

République Algérienne Démocratique et Populaire
Ministère de L'Enseignement Supérieur et
De la Recherche Scientifique
Université Abderrahmane Mira – Bejaïa-



Faculté des Lettres et des Langues
Département de français

Mémoire de master

Option : Sciences du langage.

**Vers un prototype de segmentation des termes médicaux à composition
savante.**

Présenté par :

M^{elle} MOHELLEBI Sara.
M^{elle} MEZHOUD Amina.

Le jury :

Président : Mme. TOUATI Radia ,
Directeur : Mr. BEDDAR Mohand,
Examineur :Mr. SEGHIR Atmane,

2019/2020

Remerciements

Merci à dieu le tout puissant qui nous a donné du courage, et nous a maintenu en bonne santé pour pouvoir réaliser ce travail.

Nous tenons à remercier notre directeur de recherche BEDDAR Mohand pour ses conseils, ses orientations et remarques pertinentes, mais surtout pour sa disponibilité malgré cette crise sanitaire.

Nos vifs remerciements pour les membres du jury pour l'intérêt qu'ils ont porté à notre modeste travail en acceptant de l'examiner et l'évaluer.

Nous remercions également notre ami AMARI Nassim qui nous a soutenues matériellement, et HADDAD Mohamed Abdelmalek pour ses conseils et aide dans la partie informatique. Nous les remercions également pour tous les encouragements.

A nos familles MOHELLEBI/ABID et MEZHOUH/MAKHLOUFI, nous vous remercions profondément pour vos encouragements et votre soutien.

Dédicaces

Je dédie ce modeste travail,

A toi maman, ma guerrière, femme aux yeux couleur de force,

A vous Nacer-eddine et Yamina, ma source de courage et de persévérance,

A toi Abdelmalek, mon très cher,

A Marouan et Melissa, mes meilleurs amis,

A toute ma famille,

A tous mes amis et famille que je n'ai pas cités ici,

A toi tonton Chérif Abid, mon deuxième papa, que dieu bénisse ton âme.

Sara.

Dédicaces

Je dédie Ce modeste travail à

Mes chers parents: Khaled, Sadika

Que nulle dédicace ne puisse exprimer mes sincères sentiments, pour leur encouragement,
pour leur patience illimitée, et leurs grands sacrifices

Mes chers frères Zidane et Hakim

Ma chère sœur : Sarah

A tous mes amis (e) :

Amari Nassim

Zenati Adeline

Khireddine Yacine

A tous les étudiants de la promotion 2019/2020.

A tous ceux qui, par mot, m'ont donné la force et le courage dans la réalisation de mon
travail.

Amina.

Sommaire d'abréviations

ALPAC= Automatic language processing advisory committee.

CL=Computational linguistics.

CRSTDLA=Centre de recherche scientifique et technique pour le développement de la langue arabe.

DIU=Dispositif intra-utérin.

IBM=International business machines corporation.

MIT=Massachussets institute of technology.

NLP= Naturel languageprocessing.

OMS= Organisation mondiale de la santé.

TAL= Traitement automatique des langues.

TLFI= Trésor de la langue française informatisé.

VIH= Virus de l'immunodéficience humaine.

Sommaire des figures

Figure 1 : Création de l'interface par QtDesigner.

Figure2 : Fonctionnement de l'application.

Figure3 : Structure de l'application.

Figure4 : Création de l'interface de l'application.

Figure5 : Algorithmes représentants de création de l'application.

Figure6 : Création de la fenêtre d'ajout de mots.

Figure7 : Création de la fenêtre d'ajout de préfixes

Figure8 : Création de la fenêtre des radicaux.

Figure9 : Création de la fenêtre des suffixes.

Figure10 : Fenêtre principale.

Figure11 : Fenêtre d'ajout de mots.

Figure12 : Fenêtre d'ajout de préfixes.

Figure13 : Fenêtre d'ajout de suffixes.

Figure14 : Fenêtre d'ajout de radicaux.

Figure15 : Fenêtre principale affichant les résultats.

Sommaire des tableaux

Tableau 1 : Segmentation des composés savant/hybrides

Tableau2 : Dictionnaires des termes.

Tableau3 : Dictionnaires des préfixes.

Tableau4 : Dictionnaires des radicaux.

Tableau5 : Dictionnaires des suffixes.

Sommaire

Introduction générale	12
Présentation du sujet.....	Erreur ! Signet non défini.
Motivation et choix du sujet.....	Erreur ! Signet non défini.
Problématique.....	Erreur ! Signet non défini.
Hypothèse.....	Erreur ! Signet non défini.
Méthodologie et corpus.....	Erreur ! Signet non défini.
Conclusion	Erreur ! Signet non défini.
Chapitre I : Langue de spécialité	16
1. La langue.....	Erreur ! Signet non défini.
2. Langue générale et langue de spécialité.....	17
3. Qu'est-ce que la terminologie ?.....	19
3.1. Définition du terme.....	21
4. La terminologie et les langues de spécialités.....	22
5. Domaines médicaux.....	Erreur ! Signet non défini.
5.1.1. La psychiatrie.....	Erreur ! Signet non défini.
5.1.2. La neurologie.....	Erreur ! Signet non défini.
5.1.3. La cardiologie.....	Erreur ! Signet non défini.
Conclusion.....	22
Chapitre II : Procédés de formation des termes médicaux	24
Première partie.....	25
1. La lexicologie.....	25
1.1. Aperçu historique de la lexicologie.....	26
1.2. La notion de mot.....	26
1.2.1. Le mot graphique.....	26
1.2.2. Le mot linguistique (lexème).....	26
2. L'origine de la terminologie médicale.....	26
3. La morphologie.....	27
3.1. Le morphème.....	27
3.1.1. Morphèmes grammaticaux.....	27
3.1.1.1. Les morphèmes liés.....	27
3.1.1.1.1. Morphèmes dérivationnels.....	28

3.1.1.1.2. Morphèmes flexionnels	28
3.1.1.2. Les morphèmes libres	28
3.1.2. Morphèmes lexicaux	28
Deuxième partie.....	28
1. Formation du lexique français	28
1.1. Les contextes de formation.....	28
1.1.1. Le contexte de spécialisation.....	29
1.1.2. Le contexte de vulgarisation.....	29
1.1.3. Le contexte du quotidien :	29
2. Les différents procédés de formation du lexique français, et leur utilisation dans la terminologie médicale	29
2.1. L'emprunt	29
2.1.1. L'emprunt direct :	29
2.1.2. Le calque	30
2.1.3. L'emprunt sémantique.....	30
2.2. Le xénisme ou pérégrinisme.....	30
2.3. Le néologisme, et néologisme sémantique	31
2.4. La siglaison.....	31
2.5. L'acronyme.....	32
2.6. La troncation.....	32
2.7. La dérivation.....	33
2.7.1. La suffixation	33
2.7.2. La préfixation	33
2.7.3. La parasynthétisation.....	33
2.8. La composition	33
2.8.1. Les types de composition	34
Troisième partie.....	35
Approche morphologique.....	36
Conclusion.....	39
Chapitre III : Traitement Automatique des Langues	40
1. Introduction	41
2. Traitement automatique des langues, qu'est-ce que c'est ?	41
3. Les dénominations de TAL	41
4. Aperçu historique de la discipline	42

5.	Qu'est-ce que la programmation	43
5.1.	Les langages de programmation	43
5.1.1.	Le java	43
5.1.2.	Le Python	43
6.	Champ de recherche en TAL.....	44
6.1.	Traitement documentaire	44
6.1.1.	La traduction automatique.....	44
6.1.2.	Correction orthographique	44
6.1.3.	Résumé automatique des textes.....	44
6.1.4.	Génération automatique des textes.....	44
6.2.	Le traitement syntaxique	44
6.2.1.	La racinisation/stemmatisation.....	44
6.2.2.	La morphologie	45
6.2.3.	La lemmatisation	45
6.2.4.	Étiquetage morpho-syntaxique.....	45
6.3.	Extraction de l'information	45
6.3.1.	Reconnaissance d'entités nommées	45
6.3.2.	Fouille de texte	45
6.3.3.	Recherche d'informations	45
7.	Les difficultés de TAL	45
7.1.	Ambiguïté du langage naturelle.....	45
7.2.	L'implicite	46
	Conclusion.....	46
	Chapitre III	47
	Formalisation linguistique	47
1.	Introduction	48
2.	Définition de la modélisation linguistique	48
3.	La logique du fonctionnement du code du prototype.....	48
4.	Modèle de segmentation des termes à composition savante et hybride.....	49
4.1.	Dictionnaire des termes	50
4.2.	Dictionnaire des préfixes	50
4.3.	Dictionnaire des radicaux	51
4.4.	Dictionnaire des suffixes	51

5.	Présentation de l'application	52
6.	Outils informatique utilisés	52
6.1.	Le langage de programmation Python.....	52
6.2.	PyCharm	53
7.	Résultats	53
7.1.	Programme principal	53
8.	Description de l'interface :.....	74
8.1.	Fenêtre principale	74
8.2.	Fenêtre d'ajout de mots	75
8.3.	Fenêtres d'ajout de morphèmes et leurs définitions	76
8.4.	Fenêtre principale affichant les résultats	77
	Conclusion	77
	Conclusion générale	79
	Bibliographie	81
	Annexe	86

Introduction générale.

La linguistique et la terminologie sont deux disciplines carrefours , étant donné qu'elles travaillent sur la langue. Cependant, la terminologie consiste en l'étude du recueil, de l'utilisation voire la création des termes, qui sont propres un à vocabulaire caractérisant les différentes sciences existantes dans le monde scientifique.

La terminologie fait que la langue ait des spécificités uniques d'une spécialité à une autre, et d'une science à une autre, c'est pour cela que nous parlions d'une langue de spécialité, ou langue spécialisée.

Le débat se lève à propos de cette notion de langue de spécialité, spécialisée, ou encore langues spécialisées. Dans notre travail de recherche, nous adopterons celle de langue de spécialité.

La médecine est une science qui a pour objet la protection et le rétablissement de la santé de l'être humain, elle a sa propre terminologie qui diffère de celles des autres sciences, car elle requiert un ensemble de termes spécifiques ; et non de mots « grammaticaux ». Ils sont au fait, des mots purement « lexicaux », ils peuvent être que de nature adjectivale, nominale, adverbiale ou encore verbale, bien spécifiques. Le terme, étant un concept, a de différentes définitions qui sont proposées depuis les années 30, et ont évolué au fil du temps.¹

La langue française représente un lexique qui est formé à partir de plusieurs procédés, dont la composition. Cette dernière se divise en deux catégories, la composition populaire, ainsi que savante. Pour notre projet de recherche, nous nous focaliseront sur la composition savante, car c'est le procédé le plus utilisé dans la terminologie médicale.

Il s'avère parfois difficile de comprendre les termes composés en médecine, c'est pour cela que nous faisons recours à la segmentation. A cet effet, une application de traitement automatique des langues nous semble utile pour simplifier la compréhension de ces termes.

Le TAL², est un domaine pluridisciplinaire qui fait appel à l'informatique ainsi qu'à la linguistique. En vu l'ambiguïté présente dans cette dernière, le TAL tente à les résoudre en informatisant la langue ; par la création de plusieurs applications qui aident les utilisateurs sur tous les plans ; traduction, conjugaison et autres.

² Traitement automatiques des langues.

Notre thématique de recherche s'inscrit dans deux domaines, en lexicologie en premier lieu ; une discipline qui s'intéresse au sens des unités lexicales ainsi que leurs formes. Et en traitement automatique des langues en deuxième lieu. Ce dernier est relativement récent, il tente d'informatiser les langues, et rendre l'apprentissage, la compréhension, ainsi que les différentes tâches linguistiques plus accessibles.

Notre thème de recherche se base essentiellement sur l'analyse morphologique des mots formés par la composition savante. Nous avons choisi la langue française comme langue support parce que le domaine de médecine est exercé en langue française en Algérie.

Le domaine du TAL n'est pas très exploité dans notre université, et fait partie des initiatives récentes de notre département, c'est pour cette raison que nous nous sommes intéressées à ce sujet, notre motivation découle alors de ce principe.

Notre but ultime derrière le choix de ce sujet, est bien d'apporter une nouveauté, éveiller l'intérêt des autres étudiants concernant ce domaine, mais également proposer quelque chose de tangible, qui sera sous forme d'une application.

Nous voulons parvenir, à travers notre étude, à créer un prototype de segmentation de termes médicaux à composition savante.

Pour cette recherche, nous nous sommes basées sur la problématique suivante :

- Est-t-il possible de réaliser une application de traitement automatique des langues permettant de segmenter des termes médicaux à composition savante ?
-

Notre hypothèse se résume en un point important

- Il aurait difficile de segmenter automatiquement des termes médicaux à composition savante compte tenu de leur structure morphologique.

Afin de réaliser notre travail, nous nous servirons de thèses d'articles et de cours en médecine pour dégager tous les termes utilisés, ces derniers constitueront notre corpus. Les termes seront segmentés pour pouvoir dégager les morphèmes qui les constituent.

L'étude que nous mènerons se fera en quatre chapitres. Le premier chapitre sera consacré à la langue de spécialité, ses caractéristiques et autres. Le deuxième chapitre sera consacré aux procédés de formation des termes médicaux, dans lequel nous mettrons l'accent sur la composition savante. Le troisième chapitre sera consacré au domaine de TAL dans lequel nous détaillerons tous ses concepts clé. Le quatrième chapitre sera consacré à la formalisation linguistique, où nous élaborions le prototype en question, qui nous permettra de segmenter le corpus recueilli.

Nous terminerons avec une conclusion générale, qui proposera des perspectives, qui pourront intéresser les autres étudiants pour en faire un autre projet de recherche.

Bien que le traitement automatique des langues a connu une évolution dans son usage dans le monde scientifique, mais il demeure méconnu et encore récent en notre département.

Dans notre travail de recherche nous tenterons d'analyser et segmenter les termes médicaux à composition savante d'une manière automatique.

Chapitre I :

Langue de

spécialité.

Chapitre I : Langue de spécialité

Sous ces trois formes : vocale, graphique et gestuelle, la langue est l'outil utilisé depuis notre première ère afin de communiquer, recevoir et transmettre des messages, voire, sous une sphère moderne de partager les savoirs.

D'un point de vue linguistique, et en vue de son perpétuel évolution, la langue ne peut être définie tel un élément homogène. En effet, c'est un système complexe de signes linguistiques mettant à disposition les outils nécessaires qui permettent la communication. Selon André Martinet « *Une langue est un instrument de communication selon lequel l'expérience humaine s'analyse* »³.

La langue ne permet pas seulement la communication dans son sens restreint, mais elle véhicule également les savoirs.

Moyennant la langue, les différents domaines et spécialités transmettent leurs savoirs, et s'approprient ainsi un lexique bien déterminé et spécifique. La langue diffère donc d'un domaine à un autre, d'où la dénomination de langue de spécialité.

1. Langue générale et langue de spécialité

Lorsque nous parlons de la langue générale nous évoquons cet ensemble de normes et d'unités linguistiques qui régissent les constructions phrastiques, que les locuteurs utilisent afin d'exprimer leurs intentions et idées. Elle regroupe tous les points linguistiques qui sont propre à toutes les langues du monde, qui sont l'orthographe, la grammaire, la syntaxe, la conjugaison, le vocabulaire, la ponctuation, les sigles, les abréviations et autres.

