'avoir acceptée de nous encadrer, Un très grand merci pour sa compréhension, sa disponibilité, sa gentillesse et son temps précieux qu'elle nous a accordé afin de pouvoir soutenir notre mémoire.*

*Nous tenons particulièrement à exprimer nos profonds remerciements à notre co-encadrante Mme. **Aloui Soraya** a su nous guider pour mener à bien nos travaux de recherche. Qu'elle trouve ici le témoignage de nos profondes gratitude à son égard et l'estime respectueuse que nous lui porte.*

*Nous souhaitons également faire part de nos reconnaissances à tous les enseignants qui nous ont éclairés la voie du savoir durant notre cursus. Un grand merci également à l'ensemble du personnel pédagogique, technique et administratif du département.*

*Nous tenons également à exprimer nos remerciements aux membres du jury, qui ont accepté d'évaluer notre travail.*

*Nous remercions infiniment Mme **El Bouhissi Houda** d'avoir accepté de présidence mémoire. Nous la remercions pour le temps qu'elle a consacré pour lire et évaluer notre travail et pour son déplacement afin de participer au jury.*

*Nous remercions également Mr. **Akilal Karim** d'avoir accepté sans hésitation de rapporter notre travail. Nous le remercions aussi d'avoir consacré du temps pour lire et examiner notre mémoire.*

*Nous remercions également Mr **Y. Mechref**, Mr **F.Haddadi** et Mr **M. Houde** pour leurs gentillesse et leurs aides précieuses pour la traduction des mots au tamazight, et pour élucider les mots ou les phrases en amazighe que nous ne comprenions pas.*

*Nous clôturons cette liste de remerciements avec une pensée très affectueuse pour nos familles, qui nous ont toujours soutenus. Merci à nos parents, nos frères, nos sœurs, pour leurs encouragements, même à distance, et leur motivation sans faille.*

*Nous exprimons notre profonde gratitude à tous celles et ceux qui nous ont apporté leurs soutiens, leur amitié ou leur expérience tout au long de ce travail*

## Table des matières

<b>Liste des tableaux</b> .....	<b>V</b>
<b>Liste des figures</b> .....	<b>VI</b>
<b>Liste des abréviations</b> .....	<b>VII</b>
<b>Introduction général</b> .....	<b>10</b>
<b>Chapitre 1 : Introduction aux ontologies</b> .....	<b>13</b>
1.1. Introduction .....	13
1.2. C'est quoi une ontologie et quelles sont ses composants !.....	13
1.2.1. Apparition d'ontologie et ses notions .....	13
1.2.2. Définitions.....	14
1.2.3. Les rôles des ontologies .....	15
1.2.4. Les constituants d'une ontologie .....	15
1.2.6. Classifications d'ontologies .....	17
1.2.7. Les approches de construction d'ontologies .....	19
1.3. Les langages et les logiciels pour servir les ontologies !.....	22
1.3.1. Présentation de XML .....	22
1.3.2. Structure d'un fichier XML.....	22
1.3.3. Présentation de RDF et de RDFS.....	23
1.3.4. Présentation du langage d'ontologie OWL.....	23
1.3.5. Des éditeurs d'ontologies.....	25
1.4. Conclusion .....	26
<b>Chapitre2 : Les caractéristiques de la langue tamazight.</b> .....	<b>28</b>
2.1. Introduction .....	28
2.2. Particularité de la langue Tamazight .....	28
2.3. Les catégories du mot .....	29
2.3.1. Le nom.....	29
2.3.2. L'adjectif.....	32
2.3.3. L'adverbe .....	33
2.3.4. Le verbe.....	33
2.4. La structure morphologique d'un mot Tamazight .....	33
2.4.1. Dérivation.....	34
2.4.2. Composition .....	34
2.4.3. Expressivité .....	35
2.5. Conclusion .....	35

<b>Chapitre 3 : La traduction automatique et ses approches.....</b>	<b>37</b>
3.1. Introduction .....	37
3.2. Un bref historique .....	37
3.3. Les différentes approches .....	38
3.3.1. L’approche experte.....	38
3.3.2. L’approche empirique .....	42
3.3.3. L’approche statistique .....	45
3.4. Critères de qualité des traductions .....	47
3.4.1. Évaluation manuelle des traductions.....	47
3.4.2. Évaluation automatique des traductions.....	48
3.5. Conclusion .....	48
<b>Chapitre 4 : La réalisation d’un traducteur .....</b>	<b>50</b>
4.1. Introduction .....	50
4.2.La démarche suivie .....	50
4.2.1. Diagramme de cas d’utilisation.....	51
4.2.2. Diagramme d’interaction.....	51
4.2.3. Description textuelle .....	53
4.3. Outils de développement .....	53
4.3.1. Androïd Studio .....	54
4.3.2. Jena (framework).....	54
4.3.3. AndroJena.....	54
4.3.4. Logiciel protégé (voir chapitre 01).....	55
4.4. Langages de développement d’applications Androïd.....	55
4.4.1. XML .....	55
4.4.2. Les types d’applications Androïd :.....	55
4.5. Aspects conceptuels de l’élaboration d’une ontologie sous protégé .....	55
4.5.1. Constitution du corpus .....	55
4.5.2. Implémentation.....	56
4.5.3. Éléments de l’ontologie.....	56
4.5.4. Intégration de l’ontologie dans Android studio.....	60
4.5.5. Développement de l’application sous Android studio .....	60
4.6. Conclusion .....	67
<b>Conclusion et perspectives.....</b>	<b>69</b>
<b>Bibliographie.....</b>	<b>73</b>

## Liste des tableaux

Tableau 1 : La comparaison entre les trois sous-langages d'OWL. ....	24
Tableau 2 : Les lettres de la langue amazighe. ....	29
Tableau 3 : Les parties de mot "argaz". ....	29
Tableau 4 : Quelques exemples de mots convertis du masculin au féminin. ....	30
Tableau 5 : Quelques exemples de noms des nombres convertis du masculin au féminin. ....	30
Tableau 6 : Exemples du pluriel externe. ....	31
Tableau 7 : Exemples de pluriel interne. ....	31
Tableau 8 : Exemples de pluriel mixte. ....	31
Tableau 9 : États de nom et leurs catégories. ....	32
Tableau 10 : Un exemple illustrant les variations de l'adjectif. ....	32
Tableau 11 : Déterminants invariables. ....	33
Tableau 12 : Formes figées, composées du nom et/ou d'éléments grammaticaux. ....	33
Tableau 13 : Un exemple illustrant les parties de verbes " ččiy " ....	33
Tableau 14 : Exemples de dérivation verbale. ....	34
Tableau 15 : Exemples de dérivation nominale. ....	34
Tableau 16 : la juxtaposition et le syntème de mot. ....	34
Tableau 17 : les opérations qui rendent le mot expressif. ....	35
Tableau 18 : Avantages et inconvénients de l'approche experte. ....	42
Tableau 19 : Exemple illustrant l'approche par analogie ou basée sur l'exemple. ....	44
Tableau 20 : Les forces et les limites des approches statistiques. ....	46
Tableau 21 : La description textuelle. ....	53

## Liste des figures

Figure 1 : Processus de développement et le cycle de vie de METHONTOLOGY .....	20
Figure 2 : Le triangle de Vauquois .....	38
Figure 3 : Architecture du système de TA par transfert .....	40
Figure 4 : Un exemple de TA avec le système de transfert.....	40
Figure 5 : Architecture du système de traduction automatique par langue pivot .....	42
Figure 6 : Architecture d'un système statistique de traduction automatique .....	46
Figure 7 : Diagramme de cas d'utilisation. ....	51
Figure 8: diagramme d'interaction. ....	52
Figure 9: La relation d'hierarchie sous protégè.....	57
Figure 10 : La relation de disjonction sous protégè. ....	58
Figure 11: la relation rdfs:comment sous protégè.....	58
Figure 12: La relation d'équivalence sous protégè.....	59
Figure 13: Statistiques de notre application donner par protégè .....	59
Figure 14: Les bibliothèques Androjena.....	60
Figure 15: Exemple de la traduction d'un nom féminin déduit à partir de son masculin. ....	61
Figure 16: Exemple de la traduction d'un nom en pluriel externe déduit à partir de son singulier. ....	62
Figure 17:Exemple de la traduction d'un nom en pluriel interne déduit à partir de son singulier. ....	62
Figure 18:Exemple de la traduction d'un nom en état d'annexion déduit à partir d'un état libre. ....	63
Figure 19: La forme verbale en Tamazight .....	63
Figure 20:Exemple de la traduction d'un verbe à l'infinitif et, sa conjugaison à l'impératif. ....	64
Figure 21:Exemple de la traduction d'un verbe conjuguais au future. ....	64
Figure 22: Exemple de la traduction d'un verbe conjuguais au passé composé.....	65
Figure 23Exemple de la traduction de quelques phrases simples. ....	65
Figure 24:Exemple de la traduction d'un nom avec un pronom affixe.....	66
Figure 25: Exemple de la traduction d'une préposition liée un pronom affixe.....	67

## Liste des abréviations

A-score	Adequacy score
DAML	Digital Asset Modeling Language
DTD	Document Type Definition
F-score	Fluency score
GUI	Graphical User Interface
I-score	Informativeness score
IA	Intelligence Artificielle
IBM	International Business Machines
IC	Ingénierie des Connaissances
KIF	Knowledge Interchange Format
LIPN	Laboratoire d'Informatique de Paris Nord
LOOM	Large Object-Oriented Memory.
OCML	Operational Conceptual Modelling Language
ODE	Ontology Design Environment
OIL	Ontology Interchange Language
OilEd	Ontology Interchange Language Editor
OWL DL	Ontology Web Language Description Logics
OWL	Ontology Web Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schéma
RTO	Ressource Termino-Ontologique
SMI	Sports Medicine Institute
TA	Traduction Automatique

TAL	Traitement Automatique du Langage
TIA	Terminologie et Intelligence Artificielle
TOVE	Toronto Virtual Enterprise
URI	Universal Resource Identifiers
XML	eXtensible Markup Language
W3C	World Wide Web Consortium
WebOde	web Ontology Design Environment
WebOnt	Web Ontology

# Introduction générale

## **Introduction général**

Selon le mythe de la tour de Babel, l'existence de plusieurs langues sur notre planète résulte d'une punition que Dieu affligea aux hommes pour les punir de leur orgueil. Nemrod, le "roi-chasseur" régnant sur les descendants de Noé, eut l'idée de construire à Babel (Babylone) une tour assez haute pour que son sommet atteigne le ciel (le trône de Dieu). Mais Dieu fit échec à cette entreprise en introduisant la "confusion" (la diversité) des langues. Ainsi les hommes ne parvinrent plus à communiquer, la construction de la tour fût abandonnée et les hommes se dispersèrent partout sur la Terre.

Bien que ce ne soit qu'un mythe, l'existence de plusieurs langues est un fait, et crée un véritable obstacle entre les êtres humains désirant communiquer. Environ 7000 langues sont recensées sur notre planète et tout le monde peut un jour se heurter à cette barrière. Le seul moyen de franchir cet obstacle est le recours à la traduction. Les besoins en traduction ont donc considérablement augmentés depuis ces dernières décennies et la nécessité de pouvoir traduire automatiquement, et à tout moment, toutes sortes d'informations est devenue indéniable.

La traduction automatiquement connaît un essor considérable sur le Web, avec plusieurs systèmes en plusieurs langues. Mais nous avons remarqué une injustice envers notre langue maternelle, qui est tamazight où les moyens pour la traduction vers d'autres langues sont rares. C'est ce qui nous a fait penser à créer un traducteur Tamazight-français.

Ce qui nous a le plus incité à créer ce traducteur est le mouvement de contestation initié le 4 décembre 2017 aux facultés Aboudaou et Targa Ouzemmour par les étudiants en langue tamazight de l'université Abderrahmane Mira de Béjaïa, pour assurer une formation de qualité pour les étudiants de la langue Amazigh, ainsi qu'à l'élargissement de son champ de recherche en dotant cette langue d'instituts et d'institutions qui travaillent pour sa promotion, mais aussi sa généralisation immédiate.

Pour toutes ces raisons, l'objectif principal de ce travail est de contribuer à la construction d'un traducteur unidirectionnel Tamazight-Français basé sur une ontologie.

Pour atteindre notre objectif, nous dresserons un plan simple et cohérent à nos yeux, composé de quatre chapitres :

Le premier chapitre est consacré aux ontologies, les classifications existantes, les stratégies de construction, les formalismes et les langages utilisés, et est conclu par la présentation de quelques éditeurs d'ontologie.

Dans le deuxième chapitre, nous présentons les concepts de la langue tamazight et ses principales caractéristiques, où nous allons voir toutes les notions nécessaires et basiques pour connaître cette langue.

Dans le troisième chapitre, nous présenterons d'abord la traduction automatique, ainsi que ses deux approches (empirique et experte). Nous donnerons ensuite les critères de qualité des traductions.

Le quatrième chapitre servira à la conception et à la réalisation de notre application où nous allons présenter quelques diagrammes UML et notre ontologie des mots tamazight ainsi que les outils et les langages de programmation utilisés. Nous détaillons, par la suite, le fonctionnement de notre application à travers quelques captures d'écran.

Nous terminons ce manuscrit par une conclusion sur nos travaux ainsi que par des perspectives de recherche à court et à long terme.

# **Chapitre 1 :**

## **Introduction aux ontologies.**

## **Chapitre 1 : Introduction aux ontologies.**

### **1.1. Introduction**

Les ontologies ont aujourd'hui une place majeure dans la représentation et la modélisation des connaissances dans un domaine donné. Elles sont utilisées pour formaliser les connaissances d'un domaine et ainsi ajouter une couche sémantique aux systèmes et aux applications informatiques afin de pouvoir les exploiter par des machines et les munir de sémantique.

Les ontologies permettent de représenter de manière explicite les connaissances d'un domaine au moyen d'un langage formel afin qu'elles puissent être manipulées automatiquement et partagées aisément. Leur utilisation est largement répandue et ce dans divers domaines de recherche (l'ingénierie des connaissances, la recherche d'information, le traitement du langage naturel, les systèmes d'information coopératifs, l'intégration intelligente d'information, et la gestion des connaissances).

Dans ce chapitre, nous allons présenter les différentes notions liées aux ontologies.

- Dans la première partie : nous commençons d'abord par donner quelques définitions courantes des ontologies avant de décrire leurs différents composants et leur typologie.
- Dans la deuxième partie : Nous présentons les différents langages de représentation et de manipulation des ontologies.

Enfin, nous terminons par une conclusion.

### **1.2. C'est quoi une ontologie et quelles sont ses composants**

#### **1.2.1. Apparition des ontologies et de leur notions**

Depuis plus de 2300 ans, des philosophes grecs ont utilisé le terme « Ontologie », ce mot est composé de « **ontos** » qui signifie l'être, l'existant et « **logos** » qui veut dire l'étude, le discours, univers, d'où sa traduction par « l'étude de l'être » [1].

Ensuite, Les ontologies sont apparues dans des contextes différents touchant à la philosophie, où le terme a été initialement introduit à l'intelligence artificielle, et à l'ingénierie des connaissances (IC) en particulier.

Dans le premier usage, l'ontologie est un terme qui implique une branche de la philosophie qui traite la nature et l'organisation de la réalité. Elle est la science de l'être dans laquelle on cherche à représenter le monde [1].

Dans le domaine de l'intelligence artificielle (IA), une "ontologie" n'est pas une science, mais un objet, le fruit d'une tentative de formulation exhaustive et rigoureuse de la conceptualisation d'un domaine. Neches et ses collègues [2] furent les premiers à proposer une définition à savoir :

*« Une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire »* [3].

Dans le domaine IC, la notion d'ontologie a été introduite dans les années 1990 avec les travaux de Tom Grüber pour la résolution automatique des problèmes [3].

### 1.2.2. Définitions

Malgré l'existence de plusieurs définitions pour les ontologies, celle qui reste une référence et certainement la plus populaire, parmi celles qui existent dans la littérature est celle de Tom Grüber [2] *« une ontologie est une spécification explicite d'une conceptualisation »*. Une **conceptualisation** est une vue abstraite des entités réelles du monde que l'on veut représenter et de leurs relations ; c'est un modèle abstrait des phénomènes réels du monde. L'expression **spécification explicite** signifie que la conceptualisation est faite de façon non ambiguë dans un langage concret ; les concepts et les contraintes utilisés pour représenter le domaine de connaissances doivent être explicitement définis [3].

Cette définition a été légèrement modifiée par Borst [3] comme *« spécification formelle d'une conceptualisation partagée »*. Ces deux dernières définitions sont regroupées dans celle de Studer et ses collègues [4]. Ils définissent une ontologie comme *« une spécification formelle et explicite d'une conceptualisation partagée »*. Le terme **formel** fait référence au fait qu'une ontologie doit être compréhensible par les machines. **Partagée** indique que l'ontologie supporte la connaissance consensuelle partagée, et elle n'est pas restreinte à certains individus mais acceptée par une communauté d'utilisateurs afin d'en garantir une utilisation plus large [3].

Selon Uschold, [5] *une ontologie peut prendre différentes formes, mais elle inclura nécessairement un vocabulaire de termes et une spécification de leur signification. Cette dernière inclut des définitions et une indication de la façon dont les concepts sont reliés entre eux, les liens imposant collectivement une structure sur le domaine et contraignant les*

*interprétations possibles des termes*. Une ontologie est vue comme un ensemble de concepts et de propriétés caractérisant les objets d'un domaine du monde réel. Ces concepts sont organisés sous la forme d'un graphe et structurés par des relations hiérarchiques (ou taxonomiques) et associatives (ou transversales). En plus des concepts et relations, elle définit des **règles** ou **contraintes** sur leurs définitions et interprétations permettant d'effectuer des déductions [3].

Roche a résumé ces différentes définitions : *définie pour un objectif donné et un domaine particulier, une ontologie est pour l'ingénierie des connaissances une représentation d'une modélisation d'un domaine partagé par une communauté d'acteurs. Objet informatique défini à l'aide d'un formalisme de représentation, elle se compose principalement d'un ensemble de concepts définis en compréhension, de relations, et de propriétés logiques* [6].

### 1.2.3. Les rôles des ontologies

Les ontologies sont utilisées pour résoudre deux grands problèmes qui sont l'interopérabilité et la réutilisation [7] :

- **L'interopérabilité** : Les ontologies présentent une solution face aux ambiguïtés des langages utilisés, car elle assure la cohérence en éliminant la confusion terminologique concernant les termes utilisés.
- **La réutilisation** : ce dernier utilise le moyen de réutilisation des concepts et des méthodes déjà établis.

### 1.2.4. Les constituants d'une ontologie

D'après les définitions précédentes, les ontologies fournissent un vocabulaire commun d'un domaine et définissent la signification des termes et des relations entre eux. La connaissance dans les ontologies est principalement formalisée en utilisant les cinq types de composants ,à savoir [3] :

- **Un concept** (termes ou classe au sens des standards du Web Sémantique) représente un ensemble d'objets et leurs propriétés communes ; il peut représenter un objet matériel, un événement, une notion profonde , ou une idée [1].

Un concept a généralement une **intention** et/ou une **extension** [1].

- Son intension : désigne la sémantique, c'est-à-dire l'ensemble des attributs et propriétés définissant un concept.
- Son extension : l'ensemble des **objets** qu'englobe un concept

Ces concepts peuvent être classifiés en plusieurs niveaux [1] :

- Niveau d'abstraction (concret comme livre ou abstrait comme Piété)
  - Niveau d'atomicité (élémentaire ou composée).
  - Niveau de réalité (réel ou fictif).
- **La relation ou la propriété** est un lien (binaire, tertiaire, etc.) entre des entités, et qui peut généralement être doté d'une valeur [3].