Selon Picht et Draskau, une langue générale est « *General reservoir on which the LSP's of various special areas draw* » [Un réservoir général où puisent les différentes langues de spécialité]⁴. Elle n'est donc pas un ensemble réduit à des règles ou à un outil de communication, mais elle inclut également des sous-codes qui sont employés selon le contexte du besoin d'usage des locuteurs ou spécialistes, nous sous-entendons par ceci, les langues de spécialité.

Dans notre monde actuel, les sciences triomphent et imposent par conséquent la maîtrise des langues de spécialité. Une langue de spécialité est en effet la base de tout travail scientifique. Pour un échange sans ambiguïté entre les scientifiques, un comportement langagier particulier leur est nécessaire. Elle est donc une langue qui est réservée exclusivement à un groupe de spécialistes d'un domaine donné ; elle est généralement créée par eux, et pour eux, et évidemment destinée à l'exercice du domaine en question.

Une langue de spécialité est une forme langagière que partagent les scientifiques afin de communiquer entre eux, ce n'est donc passeulement une liste de terme à connaître, mais c'est surtout savoir les enchaîner dans le but de créer des discours cohérents ayant un sens.

³MARTINET André, 1960, *Eléments de linguistique générale*, Paris, Armand Colin. p. 20.

⁴Picht, H & Draskau, 1985, *Terminology: an introduction*. Guildford: University of Surrey, Department of Linguistic and International Studies. P3.

Chapitre I : Langue de spécialité

Pour une définition générale, la langue de spécialité est pour Galisson et Coste une « *Expression générique pour désigner les langues utilisées dans des situations de communications (orales ou écrites) qui impliquent la transmission d'une information relevant d'un champ d'expérience particulier* »⁵.

Elle est de ce fait une forme linguistique bien particulière dans une situation de communication bien particulière qui relève d'un besoin d'exposition d'un fait, comme le souligne Maria Teresa Cabré : « *Une langue donnée est donc constituée par un ensemble diversifié de sous-codes que le locuteur emploie en fonction de ses modalités dialectales et qu'il sélectionne en fonction de ces besoins d'expression et selon les caractéristiques de chaque situation de communication [...]* »⁶.

La langue de spécialité ne date pas seulement d'aujourd'hui, mais remonte, selon J. Binon & S. Verlande⁷, aux années 60. Elle désigne l'utilisation d'une langue pour une profession bien particulière.

Son but réside dans sa démarcation de la langue générale pour éviter toute ambiguïté et donner une empreinte propre à elle. Cette dernière s'oppose donc selon certains linguistes à la langue usuelle qui sert seulement aux productions quotidiennes.

Une langue de spécialité se démarque ainsi de la langue générale en fonction des situations de communications et par rapport au niveau de l'usage. C'est vrai qu'elle puise de la langue générale, mais elle possède des unités linguistiques spécifiques propres à divers domaines. Elle a une spécificité qui est la monosémie au niveau sémantique ; elle est dépourvue de tout trait connotatif, avec un lexique mono-référentiel.

Une autre caractéristique de la langue de spécialité est l'enrichissement et le renouvellement rapide du lexique conditionné par des besoins langagiers relatifs à de nouvelles réalités physiques imposées, elle peut également s'ouvrir à d'autres paramètres pour l'enrichissement tels que les emprunts. C'est justement le cas actuellement avec l'apparition de la pandémie du Covid-19 déclarée le 11 mars 2020 par l'OMS, où nous assistons à la naissance de nouveaux néologismes tels que le terme « quatorzaine » qui est construit sur le modèle du mot de base « quarantaine ». Celui-ci désigne l'isolement des personnes affectées par la peste pendant 40 jours, contrairement à « quatorzaine » qui désigne l'isolement des personnes porteuses du virus Covid-19 pendant 14 jours. Ce qui est également à noter, est le fait que la langue de spécialité favorise l'utilisation de la forme nominale, ce qui est le cas de la terminologie de médecine, où l'on souligne également l'utilisation d'une morphosyntaxe spécialisée que nous développerons dans le chapitre suivant.

Une langue de spécialité n'est pas donc à soustraire de la langue générale parce qu'elle est intégrée dans cette dernière et elle se sert du même système de normes, et ne se limite pas

⁵R. Galisson & D. Coste, 1976, *Dictionnaire de didactique des langues*, Hachette, Paris. P511.

⁶CABRE Maria Térésa, 1998, *La terminologie. Théorie, méthode et applications*, Armand Colin Paris p 115.

⁷BINON Jean & VERLANDE Serge, 1999, « *la contribution de la lexicologie pédagogique à l'apprentissage d'une langue étrangère ou seconde, études de la linguistique appliquée* ». p40

Chapitre I : Langue de spécialité

seulement à un lexique propre à elle, « [...] c'est donc au niveau de l'usage que se manifeste la spécificité des langues de spécialité »⁸. La langue de spécialité véhicule à travers son lexique des connaissances des différentes sciences et spécialités, tandis que la langue générale s'utilise pour un contexte non scientifique. C'est à ce niveau que celles-ci se distinguent. Les langues de spécialités sont finalement pour Lerat : « La notion de langue spécialisée est [plus] pragmatique : c'est une langue naturelle considérée en tant que vecteur de connaissances spécialisées ».

2. Qu'est-ce que la terminologie ?

En premier lieu, la terminologie est une discipline linguistique qui a marqué sa naissance dans la première moitié du XX^{ème} siècle par plusieurs chercheurs, parmi eux, Eugen Wüster(1898-1977) qui se nomme comme père fondateur de la terminologie moderne.

En 1931, Wüster a exposé sa première théorie dans sa thèse qui s'intitule « *La normalisation linguistique internationale en technologie, en particulier en électronique* »⁹. Toutes ses études se sont penchées sur la technologie et la linguistique.

La théorie wüstérienne repose sur plusieurs approches importantes qui sont : l'approche onomasiologique, l'approche normalisatrice et l'idéal de biunivocité.

L'optique wüstérienne ne s'appuie pas uniquement sur ces approches, mais aussi sur la notion du sens et le concept, que confirme Wüster dans sa citation : « *le domaine des concepts est considéré par la terminologie comme étant indépendant de celui des dénominations.* »

Nous comprenons que cette théorie de la terminologie se démarque du même principe que la théorie saussurienne où le signe possède un signifiant et un signifié, c'est ce que nous constatons pour cette optique de Wüster concernant le terme qui regroupe un concept et un mot.

Suite à cette théorie Wüstérienne, plusieurs écoles ont mis le point sur la terminologie, comme l'école de Vienne, l'école de Prague, l'école Russe, l'école Canadienne et l'école Manchester. Ces dernières ont travaillé d'avantage sur la théorie Wüstérienne pour la développer.

En 1990, les chercheurs de la terminologie, entre le camp « classique » héritiers de Wüster et le camp de « préfix-socio », ont créé un lien étroit entre la terminologie et la linguistique. Cette branche est également considérée comme un champ interdisciplinaire selon Laurén et Al¹⁰. Ces derniers citent les disciplines qui sont en relation avec la terminologie,

⁸ CABRE Maria Térésa, 1998, *La terminologie. Théorie, méthode et applications*, Armand Collin Paris. P111.

⁹WÜSTER, Eugen, 1931, *Internationale Sprachnormung in der Technik besonders in der Elektrotechnik (Die nationale Sprachnormung und ihre Verallgemeinerung)*. (Normalisation internationale des langues dans les technologies, en particulier dans Génie électrique (La normalisation en langue nationale et sa généralisation), VDI Verlag, p431

¹⁰LAUREN, Christer & PICHT Heribert, 2001, « Terminologie aus linguistischer Sicht ». *Terminology Science & Research*. 200112(1-2) p30-40.

Chapitre I : Langue de spécialité

comme la sémiologie, l'informatique, les sciences de l'information et la documentation de la philosophie (en particulier l'épistémologie et la théorie de la science), et la normalisation et la planification linguistique.

Vu que la terminologie est un champ d'étude interdisciplinaire alors elle est une discipline scientifique, comme le confirme Sager en sa citation :

« *Il n'existe pas de corpus substantiel de littérature pour appuyer l'idée de la terminologie comme discipline à part entière, et il n'y aura probablement jamais. Tout ce qu'on peut dire d'important sur la terminologie peut être dit de façon plus appropriée dans le cadre de la linguistique, des sciences de l'information ou de la linguistique computationnelle* »¹¹

En deuxième lieu, la terminologie porte un intérêt dans le monde moderne¹² où nous observons plusieurs définitions, qui sont associées à ce terme par des théoriciens et des dictionnaires généralistes. Le dictionnaire Le Robert en ligne¹³ donne la définition suivante : « *Ensemble de désignations et de notions appartenant à un domaine spécial (science, technique, etc.) Vocabulaire didactique d'un groupe social.* » et également cette brève définition : « *Etude des systèmes de termes et de notions.* ». Un autre dictionnaire de référence qui est le Trésor de la langue française informatisé (TLFI)¹⁴ donne deux définitions distinctes à la notion de terminologie, la première : « *Ensemble de termes relatifs à un système notionnel élaboré par des constructions théoriques, par des classements ou des structurations de matériaux observés, des pratiques sociales ou d'ensembles culturels.* », la deuxième : « *« Art de repérer, d'analyser et au besoin, de créer le vocabulaire pour une technique donnée, dans une situation concrète de fonctionnement de façon à répondre aux besoins d'expression de l'utilisateur ».*

Alors, la seule signification que nous pouvons associer à ce lemme « terminologie » est la suivante : « *vocabulaire spécifique, propre à un domaine de spécialité* ».

En somme, Le TLFI définit la notion de terminologie comme un ensemble et une étude des termes d'un domaine de spécialité précis, dans une situation concrète alors que le dictionnaire Le Robert se focalise quant à lui sur l'étude des termes d'une manière générale de toutes les spécialités confondues.

Un troisième dictionnaire généraliste LE GRAND Larousse¹⁵ propose quatre différentes définitions de la terminologie qui sont :

- a) « *Ensemble de termes spécifiques d'une technique, d'une science, d'un art, d'un domaine défini d'activité, d'un type d'organisation* ».
- b) « *Vocabulaire particulier qu'emploie un autre, qu'adopte un groupe social* ».

¹¹ JUAN C. Sager, 1990, « *A practical Course in Terminology Processing* », John Benjamins, Amsterdam, p 246.

¹² BEDDAR Mohand, 2013. Thèse de doctorat « *Vers un prototype de traduction automatique contrôlée français/arabe appliquée aux domaines à sécurité critique* », Sciences du langage, Université FRANCHE-COMTE.

¹³ <https://dictionnaire.lerobert.com/> 27/06/2020 à 14h16.

¹⁴ <https://www.cnrtl.fr/definition/terminologie/substantif> 28/06/2020 à 18h05.

¹⁵ Le Grand Larousse de la langue française, 1978. Paris: Larousse.

Chapitre I : Langue de spécialité

- c) « *Ensemble des activités pratiques relatives au recueil, à la distribution, à la traduction, et à la diffusion des unités terminologiques* ».
- d) « *Science qui a pour objet l'étude théoriques des dénominations des concepts et d'objet dans les domaines spécialisés du savoir et de leur fonctionnement social dans une langue ou plusieurs langues, des relations des unités terminologiques avec la logique, la linguistique, la lexicographie, la traduction, la documentation et l'informatique* ».

Nous constatons que le mot « Terminologie » se définit comme une discipline qui étudie et qui englobe les différents termes techniques qui appartiennent à un domaine spécifique comme la terminologie médicale.

En revanche, le TLFi définit le mot *terminologie* d'une manière plus riche et plus précise, dont la première définition correspond à la même définition donnée par le dictionnaire le Robert. Ce qui englobe: la terminologie grammaticale, mathématique, linguistique, philosophique, scientifique, technique, la terminologie de la musique, la terminologie des sciences exactes, des sciences sociales et la terminologie de la médecine qui est notre support d'étude.

Nous constatons que les quatre définitions présentées par le Grand Larousse diffèrent les unes des autres et se distinguent des définitions des autres dictionnaires. Elles indiquent bien le vocabulaire spécialisé d'un domaine particulier, scientifique ou technique, d'un auteur ou d'un groupe social.

Enfin, le terme de terminologie est une notion plus complexe à définir au début, elle sert à dégager les notions et les réalités propres à un domaine spécifique.

La création de chaque terminologie se focalise sur un concept si important qui est « les langues de spécialités » dont ces dernières englobent la lexicologie, la langue et le terme¹⁶.

3.1. Définition du terme

Nous pouvons définir le terme à partir de la notion du mot ; mais pouvons-nous dire que le mot « terme » marque son émergence à partir du contexte de la terminologie ?

Après une étude bien approfondie, nous nous sommes rendu compte qu'une définition bien précise du terme ne peut se faire sans la définition de la notion de mot.

Le mot est défini comme une petite unité lexicale ayant un sens « morphème », c'est une unité de première articulation selon André Martinet. Il s'agit d'une suite de lettres séparées par des espaces (ou des signes graphiques). Mais Le TLFi le définit comme étant : « *un mot ou un ensemble de mots ayant dans une langue donnée, une signification précise et exprimant une idée définie* ». Il est défini également par ce qui suit « *Nom qui appartient à un système structuré, en se distinguant de tous les autres éléments du système, et qui dénote dans une langue donnée des classes ou des unités à l'intérieur de ces classes* »¹⁷

¹⁶<http://www.culture.fr/Ressources/FranceTerme/Ou-est-ce-que-la-terminologie> 27/06/2020 à 22h38.

¹⁷<http://tice-lt.info/termscience/mod/book/view.php?id=90> 28/06/2020 à 13h11.

Chapitre I : Langue de spécialité

4. La terminologie et les langues de spécialités

Rostislav Kocourek, est le linguiste chercheur le plus influant dans l'étude de la notion des langues de spécialité dans son fameux ouvrage : « *La langue française de la technique et de la science* » Il a accordé la définition suivante aux langues de spécialité :

« *La langue de spécialité est une variété de langue , à dominante cognitive , dont les textes commutatifs d'émotivité, de subjectivité et de métaphoricité contrôlées ,et délimitées de manière externe, ont pour but de signifier et de communiquer , au sein d'une collectivité restreinte , le contenu thématique , raisonné et circonstancié et dont les ressources qui sous-tendent ces textes sur tous les plans linguistiques , sont marquées par des caractères graphiques , par des tendances syntaxiques et surtout par un ensemble rapidement renouvelable des unités lexicales qui requièrent et reçoivent dans les textes , une précision sémantique métalinguistique »¹⁸.*

De ce fait, une langue de spécialité est une langue d'un domaine spécifique qui se caractérise par des termes spécifiques à ce domaine, c'est ce qu'Alain Rey confirme dans le passage suivant :

« *Ce qui rend « spéciale » une langue de spécialité ce n'est pas seulement son vocabulaire (en discours), son lexique (en système) sa terminologie sur le plan (cognitif), mais c'est avant tout cela, en tant que représentant langagier d'une cohérence conceptuelle (sic). c'est ici que la terminologie et le langage de spécialité coïncident au plus près sans nier les spécificités possibles , mais partielles , d'autres composantes , syntaxiques , stylistiques , rhétoriques, etc »¹⁹.*

Vu que notre corpus se base sur la terminologie médicale, nous allons définir trois domaines médicaux auxquels appartiennent tous les termes que nous avons recensés.

Conclusion

Dans ce chapitre nous avons distingué la langue générale de la langue de spécialité en ayant souligné les spécificités de chacune. Nous avons conclu donc que la langue de spécialité est réservée seulement aux spécialistes, contrairement à la langue générale qui est utilisée par des non-spécialistes, des locuteurs ordinaires.

Ce qui caractérise une langue de spécialité est sa terminologie. Wüster Eugen est parmi les premiers à avoir travaillé sur cette discipline de la linguistique. Plusieurs définitions sont avancées afin de mieux la cerner.

Nous avons évoqué et détaillé la terminologie médicale car elle est le pilier de notre sujet de recherche. Elle s'agit de l'étude des termes spécifiques de la médecine que l'ensemble du personnel paramédical et médical utilise et maîtrise surtout.

¹⁸ROSTISLAV Kocourek, 1991 « *La langue française de la technique et de la science* » Brandstetter , p42.

¹⁹CELLARD Jacques, ALAIN Rey, 1991, « *Dictionnaire du français non conventionnel* », Titolo Collona. Paris. Hachette

Chapitre I : Langue de spécialité

Les termes médicaux sont créés par de différents procédés de formations de lexique, parmi eux, les emprunts les néologismes, la troncation, la dérivation, la composition savante. Tous ces procédés seront développés dans le chapitre qui suit.

Chapitre II :
Procédés de
formation des
termes
médicaux.

Chapitre II : Procédés de formation des termes médicaux

Première partie

1. La lexicologie

Ferdinand de Saussure définit la linguistique comme une science qui étudie la langue et le langage humain d'une manière profonde. Elle est également définie comme tout ce qui appartient au lexique, relève du lexique ²⁰

Cette science regroupe plusieurs branches comme la phonétique, la phonologie, la morphosyntaxe. En effet, le mot lexicologie contient deux radicaux grecs: *lexicon* – vocabulaire et *logos* – étude, la traduction littérale du terme lexicologie étant « l'étude raisonnée des mots »²¹, c'est l'étude de la signification des unités qui constitue le lexique d'une langue.

Le lexique est défini par Marie Françoise Mortureux comme : « *l'ensemble des lexèmes, des morphèmes lexicaux d'une langue et des règles de leur fonctionnement.* »²²

Alors c'est l'ensemble des mots et des locutions figées d'une langue, il s'agit de tous les morphèmes lexicaux.

La lexicologie est à la croisée de toutes les autres disciplines de la linguistique comme la phonologie qui s'occupe de la forme des mots, la syntaxe qui touche à son tour la combinaison entre les unités lexicales, la morphologie où nous distinguons dans cette branche les morphèmes lexicaux et les morphèmes grammaticaux. Ces derniers portent l'intérêt de la lexicologie, enfin la sémantique qui sert à décrire le sens des mots et des syntagmes lexicaux.

Jacqueline Picoche affirme que la lexicologie : « *Peut être définie par rapport aux disciplines plus vastes dont elle n'est qu'une partie : la sémantique dont l'objet est l'étude des significations linguistiques, elle-même branche de la sémiologie qui traite des codes de signe en général.* »²³.

La lexicologie est une branche qui porte deux sens différents, le premier sens appelé restreint, il est lié à la sémantique, ce qui veut dire que la lexicologie fait partie de la sémantique qui étudie le sens des unités lexicales ; quant au deuxième sens qui est le sens large, là la lexicologie s'occupe de la formation des unités lexicales et aux relations existantes entre la syntaxe et le lexique.

L'école russe considère la lexicologie comme branche de la linguistique et qui a pour objectif l'étude de « mot » tout en précisant son origine, ses particularités de sens, les voies de son apparition, les procédés de sa formation et aussi son rôle dans un discours

²⁰<https://cnrtl.fr/definition/lexicologie> 18/08/2020 à 19h29.

²¹MARTIN, R., BERTHET, F. 1998, *Introduction à la lexicologie, sémantique et morphologie*, Dunod, p

²²MORTUREUX Marie Françoise, 1997, *La lexicologie entre langue et discours*, SEDES, p

²³PICOCHÉ Jacqueline, 1977 *Précis de lexicologie française*, collection « Nathan Université », Paris, p

Chapitre II : Procédés de formation des termes médicaux

1.1. Aperçu historique de la lexicologie

Les débuts de la lexicologie sont rattachés aux travaux publiés à titre posthume de Ferdinand De Saussure en « *cours de linguistique générale* »

Georges Matoré est le pionnier de la lexicologie en tant que discipline autonome dans son ouvrage intitulé « *la méthode en lexicologie* » apparue en 1953, il a pour objectif de distinguer « la sémantique » qui étudie « les valeurs successives des mots », « la lexicographie » qui est l'étude analytique des faits de vocabulaire, discipline linguistique et la lexicographie « discipline sociologique qui envisage des groupes de mots considérés statiquement du point de vue notionnel »²⁴

Georges Matoré a publié sa thèse en 1951 qui s'intitule « le vocabulaire et la société » sous Louis Philippe où il s'est intéressé à la lexicologie synchronique et sociale en réalisant que cette branche est extralinguistique.

1.2. La notion du mot

1.2.1. Le mot graphique

Le mot graphique s'agit de l'unité lexicale qui se situe entre deux blancs, un seul mot graphique peut renvoyer à plusieurs mots qui sont différents grammaticalement ; prenons l'exemple de Rey qu'il a emprunté dans représentation de petit Robert 1976: « nous avons mangé hier des pommes de terre ». Dans exemple nous distinguons huit mots graphiques « nous /avons /manger/ hier/ des/ pommes/ de/ terre/ ». Chaque mot qui se situe entre deux espaces sera considéré comme mot graphique, contrairement au mot linguistique.²⁵

1.2.2. Le mot linguistique (lexème)

Ce type de mot est différent du mot graphique, nous prenons le même exemple précédent de Rey : « nous avons mangé hier des pommes de terre », le linguiste va compter 05 mots linguistiques (lexicales) « nous/ avons manger/hier/ des/pommes de terre/ ». Parce-que chacun des groupes que nous analysons sera considéré comme un seul mot, comme une seule unité linguistique, alors nous pouvons même remplacer le groupe verbal « avons mangé » par « mangeons », comme nous pouvons remplacer tout l'énoncé par un autre qui lui sera équivalent du type « nous/mangeons / ce soir /des bananes/ ».

2. L'origine de la terminologie médicale

Les tout premiers textes et documents médicaux remontent au quatrième siècle avant J.-C. Cette période-là est caractérisée par les termes médicaux du grec. Le système de médecine est basé à cette époque-là sur la théorie des humeurs par Hippocrate.

Par la suite, une encyclopédie médicale « *De Medecina* » est écrite en latin tout en gardant tous les termes grecs auxquels il n'y avait pas d'équivalents en langue latine.

²⁴MATORÉ Georges, 1953, *La méthode en lexicologie*, Didier, Paris, p

²⁵POLUGUERE Alain, 2003, « *lexicologie sémantique lexicale : notions fondamentales* », PUM, p42-45.

Chapitre II : Procédés de formation des termes médicaux

A partir de la renaissance, la médecine connaît un tournant remarquable : la langue grecque recule face à la langue latine, et c'est là que la période médicale latine débute réellement. Tous les travaux en médecine dans cette période sont traduits au latin, et s'enrichissent avec le temps par les autres langues.

La terminologie médicale se sert également de la lexicologie, car certains termes puisent leurs règles de formation dans les procédés de formation du lexique français, ayant une morphologie particulière.

3. La morphologie

La morphologie est une science qui s'intéresse aux normes qui régissent la formation des mots, son objet d'étude est le morphème. Pour G. Mounin, la morphologie est définie comme une : « *Étude des formes sous lesquelles se présentent les mots dans une langue, des changements dans la forme des mots pour exprimer leurs relations à d'autres mots de la phrase, des processus de formation de mots nouveaux, etc.* »²⁶.

La morphologie s'intéresse essentiellement à la forme et à la composition des mots, ces derniers sont formés à partir d'entités pourvus de sens qui sont les morphèmes.

3.1. Le morphème

Le morphème est la plus petite unité d'analyse grammaticale est de signification de la langue. Pour obtenir des morphèmes nous faisons recours à la segmentation qui permet d'appliquer les deux tests grammaticaux, qui sont la commutation et la distribution. La condition nécessaire pour qu'une partie d'un mot puisse constituer un morphème est qu'elle puisse être remplacée par un autre élément. Faisons référence à la théorie structuraliste américaine par Leonard Bloomfield, le morphème est l'unité minimale pourvue de sens et non décomposable, cette définition est ensuite complétée par la théorie fonctionnaliste en ajoutant le critère de pertinence

Selon André martinet, il y a deux grandes classes de morphèmes : les morphèmes grammaticaux (grammèmes) et les morphèmes lexicaux (lexèmes).

3.1.1. Morphèmes grammaticaux

Ce sont des unités qui appartiennent à des listes fermées de la grammaire doués d'un sens de cette dernière. Nous distinguons deux types de morphologie grammaticale :

3.1.1.1. Les morphèmes liés

Ce sont des forment qui ne peuvent pas fonctionner et apparaitre de façon autonome dans une phrase. Ils sont toujours liés à une base. Ces même morphème de divisent en deux sous catégories :

²⁶ MOUNIN, G., 1974, *Dictionnaire de la linguistique*, Paris, PUF.

Chapitre II : Procédés de formation des termes médicaux

3.1.1.1.1. Morphèmes dérivationnels

Ce sont des éléments qui peuvent être postposés ou antéposés par rapport à une base. Par exemple, la préfixation ne change pas la catégorie grammaticale d'un mot contrairement à la suffixation

- vivifier + (préfixe re-)= revivifier. (v)
- Réanimer (v) + (suffixe -tion)= réanimation (adv).

3.1.1.1.2. Morphèmes flexionnels

Ils ne servent pas à créer de nouveaux mots mais plutôt de créer de nouvelles formes différentes les unes les autres. Nous nous entendons par les nouvelles formes les désinences et les marques du genre et du nombre.

- défibrillateur +-s du pluriel= défibrillateurs : changement du singulier au pluriel.

3.1.1.2. Les morphèmes libres

Ce sont une liste des morphèmes autonomes : les morphèmes amalgamés et portemanteaux, les articles, les pronoms, les prépositions. Cette même liste se divise en deux sous catégories.

Les morphèmes portemanteaux c'est la fusion de plusieurs valeurs grammaticales dans un seul morphème. Exemple :

- L'adjectif possessif : *Mon*. Il indique à la fois le nombre singulier de l'objet de possession, le genre masculin et la première personne du singulier « possesseur de l'objet »

Les morphèmes amalgamés c'est la fusion de deux morphèmes à la fois. Exemple :

- à +les= aux.

3.1.2. Morphèmes lexicaux

Ils sont appelés également radicaux. C'est un inventaire dans lequel sont recensés toutes les unités pourvues d'un contenu significatif, mais qui sont dépendantes des morphèmes grammaticaux afin qu'elles soient logiques et actualisées.

Deuxième partie

1. Formation du lexique français

1.1. Les contextes de formation

Les contextes d'échange linguistique diffèrent par rapport à ce qu'ils transmettent comme connaissances. Ils donnent et enrichissent donc le lexique en fonction du besoin. Ce dernier est créé généralement à partir de trois contextes :

Chapitre II : Procédés de formation des termes médicaux

1.1.1. **Le contexte de spécialisation** : dans lequel les spécialistes d'un domaine donné trouvent le besoin de communiquer avec leur propre vocabulaire.

1.1.2. **Le contexte de vulgarisation** : dans lequel les spécialistes s'adressent à de non-spécialistes, comme pour un médecin s'adressant à son patient. Un vocabulaire parallèle né donc de ce contexte. Exemple :

a. *Fièvre* pour *hyperthermie*.

b. *Coronavirus* pour *SARS-CoV-2*.

1.1.3. **Le contexte du quotidien** : ce sont toutes les situations du quotidien qui influencent la création de mots.

2. Les différents procédés de formation du lexique français, et leur utilisation dans la terminologie médicale

La formation du lexique repose sur plusieurs mécanismes, soit utiliser les stocks lexicaux de la langue source et étendre les sens, combiner ces stocks existants, ou encore, se servir d'autres langues pour un enrichissement.

2.1. L'emprunt

La langue ne se contente pas seulement de ses ressources linguistiques pour enrichir le lexique, mais elle fait aussi appel à d'autres procédés comme l'emprunt. « *L'emprunt à une autre langue, contemporaine ou ancienne (langue source), est l'un des moyens que possèdent les langues pour enrichir leur vocabulaire* »²⁷. C'est un procédé par lequel une langue emprunte un mot ou une expression d'une langue étrangère, sans la traduction, ou traduit sémantiquement dans certains cas. La séquence empruntée s'adapte directement aux règles morphosyntaxiques et phonétiques de la langue emprunteuse (langue d'accueil). Les emprunts viennent afin de combler un vide lexical pour désigner de nouvelles réalités sociolinguistiques.

Les guerres, les commerces et la colonisation, sont les principales causes des emprunts linguistiques. C'est donc dû aux contacts des langues. En plus de ces contextes, la situation géographique joue également un rôle dans l'adoption de ce procédé. Le français par exemple a emprunté principalement à des pays voisins : l'Espagne, l'Italie, et la Grande-Bretagne.

L'emprunt n'est pas un procédé qui puise seulement de nouveaux mots d'une autre langue ; l'adaptation se fait sous trois formes :

2.1.1. L'emprunt direct : le mot emprunté est repris tel qu'il est sans modification ou adapté aux règles de la phonétique et l'orthographe. Les exemples suivants, sont pris de l'anglais et du français.

Exemple : Emprunt sans modification : *botox*.²⁸

²⁷ APOTHELOZ Denis, 2002, *La construction du lexique français: principes de morphologie dérivationnelle*, Editions OPHRYS. P 16.

²⁸ Toxine botulique.

Chapitre II : Procédés de formation des termes médicaux

Emprunt avec adaptation phonique et orthographique : Conebeam²⁹

2.1.2 Le calque : c'est la traduction littérale d'un mot ou d'un groupe de mots. « Une façon adoucie et également peu perceptible d'adopter un mot ou un tour étranger, c'est de le décalquer, c'est-à-dire de créer un équivalent indigène sur la base d'une correspondance lexicale antérieure »³⁰

Exemple: Delivery room >salles d'accouchement.

2.1.3. L'emprunt sémantique : c'est l'ajout d'un sens d'un mot étranger pour le mot de la langue d'arrivée. « l'influence la moins perceptible qu'une langue exerce sur une autre, se trouve sans doute dans l'emprunt du sens »³¹

Exemple : réaliser > «se rendre compte» > to realize.

La langue médicale a connu ces dernières années une très grande influence de l'anglais, cela revient au progrès très actif dans les recherches en médecine dans les pays anglophones.