Les relations sur les concepts sont [8] :

- **La subsomption** : dite taxonomique, elle lie un élément supérieur (le concept plus général) et un élément inférieur (le concept spécifique). La relation de subsomption « est-un » (is-a) a un statut particulier car elle structure la hiérarchie ontologique. À ce titre, elle est implicite. Par exemple Homme « **est-un** » (**is-a**) Humain [3].
  - **L'équivalence** (synonymie) : pour exprimer que deux concepts représentent la même chose, deux concepts sont équivalents s'ils ont la même extension. Par exemple, Homme et humain.
  - **La disjonction** (ou bien incompatibilité) : deux concepts sont disjoints si leurs extensions sont disjointes. Par exemple, Homme et Femme.
  - **Les relations de méronymie** (relation composé-composant de type « *partie de* ») C'est une relation qui lie un couple de concepts dont l'un est une partie de l'autre. Dans une relation de méronymie, les propriétés du tout ne sont pas obligatoirement transmises à ses parties.
  - **Associatives/transversales** sont des relations hiérarchiques et partitives sur les concepts.
- **Individus / Instances** l'individu est une instance de concept. Les individus sont considérés comme des composants d'une ontologie.
  - **Axiomes/Fonctions**
    - **Axiome** : Ce sont des contraintes ou des règles sur des concepts ou des instances de l'ontologie considérées comme toujours vraies [8].

- **Fonction** : C'est un héritage de la relation, qui définit le nième élément de la relation en fonction de n-1 éléments précédents [8].

### 1.2.5. Classifications d'ontologies

Différentes typologies ont été proposées pour distinguer les multiples ontologies existantes, on peut distinguer quatre typologies selon [7] :

- Le niveau de complétude.
- L'objet de conceptualisation.
- Le niveau du formalisme.
- Le niveau de granularité.

#### 1.2.5.1. Classification selon niveau de complétude

On peut définir trois niveaux [7] :

- **Niveau sémantique** : Tous les concepts, caractérisés par un terme/libellé, doivent respecter les quatre principes différentiels :
  - Communauté avec l'ancêtre.
  - Différence, spécification, par rapport à l'ancêtre.
  - Communauté avec les concepts frères, situés au même niveau.
  - Différence par rapport aux concepts frères.
- **Niveau-Référentiel** : Les concepts référentiels ou formels, se caractérisent par un terme/libellé dont la sémantique est définie par une extension d'objets.
- **Niveau-Opérationnel** : Les concepts du niveau opérationnel ou computationnel sont caractérisés par les opérations qu'il est possible de leur appliquer pour générer des interfaces ou engagement computationnel.

#### 1.2.5.2. Classification selon l'objet de conceptualisation

En fonction de leur portée et des objectifs visés, différents types d'ontologies sont développés, et qui sont [7] :

- **Les ontologies de type taxonomie** : des concepts structurés par des relations de subsomption.

- **Les ontologies de type thésaurus** : c'est un réseau de termes organisé avec plusieurs types de relations.
- **Les ontologies du domaine** : Les ontologies de domaine est le type d'ontologies le plus courant en ingénierie ontologique. Elles sont définies de deux manières :
  - Des ontologies spécifiques à un domaine particulier.
  - Des ontologies de tâches décrivent un vocabulaire en relation avec une tâche ou une activité d'un domaine. Généralement, ces ontologies ne sont pas réutilisables et possèdent une portée limitée.
- **Les ontologies applicatives** : Ces ontologies contiennent des connaissances du domaine nécessaire à une application donnée, elles sont spécifiques et non réutilisable.
- **Les ontologies de haut niveau (top-ontologies ou upper-level ontologies)** : ces ontologies expriment l'étude des catégories des choses qui existent dans le monde, tels que les entités, les événements, les états, les processus, les actions, le temps, l'espace, les relations et les propriétés.
- **Les ontologies génériques (méta-ontologies ou core-ontologies)** : décrivent des concepts génériques valables dans différents domaines mais moins abstraits que ceux décrits dans les ontologies de haut niveau.
- **Les ontologies de représentation** : Ce type est un cas particulier d'ontologies supérieures (**haut niveau**), elles fournissent les primitives nécessaires pour décrire les concepts des autres types d'ontologies.
- **Les ontologies géographiques** : Ces ontologies sont typiquement élaborées par de grands organismes de normalisation. Elles décrivent les concepts qui caractérisent l'espace comme le point, la ligne (Les ontologies spatialisées ou spatio-temporelles). Une composante temporelle est souvent nécessaire en complément pour la modélisation de l'information géographique.

### 1.2.5.3. Classification selon leur granularité (le niveau de détail)

Selon le niveau de description utilisé, on peut distinguer deux types d'ontologies [7] :

- **Granularité fine** : Elle permet de décrire un vocabulaire plus riche et plus détaillé des concepts pertinents d'un domaine ou d'une tâche.

- **Granularité large** : Elle permet de décrire des vocabulaires moins détaillés, elle est utilisée souvent pour représenter les ontologies de haut niveau (exprimées en termes de scénarios, événements, temps et objets) car les concepts qu'elles traduisent sont normalement raffinés ultérieurement dans d'autres ontologies de domaine ou d'application.

#### 1.2.5.4. Classification selon le niveau de formalisme utilisé

On peut distinguer quatre types d'ontologies selon le formalisme utilisé [7]:

- **Les ontologies Informelles** : exprimées en langage naturel.
- **Les ontologies Semi-informelles** : exprimées dans une forme restreinte et structurée de la langue naturelle (en utilisant des modèles).
- **Les ontologies Semi-formelles** : exprimées dans un langage artificiel défini formellement.
- **Les ontologies Formelles** : exprimées dans un langage artificiel possédant une sémantique formelle. L'intérêt d'une ontologie formelle est la possibilité d'effectuer des vérifications sur l'ontologie : complétude, non redondance, consistance, cohérence, etc.

#### 1.2.6. Les approches de construction d'ontologies

La méthodologie varie en fonction de l'usage qu'on veut faire de l'ontologie mais aussi des ressources disponibles à utiliser pour la construction. Les domaines d'applications des ontologies étant variés, les approches restent différentes et peuvent être manuelles ou semi automatiques. On peut distinguer quatre grandes catégories d'approches [7]:

##### 1.2.6.1. Les approches de construction d'ontologies à partir de zéro

Nous présentons dans cette sous-section quelques approches considérées comme parmi les plus représentatives

- **L'approche proposée par Uschold et King [9]**

C'est une approche basée sur l'expérience acquise lors du développement de l'ontologie avec un processus de construction qui comprend quatre phases :

- La phase d'identification du but pour l'ontologie.

- La phase de développement.
- La phase d'évaluation et d'analyser d'ontologie résultante
- La phase de documentation.

Mais cette approche n'est pas meilleure car il n'y a pas d'indication sur la manière de déterminer les concepts clefs de l'ontologie.

- **L'approche TOVE (Toronto Virtual Enterprise)**

Comme son nom l'indique elle est utilisée dans le cadre du projet TOVE [7] dans le domaine de l'entreprise. Elle vise à construire les ontologies en respectant les étapes suivantes [7] :

- Identification des problèmes (scénarii).
- Formulation de questions informelles (basées sur les scénarii) auxquelles l'ontologie doit permettre de répondre.
- Spécification d'une terminologie à partir des termes apparaissant dans les questions.
- Spécification formelle des axiomes et des définitions pour les termes de la terminologie.
- Evaluation de la complétude de l'ontologie.

Cette méthodologie reste limitée car ni les différentes étapes ni les techniques utilisées ne sont décrites précisément.

- **La méthode METHONTOLOGY [10]**

C'est une méthodologie mise au point par l'équipe du laboratoire de l'intelligence artificielle de l'Université polytechnique de Madrid [2]. Cette méthode inclut :

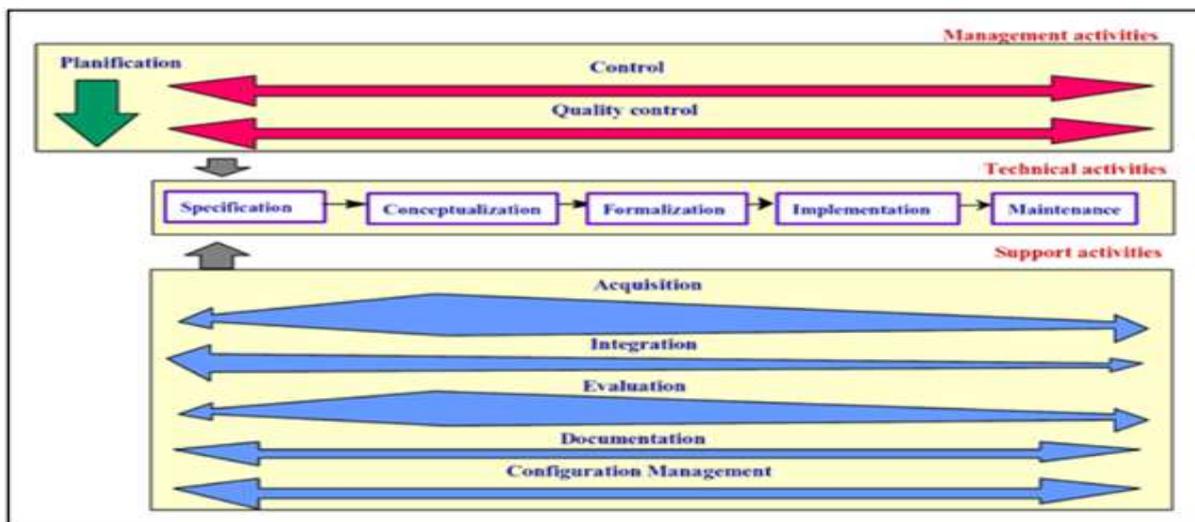


Figure 1 : Processus de développement et le cycle de vie de METHONTOLOGY [7].

- L'identification du processus de développement de l'ontologie.
- Le cycle de vie basé sur l'évolution de prototypes.
- Les techniques de gestion de projet (planification, assurance qualité), de développement (spécification, conceptualisation, formalisation, implémentation, maintenance) et des activités de support (intégration, évaluation, documentation)

### **1.2.6.2. Les approches de construction d'ontologies à partir de textes**

Les méthodes et les outils d'extraction de connaissances à partir de texte sont devenues robustes avec les avancées en TAL. L'analyse et le traitement des documents textuels deviennent ainsi plus simples. Ces approches combinent souvent des techniques linguistiques, statistiques et/ou d'apprentissage automatique pour extraire des connaissances ontologiques à partir de textes. Le développement des supports électroniques a aussi rendu ces données textuelles plus accessibles et ont ainsi facilité leur exploitation par des outils de TAL [7].

Nous présentons dans ce qui suit quelques méthodologies considérées comme représentatives de cette catégorie dans la littérature.

- **TERMINAE**

C'est un outil de construction semi-automatique d'ontologies à partir de textes développés par un groupe de recherche *Terminologie et Intelligence Artificielle* (TIA) au LIPN [8]. Son processus comprend 5 phases :

- La description des besoins.
- La constitution du corpus.
- L'étude linguistique.
- La normalisation ou conceptualisation.
- La dernière étape est celle de la formalisation.

Une des limites principales qu'on peut noter pour ces méthodes est qu'elles peuvent fonctionner pour des tâches de construction où un corpus textuel suffisant est disponible pour faire fonctionner les outils de TAL.

### **1.2.6.3. Les approches basées sur la réutilisation de RTO existantes**

De nouvelles approches s'intéressent à l'exploitation de l'ensemble des sources pour l'enrichissement d'ontologies existantes (modèle formel) ou le développement de nouvelles

ontologies. Pour cela Farquhar et ses collègues [11] ont proposé un ensemble d'outils partagés par des groupes distants sur un serveur dédié comme *Ontolingua* [8].

#### 1.2.6.4. Les approches basées sur le « crowd-sourcing »

Le « crowd-sourcing » consiste à externaliser des tâches traditionnellement effectuées par un agent désigné (comme un employé ou un entrepreneur) en faisant appel à l'intelligence et au savoir-faire d'un grand nombre de personnes [8].

Il existe une multitude de méthodes d'ingénierie ontologique mais l'absence d'une démarche commune de construction d'ontologies reste un problème ouvert qui ralentie le développement d'ontologie ,à l'intérieur et entre les équipes [8].

### 1.3. Les langages et les logiciels pour servir les ontologies

On présentera donc, dans une première partie, les langages XML et RDF et OWL. Dans la deuxième partie, nous proposons quelques exemples de logiciels en lien direct avec OWL.

#### 1.3.1. Présentation de XML

XML (eXtensible Markup Language) : est un métalangage permettant de représenter un document texte de manière arborescente en utilisant un système de balisage. Il est recommandé par le W3C depuis le 10 février 1998. Il facilite l'échange, le partage et la publication des données à travers le web.

#### 1.3.2. Structure d'un fichier XML

Un document XML est composé de deux structures : logique et physique.

La représenté physique est sous forme d'un fichier texte structuré en 4 parties :

- 1) **Prologue** : une déclaration qui identifie le document comme un document XML
- 2) **Un élément racine** : cet élément doit contenir tous les autres éléments du document et ne peut apparaître qu'une seule fois dans un document XML.
- 3) **Une déclaration de type du document** : un document XML bien formé est dit « valide » lorsqu'il est conforme à la déclaration de type de document (DTD<sup>1</sup>, XML Schema, etc.) qui l'accompagne.
- 4) Et enfin, la dernière partie d'un fichier XML est **l'arbre des éléments**

---

<sup>1</sup> DTD (document type définition) c'est un document écrit en XML (des balises) pour définir précisément la grammaire et les contraintes que doit respecter un document pour être valide.

### 1.3.3. Présentation de RDF et de RDFS

RDF (Resource Description Framework) c'est un modèle de données qui permet de représenter un ensemble de ressources reliées par des liens étiquetés. Sa structure fondamentale est un multi-graphe orienté étiqueté composé de triplet : le nœud de départ (Sujet) prédicat (propriété), nœud cible (Objet) [12].

- Ressource (Sujet) : une entité d'informations pouvant être référencée par un identificateur. Cet identificateur doit être une URI (Universal Resource Identifiers).
- Propriété (prédicat) : un aspect spécifique, un attribut, une caractéristique ou relation utilisée pour décrire une ressource...
- Valeur (Objet) : un littéral (simple chaîne de caractères) ou une ressource.

RDF Schéma est un langage recommandé par le W3C qui étend RDF avec un vocabulaire de termes et les relations entre ces termes. Il permet la définition :

- Des classes et des propriétés (rdfs:Class, rdfs:Property).
- Les sous-classes, les superclasses, les sous-propriétés, et les super-propriétés (rdfs:subClassOf, rdfs:subPropertyOf).
- Le domaine de définition et le domaine image des propriétés (rdfs:domain, rdfs:range).

### 1.3.4. Présentation du langage d'ontologie OWL

L'impossibilité de raisonner et de mener des raisonnements automatisés sur les modèles de connaissances établis à l'aide de RDF/RDFS a mené le groupe de travail sur le Web Sémantique du W3C à proposer le langage OWL, comme une extension du vocabulaire de RDF pour pallier aux insuffisances de ce dernier [8].

- **OWL « Ontology Web Language »** : langage employé pour représenter explicitement les sens des termes des vocabulaires et les relations entre ces termes. OWL vise également à **faciliter le raisonnement d'agents logiciels** qui accèdent à des **ressources** sur le **Web**, d'une part en les structurant d'une façon compréhensible et standardisée, et d'autre part en leur ajoutant des méta-informations.
- **Le rôle d'OWL**
  - OWL offre aux machines une plus grande capacité d'interprétation du contenu web que RDF et RDFS

- Il permet de définir des classes de manière plus complexe en utilisant les opérateurs ensemblistes (et / ou logique)
- Il permet de définir certaines caractéristiques des propriétés : transitivité ; unicité, propriété inverse. Il définit la restriction de cardinalité sur les propriétés
- Il tient compte de l'aspect diffus des sources de connaissances et permet à l'information d'être recueillie à partir de sources distribuées, notamment en permettant la mise en relation **de plusieurs ontologies** et l'importation des informations provenant explicitement d'autres ontologies.

Les ontologies OWL se présentent, généralement, sous forme de fichiers texte et de documents OWL.

- **Types de langages OWL**

Le groupe de travail WebOnt du W3C a doté OWL de trois sous-langages de plus en plus expressifs et c'est selon les besoins, qu'on choisit le langage qui convient [12] :

*Tableau 1 : La comparaison entre les trois sous-langages d'OWL.*

<b>OWL Lite</b>	<b>OWL DL</b>	<b>OWL Full</b>
<ul style="list-style-type: none"> <li>- Le plus simple et facile à programmer.</li> <li>- Il est adapté aux migrations rapides.</li> <li>- Il a des raisonnements complets et rapides.</li> <li>- Il a un sous-ensemble réduit des constructeurs disponibles.</li> <li>- Peut être vu comme une extension d'une vue restreinte de RDF.</li> </ul>	<ul style="list-style-type: none"> <li>- Utilise la logique de description DL.</li> <li>- Plus complexe et plus lents qu'OWL Lite.</li> <li>- Permettant une expressivité élevée.</li> <li>- Garantit la complétude du calcul et leur décidabilité.</li> <li>- Il introduit des restrictions comme la séparation des types.</li> <li>- Peut être vu comme une extension d'une vue restreinte de RDF</li> </ul>	<ul style="list-style-type: none"> <li>- La version la plus complexe et la plus lente d'OWL</li> <li>- Permettant une expressivité maximale.</li> <li>- Il a un haut niveau de capacité de description.</li> <li>- Il ne garantit pas la complétude et la décidabilité des calculs liés à l'ontologie.</li> <li>- Des mécanismes permettent à une ontologie d'augmenter la signification du vocabulaire prédéfini (RDF ou OWL).</li> <li>- Peut être vus comme une extension d'une vue restreinte de RDF.</li> </ul>

### 1.3.5. Des éditeurs d'ontologies

De nombreux outils permettent aujourd'hui d'éditer des ontologies. Parmi ceux-ci, quelques-uns essaient de guider leur utilisateur dans l'élaboration de l'ontologie en suivant une méthodologie de conception plus ou moins complète.

On distingue deux familles d'outils [13] :

- Les outils de construction d'ontologie dépendants de formalisme de représentation.
- Les outils de construction d'ontologie indépendants de formalisme de représentation.

#### 1.3.5.1. Les outils dépendants de formalisme de représentation :

Permettent de spécifier les ontologies au niveau symbolique auquel le créateur et l'utilisateur de l'ontologie doivent se plier. On trouve les plus anciens historiquement [14].

- **Ontolingua :**

Un serveur d'édition d'ontologies qui utilise le langage de représentation de connaissances KIF.

- **OntoSaurus :**

OntoSaurus est composé de deux serveurs :

- Un serveur utilisant LOOM.
- Un serveur de navigation.

- **WebOnto :**

Une application Web pour naviguer et développer les ontologies. Il est implémenté avec le langage OCML.

- **OilEd (Oil Editor) :**

Un éditeur d'ontologies utilisant le formalisme OIL développé sous la responsabilité de l'université de Manchester.