2.2. Le xénisme ou pérégrinisme

Le xénisme est la forme première d'un terme qui vient d'une langue étrangère sans qu'il subisse de changement. C'est un mot étranger, utilisé pour faire référence à une réalité étrangère. « Pour L.Guilbert « ces mots ont leur valeur propre par leur forme étrangère : ils sont en quelque sorte cités comme témoins de la réalité évoquée » »³²

C'est au degré de l'intégration que se distingue le xénisme de l'emprunt : en effet, le xénisme qui n'est pas utilisé par une grande communauté linguistique, et qu'il n'est pas intégré dans leur réalité. Le passage du xénisme à l'emprunt se fait quand le mot en question se fait par son intégration, et son adaptation par certaines normes de la langue d'accueil, tel que l'influence phonétique de la langue d'accueil. La phase intermédiaire entre le xénisme et l'emprunt est appelé pérégrinisme. « Le xénisme désigne la première occurrence, celle où le mot est encore perçu comme étranger, tandis que le pérégrinisme désignerait un stade ultérieur, une première phase d'intégration du mot, qui n'est plus perçu comme étranger »³³

A titre d'exemple, l'expression anglaise « *Laughing out loud* » connu sous l'acronyme de « *LOL* » est un xénisme, car à vrai dire ce n'est pas un emprunt, il reflète toujours une réalité étrangère malgré sa vaste utilisation.

En médecine, les xénismes ne risquent pas d'être utilisés car la langue médicale relève d'un langage scientifique, et donc qui nécessite des mots qui sont compréhensibles par toute

²⁹ Expression anglaise désignant un appareil de radiographie pour l'examen des dents.

³⁰DEROY Louis, 1956, *L'emprunt linguistique*, Chapitre IX. Les degrés de la pénétration, Liège, Presses universitaires de Liège, Les Belles Lettres, Bibliothèque de la faculté de philosophie et lettres de l'université de Liège.

³¹DEROY Louis, 1956, *L'emprunt linguistique*, Chapitre IX. Les degrés de la pénétration, Liège, Presses universitaires de Liège, Les Belles Lettres, Bibliothèque de la faculté de philosophie et lettres de l'université de Liège.

³²FARMERIE Etienne, 1998, *Le latin et le grec d'Appien: contribution à l'étude du lexique d'un historien grec de Rome*, Librairie Droz. P366.

³³SUCHET Myriam, 2009, *Outils pour une traduction postcoloniale: littératures hétérolingues*, Archives contemporaines. P58.

Chapitre II : Procédés de formation des termes médicaux

la communauté scientifique mais également par les patients ; et surtout qui aient une référence dans les dictionnaires médicaux.

2.3. Le néologisme, et néologisme sémantique

Le développement de la médecine nécessite en parallèle la création de nouveaux termes afin de pouvoir désigner de nouvelles réalités découvertes mais aussi communiquer sur les nouveaux acquis. Le néologisme est donc le procédé auquel le domaine médical fait très souvent recours. Les néologismes ne concernent pas seulement le contexte professionnel ; il est aussi utilisé dans les dialogues médecins/patients. Ces derniers possédant pas forcément une connaissance du domaine, les médecins alternent dans leurs discours avec les patients entre la langue de spécialité et la langue générale, ce qui peut entraîner la création de nouveaux mots pour leur faciliter la compréhension. Dans les deux exemples suivants :

- Le *HHV-3* est un virus très contagieux, responsables de petites épidémies. (1)
- La *Varicelle* est un virus très contagieux, responsable de petites épidémies. (2)

Dans l'exemple (1), le terme médical relevant de la terminologie « *HHV-3* » est simplifié dans l'exemple (2). Les deux termes sont différents, mais nous soulignons un néologisme sémantique : les deux termes signifient la même chose.

Le néologisme en médecine est principalement créée à partir des discours médecins/patients. Il est à noter qu'il est principalement dû à trois facteurs :

- La non maîtrise de la terminologie médicale par les patients.
- Le besoin d'expression des patients pour leur médecin afin de leur expliquer leurs maux
- Le besoin des médecins de simplifier le diagnostic pour leurs patients.

le néologisme ne relève pas seulement d'un processus de formation de mots nouveaux, ce procédé est également influencé par le contexte par lequel il est créé « *La formation du néologisme n'est pas une opération purement linguistique, qu'elle est à la fois un phénomène de la langue et un phénomène de société* »³⁴.. Précédemment mentionné, il y a trois contextes qui font que de nouveaux mots naissent.

2.4. La siglaison

Le principe de la siglaison est d'additionner les initiales de différents mots afin d'en créer de nouveaux, non séparés par des points à l'écrit. C'est l'un des procédés qui fait partie de l'abrégement des mots. « *La siglaison consiste à former des mots nouveaux avec des sigles c'est-à-dire en additionnant les initiales de plusieurs mots* »³⁵. En langue française, les sigles sont invariables, sauf dans certains cas où ils se transforment en nom communs. Une fois ancrés dans la langue courante, ils peuvent être dérivés et être aussi variables.

En médecine, il existe certains mots trop long et très complexe, la difficulté de ces mots réside dans leur lecture et écriture. C'est pour cette raison que les praticiens font très souvent recours au procédé de la siglaison par souci d'économie communicationnelle.

³⁴真華·徐(Zhen Hua Xu), 2001, *Le néologisme et ses implications sociales*, Harmattan.P 15.

³⁵M Bénaben, 2002, *Manuel de linguistique espagnole*, Editions OPHRYS. 289p.

Chapitre II : Procédés de formation des termes médicaux

La siglaison est considérée comme un procédé propre aux langues de spécialité car, si nous le remarquons, dans le domaine médical plus précisément, tous les énoncés longs sont exprimés en siglaison.

Nous citerons ici quelque exemple de siglaison : invariables, variables, dérivationnels,

- Siglaison invariables: AVC->accident vasculaire cérébral.
- Siglaison variable : une MST->des MST.
- Siglaison dérivationnel : éremiste³⁶.

2.5. L'acronyme

L'acronyme est la réduction d'une expression aux lettres initiales« *Le sigle qui s'épelle ou l'acronyme qui se prononce comme une unité phonique (...)* »³⁷. La différence entre un sigle et un acronyme, c'est la prononciation du mot : un acronyme est prononcé d'une seule émission tel un mot ordinaire : La chaîne des syllabes se prononce comme un mot simple, contrairement à un sigle.

Prenons l'exemple de DIU³⁸, ce mot est un acronyme vu qu'il se prononce tel un mot simple, contrairement à VIH³⁹, où chaque lettre est émise seule, l'espacement entre les lettres est phonique et non graphique.

2.6. La troncation

La troncation est une forme d'abrègement, qui consiste en la suppression d'une ou plusieurs syllabes au début ou à la fin du mot.

Il est généralement motivé par l'économie du langage, au lieu de prononcer un mot long, nous faisons le choix de tronquer le mot, et il possède son entrée dans les dictionnaires.

Il existe trois types de troncation :

- **L'apocope** : qui consiste à supprimer la syllabe finale du mot comme dans : gynécologue-> gynéco.
- **L'aphérèse** : consiste à la suppression des syllabes du début du mot. Par exemple : endoscopie⁴⁰> scopie.
- **La syncope** : il s'agit de la suppression de plusieurs syllabes ou au milieu d'un mot. Par exemple : le radon⁴¹->Rn.

Les praticiens en médecine font très souvent recours à cette pratique langagière car dans leurs discours ils cherchent à faire courts, vu que la majorité des mots de cette spécialité sont longs.

³⁶Bénéficiaire du RMI.

³⁷MILLIARESSI Tatiana & RASIERR François, 2011, *De la linguistique à la traductologie: Interpréter/traduire*, Presses Université. Septentrion. 203p.

³⁸ Dispositif intra-utérin.

³⁹ Virus de l'immunodéficience humaine.

⁴⁰ Méthode d'exploration visuelle médicale industrielle de l'intérieur d'une cavité inaccessible à l'œil.

⁴¹Élément chimique, un gaz radioactif.

Chapitre II : Procédés de formation des termes médicaux

Les mots simples, dans les mots composés, les spécialistes ont tendance à tronquer les syllabes du deuxième mot. Exemple : consultation matinale-> consultation mat'.

2.7. La dérivation

La dérivation est le procédé de formation consistant à la création de nouvelles unités lexicales par l'adjonction à une base, que nous appelons radical, à une liste de morphèmes dépendants. « *Un mot dérivé est constitué d'un seul lexème, c'est-à-dire d'un morphème indépendant et ou de plusieurs morphèmes dépendants, appelés affixes.* »⁴²

Le radical est la base à laquelle nous ajoutons des affixes de nature dérivationnels ou flexionnels dans le but de former des mots nouveaux « *On appelle radical le morphème lexical qui subsiste quand tous les affixes, dérivationnels et flexionnels, ont été enlevés* »⁴³. La différence entre un radical et une racine, c'est que le terme racine est utilisé lorsque nous faisons une étude diachronique. C'est une structure d'un morphème dans état antérieur d'une langue donnée.

Nous distinguons trois types de dérivation : la préfixation, la suffixation et la formation parasynthétique.

2.7.1. La suffixation

Le principe de la suffixation est d'ajouter un morphème dérivationnel « suffixe » à un radical. La suffixation change toujours la catégorie du mot selon le suffixe ajouté. En guise d'exemple :

- Certains suffixes comme *-et* ont un rôle diminutif du sens. Exemple : une pince->une pincette.
- Certains suffixes comme *-logie* classent les mots dans des registre de spécialité, comme le cas de la médecine. Exemple : cardiologie.

2.7.2. La préfixation

Pour la préfixation, le principe est d'adjoindre un affixe postposé du radical. Contrairement à la suffixation, la préfixation ne change jamais la catégorie d'un mot.

2.7.3. La parasynthétisation

La forme parasynthétique c'est l'adjonction simultanée d'un préfixe et suffixe pour le même radical. Ce procédé permet la création de :

- Verbe, à partir d'un adjectif : plat (adj) ->aplatir.
- Adjectif à partir d'un verbe : déterminer (v)->indéterminable.

2.8. La composition

Le Trésor de la langue française a donné plusieurs définitions au mot « composition » où nous réalisons que ce mot est polysémique. D'un point de vue linguistique il s'agit de la :

⁴²DELBECQUE Nicole & M. Jean-Rémi Lapaire, 2006, *Linguistique cognitive: Comprendre comment fonctionne le langage*, De Boeck Supérieur. P 88.

⁴³APOTHELOZ Denis, 2002, *La construction du lexique français: principes de morphologie dérivationnelle*, Editions OPHRYS. P 16.

Chapitre II : Procédés de formation des termes médicaux

« Formation d'un mot par combinaison de deux mots simples reliés par un trait d'union (ou une conjonction de coordination) ou par adjonction d'un préfixe; la manière dont ce mot est composé. Processus, règles de composition (cf. nom composé) et également en ling. mod. dérivation, lexie, synapsie, synthème. »⁴⁴

Selon Jean Dubois la composition « est la formation de nouveaux mots en juxtaposant deux ou plusieurs mots ensemble (nom+verbe+adjectif). »⁴⁵

La composition est la juxtaposition de deux éléments qui existent à l'état libre. Les mots simples comme « arc-en-ciel », ou des éléments greco-latins comme « démocratie ». la composition est un groupe de mots à signification unique, il s'agit d'une unité lexicale qui renvoie à un seul référent.

Il existe plusieurs critères qui servent à repérer les mots composés, parmi eux, le critère graphique qui considère à son tour le mot composé comme toute unité dont ses éléments forment une unité graphique ininterrompue, ce critère n'englobe pas les mots en général comme les mots qui sont formés par un trait d'union (-). Un autre critère qui s'appelle le critère de l'inséparabilité qui sert à interdire l'insertion des unités constitutives le mot composé d'un quelconque élément et même l'impossibilité de modifier ou de perturber l'ordre et l'emplacement des éléments constitutifs du mot composé c'est-à-dire les mots composés sont inscrits à côté des mots simples ou des mots dérivés ; l'avant dernier critère est le critère de commutation qui consiste à mettre en équivalence chaque mot composé avec d'autres mots de même paradigme par exemple le mot « pomme de terre » appartient au paradigme de « Carotte et navet », ce qui confirme justement le critère de l'inséparabilité . La dérivation comme dernier critère est parmi les indices qui permettent l'identification et la constitution des mots composés et même l'existence des dérivés suffixaux.⁴⁶

2.8.1. Les types de composition

Nous distinguons trois types de compositions, la composition dite populaire qui se construit à partir du mot français, la composition savante qui se réalise à partir d'éléments grecs ou latins et enfin les mots composés hybrides.⁴⁷

Un mot composé est un ensemble de mots qui produit une seule signification et qui renvoie à un seul référent. La composition populaire consiste à unir deux éléments autonomes qui sont soit séparés ou reliés par un trait d'union (-), une apostrophe ou aussi par un blanc graphique, autrement dit la combinaison de deux mots qui peuvent exister à l'état libre. Ces deux mots sont inséparables, ils possèdent aussi un sens propre lorsqu'ils sont seuls, par exemple :⁴⁸

Les mots qui sont formés sur la base de la composition populaire possèdent plusieurs formes :

⁴⁴<https://www.cnrtl.fr/definition/composition/substantif> le 15/08/2020 à 18h15

⁴⁵ DUBOIS Jean, 1992, Dictionnaire de linguistique et des sciences du langage, édition Larousse, p

⁴⁶<http://www.diacronia.ro/ro/indexing/details/A17763/pdf> 18/08/2020 à 23h20.

⁴⁷<https://www.home.uni-osnabrueck.de/bschwisc/archives/formation.htm> 23/08/2020 à 21h40.

⁴⁸http://www.cilliac-arp.univ-paris-diderot.fr/media/user/najet_boutmgharine/cours-6-.pdf 23/08/2020 à 23h29.

Chapitre II : Procédés de formation des termes médicaux

- Deux noms

Exemple : canapé-lit

- Un nom +préposition +nom.

Exemple : boîte à pharmacie.

- Adjectif +nom

Exemple: premier étage.

Deux verbes

Exemple: le savoir-faire

- Une préposition ou un adverbe +un nom

Exemple : sans domicile fixe (SDF), un sans-abri.

- Un verbe et d'un nom

Exemple: cache-nez.

La langue médicale est caractérisée par la complexité de son vocabulaire. Il est essentiellement fondé sur les mots composés d'origine gréco-latine.

La troisième partie sera consacrée pour l'analyse morphologique des composés savants et hybrides afin de dégager leurs structures morphologiques.

Troisième partie

D'un point de vue morphologique, le vocabulaire médical comprend une très grande partie de termes complexes. Ceux-ci sont construits à partir de racines grecques et latines, appelés « composés savants ».

La composition savante est la formation : « (...) *des mots nouveaux en combinant entre deux formants grecs ou latin qui n'existent pas de manière autonome en français (composition savante) ou en associant ceux-ci à un mot français (composition hybride)*»⁴⁹. Les mots composés savants, sont de nouvelles formes lexicales. Elles se forment principalement par les emprunts du latin ou du grec. C'est en effet une juxtaposition de deux éléments, ou plus. Les unités grecques ou latines peuvent aussi être associées à des mots français qui sont autonomes, ce procédé est la composition hybride.

⁴⁹ GREVISSE Maurice, FAIRON Cédric, SIMON Anne-Catherine 2018, *Le petit Bon usage de la langue française: La grammaire qui aime les écrivains*, Grevisse langue française. De Boeck Supérieur. P72.