#### 1.3.5.2. Les outils indépendants du formalisme de représentation

Ces outils prennent mieux en compte l'importance du niveau des connaissances : ils proposent à leur utilisateur de créer l'ontologie de manière relativement indépendante de tout langage implémenté et prennent ensuite automatiquement en charge l'opérationnalisation de l'ontologie, en la transposant dans divers langages. Cette évolution tend à rapprocher les ontologies de leur but original

- **Protégé 2000 [3] :**

Un éditeur d'ontologies gratuit (JAVA), développé et distribué en open source par l'université en informatique médicale de Stanford (SMI). Il fournit une suite d'outils pour construire des bases de connaissance et des ontologies pour un domaine donné. Le modèle de connaissances sous-jacent **Protégé 2000** est issu du modèle des frames et consiste en une hiérarchie de *classes* qui ont des attributs (*slots*), qui peuvent eux-mêmes avoir certaines propriétés (*facets*). L'édition des listes de ces trois types d'objets se fait par l'intermédiaire de l'interface graphique (GUI) lui permettant de manipuler aisément tous les éléments d'une ontologie, sans avoir besoin d'exprimer ce que l'on a à spécifier dans un langage formel .

Il suffit juste de remplir les différents *formulaire*s correspondant à ce que l'on veut spécifier. Ce modèle autorise d'ailleurs une liberté de conception assez importante puisque le contenu des formulaires à remplir peut-être modifié suivant les besoins *via* un système de *métaclasses*, qui constituent des sortes de « patrons » de connaissance.

**Protégé 2000** est aussi une plate-forme extensible, grâce au système de plug-ins, qui permet de gérer des contenus multimédias, interroger, évaluer, et fusionner des ontologies, etc. Protégé 2000 permet aussi de créer ou d'importer des ontologies écrites dans les différents langages d'ontologies tel que : RDF-Schéma, OWL, DAML, OIL, etc. Cela est rendu possible grâce à l'utilisation de nombreux plug-ins qui sont disponibles en téléchargement pour la plupart de ces langages

**Protégé 2000** regroupe une communauté d'utilisateurs assez importante et constitue une référence pour beaucoup d'autres outils.

- **ODE et WebOde :**

C'est une plateforme de conception d'ontologies fonctionnant en ligne, l'outil fait suite à ODE (Ontology Design Environment) qui Permet de construire des ontologies au niveau connaissance, comme le préconise la méthodologie METHONTOLOGY.

- **OntoEdit (Ontology Editor) :**

Un environnement équipé des outils graphiques pour la visualisation et la construction d'ontologies indépendamment de tout formalisme.

## 1.4. Conclusion

Dans ce chapitre, nous avons abordé les différentes notions basiques qui sont très nécessaires pour connaître ce qu'est une ontologie et ses différents composants et ses types.

**Chapitre 2 :**  
Les caractéristiques de  
la langue tamazight

## **Chapitre2 : Les caractéristiques de la langue tamazight.**

### **2.1. Introduction**

Le berbère ou tamazight<sup>2</sup>, langue commune, réalisée de différentes façons selon les régions, en Afrique du Nord et au Sahel, existe bien en tant qu'unité structurale profonde (grammaire et fonds lexical communs) mais il n'est pas une langue véhiculaire (ou un outil de communication) et encore moins un standard. Il ne s'exprime et ne se pratique qu'à travers les dialectes qui le composent (chaoui, chleuh, kabyle, rifain, touareg, etc.).

Le berbère est généralement présenté comme un ensemble de dialectes (ou variétés régionales) distincts, à frontières plus ou moins hermétiques, qui sont, le plus souvent, étudiés de façon clivée voire exclusive. Les pays où le berbère est reconnu comme langue nationale sont : (Mali et Niger, en 1966 ; Algérie en 2001) ou officielle (Maroc, en 2011 ; Algérie, en 2016).

En effet, toute langue doit avoir ou bien choisir ses propres alphabets qui vont être uniques. Dans le cas du berbère, l'utilisation de graphies<sup>3</sup> différentes (arabe, latine, tifinagh) constitue aussi un obstacle à la normalisation et à la diffusion de cette langue [15].

### **2.2. Particularité de la langue Tamazight**

Actuellement, on utilise trois alphabets (arabe, latin, tifinagh) pour transcrire le berbère mais l'usage se fait dans des proportions différentes selon les pays ou même les régions. En Algérie, on emploie concomitamment les trois avec une nette prédominance du « latin » d'où dans le cadre de notre projet on va se focaliser sur l'alphabet latin [15].

L'alphabet du tamazight contient 23 lettres latines standards et 11 lettres supplémentaires. Sa forme a été fixée par le linguiste Mouloud Mammeri [16] dans les années 1960. Elle est composée de 33 lettres (29 consonnes et 4 voyelles longues (a, e, i, u) [17].

---

<sup>2</sup>Tamazight est le nom originel de la variété berbère parlée au Maroc central (Moyen Atlas) mais aussi le terme générique désignant la « langue berbère ».

<sup>3</sup>Graphies c'est toute représentation écrite d'un mot ou d'une lettre.

Tableau 2 : Les lettres de la langue amazighe.

Les lettres de tamazight avec leurs prononciations en arabe																
<b>A</b>	<b>b</b>	<b>c</b>	<b>č</b>	<b>d</b>	<b>ḍ</b>	<b>e</b>	<b>ε</b>	<b>f</b>	<b>g</b>	<b>ğ</b>	<b>y</b>	<b>h</b>	<b>ḥ</b>	<b>I</b>	<b>j</b>	<b>K</b>
ء	ب	ش	تش	د	ذ	ء	ع	اف	ga	ج	غ	ه	ح	أي	جو	ك
<b>L</b>	<b>m</b>	<b>n</b>	<b>q</b>	<b>r</b>	<b>ṛ</b>	<b>s</b>	<b>ş</b>	<b>t/tt</b>	<b>ţ</b>	<b>u</b>	<b>V</b>	<b>w</b>	<b>X</b>	<b>Y</b>	<b>z</b>	<b>ẓ</b>
ال	ام	ان	ق	ر	رر	س	ص	ت/تث	ط	u	V	و	خ	ي	ز	ززا

### 2.3. Les catégories du mot

Il existe deux catégories principales de mots en tamazight : les noms et les verbes. Les autres mots sont les adjectifs, les adverbes, les pronoms et autres morphèmes<sup>4</sup>. Tamazight, à l'instar des autres langues, a fait des emprunts auprès des langues étrangères avec lesquelles il a été en contact (latin, grec, arabe, turc, français, etc.).

**Exemples :** Le mot **akamyun** « camion ». Le verbe **urar** « jouer » est concurrencé par le verbe arabe **laeb** [15].

#### 2.3.1. Le nom :

Il est composé d'un radical accompagné d'éléments indicatifs (genre, nombre, affixes dérivationnels, etc.).

Tableau 3 : Les parties de mot "argaz".

<i>argaz</i> « homme »	
<i>a</i>	<i>rgaz</i>
↓	↓
<b>Indique que le nom est masculin</b>	Le radical

En tamazight le nom varie en genre, en nombre et en état (libre/d'annexion).

#### 2.3.1.1. Genre (masculin ou féminin) :

Le nom masculin commence généralement par la voyelle initiale : *a*, *i* ou *u*. On forme généralement le féminin préfixant et en suffixant le masculin par *t* devant le nom masculin et un autre à la fin de mot.

<sup>4</sup>morphèmes : c'est l'unité minimale porteuse de sens, isolée par la segmentation d'un mot, le plus souvent dépourvu d'autonomie linguistique.

Tableau 4 : Quelques exemples de mots convertis du masculin au féminin.

Nom berbère masculin	Signification en français	Féminin de nom berbère
<b>Igider</b>	Aigle	Tigidert
<b>Agujil</b>	Orphelin	Tagujilt
<b>Uccen</b>	Chacal	Tuccent

#### Exceptions sur les noms :

- Les noms de parenté ne comportent pas de marque de genre en particulier comme : **Egma** « **mon** frère ».
- Le nom de nombre se comporte comme adjectifs tel que :

Tableau 5 : Quelques exemples de noms des nombres convertis du masculin au féminin.

Berbère	Français	Berbère	Français
Sin	Deux .M	<b>Snat</b>	Deux. F
Amezwaru	Premier	<b>Tamezwarut</b>	Première
Imezwura	Premiers	<b>Timezwura</b>	Premières
wis sin	Le deuxième	<b>Tissnat</b>	La deuxième

- Des noms qui n'ont pas de voyelle initiale comme : *seksu* « couscous ».
- Des noms féminins qui n'ont pas de masculin comme : *tala* « fontaine ».
- Des noms masculins qui n'ont pas de féminin comme : *adfel* « neige ».

#### 2.3.1.2. Nombre (singulier et pluriel) :

Pour décrire le pluriel des noms singuliers de tamazight, on change le **a** du singulier à **i** au pluriel tandis que les voyelles initiales **i** et **u** restent inchangés.

Pour décrire le pluriel des noms féminins singuliers qui se terminent par **i** et **it** on change seulement leurs terminaisons en **a**, Et ceux qui terminent par **a** on change leur terminaison en **iwin** ou **wa**.

On distingue trois types de pluriel :

### 1) Externe :

On ajoute un suffixe au masculin (*en, an, awen, ayen, iwen, etc.*), ainsi qu'au féminin (*in, atin, awin, etc.*), sans modifier le radical [18].

Tableau 6 : Exemples du pluriel externe.

<i>Amyar</i>	<b>Vieux</b>	<i>Imyaren</i>
<i>Inzer</i>	<b>Nez</b>	<i>Inzeren</i>
<i>Tamyart</i>	<b>Vielle</b>	<i>Timyarin</i>

### 2) Interne :

Le radical est modifié par alternance vocalique <sup>5</sup>et/ou consonantique <sup>6</sup>[18] :

- transformation d'une voyelle (ou plus) en une autre.
- disparition d'une voyelle.
- apparition éventuelle d'une semi-consonne.

Tableau 7 : Exemples de pluriel interne.

<i>Azrar</i>	<b>Chaîne</b>	<i>Izurar</i>
<i>Amcic</i>	Chat	<i>Imcac</i>
<i>Igider</i>	Aigle	<i>Igudar</i>
<i>Awtul</i>	Lièvre	<i>Iwtal</i>

### 3) Mixte :

C'est la combinaison des pluriels externe et interne.

Le radical est modifié par alternance vocalique et/ou consonantique et un des suffixes est ajouté. Ce pluriel est caractérisé par la tension d'une consonne du radical [18].

Tableau 8 : Exemples de pluriel mixte.

<i>Abrid</i>	<b>Chemin</b>	<i>Iberdan</i>
<i>Tamurt</i>	Pays	<i>Timura</i>
<i>Adlis</i>	Livre	<i>Idlisan</i>
<i>Tawwurt</i>	Porte	<i>Tiwura</i>

<sup>5</sup>alternance vocalique appelé aussi gradation vocalique ou ablaut désigne un système de gradations des timbres vocaliques en indo-européen qui a encore des effets dans la langues indo-européennes modernes

<sup>6</sup>Alternance consonantique désigne les modifications qui subissent les consonnes dans certaines langues.

### 2.3.1.3. État libre/d'annexion

Le nom possède deux états : un état libre et un état d'annexion.

Tableau 9 : États de nom et leurs catégories.

Etats de noms	Des catégories des états de noms	Des exemples
<b>Libre</b>	Isole	<i>uccen</i> « chacal » <i>argaz</i> « homme »
	En début de phrase	<i>Argaz iruh yer ssuq</i> L'homme est allé au marché
<b>Annexion</b> C'est l'action d'un substantif Sur un autre, s'exprime de manière suivante :	Quand il est précédé d'une préposition	<i>Allen n wuccen</i> « Les yeux <b>de</b> chacal »
	Quand il suit un verbe	<i>Iruh wargaz yer ssuq</i> L'homme est allé au marché

Les conditions d'apparition du nom à l'état libre ou à l'état d'annexion dans l'énoncé ainsi que les règles de modification de sa première syllabe sont nombreuses et variées.

### 2.3.2. L'adjectif

L'adjectif est identifié par sa morphologie et sa syntaxe, et il varie en genre et en nombre.

Par exemple le mot rouge « **azgay** »

Tableau 10 : Un exemple illustrant les variations de l'adjectif.

<b>En genre</b>	<i>azgay</i>	<i>tazgayet</i>	Quelque chose qui est rouge
<b>En nombre</b>	<i>azgay</i>	<i>Izgayen</i>	Quelques trucs qui sont rouges
<b>En état d'annexion</b>	<i>azgay</i>	<i>Uzwiy/comme</i> <i>Uzwiy udem</i>	Son visage est toujours rouge

### 2.3.3. L'adverbe

Il existe un ensemble de (noms, prépositions et autres déterminants invariables) qui peuvent avoir un emploi adverbial.

Tableau 11 : Déterminants invariables.

aṭas	⇒	<b>Beaucoup</b>	drus	⇒	<b>Peu</b>
azekka	⇒	Demain	tameddit	⇒	Après-midi

Tableau 12 : Formes figées, composées du nom et/ou d'éléments grammaticaux.

afella	⇒	<b>Le dessus</b>	sufella	⇒	<b>Au/par-dessus</b>
ass	⇒	Jour	assa	⇒	Aujourd'hui

### 2.3.4. Le verbe

La forme verbale est composée d'une désinence <sup>7</sup>personnelle et thème verbal qui se compose lui-même d'une racine significative et d'un schème indiquant l'aspect.

Tableau 13 : Un exemple illustrant les parties de verbes " ččiy "

La forme verbale ččiy « j'ai mangé »		
Thème verbal «čči»		Désinences personnelles :y qui représente la 1ère personne du singulier.
čč (est la racine du verbe ečč« manger»)	i (qui exprime le passé composé)	

## 2.4. La structure morphologique d'un mot Tamazight

Le modèle sémitique de formation du mot (racine + schème) s'applique donc tel quel en berbère et il n'est pas sûr que l'analyse du système morphologique berbère selon ce modèle soit totalement conséquente [15].

- **racine** est une suite ordonnée de consonnes radicales.
- **schème** est constitué de divers éléments indicatifs.

<sup>7</sup>Désinence : élément variable qui s'ajoute au radical d'un mot pour marquer la personne, le nombre, la voix, le mode, le temps d'un verbe.

Tamazight est une langue générative, les noms et les verbes sont dérivés d'une racine. Nous pouvons engendrer plusieurs mots différents à l'aide d'un schème et à partir d'une même racine. Il existe différents procédés de formation lexicale : la dérivation (la plus courante et la plus productive), la composition et l'expressivité [15].

### 2.4.1. Dérivation

Il existe deux procédés principaux de dérivation lexicale :

#### 1) Dérivation verbale :

Elle se fait en préfixant la base par un morphème dérivationnel (s-, m-, n-, t- ou une de leurs combinaisons) [15].

Tableau 14 : Exemples de dérivation verbale.

<i>awal</i> « mot »	⇒	Base nominale	⇒	<i>Siwel</i> « parler »
<i>aker</i> « voler »	⇒	Base verbale	⇒	<i>Ittwaker</i> « être volé »

#### 2) Dérivation nominale :

Elle sert à créer des mots tels que le nom verbal, le nom, le nom d'animé, le nom d'instrument, etc., par préfixation d'un morphème (m-, s-, ms-, etc.) [15].

Tableau 15 : Exemples de dérivation nominale.

<i>abrid</i> « chemin »	⇒	<i>amsebrid</i> « passant »	⇒	Nom d'animé
<i>aḍen</i> « être malade »	⇒	<i>amaḍun</i> « malade »	⇒	Nom patient

### 2.4.2. Composition :

Consiste à la formation d'une nouvelle unité lexicale à partir de deux unités lexicales autonomes [15], elle se fait généralement par la juxtaposition de mots ou sous forme de synthème<sup>8</sup>.

Tableau 16 : la juxtaposition et le synthème de mot.

<b>Juxtaposition</b>	Nom-nom	Le nom est « <i>ifer</i> »	<i>iferzizwi</i> <i>ifer-tizizwit</i>	Mélisse Feuille-abeille
	Verbe-nom	Le verbe est « <i>Magar</i> »	<b>Magritij</b> [19] magar + <i>iṭij</i> »	Tournesol

<sup>8</sup>Synthème est une concaténation des monèmes (unité de langage de la plus petite taille possible) qui peuvent être analysés en plusieurs unités significatives.

		(Se rencontrer)	(rencontrer + soleil)	
			<b>Mager aman</b> <b>Amagraman [19]</b> "va à la rencontre de l'eau" Magar Aman	Inule visqueuse
<b>Synthème</b>	<i>tawetžit n uđar</i>			Cheville

### 2.4.3. Expressivité :

C'est le fait de rendre les choses (abstraites ou concrètes) plus expressives en les nommant de manière suggestive par des effets lexicaux (mouvement, bruit, son, musicalité, etc.)

En ce qui concerne le lexique, l'expressivité est un moyen morphologique qui permet d'attribuer un trait particulier à un mot en lui modifiant la forme et en ajoutant à son sens, une valeur augmentative, diminutive, affective, méliorative, péjorative, etc. [15].

*Tableau 17 : les opérations qui rendent le mot expressif.*

- <b>Affixer un ou deux phonèmes</b> <sup>9</sup> (b, c, f, h, ħ, ε, x, hn, xn, etc.) <b>au mot</b>	ađad « <b>Doigt</b> »		tadađet « <b>Petit doigt</b> ».
	<i>Inzer</i> « <b>Nez</b> »		<i>Funzer</i> « <b>Saigner du nez</b> ».
- <b>Utiliser l'onomatopée</b> <sup>10</sup>	<i>biđbiđ</i>		Vanneau
- <b>Passer du féminin au masculin et inversement :</b>	<i>Afus</i> « <b>Main</b> »	<i>Tafust</i> « <b>Petitemain</b> »	Pour Le diminutif
	<i>Argaz</i> « homme »	<i>Targazt</i> « femmelette »	Pour le diminutif/dépréciatif
- <b>Dupliquer /redoubler une consonne ou une syllabe :</b>	<b>Ferfer</b> « voleter »		

### 2.5. Conclusion

Dans ce chapitre, nous avons évoqué les spécificités majeures de la langue Tamazight et qui peuvent expliquer la différence au niveau du traitement automatique entre Tamazight et les autres langues

<sup>9</sup>Phonèmes c'est le son de prononciation (sonore) des alphabets berbères.

<sup>10</sup>onomatopée (mot féminin issu du grec) est une catégorie d'interjections émises pour simuler un bruit particulier associé à un être, un animal ou un objet.

**Chapitre 3 :**  
La traduction  
automatique et ses  
approches.

## **Chapitre 3 : La traduction automatique et ses approches.**

### **3.1. Introduction**

Ces dernières années, les besoins en systèmes de traduction automatique augmentent de plus en plus et cela à cause des rapprochements des peuples grâce aux voyages et aux réseaux sociaux. En effet, plus de 6800 langues sont recensées sur notre planète et tout le monde peut un jour se heurter à cette barrière.

La traduction est une tâche complexe. Traduire ne consiste pas simplement à prendre chaque mot d'une phrase et à le remplacer par sa traduction dans une autre langue. Il faut faire face à de nombreuses difficultés comme les ambiguïtés qui doivent être levées avant de pouvoir prendre une décision quant à la bonne traduction des mots.

Si par le passé, faire traduire un texte par une machine relevait de l'utopie, les avancées technologiques et les efforts constants de nombreux chercheurs en Traduction Automatique (TA) font que cela paraît aujourd'hui possible. La TA désigne le fait de faire traduire un texte d'une langue source vers une langue cible par une machine sans aucune intervention humaine

Dans ce chapitre, nous présentons la Traduction Automatique (TA) et ses deux grandes approches la caractérisant : premièrement l'approche que nous allons utiliser dans notre application dite l'approche experte qui repose sur des connaissances. Ensuite on verra la 2ème approche dite l'approche empirique qui tire ses connaissances à partir de grandes quantités de données.

### **3.2. Un bref historique**

Les premières idées réelles pour la Traduction Automatique sont issues du Français Georges Artsrouni et du Russe Petr Trojanskij en 1933. Ces dernières sont appliquées en janvier 1954 dans un projet de l'université de Georgetown et par IBM où Quarante-neuf phrases russes sont traduites automatiquement en anglais par un premier système qui supporte un vocabulaire de seulement deux cent cinquante mots et utilise en tout et pour tout six règles de grammaire [20]. L'existence de plusieurs langues rend la Traduction Automatique très lente, moins performante, et deux fois plus coûteuse que les traductions réalisées par des experts humains [20].

### 3.3. Les différentes approches

Avant d'aborder la traduction **automatique** et ses approches, voyons le processus de la traduction naturelle d'un texte qui peut être déroulé en trois phases :

- 1) Lire et comprendre le texte source.
- 2) Transformation vers la langue cible.
- 3) Génération d'un texte dans la langue cible.

Ces trois étapes d'analyse, de transfert et de génération sont illustrées par le triangle de Vauquois (figure 1).

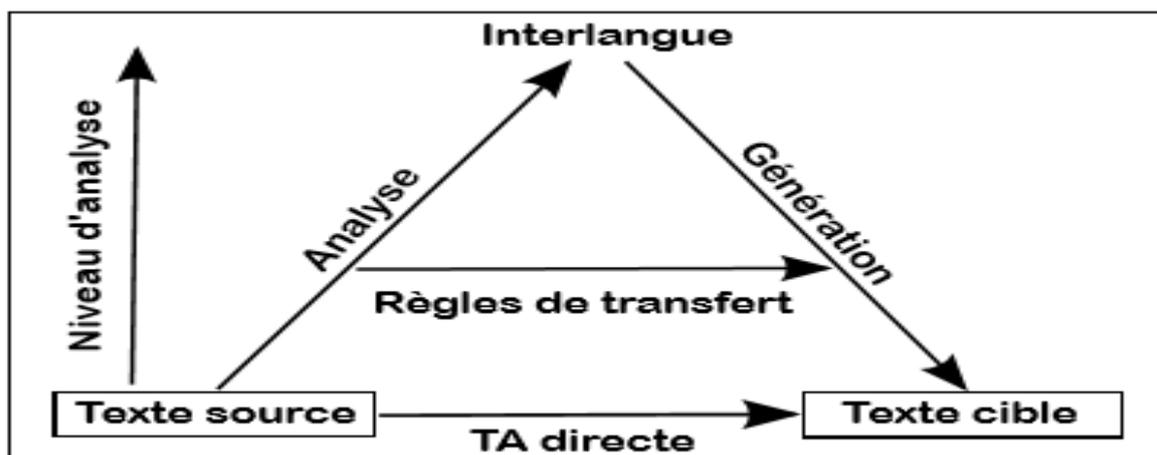


Figure 2 : Le triangle de Vauquois [20].

L'ajout de notion **Automatique** à la traduction consiste à garder et à appliquer ces phases sans aucune intervention humaine. Sur ce point, on distingue deux approches dites **Experte** et **Empirique**. Chacune des approches envisagées, qu'elle soit experte ou empirique, traite ces étapes de façon particulière avec des niveaux d'analyse, de transfert, et de génération différents. Nous décrivons dans les sections suivantes les principales méthodes de TA issues de chacune des deux approches [20].

#### 3.3.1. L'approche experte

Les méthodes expertes, comme leur nom l'indique, nécessitent l'intervention de spécialistes de la langue et les connaissances d'experts humains. La traduction selon cette approche se déroule en trois phases similaires à celles exposées dans le triangle de Vauquois :

- 1) **Analyse** : la phrase source est segmentée en unités liées les unes aux autres. Selon le niveau d'analyse, ces unités peuvent être morphosyntaxiques, grammaticales, ou sémantiques.
- 2) **Transfert** : Les unités de la phrase source sont transformées en unités équivalentes de la langue cible en utilisant des règles et des dictionnaires bilingues.
- 3) **Génération** : transformation des unités cibles en texte cible, souvent avec un réordonnement des termes.

Les trois méthodes principales dérivées de cette approche sont la TA directe, la TA à base de règles de transfert ,et la TA fondée sur l'inter-langue [20].

#### **3.3.1.1. TA directe**

Dans cette méthode nous n'utilisons pas de représentation intermédiaire de la phrase source (résultat de l'étape d'analyse) pour générer la traduction à partir de ces représentations. La phrase source est segmentée souvent par mots et le système exécute la traduction de façon partielle (mot à mot), le transfert de ces mots repose sur un dictionnaire bilingue qui va spécifier un ensemble de règles permettant de les traduire. Une fois que ces mots sont traduits, ils sont ensuite réordonnés à l'aide de simples règles afin de générer une traduction.

Les systèmes de traduction directe sont des systèmes bilingues et unidirectionnels : ils traitent une seule paire de langues dans une seule direction de traduction [21].

#### **3.3.1.2. TA à base de règles de transfert**

Dans cette approche la phase d'analyse se fait par un analyseur syntaxique et d'une grammaire qui donne une représentation arborescente. Quant au transfert, le passage est assuré par des tables bilingues, contenant les règles de transfert (syntaxiques, lexicales ou encore sémantiques) entre les nœuds de l'arbre source et ceux de l'arbre cible. Enfin, une grammaire de la langue cible va générer la phrase cible à partir de cette représentation.

Le schéma suivant (figure 2) illustre cette méthode.

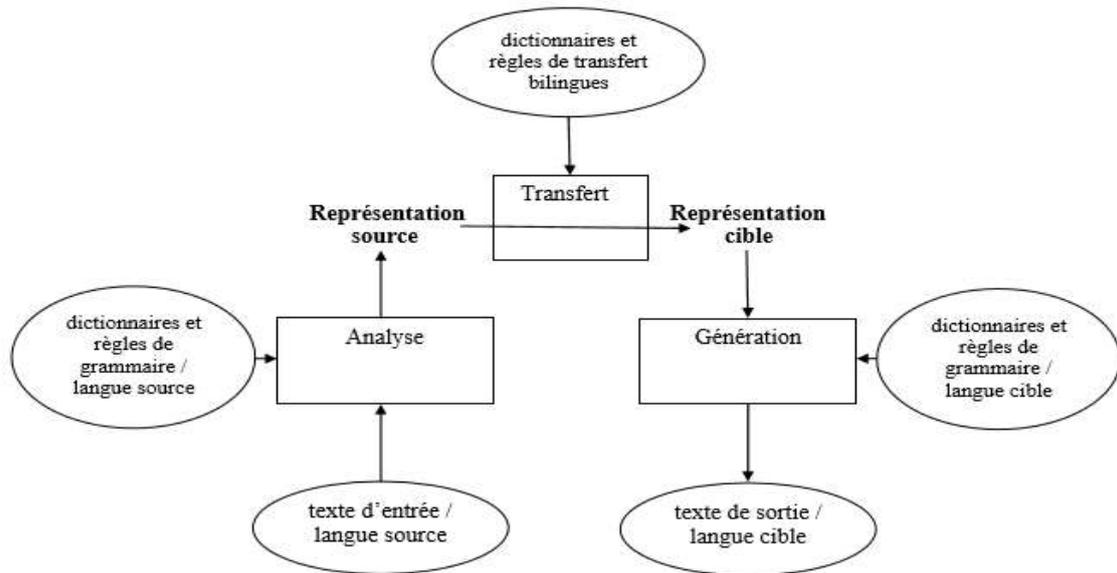


Figure 3 : Architecture du système de TA par transfert [22].

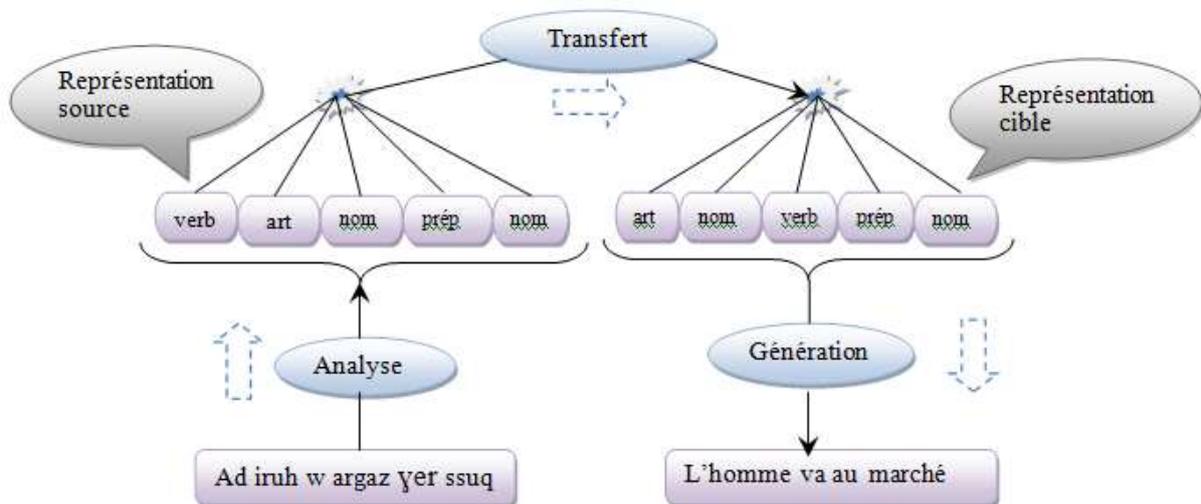


Figure 4 : Un exemple de TA avec le système de transfert.

La figure 3 illustre un exemple simple de TA d'une phrase dans la langue tamazight vers la langue français à base de transfert, ou la phrase source (Adiruh wargaz yer ssuq) passe premièrement par l'étapes d'analyse qui nous donne une représentation arborescente (représentation source) (verbe ,article, nom ,préposition ,nom) grace aux règles de grammaire de la langue source, vient ensuite, le passage qui va nous donner une représentation cible (article, nom ,verbe, préposition ,nom) qui doit être compatible avec les règles de la langue cible pour avoir une phrase correcte dans la phase de génération (L'homme va au marché).

L'inconvénient majeur de cette méthode réside donc dans la difficulté de formaliser toute la grammaire réelle d'une langue, ainsi que les grammaires de transfert.

### **3.3.1.3. TA fondée sur une inter-langue**

Cette approche, comme son nom l'indique, décrit le cas idéal d'une représentation intermédiaire indépendante des langues source et cible dites la représentation pivot. Elle consiste à «*enconvertir*» la phrase source vers la représentation pivot puis à «*déconvertir*» la phrase cible à partir de cette représentation pivot la traduction se fait sans passer par l'application de règles de transfert, ou la sortie du composant d'analyse du texte source serait directement l'entrée du composant de génération du texte cible.

L'avantage c'est la possibilité de l'appliquer dans un environnement multilingue. Pour couvrir tous les sens de traduction entre  $n$  langues, nous n'avons besoin que de  $n$  modules d'enconversion et de  $n$  modules de déconversion [21].

Mais la complexité de cette méthode réside dans l'obligation de construire un vocabulaire pivot pour représenter tous les concepts possibles de toutes les langues et les liens des concepts entre deux langues.

Nous verrons ci-dessous un schéma (figure 4) qui résume les étapes de cette approche

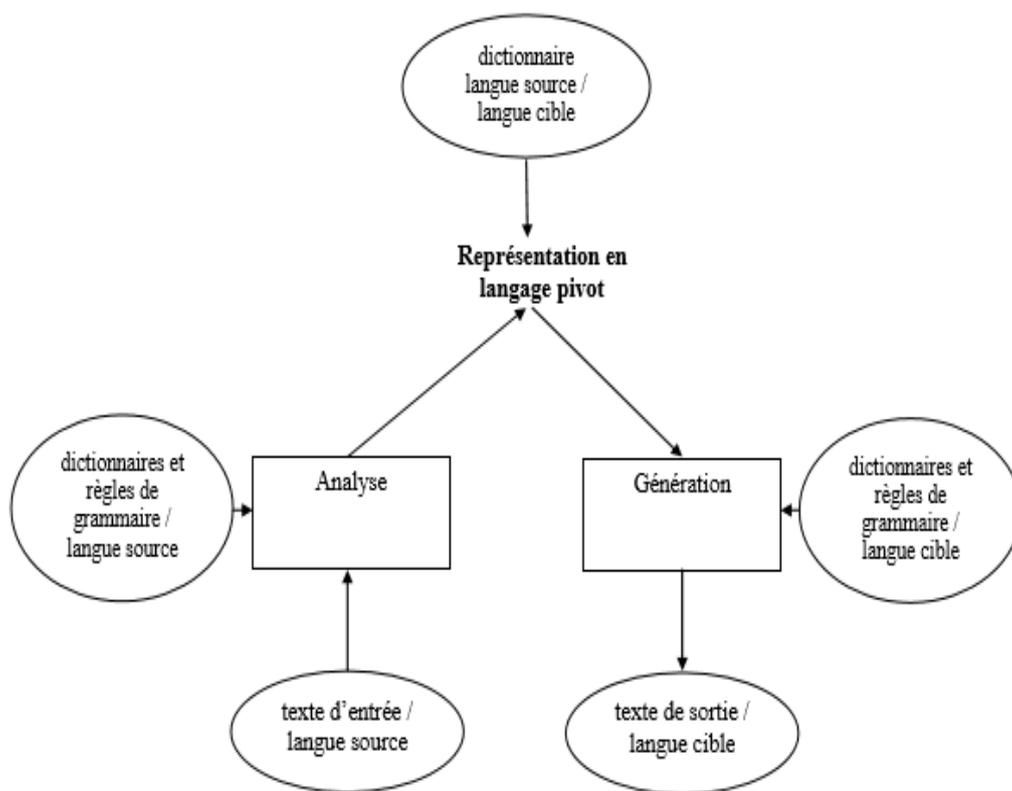


Figure 5 : Architecture du système de traduction automatique par langue pivot [22].

### 3.3.1.4. Avantages et inconvénients de l'approche experte

Tableau 18 : Avantages et inconvénients de l'approche experte.

Pré requis	Entrées de dictionnaire et des règles linguistiques
<b>Avantages</b>	La qualité de la traduction est très satisfaisante surtout lorsqu'il s'agit d'une traduction spécialisée.
<b>Inconvénients</b>	L'élaboration des règles linguistiques est couteuse en termes de temps et d'effort.

### 3.3.2. L'approche empirique

Le but de cette approche repose sur l'élaboration d'un nouveau type d'architectures qui ne nécessite aucune connaissance à priori pour mettre en place un système de traduction mais provient directement de données réelles. Sur ce point, on distingue deux approches : l'approche par **analogie** et l'approche **statistique**, qui tirent profit des données brutes contrairement aux méthodes de l'approche experte.

### 3.3.2.1. L'approche par analogie ou basée sur l'exemple

Comme son nom l'indique, cette approche est basée sur la réutilisation des exemples existants où les traducteurs s'en servent comme modèles pour trouver la traduction de ce qu'il recherchent.

Pour développer un système par analogie, il est nécessaire, tout d'abord, de construire une base d'exemples de traduction à partir de grandes collections de textes traduits auparavant par des humains dans un domaine donnée. Ces exemples peuvent être des couples bilingues (source - cible) de phrase, de séquence de mots, une expression, ou un mot.

Le processus de traduction d'une phrase source est basé sur la correspondance entre cette phrase et la base d'exemples. Le processus de TA fondée sur les exemples se décompose en trois étapes :

- 1) **Correspondance** (*matching*) cette étape consiste à utiliser l'ensemble des méthodes de sélection pour le calcul de similarité entre la phrase source à traduire et les phrases « exemples » les plus proches dans la base d'exemples. Il existe plusieurs méthodes de sélection : celles basées sur la capacité de remplacer des mots correspondants entre deux phrases, sur des listes de synonymes, ou sur la similarité au niveau syntaxique ou sémantique, ou encore sur des informations statistiques.
- 2) **Aligner les phrases** : la phrase à traduire et les phrases exemples diffèrent sur un ou plusieurs mots. L'alignement est utilisé pour identifier les parties des phrases qui peuvent être réutilisées dans la traduction de la phrase source. Les autres parties qui diffèrent entre la phrase source et les phrases exemples peuvent être traduites en utilisant un dictionnaire bilingue.
- 3) **Combiner** : dans cette dernière étape, les traductions des parties réutilisées et non réutilisées sont assemblées d'une manière logique pour créer la traduction de la phrase source entière. La traduction est extraite à partir de véritables exemples et la qualité de traduction est garantie.

#### Exemple :

Nous allons mieux comprendre ces trois étapes à partir de ce simple exemple ci-dessous.

Tableau 19 : Exemple illustrant l'approche par analogie ou basée sur l'exemple.

Phrase à traduire :	Adiruh wargaz yer ssuq		
Les étapes à suivre	1-trouver les correspondances	<b>ssuq</b> yeseεa igumma Iwid <b>wargaz</b> azemmur –is <b>adiruh w</b> ameyar <b>yer</b> w axxam	
	2-aligner les phrases	<b>ssuq</b> yeseεa igumma	Sur <b>le marché</b> il y a des fruits
		I wid <b>w argaz</b> azemmur–is	<b>L'homme</b> a apporté ses olives
	<b>Adiruh w</b> ameyar <b>yer</b> w axxam	<b>Le</b> vieil homme <b>va à</b> la maison	
3-combiner	Adiruh wargaz yer ssuq		<b>L'homme va au marché</b> Remarque :(au=le+à)

Le tableau ci-dessous explique les trois étapes de système interligne de la TA, où la phrase source c'est : « Ad iruh wargaz yer ssuq ». La 1ere étape consiste à sélectionner toutes les phrases les plus proches dans la base d'exemples, et faire l'alignement entre ces phrases et leurs traductions pour voir la traduction des mots qui sont en commun, pour enfin réunir la traduction de chaque mot pour avoir la phrase cible de la phrase source dans a dernière étape.

### 3.3.2.2. Les avantages de cette approche sont nombreux :

- Elle s'appuie sur une collection d'exemples réels qui implique une meilleure qualité de traduction.
- Cette approche est un moyen simple et efficace pour améliorer les performances d'un système de TA.

L'inconvénient majeur de cette approche dont être ressenti avec les langues moins dotées comme les langues pour lesquelles des ressources langagières de ce genre ne sont pas largement disponibles [22].

### 3.3.3. L'approche statistique

L'approche statistique de la TA consiste à dériver la traduction la plus probable d'une phrase source.

Le but de la traduction statistique est de trouver la phrase  $\hat{E}$ , qui maximise  $P(e|f)$  étant donnée une phrase source  $f$ . Mathématiquement :  $\hat{E} = \operatorname{argmax}_E P(e|f)$ .

D'après le théorème de Bayes :  $P(e|f) = P(f|e).P(e)/P(f)$

Puisque le dénominateur  $P(f)$  est indépendant de  $P(e)$  la maximisation devient alors:

$$\hat{E} = \operatorname{argmax}_E P(e|f) = \operatorname{argmax}_E P(e) * P(f|e).$$

$P(e)$  est appelé le modèle de la langue cible, tandis que  $P(f|e)$  est appelé un modèle de traduction. Ainsi, la traduction statistique repose sur deux modèles :

- Modèle de traduction qui propose pour chacun des segments de la phrase source, des traductions candidates en langue ciblent.
- Modèle de langue qui se charge de capter les contraintes imposées par la syntaxe de la langue ciblent par une ou plusieurs fonctions probabilistes.

La traduction statistique repose sur deux modèles:

- Modèle de traduction noté  $P(f|e)$  qui propose pour chacun des segments de la phrase source ( $f$ ), des traductions candidates ( $e$ ) en langue cible avec une probabilité  $P$ .
- Modèle de langue, noté  $P(e)$ , qui se charge de capter les contraintes imposées par la syntaxe de la langue cible par une ou plusieurs fonctions probabilistes.