Chapitre II : Procédés de formation des termes médicaux

Les mots composés savants ou hybrides sont motivés, les constituants du mot analysé et segmenté ont un sens ; ils ne sont pas arbitraires. Afin de dégager les normes de constructions de ces composés, nous procéderons par une segmentation.

La segmentation est une tâche de l'analyse morphologique. C'est une opération qui permet de diviser une construction afin de connaître ses constitutions. Après la segmentation, vient la substitution, celle-ci consiste à remplacer un élément X dans une construction donnée par un élément Y sans que cette dernière soit agrammaticale.

Approche morphologique

Nous verrons dans cette partie, comment est-ce qu'un composé savant ou hybride peut être segmenté, et quelles sont les règles de formations de ces mots.

Pour rappel, la morphologie est selon Bécharde : « *l'étude des formes que peuvent prendre les mots par la description des mécanismes qui d'une part établissent leur structure et de l'autre leur confèrent des catégories grammaticales(...)* ». ⁵⁰Le procédé d'analyse morphologique est donc la décomposition d'un mot, afin de dégager ses constituants.

On s'appuyant sur la citation précédente de Maurice Grevisse, un mot formé par le procédé de la composition savante, est constitué de deux bases grecques ou latines. Il peut être construit également à partir de plus de deux bases. Dans un premier temps nous recenserons quelques termes médicaux tirés de quelques thèses en cardiologie, neurologie et psychiatrie, afin de pouvoir les segmenter, dans un deuxième temps nous essayerons de dégager leurs règles de formation. Enfin déterminer leurs natures (composés savants ou hybrides).

⁵⁰ BECHARD Hervé, 1992, *phonétique et morphologie du Français moderne et contemporain*, 1ère édition, PUF.P93.

Chapitre II : Procédés de formation des termes médicaux

Le composé savant	La première base	La deuxième base	La troisième base	Catégorie
Antibiothérapie	Anti-	-Bio-	-thérapie	Composé hybride
Apolipoprotéine	Apoli-	-Protéine	/	Composé hybride.
Arythmie	A-	-Rythmie	/	Composé hybride
Athérosclérose	Athéro-	-Sclérose	/	Composé hybride
Bradycardie	Brady-	-Cardie	/	Composé savant
Cardiomyopathie	Cardio	-Myo-	-Pathie	Composé savant
Cardiopathie	Cardio	-Pathie	/	Composé savant
Dyslipidémie	Dys-	-Lipidémie	/	Composé hybride
Echocardiographie	Echo	-Cardio-	Graphie	Composé hybride
Endothélium	Endo-	-Thélium	/	Composé savant
Fibrinogène	Fibrino-	-Gène	/	Composé hybride
Hémorragie	Hémo-	-Rragie		Composé hybride
Hépatomégalie	Hépto-	-Mégalie	/	Composé savant
Hyperlipidémie	Hyper-	Lipidémie	/	Composé hybride
Iatrogénie	Iatro-	Génie	/	Composé savant
Microangiopathie	Micro-	-Angio-	-pathie	Composé savant
Néphropathie	Néphro-	-Pathie	/	Composé savant
Physiopathologie	Physio-	-Patho-	-logie	Composé savant
Rétinopathie	Rétino-	-Pathie	/	Composé hybride
Scintigraphie	Scinti-	-Graphie	/	Composé hybride
Tachycardie	Tachy-	-Cardie	/	Composé savant
Electrochoc	Electro	Choc		Composé hybride

Tableau 1 : Segmentation des composés savants/hybrides

Nous confirmons à partir de cette segmentation, la citation que nous venons d'avancer : un mot composé savant est constitué de deux bases, mais comme il peut être aussi constitué

Chapitre II : Procédés de formation des termes médicaux

de plus de deux bases tel que le premier exemple. Un composé hybride est constitué d'une base grecque et/ou latine juxtaposé à un mot français.

Nous nous sommes servis du dictionnaire étymologique afin de distinguer la forme hybride de la forme savante.

Nous déduisons à partir de ce tableau, que les règles de constitution des mots composés savants ayant deux bases ou plus sont comme suit :

- Base grecque+base grecque.
Exemple : cardiopathie.
- Base grecque+ base grecque+base grecque
Exemple : cardiomyopathie.
- Base latine+base latine.
Exemple : hémorragie.

Pour ce qui concerne les composés hybrides, les règles de constitutions sont comme suit :

- Base française+ base grecque.
Exemple : rétinopathie.

D'autres formes de constitutions sont possibles en associant une base latine à une base française ou grecque.

Il existe également des composés savants et hybrides qui sont constitués de plus de trois composants, à titre d'exemple : Dysbêtalipoprotéïnémie, Hyperhomocystéïnémie, et l'un des mots les plus longs en médecine qui est Oesophagogastroduodenoscopie⁵¹. Est donc la règle de formation de ces composés est comme suit :

- Base+base+base⁺~.⁵²

Les morphèmes constitutifs de ces mots composés ne sont pas des affixes, autrement dit, une base ou le morphème peut être à la fois des affixes mais aussi un radical. C'est un type de construction à interfixation.

Après avoir segmentés ces mots, nous allons appliquer la substitution, pour affirmer qu'en changeant les morphèmes d'un mot composé, nous pouvons en créer un autre.

A titre d'exemple, le composé hybride suivant « antibiothérapie », est composé de deux bases grecques plus une base française. En remplaçant le suffixe « thérapie » par le suffixe « gramme » nous obtenons le composé « antibiogramme »⁵³.

⁵¹ Procédé médical pour visualiser l'appareil gastro-intestinal au moyen d'une sonde.

⁵² La dernière règle de constitution, signifie qu'il peut y avoir plus de trois bases.

⁵³ Technique de laboratoire visant à tester la sensibilité d'une souche bactérienne vis-à-vis d'un antibiotique.

Chapitre II : Procédés de formation des termes médicaux

Prenons un autre exemple qui figure dans le tableau « hyperlipidémie », est composé d'une base grecque et une base française. En remplaçant le préfixe « hyper » par le préfixe « hypo » nous obtenons le composé « hypolipidémie ».

Pour un dernier exemple, nous allons changer le radical et le suffixe du mot « cardiomyopathie » en gardant le préfixe, nous obtenons donc le mot « cardiofréquence-mètre »⁵⁴

Nous déduisons qu'avec le procédé de substitution, d'autres mots peuvent se créer.

Conclusion

Nous avons à travers ce chapitre survoler les deux disciplines : la lexicologie et la morphologie qui sont en étroite relation avec la terminologie médicale, car la quasi-totalité des termes médicaux sont formés à partir des procédés de formation du lexique français.

Dans la première partie nous avons retracé brièvement l'évolution de la terminologie médicale, ensuite nous avons cerné les points les plus importants de la lexicologie, et enfin développé la thématique de la morphologie, et comment est-ce que les mots sont formés.

La deuxième partie est consacrée au développement des différents procédés de formation du lexique français qui sont également utilisés pour la formation de la terminologie médicale. Nous avons d'abord énuméré les trois contextes de formation, et expliquer par la suite chacun des procédés.

Dans la troisième partie, nous avons d'abord recensé un certains nombres de termes à composition savante et hybride pour étudier leurs formes. Ensuite nous les avons segmentés en s'appuyant sur une approche morphologique afin de dégager leurs règles de formations. Et enfin appliquer la substitution pour confirmer la pertinence de notre segmentation.

Dans le dernier chapitre, nous essayerons d'élaborer une application pouvant segmenter ces termes d'une manière automatique.

⁵⁴ Appareil permettant de déterminer la fréquence cardiaque instantanée.

Chapitre III :
Traitement
Automatique
des Langues.

Chapitre III : Traitement Automatique des Langues

1. Introduction

Un domaine de recherche relativement récent ; le traitement automatique des langues est un carrefour où s'entrecroise deux disciplines : l'informatique et la linguistique. L'objectif principal de ce domaine est l'automatisation des activités langagières par la création d'applications permettant de traduire des données linguistiques en un logiciel.

Notre travail de recherche s'inscrit d'une part dans le domaine de TAL, et d'autre part dans le domaine de la lexicologie. Dans ce chapitre nous allons mettre en lumière comment est-ce que le domaine est apparu, nous aborderons, ensuite son champ de recherche et applications, et en dernier lieu les problèmes auxquels les développeurs font face.

2. Traitement automatique des langues, qu'est-ce que c'est ?

C'est un domaine multidisciplinaire qui applique des techniques et programmes informatiques sur le langage humain. C'est un champ de recherche et de développement visant à modéliser, à l'aide d'un programme informatique, les compétences humaines en matière de production et compréhension des énoncés linguistiques.

Le traitement automatique des langues vise donc la création et la configuration de logiciels dotés d'une intelligence artificielle capable d'automatiser toute donnée linguistique.

Le processus d'automatisation repose sur la transformation d'une donnée de nature linguistique pour qu'elle soit exécutable par la machine. C'est en effet un traitement d'une donnée X aboutissant à une donnée Y qui a pour but; et selon l'application en question, de corriger, traduire, conjuguer ou encore donner un synonyme, etc....

Le TAL devient aujourd'hui un domaine d'une grande importance. Des laboratoires à travers le monde se consacrent à l'étude et à la recherche, tels que le Laboratoire du Centre de Recherche Scientifique et Technique pour le Développement de la Langue arabe en Algérie, plus exactement à Alger, Bouzareah.

3. Les dénominations de TAL

Le traitement automatique des langues, connu sous l'acronyme de TAL, connaît depuis son apparition plusieurs appellations, à titre d'exemple : l'ingénierie linguistique, traitement automatique des langues, nouvelles technologie d'analyse d'information. Et encore lors de sa traduction vers l'anglais nous obtenons d'autres appellations et acronymes tels que NLP⁵⁵et CL⁵⁶. Ce que nous déduisons à partir de ces différentes dénominations, c'est la difficulté de placer cette discipline dans un contexte précis, car nous pouvons désigner à la fois une technologie nouvelle, un domaine scientifique et un champ de recherche.

⁵⁵Natural Language Processing.

⁵⁶Computational linguistics.

Chapitre III : Traitement Automatique des Langues

4. Aperçu historique de la discipline

Les recherches et les travaux en traitement automatique des langues débutent dès les années 50 aux Etats-Unis. La sphère politique de la guerre froide a fait que l'on accorde une importance à ce domaine. La traduction automatique est parmi les premières thématiques développées les années 40. Les travaux relatifs à cette thématique ont débuté bien avant l'ingénierie linguistique.

La naissance de l'informatique revient à l'avènement du mathématicien Alan Turing, lorsqu'il a proposé la « Machine de Turing » en 1936. Il est à noter que les premières applications relevant de ce champ sont en relation avec le traitement automatique des langues.

« Computing Machinery and Intelligence »⁵⁷ est l'article d'Alan Turing dans lequel il explique le concept de « Test Turing »⁵⁸ qui a pour but d'évaluer le degré d'intelligence d'une machine et sa capacité d'imiter ou de remplacer la conversation humaine, et donc mesurer à quel degré ses réponses ressemblent à celles que fournit un humain, et non la capacité de répondre d'une manière correcte à une question. Le principe de ce test est de mettre un sujet humain face à une machine dans un échange de messages textuels afin de déterminer s'il s'agit d'un autre sujet humain ou une machine par un évaluateur. Ce test, est au fil du temps, devenu un concept très important dans le domaine de l'intelligence artificielle. En 1952, un colloque est dirigé par Yehoshua Bar-Hillel sur la traduction automatique organisé au MIT dans lequel il conclue qu'il est préférable de faire une traduction humaine assistée par une machine, qu'une traduction entièrement automatisée. Le rapport de l'ALPAC suggère de son côté de favoriser la recherche fondamentale plutôt que les applications.

Deux ans après, une traduction automatique à caractère informatique absolue est devenue possible grâce à l'expérience de Georgetown, réalisée par la collaboration de l'université de Georgetown et par la société IBM. Les concepteurs ont donc déterminé une période qui ne s'étale pas sur plus de trois ans, pour qu'enfin la traduction automatique ne soit plus un problème.

Vers la fin des années 1960, Terry Winograd, crée le programme SHRDLU, permettant à son utilisateur de converser avec un ordinateur en anglais afin d'accomplir certaines tâches, et donc exécuter des ordres en langage humain. Cependant, ses capacités restent restreintes

Sur la période qui s'étale entre 1970 et 1980, les développeurs ont démarré la création des « ontologies conceptuelles » pour une structuration d'informations en données compréhensible par l'ordinateur.

A partir de l'an 2000 à ce jour, le traitement automatique des langues est dans la mesure d'imiter le langage naturel, offrant la possibilité aux modèles d'apprentissage de

⁵⁷https://fracademic.com/dic.nsf/frwiki/70859_04/06/2020 à 23h31.

⁵⁸http://www.artificiel.net/test-de-turing_04/06/2020 à 21h21.

Chapitre III : Traitement Automatique des Langues

traiter une grande infinité d'informations. En guise d'exemple de modèles d'intelligence artificielle, ceux créés par Microsoft, tel que Word.

5. Qu'est-ce que la programmation

D'une manière générale, la programmation, appelée aussi le codage, est l'ensemble des activités permettant de créer des programmes informatiques dans le but de développer des logiciels qui sont exécutables par une machine « ex : l'ordinateur », autrement dit, c'est l'automatisation de certaines tâches à l'aide d'une machine. Dans le contexte du traitement automatique des langues nous parlons de tâches qui sont reliées à l'activité langagière humaine.

Afin d'écrire un programme informatique, des langages de programmation sont nécessaires pour la réalisation des logiciels.

5.1. Les langages de programmation

Un langage de programmation est un code avec lequel le développeur formule des algorithmes et crée des programmes informatiques qui les interprètent, et s'exécutent par un logiciel. Un langage de programmation fonctionne avec le même principe qu'un langage humain naturel ; il est composé de sa propre grammaire, vocabulaire et alphabet.

Un langage de programmation permet au programmeur de communiquer à la machine ; il lui donne des instructions, mais lui permet également de communiquer avec d'autres programmeurs étant que les programmes sont de nature écrits

En somme, un langage de programmation permet d'interpréter à chaque instruction, une action. Les langages les plus utilisés sont : le Java, le JavaScript, le langage C, le C++, le PHP, le Python.

5.1.1. Le java

Java est un langage de programmation et une plateforme informatique créée en 1995 par James Gosling et Patrick Naughton employés de Sun Microsystems. A l'origine, Java est un langage informatique qui permet de dynamiser les pages web, mais maintenant, Java est un langage de programmation, très évolué et très puissant, qui permet de développer des applications de toutes sortes. Le langage Java est très facile à apprendre et sa syntaxe ressemble au langage C car il s'inspire de ce dernier, mais il est beaucoup plus simple. L'avantage de Java, c'est qu'il est intégré à tous les systèmes d'exploitation tels que Linux et Windows.

5.1.2. Le Python

Python est un langage de programmation interprété. Ce type de langage sera développé dans le chapitre pratique, car c'est avec ce langage que notre prototype est réalisé.

Chapitre III : Traitement Automatique des Langues

6. Champ de recherche en TAL

Toutes les applications TAL nécessitent d'une part : la conception d'interfaces ergonomiques⁵⁹ , d'autre part la possibilité de traiter d'une manière intelligente les informations qui sont sous format textuelles.