Le but de la traduction statistique est de trouver la phrase  $\hat{E}$ , qui maximise  $P(e|f)$  étant donnée une phrase source  $f$ .

Mathématiquement :

D'après le théorème de Bayes :  $P(e|f) = P(f|e).P(e)/P(f)$  donc  $\hat{E} = \operatorname{argmax}_E P(e|f)$ .

Puisque le dénominateur  $P(f)$  est indépendant de  $P(e)$  la maximisation devient alors:

$$\hat{E} = \operatorname{argmax}_E P(e|f) = \operatorname{argmax}_E P(e) * P(f|e).$$

La figure 6 ci-dessous illustre l'architecture d'un système de traduction statistique.

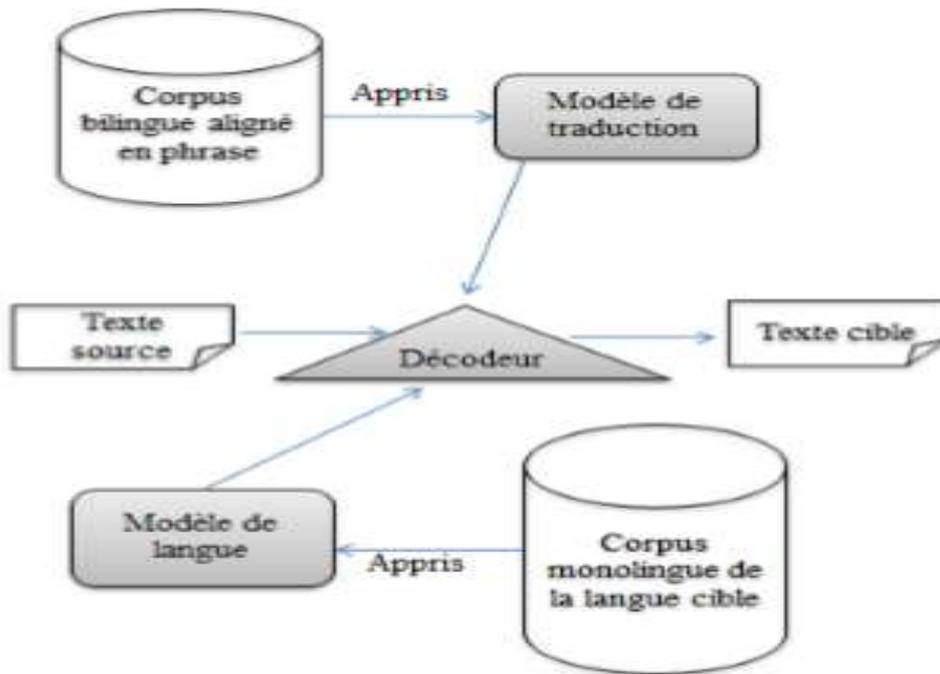


Figure 6 : Architecture d'un système statistique de traduction automatique [23].

Les méthodes statistiques peuvent être basées sur :

- Des corpus : Apprentissage automatique des “règles” de traduction à partir d'un corpus bilingue.
- Des exemples : Réutilisation des exemples de traductions existantes comme base pour la nouvelle traduction.

La table ci-dessous regroupe certains points forts et limites des approches statistiques.

Tableau 20 : Les forces et les limites des approches statistiques.

Pré requis	Mémoires de traduction (des corpus parallèles)
Avantages	<ul style="list-style-type: none"> <li>- La qualité de la traduction s'améliore automatiquement au fur et à mesure que la base d'apprentissage s'enrichit.</li> <li>- Peu de ressources linguistiques nécessaires.</li> <li>- Rapide à implémenter et facile à maintenir.</li> <li>- Portabilité : Facilité à étendre vers une nouvelle langue.</li> <li>- Possibilité d'évaluer de façon expérimentale des systèmes pratiques et des hypothèses scientifiques.</li> </ul>
Inconvénients	Difficulté de construire des ressources linguistiques suffisantes en quantité et en qualité.

### **3.4. Critères de qualité des traductions**

Comment juger si une traduction est de bonne ou de mauvaise qualité ?

Ce jugement se porte essentiellement sur deux critères.

#### **1) D'intelligibilité :**

- Il permet de rendre compte de la bonne construction de la traduction dans la langue cible.
- Il indique si la traduction est une phrase qui a un sens dans la langue cible, indépendamment de la phrase source qu'elle traduit.

#### **2) Fidélité :**

- Il mesure le degré de similitude entre la phrase source et la phrase cible du point de vue du contenu sémantique.

Un moyen d'évaluer ces deux critères d'intelligibilité et de fidélité est de leur attribuer un score, soit en utilisant le jugement d'experts humains, soit en utilisant des mesures automatiques. Ces scores vont permettre de dire ensuite si oui ou non la traduction est de bonne qualité [20].

#### **3.4.1. Évaluation manuelle des traductions**

L'idée la plus naturelle est de demander à un expert ou un ensemble d'experts d'évaluer la qualité des traductions. Souvent, l'on effectue cette évaluation phrase par phrase à l'issue desquelles des scores lui sont attribués. Comme :

- Le A-score (Adequacy score)
- F-score (Fluency score)
- I-score (Informativeness score)

#### **Les inconvénients de ces scores :**

- Ils sont extrêmement coûteux en termes de temps et de coûts financiers.
- Ces évaluations sont largement subjectives et ne sont pas reproductibles.

Il peut donc être délicat d'utiliser ces évaluations pour comparer des systèmes ou pour juger de l'évolution des performances. Afin de résoudre ces problèmes, des techniques d'évaluation automatique ont été développées [20].

### **3.4.2. Évaluation automatique des traductions**

Même si l'évaluation humaine reste le moyen le plus fiable de juger de la qualité d'une traduction, les mesures d'évaluation automatiques permettent à moindre coût de traiter un grand nombre de données en un temps limité.

Elles affectent un score à une ou plusieurs paires de traduction en comparant une hypothèse de traduction à une traduction dite de référence produite par un expert.

Chaque mesure emploie ensuite des techniques de comparaison qui lui sont propres. Ainsi, grâce à ces scores attribués automatiquement aux sorties des systèmes, il est possible de comparer rapidement et de classer les performances de différents systèmes de TA lors des campagnes d'évaluation. Ces mesures automatiques constituent également de bonnes fonctions objectives à maximiser lors du paramétrage d'un système dans le but d'améliorer ses performances.

### **3.5. Conclusion**

Ce chapitre décrit les approches principales de la traduction automatique à savoir : l'approche experte et l'approche empirique. En outre, il met en évidence les prés requis, les forces, et les faiblesses de chacune d'elles.

Dans notre cas nous allons traiter une langue peu dotée qu'est tamazight qui souffre de la rareté des ressources linguistiques surtout en matière de textes parallèles et comparables. Un système de TA avec l'approche experte s'avère être le plus adéquat plus exactement le système TA direct.

# **Chapitre 4 :**

## La réalisation d'un traducteur

## **Chapitre 4 : La réalisation d'un traducteur**

### **4.1. Introduction**

Notre objectif est de concevoir et de réaliser un traducteur basé sur une ontologie de la langue Tamazight. Dans le chapitre 1, nous avons vu qu'une ontologie peut être définie comme un réseau sémantique visualisable par un graphe orienté. Elle regroupe également un ensemble de concepts qui décrivent un domaine. Ces concepts sont liés entre eux par des relations taxinomiques d'une part (classes, attributs, facettes, instances), et sémantiques d'autre part (synonymie, antonymie hyperonymie, hyponymie...). L'intérêt de ces réseaux réside dans leur clarté et leur utilisation facile.

Dans ce chapitre, nous présentons tout d'abord, certains outils de développement que nous avons utilisés et la manière dont notre ontologie a été élaborée. Puis, nous expliquons comment nous avons intégré cette ontologie dans l'environnement de développement Android studio. Enfin, un programme générique exploitant cette ontologie permet une traduction des mots et certaines phrases.

### **4.2. La démarche suivie**

La conception de notre application est faite en utilisant le formalisme UML (Unified Modeling Language) qui permet de modéliser de manière claire et précise la structure et le comportement du système indépendamment du langage de Programmation.

UML offre beaucoup de diagrammes qui servent à la modélisation des systèmes, mais nous n'utilisons qu'un sous ensemble (suffisant) de ces diagrammes. Comme nous avons opté pour l'approche orientée objet, le fonctionnement du système sera décrit par un ensemble de diagrammes qui sont dessinés avec le logiciel Paccstar UML Diagram.

Nous avons commencé par l'analyse fonctionnelle de notre système en précisant les besoins des utilisateurs où tous les cas d'utilisation possibles de notre système ont été déterminés, et toutes fonctionnalités ont été décrites par des diagrammes des cas d'utilisations. Les interactions de l'utilisateur avec le système pour un tel cas d'utilisation sont décrites et schématisées par les diagrammes d'interactions en montrant leurs chronologies et scénarios de déroulement, après l'analyse fonctionnelle de notre système.

### 4.2.1. Diagramme de cas d'utilisation

Le diagramme des cas d'utilisation représente les principales fonctionnalités offertes par le système pour chaque acteur. Ce diagramme donne un bon aperçu du système. C'est le premier diagramme du modèle UML.

#### Identification des acteurs

- **Acteur principal** : c'est l'utilisateur de notre application.
- **Acteur secondaire** : c'est l'ontologie.

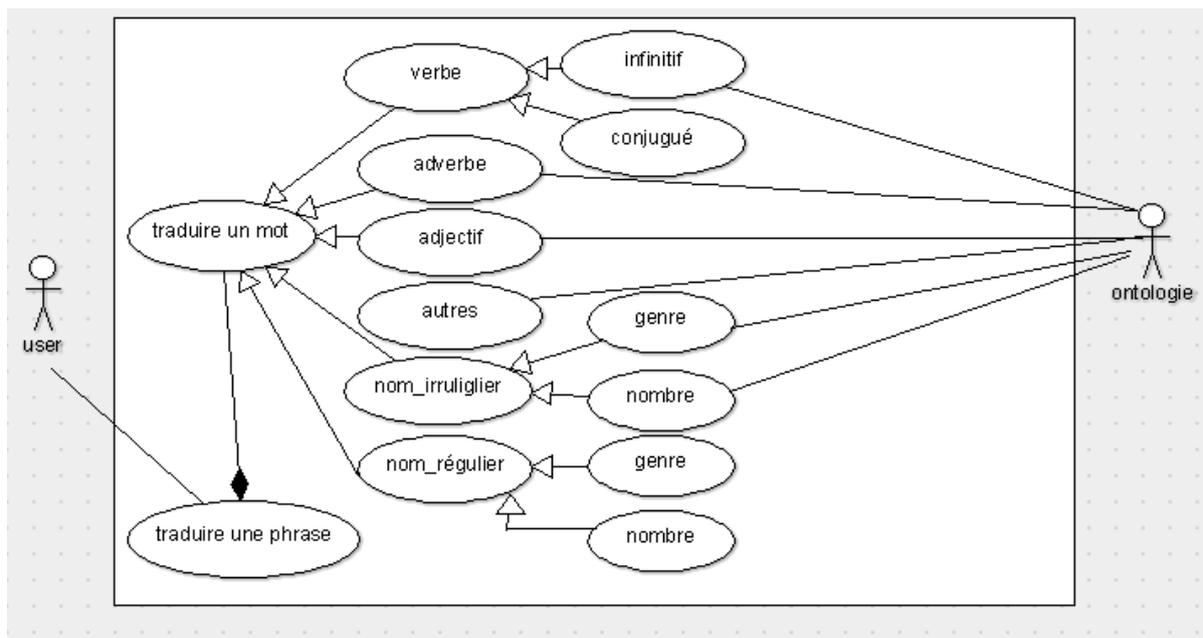


Figure 7 : Diagramme de cas d'utilisation.

### 4.2.2. Diagramme d'interaction

C'est une représentation d'un ensemble d'objets de classes différentes collaborant dans le cadre d'un scénario d'exécution du système. Dans ce diagramme les objets s'envoient des messages qui invoquent des opérations (ou méthodes) sur les objets récepteurs. Ce genre de diagramme permet de suivre visuellement les interactions dynamiques entre objets et les traitements réalisés par chacun.

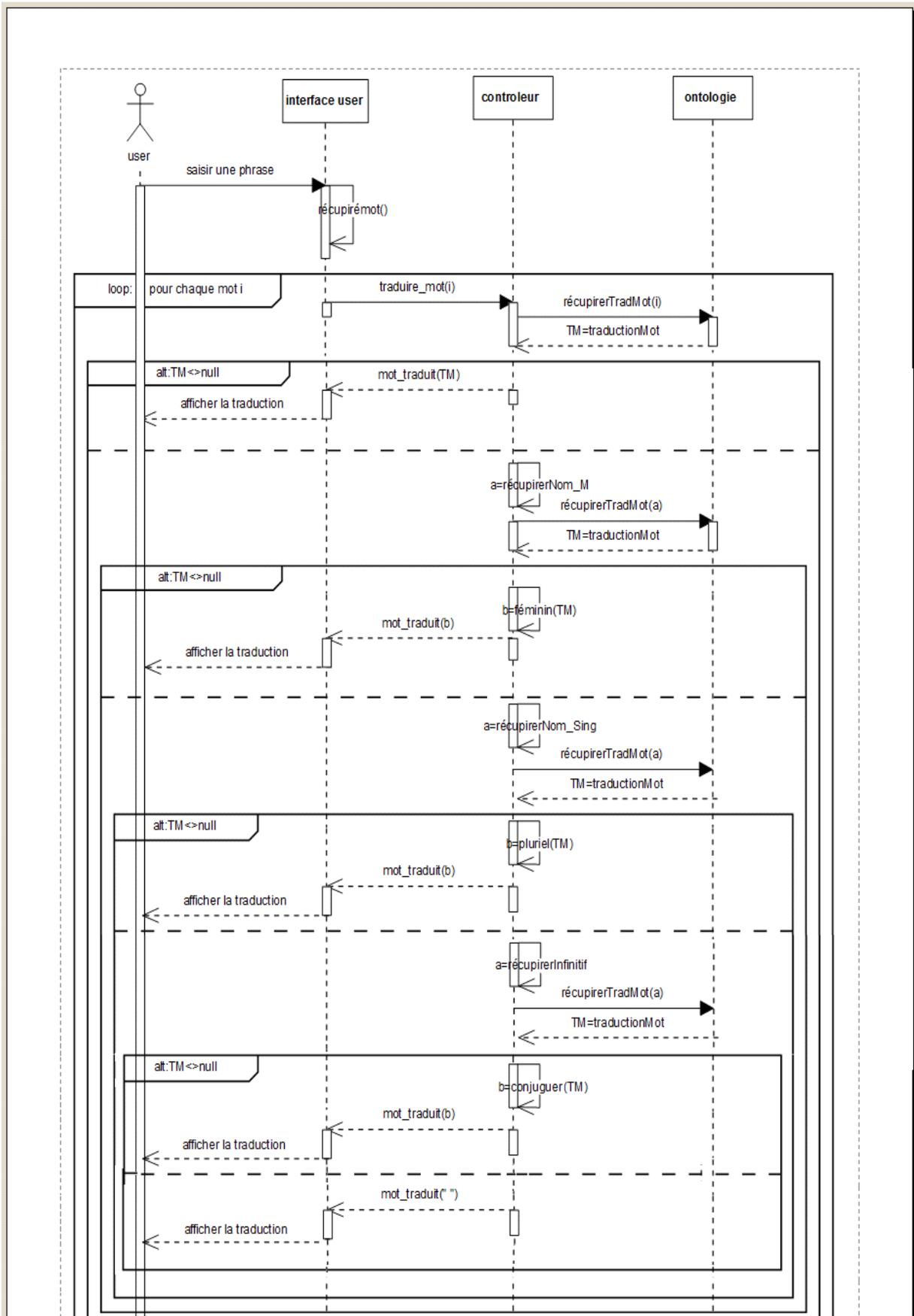


Figure 8: diagramme d'interaction.

### 4.2.3. Description textuelle

Tableau 21 : La description textuelle.

Nom du cas d'utilisation	Traduction d'une phrase
Objectif	Traduire une phrase de tamazight au français
Acteur	User, ontologie
Pré condition	Introduire une phrase par l'utilisateur
Scénario nominale	1-pour chaque mot(i) dans la phrase introduit 2-vérifier l'existence de mot(i) dans l'ontologie 3-récupérer sa traduction dans l'ontologie 4-afficher le résultat pour l'utilisateur (TM<>null)
Scénario alternatif 1	Si (TM=null) : le mot(i) n'existe pas dans l'ontologie : 1-récupérer un mot (a) masculin à partir du mot(i) 2- le système revient aux points 2 et 3 du scenario nominal 3-récupération d'un nom féminin à partir de TM (mot traduit) 4-afficher le résultat pour l'utilisateur
Scénario alternatif 2	Si (TM=null) : le mot(a) n'existe pas dans l'ontologie : 1-récupérer un mot (a) singulier à partir de mot(i) 2- le système revient aux points 2 et 3 du scenario nominal 3-récupération d'un nom pluriel à partir de TM (mot traduit) 4-afficher le résultat pour l'utilisateur
Scénario alternatif 3	Si (TM=null) : le mot(a) n'existe pas dans l'ontologie : 1- récupérer un mot (a) qu'est l'infinitif de mot(i) 2- le système revient aux points 2 et 3 du scenario nominal 3-conjuguer le mot(a) au temps et a la personne qui conviennent 4-afficher le résultat pour l'utilisateur
Scénario alternatif 4	Si (TM=null) : le mot(a) n'existe pas dans l'ontologie : La traduction du mot(i) = '' ''

### 4.3. Outils de développement

EDI pour le développement d'applications Android :«Un EDI (Environnement de développement Intégré) est un logiciel qui regroupe un ensemble d'outils permettant le

développement d'autres logiciels tels que des logiciels desktop, des applications mobiles, sites web, jeux... ou de créer des bibliothèques ou des Framework. » [24].

Il y a plusieurs EDI, certains sont polyvalents, d'autres sont utilisés pour des systèmes d'exploitation spécifiques. Parmi ces EDI les plus utilisés pour le développement Android on cite : Android Studio.

#### **4.3.1. Android Studio :**

C'est l'environnement de développement officiel pour le développement d'applications Android. Créé par Google et publié en 2014, il a remplacé les outils de développement d'Eclipse (ADT). Il est disponible gratuitement sous licence Apache. Android occupe environ 71% du marché des applications mobiles. La majorité de ces applications sont codées en Java. Parmi ses fonctionnalités principales :

- Une exécution instantanée des modifications du code dans l'application en cours d'exécution,
- Un émulateur riche qui permet de simuler des applications pour des appareils mobiles ou même des smart TV. Firebase Messaging, et Google End points pour l'intégration au cloud» [24].

#### **4.3.2. Jena (framework) :**

**Apache Jena** (ou juste **Jena**) est un framework Java gratuit et open source pour la construction des applications du web sémantique et de l'annotation sémantique (Linked data).

Il est composé de différentes APIs interagissant ensemble pour traiter les données RDF<sup>1</sup>[25].

#### **4.3.3. AndroJena :**

AndroJena est une version allégée du framework Jena pour créer des applications Web Sémantique. Pour surmonter les problèmes techniques, AndroJena a été adapté aux appareils Android [26]. AndroJena fournit plusieurs extensions :

- Le moteur de raisonnement,
- Le moteur ARQoid SPARQL.
- Le triple store TDBoid.

Bien que cet outil ne soit pas recommandé par Apache, il répond à nos besoins car il permet à la fois le raisonnement et l'interrogation dans les appareils sous contraintes basés sur Android [26].

#### **4.3.4. Logiciel protégé (voir chapitre 01).**

### **4.4. Langages de développement d'applications Android**

Pour développer une application Android, il est important de connaître les langages de programmation adéquats, à savoir Java et XML.

#### **4.4.1. XML :**

« Ce langage de balisage est utilisé dans les applications mobiles pour concevoir les interfaces graphiques. Il n'est pas indispensable pour développer une application Android, mais il facilite énormément le développement et permet plus d'organisation en séparant les interfaces du code. Ainsi, on gagne du temps et on simplifie le code de l'application, ce qui permet plus de lisibilité donc une meilleure gestion des erreurs » [27].

#### **4.4.2. Les types d'applications Android :**

« Il y a trois types d'applications mobiles : natives, web, et hybrides. Chacune a ses caractéristiques propres :

- **Application native** : installée sur un smartphone ou une tablette, elle est créée sur mesure et adaptée à chaque OS. On en crée une version spécifique à chaque type de téléphone,
- **Web-application** : site internet dont le design est adapté au mobile, créé pour être consulté à partir d'un Smartphone. La connexion web est nécessaire,
- **Application hybride** : codée en HTML, et est multiplateforme, mais avec moins de fonctionnalités » [28].

### **4.5. Aspects conceptuels de l'élaboration d'une ontologie sous protégé**

#### **4.5.1. Constitution du corpus**

D'une manière générale, la constitution d'une ontologie est une tâche très difficile ; elle est encore plus difficile dans le cas de la langue tamazight. Dans cette partie, l'objectif est de relever les exigences et les problèmes à résoudre pour la création d'une telle ontologie en tamazight. Parmi lesquelles, on cite les problèmes classiques liés à la création et à la gestion

du corpus, d'une manière générale, puis ceux liés à la création et à la gestion de corpus amazighs, en particulier.

Les concepts de notre ontologie, qui s'élèvent à plus de 1600 mots, ne sont pas tous extraits des dictionnaires existants. Nous sommes basés d'une part, sur nos connaissances de la langue en tant que natifs, d'autant plus qu'un nombre assez important d'entités sont extraites sur des sites internet, d'autre part nous avons utilisé nos connaissances acquises aux niveaux primaires et secondaire ; et nous avons demandé l'aide aux professeurs spécialisés dans la langue amazighe [29].

Pour ce qui est de la présentation, les entrées sont transcrites à l'aide des symboles latins classés selon l'ordre alphabétique des éléments constituants.

Quant aux locutions (les mots composés) qui ont toujours posé des problèmes au traitement automatique des langues, leurs éléments constituants sont séparés par des traits d'union (soulignés \_ ) pour qu'elles soient facilement reconnues et détectées par la machine en tant qu'entité formant un bloc indissociable. Toutefois, elles sont à leur tour classées selon l'ordre alphabétique de leur premier constituant.

#### **4.5.2. Implémentation**

Pour mener à terme le traitement de notre corpus, nous avons élaboré l'ontologie des mots Tamazight en utilisant le logiciel Protège (voir chapitre 1). En effet, nous l'avons jugée efficace dans la mesure où il est le plus connu et le plus utilisé des éditeurs d'ontologie. En plus d'être gratuit, Protège peut lire et sauvegarder des ontologies dans la plupart des formats d'ontologies : RDF, RDFS, OWL, etc. Outre la création collaborative de contenus, cette méthode intègre des fonctionnalités avancées comme l'import/export de contenus de type RDF/ OWL.

#### **4.5.3. Eléments de l'ontologie**

L'ontologie arabe fondée par Mustafa Jarrar [30] et l'ontologie anglaise wordnet [31] sont considérées comme les ontologies les plus populaires dans le domaine des langues. Elles sont fondées comme étant des dictionnaires monolingues informatisés, et sont utilisées dans divers domaines dont le web sémantique, d'où nous avons tiré l'idée de créer un dictionnaire pour la langue tamazight à partir de ces ontologies. Notre objectif est un peu différent par ce qu'on va l'utiliser seulement dans la traduction, en raison du manque de temps et du manque des outils de traitement de cette langue, nous avons décidé de créer notre ontologie

manuellement sous la forme d'un dictionnaire bi-langue informatisé dont l'unité de base est le concept ,elle est divisé en 5 catégorie : noms, verbes, adjectifs, adverbes, mots grammaticaux.

Ci-dessous une présentation illustrative de cette méthode. Nous présentons les relations entre les concepts que nous avons utilisés.

#### 4.5.3.1. La hiérarchie simple des mots

On parle ici des relations verticales de l'ontologie, telle que la relation d'héritage qui se résume par le mot clé (**is a** « **est un** »). Par exemple :akud **is an** isem.

La classe akud qui signifie « le temps » est une sous-classe de la classe **isem** qui signifie « le nom ».

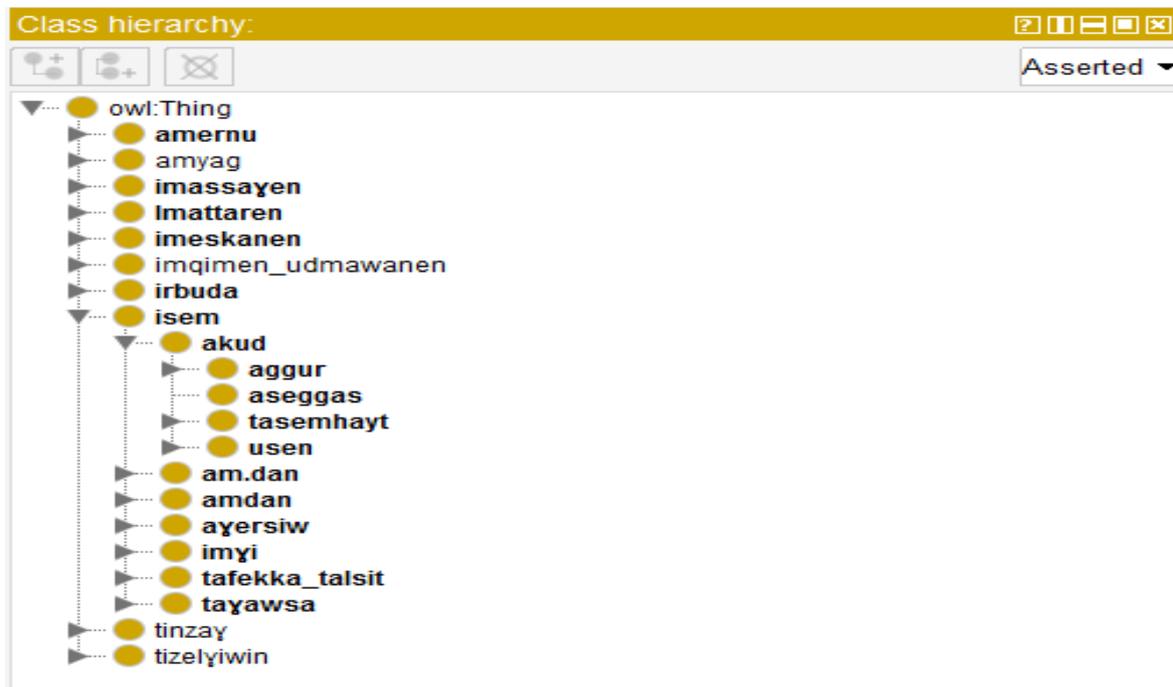


Figure 9: La relation d'hiérarchie sous protégé.

#### 4.5.3.2. Disjointes

Pour décrire la relation de disjonction il faut utiliser le mot clé « **DisjointWith** ». Par exemple : isem **DisjointWith** amyag ici le mot isem en tamazight qui signifie le nom disjoint avec le mot amyag qui signifie le verbe.

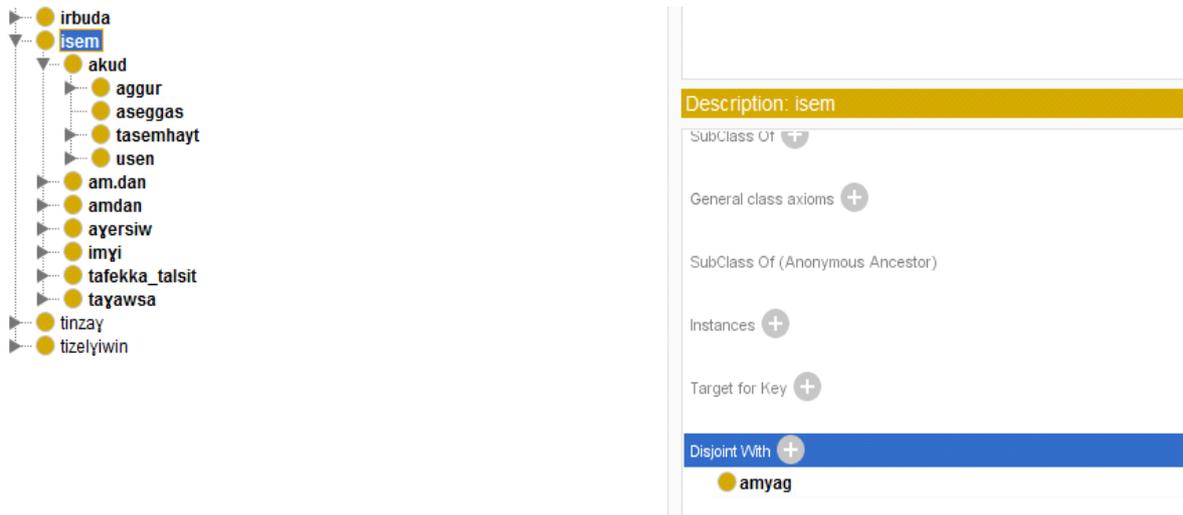


Figure 10 : La relation de disjonction sous protège.

#### 4.5.3.3. Les relations simples (Annotation)

Nous avons utilisé la partie Annotation de protège pour associer à chaque mot de Tamazight sa signification en français. La syntaxe d'écriture pour avoir cette association est écrite par le mot clé « **rdfs:comment** », par exemple :

**argaz rdfs:comment un homme.**

**argaz rdfs:comment un mari.**

Donc le mot **argaz** en tamazight a deux significations en français :un homme, et un mari.

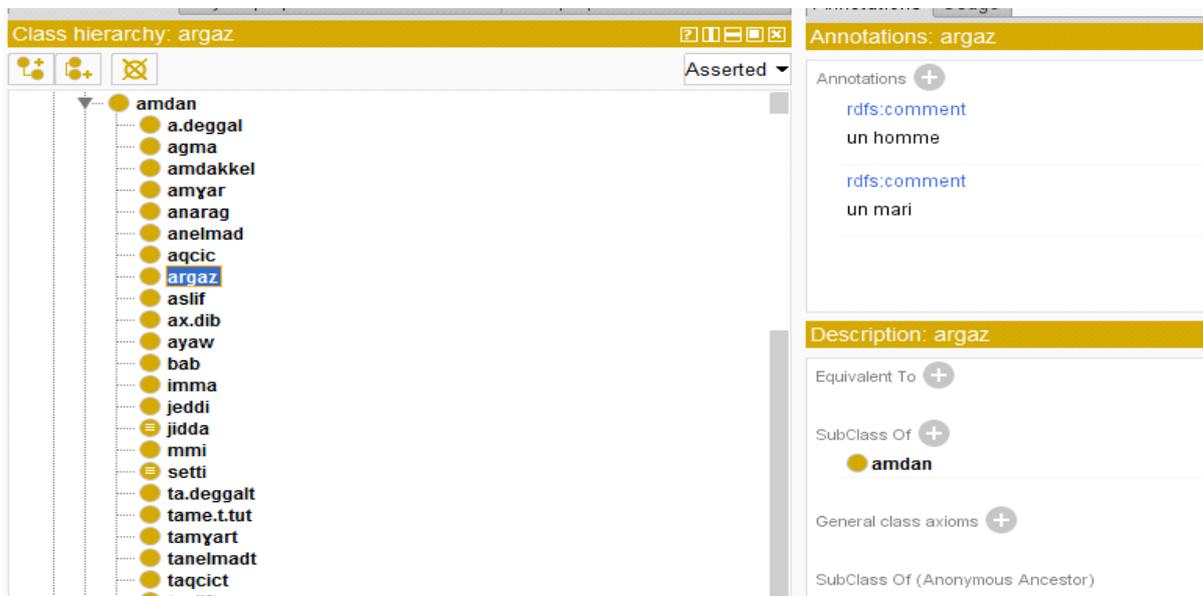


Figure 11: la relation rdfs:comment sous protège.

#### 4.5.3.4. Equivalence

C'est une relation d'égalité et pour avoir ou bien créer cette relation, on utilise le mot clé **Equivalent To**.

Par exemple : anebdu **Equivalent To** awil

Car le nom « **anebdu** » signifie la même chose que le nom « **awil** » qui est **l'été** en français.

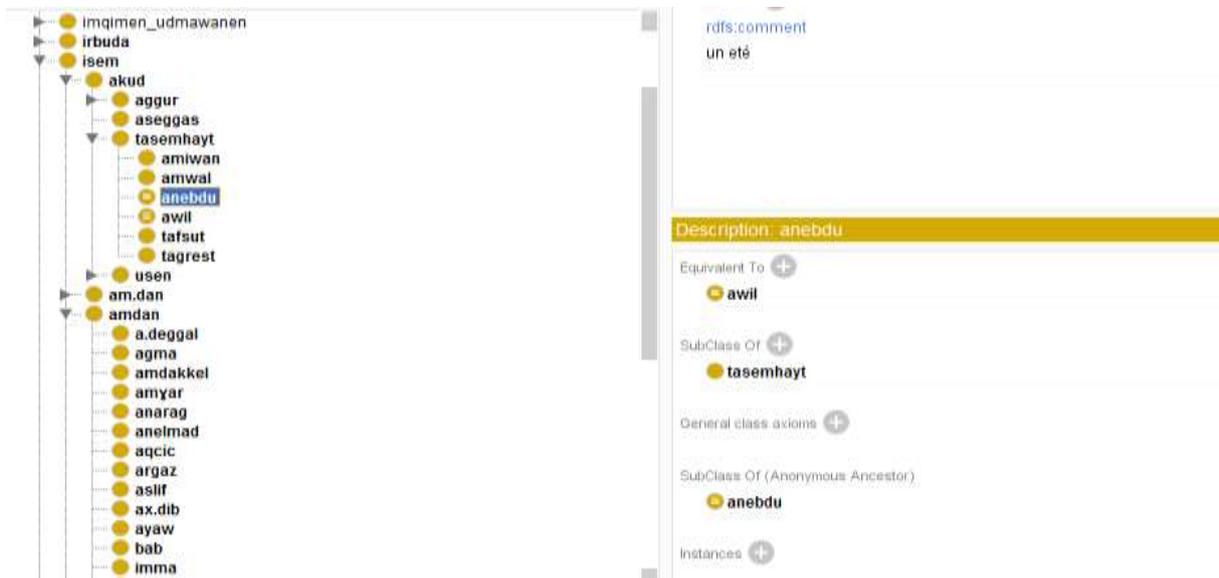


Figure 12: La relation d'équivalence sous protégé.

#### 4.5.3.5. Quelques éléments statistiques sur le contenu de l'ontologie

A l'état actuel du développement de l'ontologie, nous avons intégré au format RDF :

- Plus de 1600 concepts (180 verbes, 943 noms et les autres sont des adverbes...).
- Plus de 5400 relations (conceptuelles, ontologiques, sémantiques, etc.).
- Quelques contraintes types.

#### Metrics

Axiom	5469
Logical axiom count	1849
Declaration axioms count	1614
Class count	1614
Object property count	0
Data property count	0
Individual count	0
Annotation Property count	2

Figure 13: Statistiques de notre application données par protégé

#### 4.5.4. Intégration de l'ontologie dans Android studio

Après cela, nous avons procédé à un export en RDF de notre ontologie, facilement intégrable dans notre IDE Android studio grâce aux bibliothèques offerte par Androjena qui donne l'accès à notre ontologie.

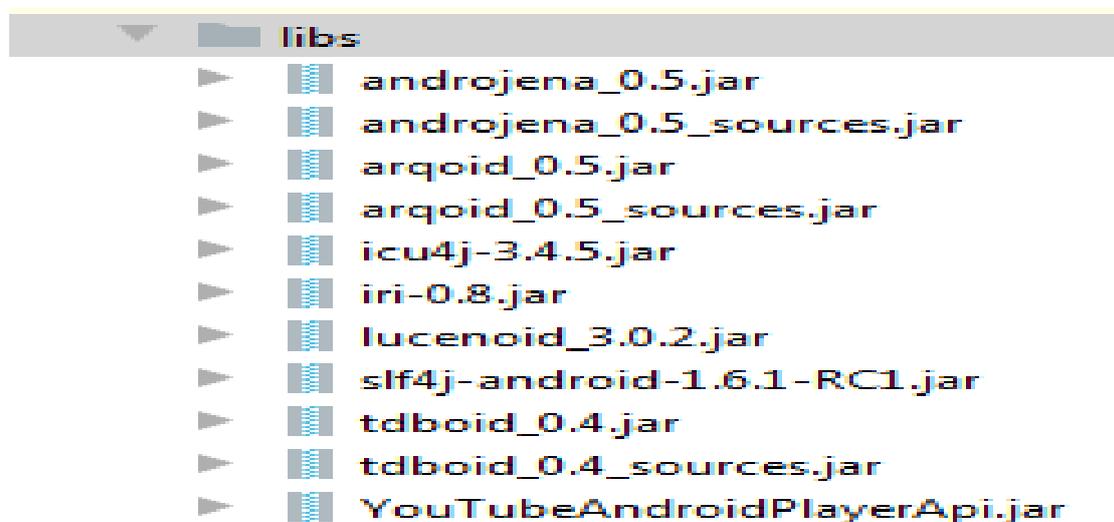


Figure 14: Les bibliothèques Androjena.

#### 4.5.5. Développement de l'application sous Android studio

La reconnaissance des mots nécessite une analyse morphologique de façon à retrouver la forme de base à partir d'une forme fléchie. Ce passage d'une forme fléchie (telle qu'elle est présentée dans le texte) à sa forme de base (telle qu'elle est enregistrée dans l'ontologie) n'est pas évident et difficilement maîtrisable pour une langue complexe comme tamazight. Nous avons réalisé certain exemplaires de ce passage grâce à des algorithmes en java.

Pratiquement parlant, les noms et les verbes sont les classes majeures qui jouent un rôle prépondérant dans toute la morphologie amazighe. Pour cela nous nous sommes basées sur ces deux catégories.

##### 4.5.5.1. La morphologie amazighe

La langue amazighe se caractérise par une morphologie complexe et productive. Elle est riche en termes de flexion, de dérivation, et de complexité qu'elle produit. Lors de la programmation nous nous sommes focalisé sur la dérivation flexionnelle.

La morphologie flexionnelle est un procédé 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 [29].

*La langue tamazight présente 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 postposer [29]. Généralement, ces marques flexionnelles permettent de distinguer : flexion des noms et flexion des verbes.*

### 1) Flexion des noms

En tamazight, 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/construit) (voir chapitre 2).