6.1. Traitement documentaire

6.1.1. La traduction automatique

Malgré son développement, la traduction automatique reste un enjeu économique et politique. C'est un champ qui nécessite des connaissances linguistiques ainsi que des connaissances dans le domaine pour lequel la traduction est faite. Il est l'un des problèmes les plus complexes depuis la toute première application en 1950.

6.1.2. Correction orthographique

Le correcteur orthographique a pour tâche :

- L'identification des mots
- La correction de l'orthographe.
- La correction grammaticale.

6.1.3. Résumé automatique des textes

Le principe du résumé automatique est l'extraction des informations les plus importantes d'un texte dans le but de le réduire et lui donner un aspect cohérent et crédible.

L'extraction s'appuie essentiellement sur :

- L'analyse thématique du corpus et sa nature.
- La concentration sur certains procédés comme la cause et la conséquence : dans un résumé, les deux procédés mentionnés sont les plus importants car ils évoquent les points de vues cités et défendus, les résultats et aussi les arguments avancés.

6.1.4. Génération automatique des textes

Il s'agit d'une production automatique d'un texte quelconque. La génération de ces textes doit être syntaxiquement et sémantiquement correcte.

6.2. Le traitement syntaxique

6.2.1. La racinisation/stemmatization

Le principe est de regrouper tous les mots appartenant à la même famille de mots. La stemmatization est moins sensible aux fautes d'orthographe que la lemmatization, car cette dernière échoue à la moindre erreur orthographique d'un mot. Le procédé de racinisation n'a pas besoin d'un contexte pour fonctionner mais seulement le mot.

⁵⁹Conceptualiser des systèmes offrant un confort aux utilisateurs.

Chapitre III : Traitement Automatique des Langues

6.2.2. La morphologie

Grâce à la morphologie, les mots sont regroupés par le type de morphologie : flexionnelle ou dérivationnelle.

6.2.3. La lemmatisation

La lemmatisation est le nom du procédé en TAL qui consiste à réduire des mots dans un texte en leur forme canonique, ou en lemme. Le lemmatiseur TreeTagger⁶⁰ a la capacité d'adaptation à de multiples langues par l'intermédiaire de ses fichiers de configuration.

6.2.4. Étiquetage morpho-syntaxique

A l'aide de l'étiquetage, chaque mot d'un texte donné est classé dans sa catégorie grammaticale.

6.3. Extraction de l'information

6.3.1. Reconnaissance d'entités nommées

Dans un texte, les noms propres sont identifiés en tant que tels. Il peut s'agir de nom de personnes, d'endroits ou de dates etc.

6.3.2. Fouille de texte

Le principe est de chercher certaines informations spécifiques d'un document donné dans des textes à nature différente. A titre d'exemple, des applications de fouille de texte se voient utiles afin éviter le plagiat dans les travaux scientifiques (thèses, mémoire...).

6.3.3. Recherche d'informations

Sous-domaine de la fouille de texte : par exemple dans le cadre médical, il existe des programmes qui facilitent la recherche d'examens médicaux ou des comptes rendu dans les dossiers des patients.

7. Les difficultés de TAL

Les difficultés que nous rencontrons dans les recherches en TAL sont principalement de deux ordres. Le premier problème concerne l'ambiguïté du langage, et la deuxième l'implicite présent dans les communications.

7.1. Ambiguïté du langage naturelle

L'un des traits caractérisant le langage naturel est l'ambiguïté et ce, sur tous les niveaux.

L'ambiguïté d'un langage naturel se manifeste par plusieurs interprétations possibles, que nous pouvons appliquer sur les entités linguistiques. Nous citons :

- L'ambiguïté sémantique d'un mot à cause de sa forme graphique. Exemple : les homographes. Un mot peut avoir plusieurs sens mais une seule forme graphique. Cela peut poser problème lors d'une traduction automatique. Un mot qui ne se traduit pas

⁶⁰<https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> 05/06/2020 à 22h36.

Chapitre III : Traitement Automatique des Langues

correctement peut nuire au sens de la phrase. Exemple, si la phrase suivante se traduit du français à l'anglais de cette manière : Cette dame dame le sol ->this lady lady the floor, le sens de la phrase devient incorrecte à l'anglais car le deuxième mot « dame » n'est pas traduit par rapport à son sens de damer.

- L'ambiguïté phonique des graphèmes pour l'encodage. Exemple : pain et pin. Cette ambiguïté peut poser problème si l'on traduit un texte oral à un texte écrit. C'est ce qu'il se passe avec les utilisateurs de recherche vocal de Google.
- L'ambiguïté de la fonction grammaticale d'un mot. Exemple : manger (nom), manger (verbe à l'infinitif). Une ambiguïté qui peut aussi poser problème lors d'une traduction.

Ce genre de problème d'ambiguïté peut être rencontré lors de l'utilisation de Google traduction.

7.2. L'implicite

L'activité langagière humaine peut être sous forme d'interaction entre deux personnes ou plus. L'interaction contient des énoncés à caractère implicite relié aux connaissances du monde. Ce qui n'est pas le cas avec une machine : elle ne dispose pas ces connaissances comme base de données ce qui rend la compréhension complète d'un énoncé implicite impossible. Si la machine ne dispose pas de cette base de connaissances donnant accès à la fois au monde et le contexte de l'énonciation, la compréhension est impossible. A titre d'exemple l'expression des sentiments peut se faire de manière implicite mais comprise par les humains, contrairement à la machine. L'expression « le dindon de la farce » veut dire implicitement que quelqu'un est victime d'une tromperie, mais la machine ne peut pas comprendre cette expression s'il n'y a pas son explication dans sa base de données.

Conclusion

Nous avons, à travers ce chapitre, vu que ce domaine est très récent, mais qu'il est en constante évolution. C'est un domaine multidisciplinaire où se rencontrent linguistique, informatique et intelligence artificielle. Le but de ce domaine est l'automatisation de toutes les activités langagières humaine et naturelle.

Le TAL est apparu les années 50 avec l'avènement de l'informatique. Son champ de recherche est diversifié, et donc nous retrouvons une multitude d'applications, cependant l'ambiguïté et l'implicite de la langue pose toujours problème.

Chapitre III

Formalisation

linguistique

Chapitre III : Formalisation linguistique

1. Introduction

L'objectif de notre projet, est de créer une application qui puisse segmenter des termes médicaux à composition savante et hybride, et donner les morphèmes qui les composent ainsi que leurs significations.

Afin de pouvoir réaliser cette application, nous procéderons par une formalisation linguistique ; nous définirons ce concept, pour mettre en place ensuite trois dictionnaires. Le premier dictionnaire comportera les préfixes avec leurs significations, le deuxième dictionnaire comportera les radicaux, et pour ce qui en est du troisième dictionnaire, il comportera les suffixes avec leurs définitions.

2. Définition de la modélisation linguistique

La modélisation ou la formalisation linguistique est une étape obligatoire dans la conception d'applications en TAL. C'est une manière de modéliser les langues naturelles et leurs fonctionnements par un raisonnement informatique.

Pour notre projet, nous formaliserons les formes des structures des termes médicaux à composition savant et hybride afin de pouvoir les segmenter automatiquement.

3. La logique du fonctionnement du code du prototype

La première étape prise, est celle du téléchargement des bibliothèques python. Ces bibliothèques sont un ensemble de fonctions qui permettent de faire des graphismes, du formatage du texte, génération de documents et autres.

La deuxième étape, est celle de la création de l'interface en utilisant QtDesigner. Celui-ci permet de créer graphiquement des interfaces ainsi que la gestion des événements

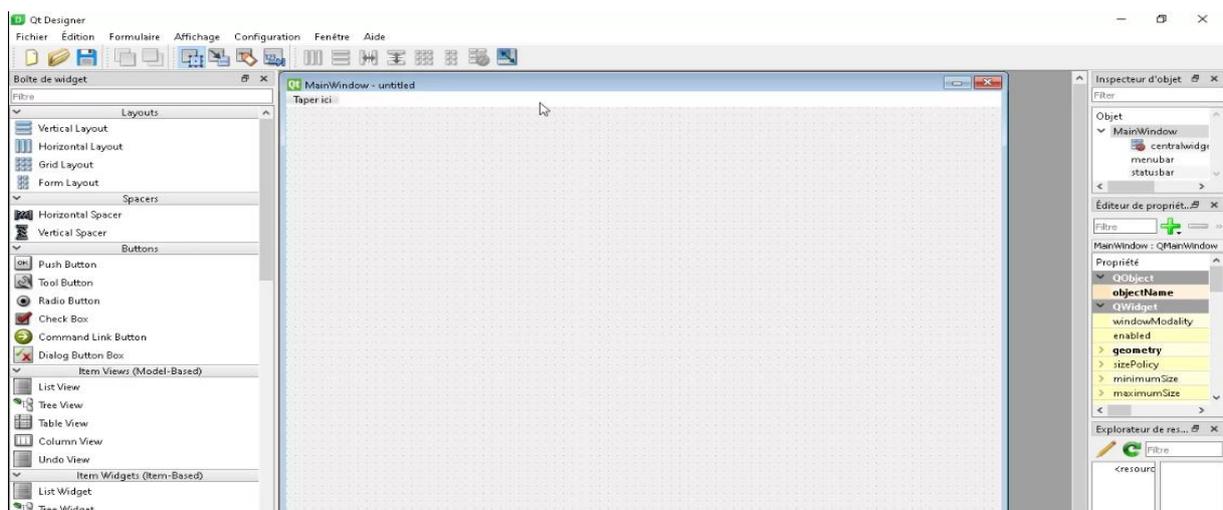


Figure 1 : Création de l'interface par QtDesigner.

Chapitre III : Formalisation linguistique

Lorsque la l'interface est créé, celle-ci est en fichier `ui` . Ce dernier est converti vers un fichier `py` . ; Il se traduit donc en un script.

La prochaine étape est l'importation des bibliothèques. Ces dernières servent à relier les boutons de l'interface qui sont générés automatiquement dans le script avec leurs fonctionnalités. Ces dernières sont : l'ajout, la suppression, la recherche, et l'affichage des messages boxes.

Dans le script d'ajout de mot, nous importons les bibliothèques qui relient les boutons avec leurs fonctions, et les bibliothèques gérant la base de données. Les trois fenêtres d'ajout de morphèmes d'un mot : la fenêtre d'ajout de préfixe, fenêtre d'ajout de radical et la fenêtre d'ajout de suffixe préalablement créés sont également importées. Ces dernières sont créées avec le même principe.

Ensuite, afin de créer une base de données, nous faisant recours à Sqlite3 et Os, grâce auxquels nous stockons les données linguistiques dans la base de données. Nous avons également importé le lemmatiseur French Stemmer afin de pouvoir segmenter un mot automatiquement même s'il n'existe pas dans la base de données. Mais malheureusement nous n'avons pas pu développer ses modules pour qu'il puisse segmenter les composés savants.

Une chaîne est créée lors de la recherche, celle-ci relie entre deux ou trois morphèmes d'un mot. Dans le cas où le mot n'existe pas, le schéma ne se crée pas, et affiche dans un message box que le mot n'existe pas.

Une fois le mot trouvé, le prototype affiche également les mots ayant le même radical. En sélectionnons l'un d'eux, la fonction Sqlite3 va fouiller dans la base de données.

Une fois terminé, nous importons la fenêtre d'ajout de mot dans le programme principal, afin d'avoir une application achevée.

4. Modèle de segmentation des termes à composition savante et hybride

Notre système fonctionne selon les deux règles suivantes :

- Base préfixale+base suffixale.

Exemple :agora+phobie=agoraphobie.

- Base préfixale+baseradicale+base suffixale

Exemple : Cardio+myo+pathie=cardiomyopathie.

Avant de mettre en place les trois dictionnaires qui recensent les morphèmes qui constituent les termes à composition savante et hybride, nous allons créer un tableau qui recense tous les termes que nous avons tirés de notre corpus. Chacun de ces termes sera suivi d'une définition. Nous comptons 27 termes qui sont :

Chapitre III : Formalisation linguistique

4.1. Dictionnaire des termes

Termes	définitions
Antibiothérapie	Substance naturelle ou synthétique qui détruit la croissance des bactéries.
Aplasie	Absence d'un organe provoqué par l'absence de développement de son ébauche embryonnaire.
Apolipoprotéine	protéines chargées de transporter les molécules hydrophobes dans le sang
Arthymie	Battement irrégulier du cœur.
Athérosclérose	Maladie touchant les artères de gros et moyen calibre.
Bradycardie	Battement de cœur inférieur à 60 battements par minute.
Cardiomyopathie	Maladie du muscle du cœur due à une augmentation de son épaisseur.
Cardiopathie	Maladie du cœur.
Dyslipidémie	Elévation du cholestérol plasmatique des triglycérides.
Dysplasie	Anomalie dans le développement des tissus, se traduisant par des malformations.
Echocardiographie	Echographie du cœur.
Endothélium	Couche de cellules épithéliales qui tapisse l'intérieur des parois du cœur.
Fibrinogène	Protéine du plasma sanguin qui se transforme en fibrine lors de la coagulation.
Hépatomégalie	Augmentation de volume du foie.
Hyperkinétie	Etat d'hyperactivité.
Hyperlipidémie	Taux élevé de lipides dans le sang.
Hyperplasie	Développement anormal d'un tissu, d'un organe.
Hypothalamus	Région située à la base du cerveau.
Iatrogénie	Trouble ou maladie consécutifs à la prise d'un médicament.
Néphropathie	Affection du rein.
Physiopathologie	Discipline de la biologie qui traite des dérèglements de la physiologie.
Traumatologie	Etude médicale des traumatismes physique.
Psychotrope	Substance capable de modifier l'activité sur système nerveux.
Rétinopathie	Lésions de la rétine chez les personnes ayant un diabète sucré depuis des années.
Scintigraphie	Méthode d'exploitation d'un organe par injection d'une substance légèrement radioactive.
Tachycardie	Maladie qui accélère le rythme cardiaque.
Traumatogène	Exposition à la mort effective ou à une menace de mort

Tableau 2 : Dictionnaire des termes

4.2. Dictionnaire des préfixes

Préfixes	définitions
Anti-	Contre
A-	Privation
Apoli-	Eloignement
Athéro-	Lésion de la surface interne des artères

Chapitre III : Formalisation linguistique

Brady-	Lenteur
Cardio-	Cœur
Dys-	Malformation
Echo-	Réflexion de son
Endo-	En dedans
Fibrino-	Protéine du plasma sanguin
Hépatho-	Foie
Hyper	Au-dessus
Hypo	Au-dessous
Iatro-	Soins médicaux
Néphro-	Rein
Physio-	Nature
Psycho-	Ame
Rétino-	Membrane interne de l'œil
Scinti-	Matériau émettant une lumière
Tachy-	Rapide
Traumato-	Choc/blessure.

Tableau 3 : Dictionnaire des préfixes.

4.3. Dictionnaire des radicaux

Radicaux
-bio-
-myo-
-cardio-
-patho-
-traumato-
-Encéphalo-

Tableau 4 : Dictionnaire des radicaux.