- **Genre :** on a intégré dans notre ontologie des noms masculins et certains noms féminins qui n'admettent pas un masculin. Quant aux les noms féminins réguliers, nous les avons traité avec un algorithme. Voici un exemple :

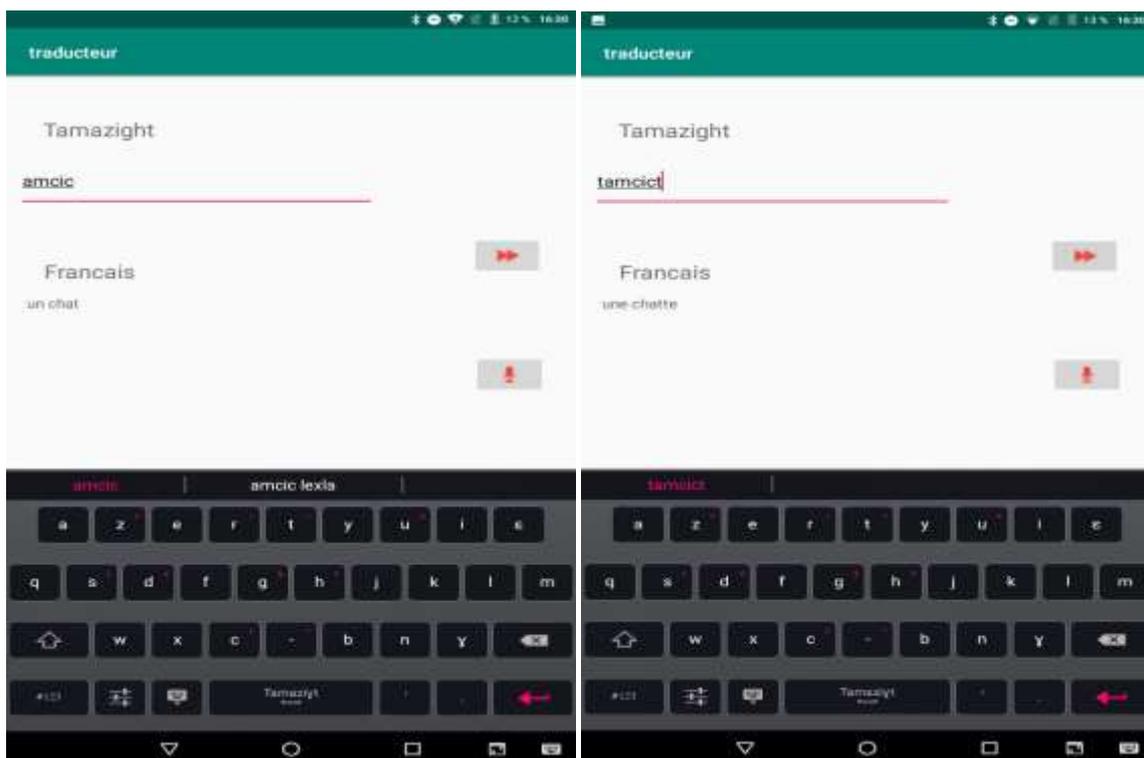


Figure 15: Exemple de la traduction d'un nom féminin déduit à partir de son masculin.

- **Nombre :** on a intégré dans notre ontologie des noms singuliers et certains noms pluriels qui n'admettent pas un nom singulier. Et comme nous l'avons dit dans le 2<sup>ème</sup> chapitre, le pluriel se divise en 3 catégories : externe, Interne ,et mixte. Nous avons réussi à créer des algorithmes pour les deux premières catégories.

### - Pluriel externe

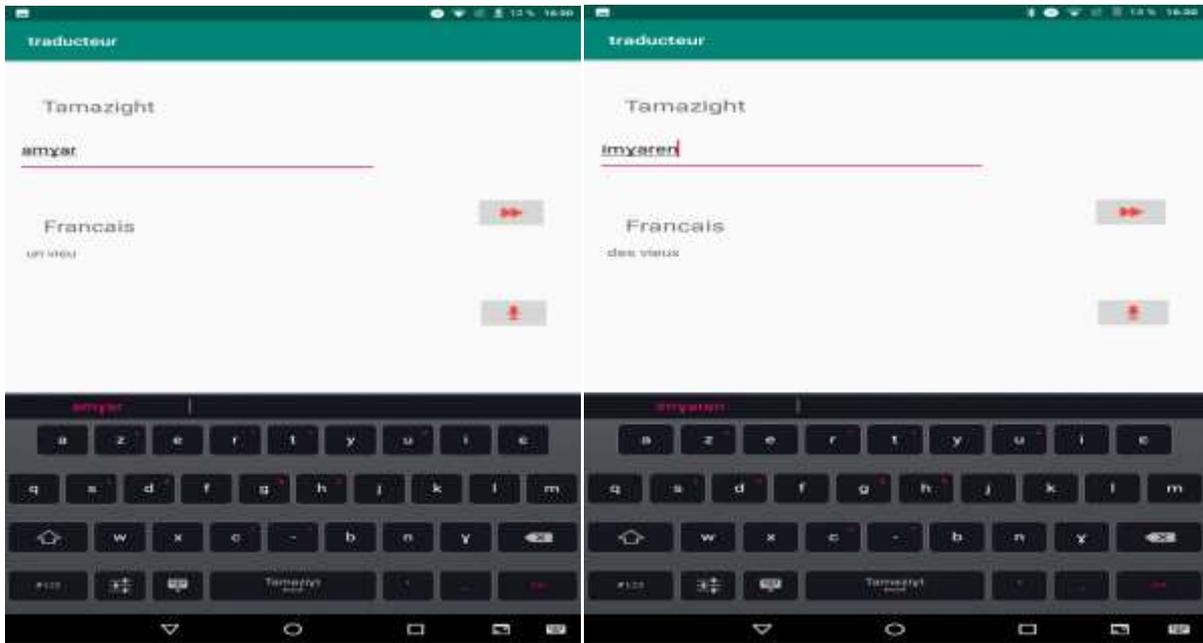


Figure 16: Exemple de la traduction d'un nom en pluriel externe déduit à partir de son singulier.

### - Pluriel Interne

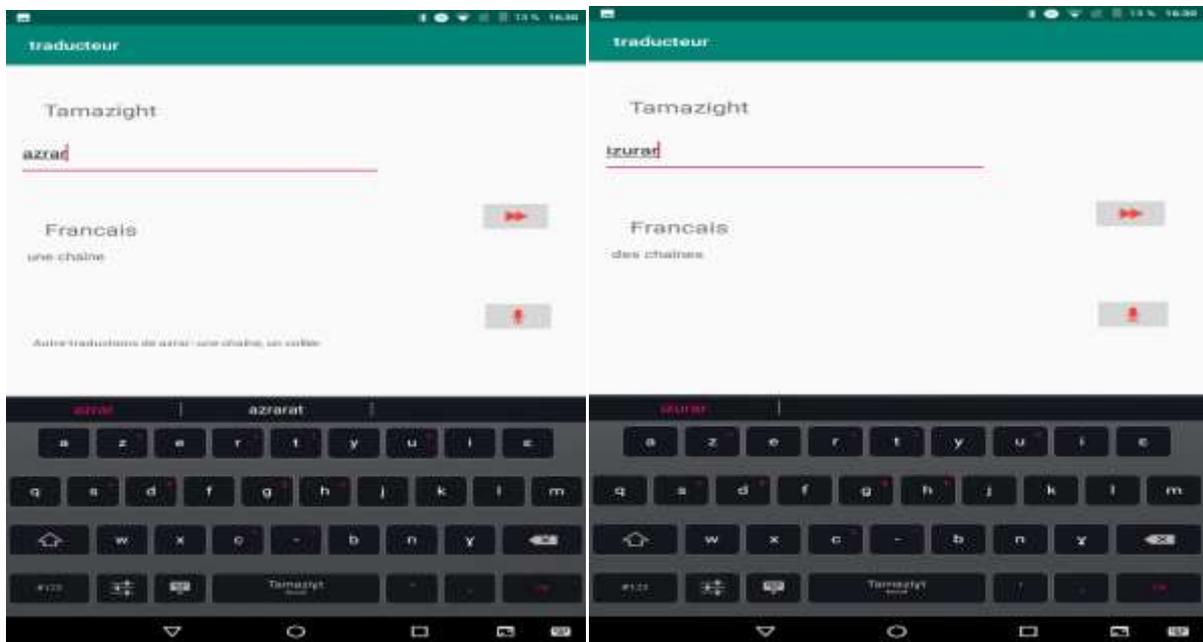


Figure 17: Exemple de la traduction d'un nom en pluriel interne déduit à partir de son singulier.

- État d'annexion :

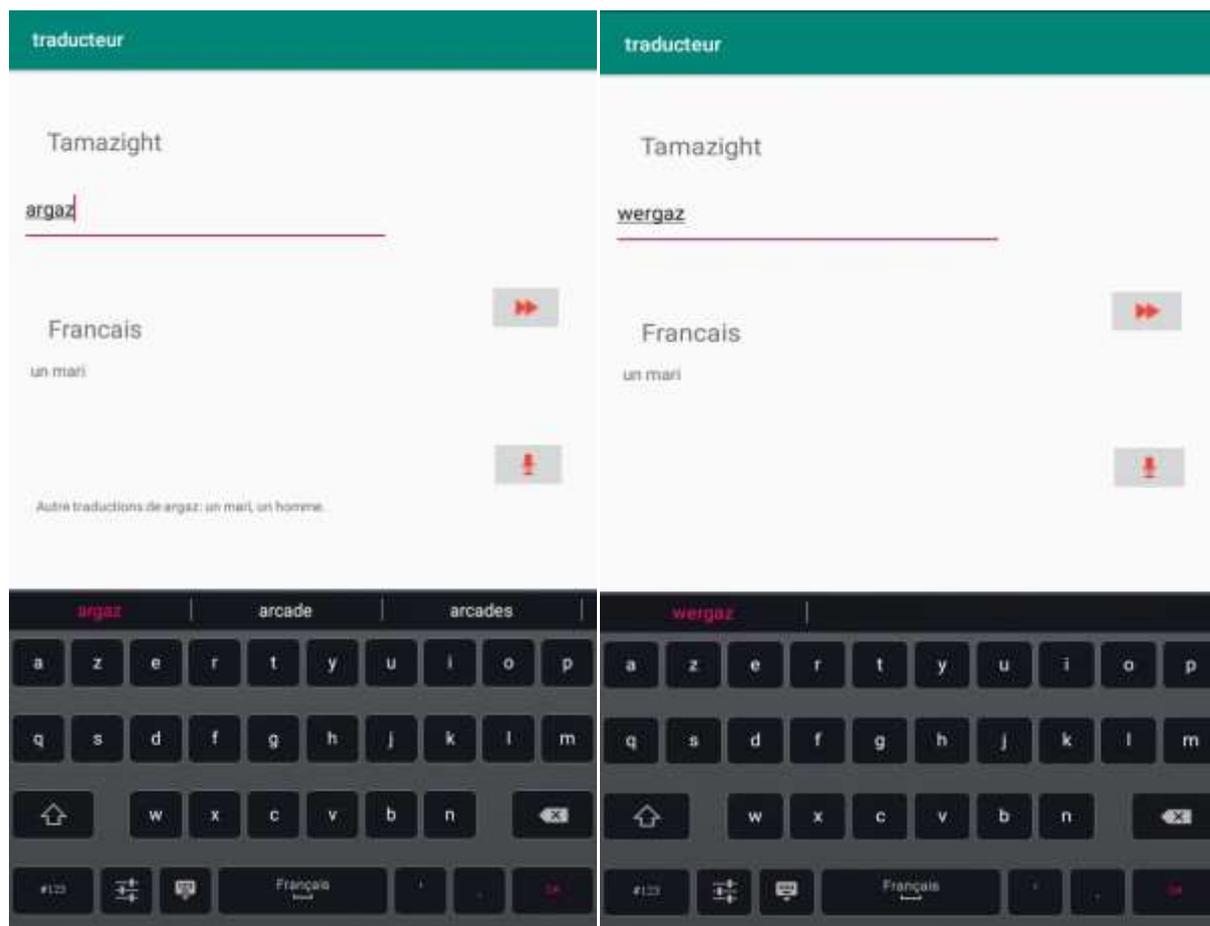


Figure 18: Exemple de la traduction d'un nom en état d'annexion déduit à partir d'un état libre.

2) Flexion des verbes

La forme simple (radical) est formée par la fusion d'une racine et d'un schème.

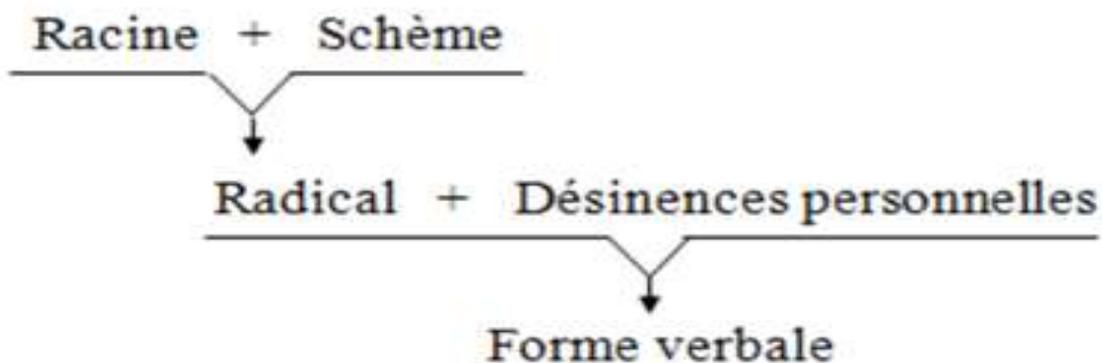


Figure 19: La forme verbale en Tamazight

Le verbe se conjugue selon 4 temps: passé, future, présent et l'impératif.

- **L'impératif**

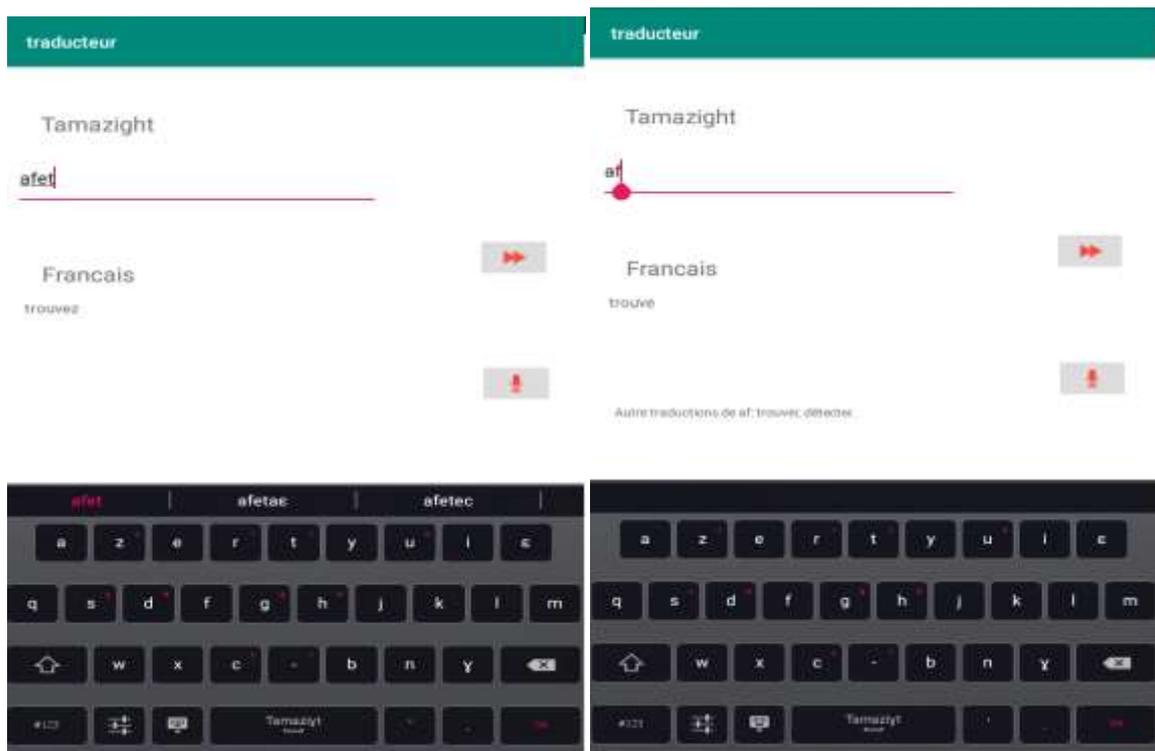


Figure 20: Exemple de la traduction d'un verbe à l'infinitif et sa conjugaison à l'impératif.

- **Le futur :**



Figure 21: Exemple de la traduction d'un verbe conjugué au futur.

- **Le passé composé :**



Figure 22: Exemple de la traduction d'un verbe conjugué au passé composé.

**La traduction de quelques phrases simples :**

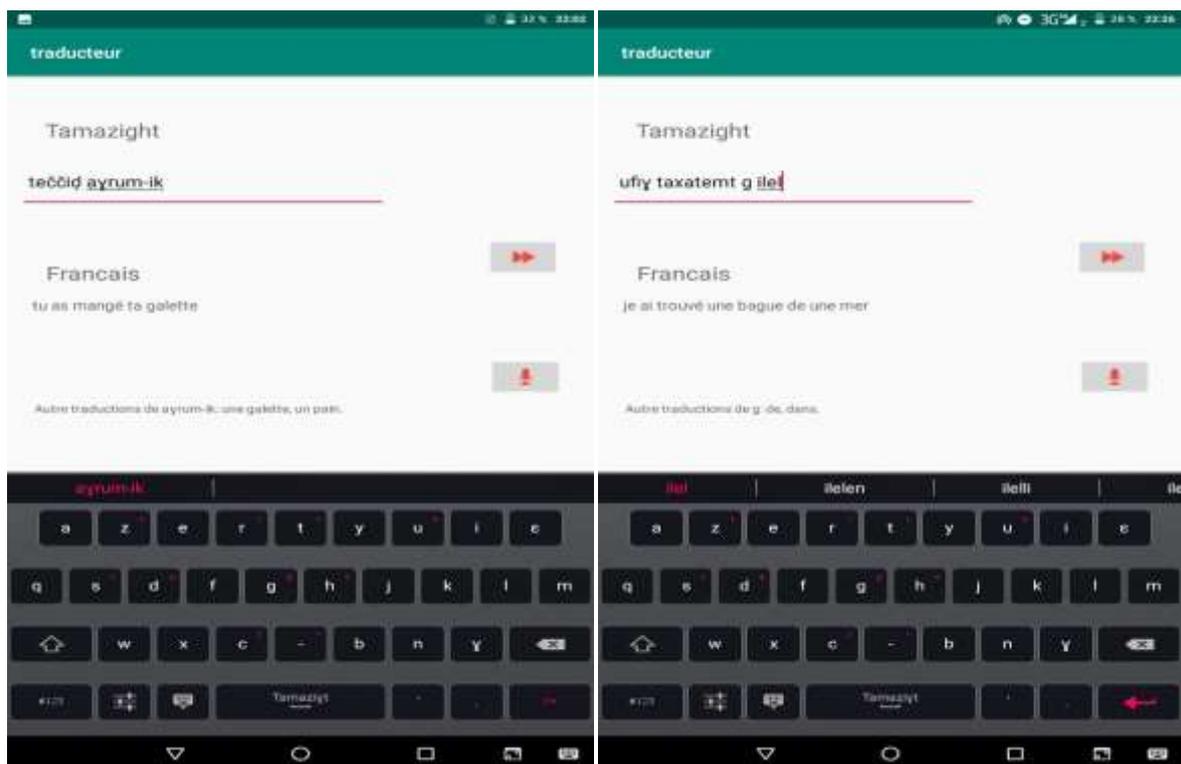


Figure 23 : Exemple de la traduction de quelques phrases simples.

### 3) Pronom personnel affixe :

Comme nous l'avons déjà cité dans le 2<sup>ème</sup> chapitre, les pronoms affixes sont rattachés à un élément de la phrase. Cet élément peut être un nom, un verbe ,ou une préposition. Nous n'avons créé des algorithmes que pour les pronoms rattachés au nom et à la préposition.

- **Affixe d'un nom**

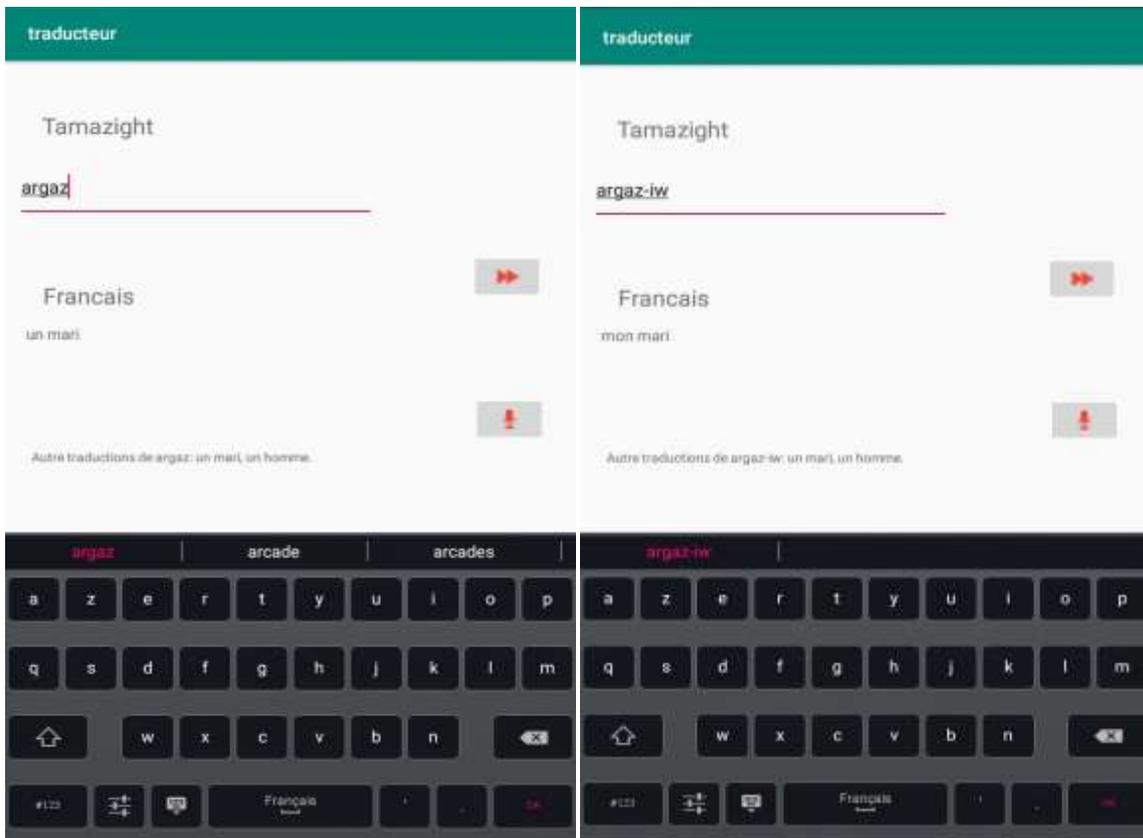
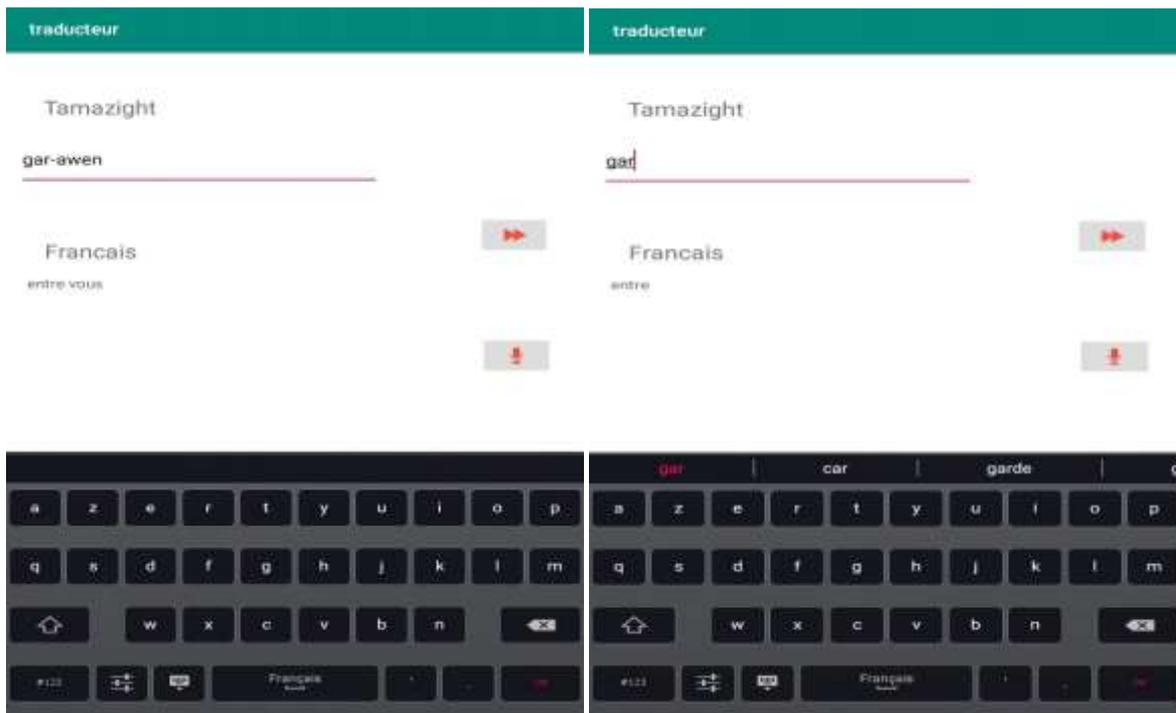


Figure 24: Exemple de la traduction d'un nom avec un pronom affixe

- **Affixe d'une préposition**



*Figure 25: Exemple de la traduction d'une préposition liée à un pronom affixe*

#### 4.6. Conclusion

Dans ce chapitre, nous avons présenté le cadre technique de réalisation de notre traducteur Tamazight – Français. L'élaboration d'une ontologie est une tâche complexe et difficile et notamment lorsqu'il s'agit de traiter des concepts d'une langue dont les références bibliographiques sont très rares. Néanmoins, nous avons pu, malgré toutes ces difficultés, élaborer une petite ontologie qui peut au moins daider à donner un exemple pour la réalisation d'un traducteur pour notre langue maternelle, qui est basée sur des concepts ayant pour corollaire la densité symbolique et culturelle. Nous avons donc intégré cette ontologie des mots, et avec l'aide de certain algorithmes que nous avons créés en java ,notre application est désormais capable de traduire autant de mots que possibles.

# Conclusion et perspectives

## Conclusion et perspectives

Le but principal convoité par ce mémoire était de concevoir un traducteur unidirectionnel (tamazight-français) qui fonctionne sous Android.

D'autres travaux de recherche seraient d'une grande utilité pour la langue amazighe. Ces travaux peuvent concerner deux champs :

Le premier concernera la création des ressources linguistiques, avec notamment l'enrichissement du corpus annoté morpho-syntaxiquement en le complétant par d'autres types d'annotations syntaxiques et sémantiques, la création d'une ontologie amazighe, la création des corpus de textes parallèles, etc.

Le deuxième champ de recherche concernerait la création d'outils de base et d'outils avancés pour le TAL. A titre d'exemple, un racineur, un analyseur morphologique ; un analyseur syntaxique partiel et syntaxique ; des outils de traduction et d'aide à la traduction ; la reconnaissance de la parole ,et la synthèse de la parole, etc.

Les principales contributions de notre travail touchent deux champs distincts : les ressources linguistiques et les méthodes et outils du TAL.

Dans le premier champ nous avons élaboré un modèle d'ontologie des mots Tamazight (langue source) et leurs significations en français (langue cible), lors de la mise en place de l'ontologie des mots Amazigh, nous avons rencontré plusieurs difficultés dont les plus importants sont :

- 1- Nous nous sommes rendues compte que des données sources, que ce soit des données brutes y compris papier, ou des données électroniques, sont très rares (Absence de sources de données, et seulement quelques publications).
- 2- D'autres problèmes sont liés à la création de nouveaux contenus :
  - Mots, concepts, etc.
  - Relations.
  - Problèmes de traduction des mots.

- 3- Absence de certains alphabets dans le logiciel que nous avons utilisé pour créer l'ontologie.
- 4- Absence des outils TAL pour cette langue.

Comme la majorité des langues qui n'ont commencé les recherches en TAL que récemment, la langue amazighe souffre encore de la pénurie d'outils et de ressources pour son traitement automatique. Malgré l'insuffisance des références traitant la langue tamazight, nous avons réussi à construire le premier corpus manuellement. Il contient environ 1600 mots. Cette ressource, même si elle est de petite taille, elle est très utile pour représenter un exemple d'une ontologie bilingue.

Quant aux méthodes et outils pour le TAL, étant donné la rareté des outils TAL de Tamazight et afin de réaliser notre traducteur nous avons besoin au minimum d'un racineur et d'un analyseur morphologique pour que notre traducteur puisse connaître certains mots dérivés. Et comme nous n'avons pas trouvé ces outils de bases, on était obligées de suivre la méthode ancienne qui est la création des algorithmes. Pour cela, nous avons réalisé plusieurs algorithmes pour le traitement de la langue tamazight, tels que celui qui traite les noms féminin les pluriels réguliers, et les verbes conjugués, etc.

Les perspectives de nos travaux sont multiples :

Notre ontologie est actuellement enregistrée sur le Smartphone ce qui conduit à une perte de temps et d'espace. Pour cela, il faut que cette dernière soit enregistrée sur un serveur et sera l'enrichir avec plus de concepts et de relations sémantiques.

On n'est pas arrivées à créer des algorithmes pour le pluriel mixte, la conjugaison au présent, et les pronoms affixe rattachés à un verbe, et des algorithmes pour reconnaître les mots composés et des phrases plus complexes. Par ailleurs, nous souhaitons aussi réduire la complexité des algorithmes qu'on avait déjà créés.

Les mots mal écrits dans les textes : dans le but de renforcer la robustesse de notre analyseur, nous aurions mis en place une série d'heuristiques qui prennent en compte l'ensemble des problèmes les plus fréquents de manière à limiter les répercussions de leurs non reconnaissance sur les tâches ultérieures d'analyse.

Nous souhaitons rendre notre traducteur compatible avec plusieurs plateformes mobiles et qu'il devienne bidirectionnel.

Cette contribution n'est qu'une esquisse qui va être développée ultérieurement tout en prenant en considération les différentes fonctions et relations lexicales. En effet ce travail étant un premier essai, n'est donc pas un modèle unique et parfait. C'est pourquoi nous restons ouverts à toutes les critiques et sommes prêts à recevoir toutes les suggestions et remarques tendant à améliorer davantage cette initiative.

# **Bibliographie**

## Bibliographie

[1]: Mohamed Akramhanini, Mondher Nacer. Conception et réalisation d'une interface graphique pour visualiser une ontologie. Mémoire de master. Université de Larbi Tébessi – Tébessa 29/05/2016.

[2]: en ligne. Consulté en juin 2020

<http://tomgruber.org/writing/AIMag12-03-004.pdf>

[3]: Souheila Khalfi. Construction d'une ontologie pour la prise en charge des patients à domicile .Mémoire de Magister. Option Systèmes distribués. Université Mentouri Constantine. Le 28/10/2009.

[4]: en ligne. Consulté en juin 2020

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.8406&rep=rep1&type=pdf>

[5]: Uschold, M., & Grüninger, M. (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11 (2), 93–155.

[6]: Roche, C., 2005. Terminologie et ontologie. Langages 157, 4862. Doi: 10.3917/lang.157.0048

[7]: Hacine Gherbi Construction d'une Ontologie pour le WEB Sémantique. Mémoire de Magister, École doctorale d'informatique, Université de Setif, 2014.

[8]: Khadim Drame. Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical. Autres [stat.ML]. Université de Bordeaux, 2014.

[9]: Uschold, M., & King, M. (1995). Towards a Methodology for Building Ontologies. In D. Skuce (Ed.), IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing, (pp. 6.1-6.10). Montreal, Canada.

[10]: Fernandez, M., Gomez-Pérez, A., Juristo, N. (1997) Methontology: From Ontological Art Towards Ontological Engineering. Proceedings of the AAAI-97 Spring Symposium Series on Ontological Engineering, Stanford, CA, USA 1997, p 33-40.

[11]: Farquhar, A., Fikes, R., Rice, J., 1997. The Ontolingua Server: a tool for collaborative ontology construction. Int J Hum-Comput Stud 46, 707–727.

[12]: Samer Abdul Ghafour., 2003-2004. Méthodes et outils Pour l'intégration des ontologies. Mémoire de stage de DEA « Documents Multimédia, Images, et Systèmes d'Information

Communicants ». Laboratoire d'Informatique en Images et Systèmes d'information LIRIS FRE2672 CNRS, INSA Lyon, UCB Lyon 1, EC Lyon ,57 p.

[13]: Henry Teguiak. Construction d'ontologies à partir de textes : une approche basée sur les transformations de modèles. Autre. ISAE-ENSMA Ecole Nationale Supérieure de Mécanique et d'Aérotechnique - Poitiers, 2012. Français.

[14]: Soraya Zaidi–Ayad. Une plateforme pour la construction d'ontologie en arabe : Extraction des termes et des relations à partir de textes (Application sur le Saint Coran). Thèse doctorat en Informatique. université de badjimokhtar –Annaba , 2012/2013.

[15]: F. Kessai. Élaboration d'un dictionnaire électronique de berbère avec annotations étymologiques. Thèse de doctorat, Université Sorbonne Paris Cité, France, 2018.

[16]: Les travaux scientifiques de Mouloud MAMMERI :

<http://mammeri100.dz/index.php/fr/un-homme-des-oeuvres/oeuvres-anthropologiques/161-les-travaux-scientifiques-de-mouloud-mammeri-dda-lmulud-a-ouvert-des-voies-a-chacun-de-les-decouvrir>

Consulté le 06/06/2020.

[17]: Alphabet berbère latin

[https://fr.wikipedia.org/wiki/Alphabet\\_berb%C3%A8re\\_latin#Sell%C3%A8s2014](https://fr.wikipedia.org/wiki/Alphabet_berb%C3%A8re_latin#Sell%C3%A8s2014)

Consulter en avril 2020.

[18]: A. Hanoteau, Essai de grammaire kabyle

[https://books.google.dz/books?hl=fr&lr=&id=i6ITAAAAQAAJ&oi=fnd&pg=PA1&dq=grammaire+de+berb%C3%A8re&ots=nw0uz3XPVh&sig=8UW3G4tOMN\\_Zv7v8sfdrCABPd-o&redir\\_esc=y#v=onepage&q=grammaire%20de%20berb%C3%A8re&f=false](https://books.google.dz/books?hl=fr&lr=&id=i6ITAAAAQAAJ&oi=fnd&pg=PA1&dq=grammaire+de+berb%C3%A8re&ots=nw0uz3XPVh&sig=8UW3G4tOMN_Zv7v8sfdrCABPd-o&redir_esc=y#v=onepage&q=grammaire%20de%20berb%C3%A8re&f=false)

Consulté en avril 2020.

[19]: J.-M. Dallet Dictionnaire kabyle-français,

<https://books.google.dz/books?id=nZdP85wiDmcC&pg=PA490&lpg=PA490&dq=amagrama n&source=bl&ots=H5J7h927mW&sig=ACfU3U3wV6l8EyIt7Mx0q915w1JefYErpA&hl=fr&sa=X&ved=2ahUKEwis45rfq-rpAhUO3BoKHQxnASYQ6AEwFnoECAoQAQ#v=onepage&q=amagraman&f=false>

Consulté en mai 2020.

[20]: CarolineLavecchia. Les Triggers Inter-langues pour la Traduction Automatique Statistique. Informatique [cs]. Université Nancy II, 2010.

[21]: Thingocdiep Do. Extraction de corpus parallèle pour la traduction automatique depuis et vers une langue peu dotée. Autre [cs.OH]. Université de Grenoble; Université de Hanoi – Vietnam, 2011.

[22]: Elena-mirabelaNavlea, La traduction automatique statistique factorisée: une application à la paire de langue français-roumain. Linguistique. Université de Strasbourg, 2014.

[23]: Imane Taghbalout, Fadoua Ataa Allah, Mohamed Elmarraki; Towards an Amazigh UNL dictionary, Recent Advances in Electrical Engineering and Educational Technologies, pp. 129-132, 2014.

[24]: Consulté le 08/08/2020

[.https://www.supinfo.com/articles/single/6135-top-10-ide-developpeurs-java](https://www.supinfo.com/articles/single/6135-top-10-ide-developpeurs-java).

[25]: Consulté le 08/08/2020

[https://fr.wikipedia.org/wiki/Jena\\_\(framework\)](https://fr.wikipedia.org/wiki/Jena_(framework))

[26]: Amélie Gyrard. Concevoir des applications Internet des Objets Sémantiques inter-domaine. Thèse de doctorat, école de l'Institut Télécom - membre de ParisTech, 24 Avril 2015.

[27]: [En ligne].Consulté le 08/08/2020

<https://fr.yeeply.com/blog/langages-de-programmation-creerune-application-android/>

[28]: [En ligne] .Consulté le 08/08/2020

<https://www.companeo.com/applicationsmobiles/guide/creation-application-mobile>.

[29]: Nejme Fatima Zahra .Traitement automatique de la langue amazighe : Construction d'outils et de ressources linguistiques pour l'analyse morpho-lexicale automatique des textes en amazighe. Thèse de doctorat .Faculté des Sciences, Maroc.14 juillet 2017

[30]: En ligne .consulté le 05 /08/2020

<http://www.jarrar.info/publications/J11.pdf>

[31]: François-Régis Chaumartin. Wordnet et son écosystème : un ensemble de ressources linguistiques de large couverture. Colloque BD lexicales, Montréal, Canada, 2007.

## **Résumé**

Comme la plupart des langues qui n'ont que récemment commencé les investigations en Traitement Automatique des Langues (TAL), la langue amazighe est peu dotée en ressources et outils du TAL. Dans ce sens, l'un des objectifs principaux de ce mémoire est de doter cette langue d'un Traducteur Automatique, ou comment traduire d'une langue source vers une langue cible sans aucune intervention humaine. Nos travaux se sont plus particulièrement concentrés sur l'utilisation des ontologies.

La création d'un traducteur est une tâche difficile et afin d'atteindre cet objectif, nous avons formé une ontologie comme étant un dictionnaire bi-langue informatisé. Ces concepts de notre ontologie ne sont pas tous extraits des dictionnaires existants. Nous avons eu recours à l'utilisation de ressources externes pour construire un corpus d'environ 1600 mots. Ensuite, nous avons élaboré des algorithmes pour l'enrichissement de ce traducteur par des mots dérivés. Enfin, un programme générique exploitant cette ontologie permet une traduction des mots et de certaines phrases.

**Mot clés :** ontologie, OWL, Traducteur Tamazight, Traduction Automatique.

## **Abstract**

Like most languages which have only recently started investigations in Language Processing (LP), the Amazigh language is poorly endowed with LP resources and tools. In this sense, one of the main objectives of this report is to provide this language with an Automatic Translator, or how to translate from a source language to a target language without any human intervention. Our work was focused more particularly on using ontology.

Creating a translator is a difficult task, and in order to attain this object, we have formed an ontology as a computerized bi-language dictionary. The concepts of our ontology are not all taken from existing dictionaries. We resorted to the use of external resources to build a corpus of around 1600 words. Next, we developed algorithms for enriching this translator with some derived words. Finally, we write a generic program exploiting this ontology to translate some words and even sentences.

**Key words:** ontology, OWL, Amazigh Translator, Automatic Translation.