4.4. Dictionnaire des suffixes

Suffixes	définitions
-thérapie	Traitement.
-plasie	Altération physiologique de caractère anarchique.
-rythmie	Rythme cardiaque.
-sclérose	Durcissement pathologique.
-cardie	Cœur.
-pathie	Maladie.
-lipidémie	Teneur du sang en lipides totaux.
-graphie	Tracé
-gène	Qui engendre
-mégalie	Augmentation
-kinétie	Mouvement
-thalamus	Cavité
-gènie	Qui engendre
-pathie	Maladie
-logie	Science
-trope	Orientation

Tableau 5 : Dictionnaire des suffixes.

Chapitre III : Formalisation linguistique

5. Présentation de l'application

Le fonctionnement de notre application repose sur la base de données sur laquelle nous avons travaillé en ajoutant tous les mots qui constituent notre corpus, les étapes se présentent comme suit :

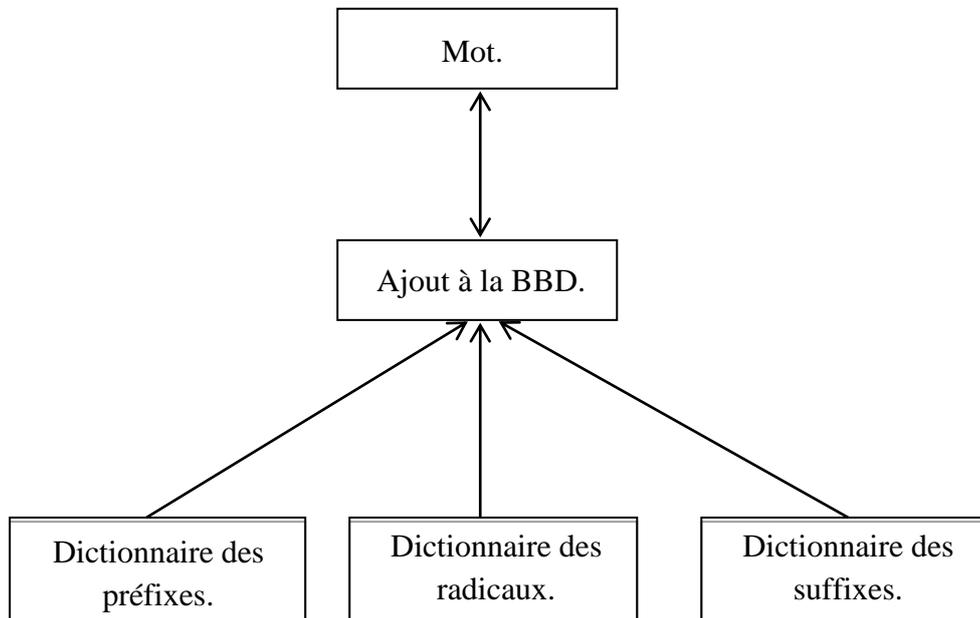


Figure 2 : fonctionnement de l'application.

Lorsque le mot est tapé dans la case « chercher », cela engendre une recherche automatique dans la base de données, dans laquelle il détecte sa présence. Dans le cas où le mot n'est pas inscrit dans la base de données, un message s'affiche indiquant que le mot recherché n'existe pas.

Une fois le mot trouvé, une consultation automatique dans les trois dictionnaires est déclenchée : il affiche en même temps les mots ayant le même radical, et aussi les deux (ou trois) morphèmes qui constituent le mot recherché.

6. Outils informatique utilisés

Afin de pouvoir réaliser une application capable de segmenter les termes en morphèmes, et de donner à chacun d'eux sa signification, nous avons choisi de travailler avec le langage de programmation Python, et l'environnement de développement PyCharm.

6.1. Le langage de programmation Python

Python est un langage de programmation interprété dont la première version est créée par Guido Van Rossum en 1991, cependant la Python Software Foundation, est créée en 2001. C'est un langage puissant et très facile à apprendre et à utiliser. Il est idéal pour le développement rapide des applications. Le développeur en Python dispose de ce que nous appelons les bibliothèques qui permettent à travailler sur des projets

Chapitre III : Formalisation linguistique

Le langage Python permet de développer des programmes complets comme les jeux, les logiciels de toutes sortes. Il est portable et fonctionne sous les différents systèmes d'exploitation.

6.2. PyCharm

PyCharm est un environnement de développement intégré utilisé pour programmer en Python. Il permet l'analyse de code et la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web.

Cet environnement est développé par l'entreprise tchèque JetBrains.

7. Résultats

Le programme que nous avons obtenu est constitué d'une fenêtre principale affichant les résultats, une fenêtre d'ajout de mots, trois fenêtres d'ajout de suffixes, radicaux et préfixes. Et enfin d'une base de données comportant les données linguistiques.

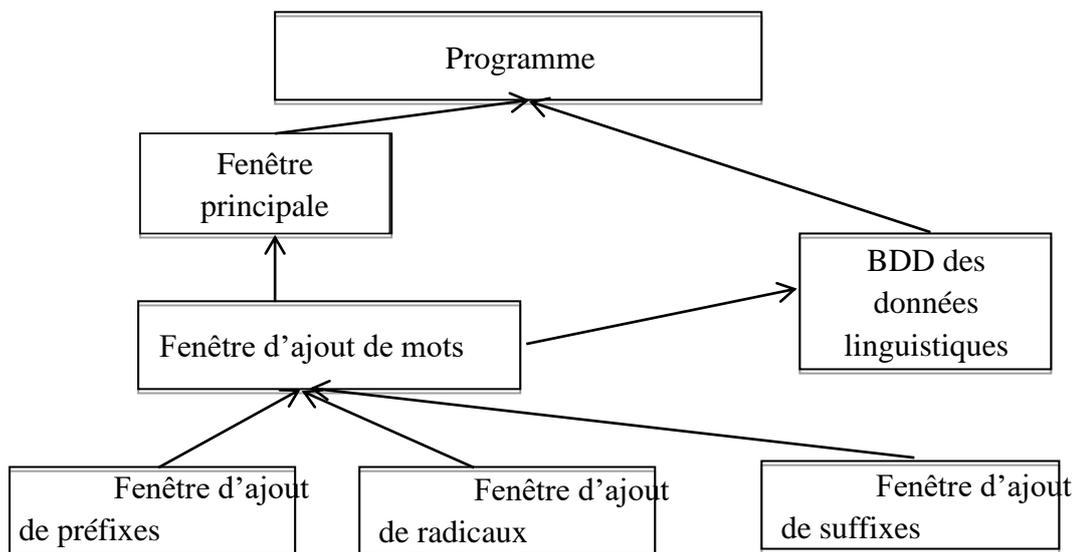


Figure 3 : structure du prototype

7.1. Programme principal

Notre prototype est un programme formé à partir d'une série d'algorithmes, ceux-ci sont une suite d'instructions. Quand ces dernières sont exécutées, une application est générée. Cette application est donc capable de segmenter les termes à composition savante et hybride.

Chapitre III : Formalisation linguistique

Ces algorithmes sont présentés dans le code source suivant :

Les lignes 1 à 75 représente la création de l'interface de l'application

```
Interface.py x Main_Window.py x Popup_mot.py x popup_Prefixe.py x popup_Racine.py x popup_Suffixe.py x
1  #-*- coding: utf-8 -*-
2
3  # Form implementation generated from reading ui file 'Interface_Tal.ui'
4  #
5  # Created by: PyQt5 UI code generator 5.9.2
6  #
7  # WARNING! All changes made in this file will be lost!
8
9  from PyQt5 import QtCore, QtGui, QtWidgets
10 #import sqlite3
11 from nltk.stem.snowball import FrenchStemmer
12
13 class Ui_MainWindow(object):
14     def setupUi(self, MainWindow):
15         MainWindow.setObjectName("MainWindow")
16         MainWindow.resize(800, 600)
17         self.centralwidget = QtWidgets.QWidget(MainWindow)
18         self.centralwidget.setObjectName("centralwidget")
19         self.List_Words = QtWidgets.QListWidget(self.centralwidget)
20         self.List_Words.setGeometry(QtCore.QRect(50, 160, 171, 311))
21         self.List_Words.setObjectName("List_Words")
22         item = QtWidgets.QListWidgetItem()
23         self.List_Words.addItem(item)
24         item = QtWidgets.QListWidgetItem()
25         self.List_Words.addItem(item)
26         self.Def_Text = QtWidgets.QTextBrowser(self.centralwidget)
27         self.Def_Text.setGeometry(QtCore.QRect(370, 160, 331, 311))
28         self.Def_Text.setObjectName("Def_Text")
```

Chapitre III : Formalisation linguistique

```
29 self.Btn_Chercher = QtWidgets.QPushButton(self.centralwidget)
30 self.Btn_Chercher.setGeometry(QtCore.QRect(70, 20, 161, 41))
31 self.Btn_Chercher.setObjectName("Btn_Chercher")
32 self.Ln_Chercher = QtWidgets.QLineEdit(self.centralwidget)
33 self.Ln_Chercher.setGeometry(QtCore.QRect(350, 20, 211, 41))
34 self.Ln_Chercher.setText("")
35 self.Ln_Chercher.setObjectName("Ln_Chercher")
36 MainWindow.setCentralWidget(self.centralwidget)
37 self.menubar = QtWidgets.QMenuBar(MainWindow)
38 self.menubar.setGeometry(QtCore.QRect(0, 0, 800, 21))
39 self.menubar.setObjectName("menubar")
40 MainWindow.setMenuBar(self.menubar)
41 self.statusbar = QtWidgets.QStatusBar(MainWindow)
42 self.statusbar.setObjectName("statusbar")
43 MainWindow.setStatusBar(self.statusbar)
44
45 self.retranslateUi(MainWindow)
46 QtCore.QMetaObject.connectSlotsByName(MainWindow)
47 self.Btn_Chercher.clicked.connect(self.ChercherRacine)
48
49 def retranslateUi(self, MainWindow):
50     _translate = QtCore.QCoreApplication.translate
51     MainWindow.setWindowTitle(_translate("MainWindow", "MainWindow"))
52     __sortingEnabled = self.List_Words.isSortingEnabled()
53     self.List_Words.setSortingEnabled(False)
54     #item = self.List_Words.item(0)
55     #item.setText(_translate("MainWindow", "Hypertension"))
56     #item = self.List_Words.item(1)
57     #item.setText(_translate("MainWindow", "Tension Artérielle "))
58     self.Btn_Chercher.setText(_translate("MainWindow", "Chercher"))
59 def ChercherRacine(self):
60     stemmer = FrenchStemmer()
61     Token=self.Ln_Chercher.text()
62     racine = stemmer.stem(Token)
63     print(racine)
64
65
66 if __name__ == "__main__":
67     import sys
68     app = QtWidgets.QApplication(sys.argv)
69     MainWindow = QtWidgets.QMainWindow()
70     ui = Ui_MainWindow()
71     ui.setupUi(MainWindow)
72     MainWindow.show()
73     sys.exit(app.exec_())
74
75
```

Figure 4 : Création de l'interface de l'application.

Chapitre III : Formalisation linguistique

Les lignes 1 à 304 représentent l'application :

```
Interface.py × Main_Window.py × Popup_mot.py × popup_Prefixe.py × popup_Racine.py × popup_Suffixe.py ×
1  # -*- coding: utf-8 -*-
2
3  # Form implementation generated from reading ui file 'Main_Window.ui'
4  #
5  # Created by: PyQt5 UI code generator 5.9.2
6  #
7  # WARNING! All changes made in this file will be lost!
8
9  from PyQt5 import QtCore, QtGui, QtWidgets
10 from Popup_mot import Ui_Form_Mot
11 from PyQt5.QtWidgets import QMessageBox
12 from nltk.stem.snowball import FrenchStemmer
13
14 class Ui_MainWindow(object):
15
16     def OpenWindow(self):
17         self.listWidget_Mots.clear()
18         self.textBrowser_Pref.clear()
19         self.textBrowser_Def_Pref.clear()
20         self.textBrowser_Suff.clear()
21         self.textBrowser_Suff_2.clear()
22         self.textBrowser_Racine.clear()
23         #self.textBrowser_Def_Raci.clear()
24         self.TextBrowser_Def_mot.clear()
25         self.window = QtWidgets.QMainWindow()
26         self.ui = Ui_Form_Mot()
27         self.ui.setupUi(self.window)
28         self.window.show()
29
30     def setupUi(self, MainWindow):
31         MainWindow.setObjectName("MainWindow")
32         MainWindow.setEnabled(True)
33         MainWindow.resize(921, 596)
34         self.centralwidget = QtWidgets.QWidget(MainWindow)
35         self.centralwidget.setObjectName("centralwidget")
36         self.Btn_chercher = QtWidgets.QPushButton(self.centralwidget)
37         self.Btn_chercher.setGeometry(QtCore.QRect(180, 30, 161, 41))
38         self.Btn_chercher.setObjectName("Btn_chercher")
39         self.lineEdit_chercher = QtWidgets.QLineEdit(self.centralwidget)
40         self.lineEdit_chercher.setGeometry(QtCore.QRect(380, 30, 281, 41))
41         self.lineEdit_chercher.setObjectName("lineEdit_chercher")
42         self.label_message = QtWidgets.QLabel(self.centralwidget)
43         self.label_message.setGeometry(QtCore.QRect(330, 510, 311, 31))
44         self.label_message.setLayoutDirection(QtCore.Qt.LeftToRight)
45         self.label_message.setText("")
46         self.label_message.setIndent(-1)
47         self.label_message.setObjectName("label_message")
48         self.Btn_Ajouter_Mot = QtWidgets.QPushButton(self.centralwidget)
49         self.Btn_Ajouter_Mot.setGeometry(QtCore.QRect(720, 510, 161, 41))
50         self.Btn_Ajouter_Mot.setObjectName("Btn_Ajouter_Mot")
51         self.listWidget_Mots = QtWidgets.QListWidget(self.centralwidget)
52         self.listWidget_Mots.setGeometry(QtCore.QRect(30, 140, 161, 351))
53         self.listWidget_Mots.setObjectName("listWidget_Mots")
54         self.label = QtWidgets.QLabel(self.centralwidget)
55         self.label.setGeometry(QtCore.QRect(30, 100, 161, 31))
56         self.label.setObjectName("label")
```

Chapitre III : Formalisation linguistique

```
57 self.label_2 = QtWidgets.QLabel(self.centralwidget)
58 self.label_2.setGeometry(QtCore.QRect(280, 250, 51, 31))
59 self.label_2.setObjectName("label_2")
60 self.label_3 = QtWidgets.QLabel(self.centralwidget)
61 self.label_3.setGeometry(QtCore.QRect(400, 320, 51, 31))
62 self.label_3.setObjectName("label_3")
63 self.label_4 = QtWidgets.QLabel(self.centralwidget)
64 self.label_4.setGeometry(QtCore.QRect(280, 400, 51, 31))
65 self.label_4.setObjectName("label_4")
66 self.textBrowser_Pref = QtWidgets.QTextBrowser(self.centralwidget)
67 self.textBrowser_Pref.setEnabled(True)
68 self.textBrowser_Pref.setGeometry(QtCore.QRect(340, 240, 161, 41))
69 self.textBrowser_Pref.setReadOnly(True)
70 self.textBrowser_Pref.setObjectName("textBrowser_Pref")
71 self.textBrowser_Racine = QtWidgets.QTextBrowser(self.centralwidget)
72 self.textBrowser_Racine.setGeometry(QtCore.QRect(450, 320, 161, 41))
73 self.textBrowser_Racine.setObjectName("textBrowser_Racine")
74 self.textBrowser_Suff = QtWidgets.QTextBrowser(self.centralwidget)
75 self.textBrowser_Suff.setGeometry(QtCore.QRect(340, 400, 161, 41))
76 self.textBrowser_Suff.setObjectName("textBrowser_Suff")
77 #self.label_5 = QtWidgets.QLabel(self.centralwidget)
78 #self.label_5.setGeometry(QtCore.QRect(510, 330, 51, 31))
79 #self.label_5.setObjectName("label_5")
80 self.textBrowser_Def_Pref = QtWidgets.QTextBrowser(self.centralwidget)
81 self.textBrowser_Def_Pref.setGeometry(QtCore.QRect(570, 230, 191, 61))
82 self.textBrowser_Def_Pref.setObjectName("textBrowser_Def_Pref")
83 self.label_6 = QtWidgets.QLabel(self.centralwidget)
84 self.label_6.setGeometry(QtCore.QRect(510, 400, 51, 31))
85 self.label_6.setObjectName("label_6")
86 self.label_7 = QtWidgets.QLabel(self.centralwidget)
87 self.label_7.setGeometry(QtCore.QRect(510, 250, 51, 31))
88 self.label_7.setObjectName("label_7")
89 #self.textBrowser_Def_Raci = QtWidgets.QTextBrowser(self.centralwidget)
90 #self.textBrowser_Def_Raci.setGeometry(QtCore.QRect(570, 310, 191, 61))
91 #self.textBrowser_Def_Raci.setObjectName("textBrowser_Def_Raci")
92 self.textBrowser_Suff_2 = QtWidgets.QTextBrowser(self.centralwidget)
93 self.textBrowser_Suff_2.setGeometry(QtCore.QRect(570, 390, 191, 61))
94 self.textBrowser_Suff_2.setObjectName("textBrowser_Suff_2")
95 self.Btn_Supprimer_Mot = QtWidgets.QPushButton(self.centralwidget)
96 self.Btn_Supprimer_Mot.setGeometry(QtCore.QRect(30, 510, 161, 41))
97 self.Btn_Supprimer_Mot.setObjectName("Btn_Supprimer_Mot")
98 self.TextBrowser_Def_mot = QtWidgets.QTextBrowser(self.centralwidget)
99 self.TextBrowser_Def_mot.setEnabled(True)
100 self.TextBrowser_Def_mot.setGeometry(QtCore.QRect(380, 120, 281, 61))
101 self.TextBrowser_Def_mot.setReadOnly(True)
102 self.TextBrowser_Def_mot.setObjectName("TextBrowser_Def_mot")
103 self.label_8 = QtWidgets.QLabel(self.centralwidget)
104 self.label_8.setGeometry(QtCore.QRect(480, 90, 91, 31))
105 self.label_8.setObjectName("label_8")
106 self.label_9 = QtWidgets.QLabel(self.centralwidget)
107 self.label_9.setGeometry(QtCore.QRect(440, 190, 171, 31))
108 font = QtGui.QFont()
109 font.setPointSize(10)
110 font.setBold(True)
111 font.setWeight(75)
112 self.label_9.setFont(font)
```

Chapitre III : Formalisation linguistique

```
113 self.label_9.setLayoutDirection(QtCore.Qt.LeftToRight)
114 self.label_9.setObjectName("label_9")
115 MainWindow.setCentralWidget(self.centralwidget)
116 self.menubar = QtWidgets.QMenuBar(MainWindow)
117 self.menubar.setGeometry(QtCore.QRect(0, 0, 921, 21))
118 self.menubar.setObjectName("menubar")
119 MainWindow.setMenuBar(self.menubar)
120 self.statusbar = QtWidgets.QStatusBar(MainWindow)
121 self.statusbar.setObjectName("statusbar")
122 MainWindow.setStatusBar(self.statusbar)
123
124 self.retranslateUi(MainWindow)
125 QtCore.QMetaObject.connectSlotsByName(MainWindow)
126 self.Btn_chercher.clicked.connect(self.chercher)
127 self.listWidget_Mots.clicked.connect(self.CompDef)
128 self.Btn_Ajouter_Mot.clicked.connect(self.OpenWindow)
129 self.Btn_Supprimer_Mot.clicked.connect(self.Delete_Mot)
130
131 #Definition des fonctions:
132 def chercher(self):
133     self.listWidget_Mots.clear()
134     import os
135     import sqlite3
136
137     db_filename = 'BDD.db'
138     schema_filename = 'BDD.sql'
139
140     db_is_new = not os.path.exists(db_filename)
141
142     with sqlite3.connect(db_filename) as conn:
143         if db_is_new:
144             print('Creating schema')
145             with open(schema_filename, 'rt') as f:
146                 schema = f.read()
147                 conn.executescript(schema)
148                 self.label_message.setText('Inserting initial data')
149
150         else:
151             self.label_message.setText('Database exists!!')
152
153     if(self.lineEdit_chercher.text()!=''):
154         Token=self.lineEdit_chercher.text().capitalize()
155
156         conn=sqlite3.connect(db_filename)
157         c=conn.cursor()
158         query="Select ecrit_raci from Mot where ecrit_mot="
159         c.execute(query,(Token,))
160         row=c.fetchone()
161         if(row!=None):
162             Racine=row[0]
163             if(Racine==None):
164                 self.listWidget_Mots.addItem(Token)
165             else:
166                 query = "Select ecrit_mot from Mot where ecrit_raci=? order by ecrit_mot"
167                 c.execute(query, (Racine,))
```

Chapitre III : Formalisation linguistique

```
168         rows = c.fetchall()
169
170         for row in rows:
171             self.listWidget_Mots.addItem(row[0])
172
173     else:
174         self.label_message.setText("Le mot n'existe pas dans la base de données")
175         stemmer = FrenchStemmer()
176         Radical = stemmer.stem(Token)
177         message = "Le radical en utilisant le module FrenchStemmer d'NLTK est :" + Radical
178         msg = QMessageBox()
179         msg.setIcon(QMessageBox.Information)
180         msg.setText(message)
181         msg.exec_()
182
183
184     else:
185         self.label_message.setText('Il faut écrire un mot')
186
187 def CompDef(self):
188     import os
189     import sqlite3
190
191     db_filename = 'BDD.db'
192     schema_filename = 'BDD.sql'
193
194     db_is_new = not os.path.exists(db_filename)
195
196     with sqlite3.connect(db_filename) as conn:
197         if db_is_new:
198             print('Creating schema')
199             with open(schema_filename, 'rt') as f:
200                 schema = f.read()
201                 conn.executescript(schema)
202                 self.label_message.setText('Inserting initial data')
203
204         else:
205             self.label_message.setText('Database exists!!')
206     mot=self.listWidget_Mots.currentItem().text()
207     conn=sqlite3.connect(db_filename)
208     c=conn.cursor()
209     query='select * from mot where ecrit_mot=?'
210     c.execute(query,(mot,))
211     row=c.fetchone()
212     Racine=row[1]
213     Definition=row[2]
214     Prefixe=row[3]
215     Suffixe=row[4]
216     if(Racine==None):
217         self.textBrowser_Racine.clear()
218         #self.textBrowser_Def_Raci.clear()
219     else:
220         QueryR="Select * from Racine where ecrit_raci=?"
221         c.execute(QueryR,(Racine,))
222         row = c.fetchone()
223
```

Chapitre III : Formalisation linguistique

```
224 |         self.textBrowser_Racine.setText(row[0])
225 |         #self.textBrowser_Def_Raci.setText(row[1])
226 |     if (Prefixe == None):
227 |         self.textBrowser_Pref.clear()
228 |         self.textBrowser_Def_Pref.clear()
229 |     else:
230 |         QueryP = "Select * from Prefixe where ecrit_pref=?"
231 |         c.execute(QueryP, (Prefixe,))
232 |         row = c.fetchone()
233 |         self.textBrowser_Pref.setText(row[0])
234 |         self.textBrowser_Def_Pref.setText(row[1])
235 |     if ( Suffixe == None ):
236 |         self.textBrowser_Suff.clear()
237 |         self.textBrowser_Suff_2.clear()
238 |     else:
239 |         QueryS = "Select * from Suffixe where ecrit_suff=?"
240 |         c.execute(QueryS, (Suffixe,))
241 |         row = c.fetchone()
242 |         self.textBrowser_Suff.setText(row[0])
243 |
244 |         self.textBrowser_Suff_2.setText(row[1])
245 |     if(Definition == None):
246 |         self.TextBrowser_Def_mot.setText("")
247 |     else:
248 |         self.TextBrowser_Def_mot.setText(Definition)
249 |
250 | def Delete Mot(self):...
251 |     import sqlite3
252 |     if (self.listWidget_Mots.currentItem() != None):
253 |         if (self.listWidget_Mots.currentItem().text() == ''):
254 |             self.label_Message.setText('il faut Selectionner un mot valide !!')
255 |         else:
256 |             try:
257 |                 token = self.listWidget_Mots.currentItem().text()
258 |                 print(token)
259 |                 db_filename = 'BDD.db'
260 |                 # schema_filename = 'BDD.sql'
261 |                 query = 'delete from Mot where ecrit_mot = ?'
262 |                 with sqlite3.connect(db_filename) as conn:...
263 |
264 |             except sqlite3.Error as error:
265 |                 print("Failed to delete record from sqlite table", error)
266 |
267 |         else:
268 |             self.label_message.setText('il faut selectionner un mot!')
269 |
270 |
271 | def retranslateUi(self, MainWindow):
272 |     _translate = QtCore.QCoreApplication.translate
273 |     MainWindow.setWindowTitle(_translate("MainWindow", "MainWindow"))
274 |     self.Btn_chercher.setText(_translate("MainWindow", "Chercher"))
275 |     self.Btn_Ajouter_Mot.setText(_translate("MainWindow", "Ajouter un mot"))
276 |     self.label.setText(_translate("MainWindow", " Mots avec le même radical"))
277 |     self.label_2.setText(_translate("MainWindow", "Préfixe"))
278 |     self.label_3.setText(_translate("MainWindow", "Radical"))
279 |     self.label_4.setText(_translate("MainWindow", "Suffixe"))
280 |     #self.label_5.setText(_translate("MainWindow", "Définition"))
281 |     self.label_6.setText(_translate("MainWindow", "Définition"))
```

Chapitre III : Formalisation linguistique

```
291 self.label_7.setText(_translate("MainWindow", "Définition"))
292 self.Btn_Supprimer_Mot.setText(_translate("MainWindow", "Supprimer "))
293 self.label_8.setText(_translate("MainWindow", "Définition du mot"))
294 self.label_9.setText(_translate("MainWindow", "Composition"))
295
296
297 if __name__ == "__main__":
298     import sys
299     app = QtWidgets.QApplication(sys.argv)
300     MainWindow = QtWidgets.QMainWindow()
301     ui = Ui_MainWindow()
302     ui.setupUi(MainWindow)
303     MainWindow.show()
304     sys.exit(app.exec_())
```

Figure 5 : Algorithmes représentant la création de l'application.

Les lignes 1 à 370 représentent la fenêtre d'ajout de mot, ou nous pouvons ajouter ou supprimer les morphèmes présents dans la base de données. Elles servent également à donner les définitions aux mots ajoutés.

```
Interface.py × Main_Window.py × Popup_mot.py × popup_Prefixe.py × popup_Racine.py × popup_Suffixe.py ×
1 ...
8
9
10 from PyQt5 import QtCore, QtGui, QtWidgets
11 from PyQt5.QtWidgets import QMessageBox
12 import sqlite3
13 import os
14 from popup_Prefixe import Ui_Form
15 from popup_Racine import Ui_Form3
16 from popup_Suffixe import Ui_Form2
17 class Ui_Form_Mot(object):
18
19     def __init__(self):
20         super().__init__()
21
22     def OpenPrefixe(self):
23         self.window = QtWidgets.QMainWindow()
24         self.ui = Ui_Form(self)
25         self.ui.setupUi(self.window)
26         self.window.show()
27
28     def OpenRacine(self):
29         self.window3 = QtWidgets.QMainWindow()
30         self.ui = Ui_Form3(self)
31         self.ui.setupUi(self.window3)
32         self.window3.show()
33     def OpenSuffixe(self):
34         self.window2 = QtWidgets.QMainWindow()
```

Chapitre III : Formalisation linguistique

```
35     self.ui = Ui_Form2(self)
36     self.ui.setupUi(self.window2)
37     self.window2.show()
38     def DeletePrefixe(self):
39         import sqlite3
40         if(self.listWidget_Pref.currentItem() != None):
41             if(self.listWidget_Pref.currentItem().text()==''):
42                 self.label_Message.setText('il faut Selectionner un préfixe valide!')
43             else:
44                 token = self.listWidget_Pref.currentItem().text()
45                 db_filename = 'BDD.db'
46                 #schema_filename = 'BDD.sql'
47                 query='delete from Prefixe where ecrit_pref = ?'
48                 Query1='delete from Mot where ecrit_pref = ?'
49
50                 with sqlite3.connect(db_filename) as conn:
51                     msg=QMessageBox()
52                     msg.setIcon(QMessageBox.Question)
53                     msg.setText('En cas de suppression,tous les mots construits avec ce préfixe seront s
54                     msg.setStandardButtons(QMessageBox.Yes|QMessageBox.No)
55                     msg.setDefaultButton(QMessageBox.No)
56                     buttonY = msg.button(QMessageBox.Yes)
57                     buttonY.setText('Oui')
58                     buttonN = msg.button(QMessageBox.No)
59                     buttonN.setText('Non')
60                     msg.exec_()
61                     if(msg.clickedButton() == buttonY):
62                         c=conn.cursor()
63                         c.execute(Query1, (token,))
64                         c.execute(query,(token,))
65                         rowIndex=self.listWidget_Pref.currentRow()
66                         self.listWidget_Pref.takeItem(rowIndex)
67
68                         self.listWidget_Pref.setCurrentRow(0)
69
70
71         else:
72             self.label_Message.setText('il faut selectionner un préfixe!')
73     def DeleteRacine(self):
74         if(self.listWidget_Raci.currentItem() != None):
75             if(self.listWidget_Raci.currentItem().text()==''):
76                 self.label_Message.setText('il faut Selectionner un radical valide!')
77             else:
78                 token = self.listWidget_Raci.currentItem().text()
79                 db_filename = 'BDD.db'
80                 #schema_filename = 'BDD.sql'
81                 query='delete from Racine where ecrit_raci = ?'
82                 Query1 = 'delete from Mot where ecrit_raci =?'
83                 with sqlite3.connect(db_filename) as conn:
84                     msg=QMessageBox()
85                     msg.setIcon(QMessageBox.Question)
86                     msg.setText('En cas de suppression,tous les mots construits avec ce radical seront s
87                     msg.setStandardButtons(QMessageBox.Yes|QMessageBox.No)
88                     msg.setDefaultButton(QMessageBox.No)
89                     buttonY = msg.button(QMessageBox.Yes)
```

Chapitre III : Formalisation linguistique

```
90         buttonY.setText('Oui')
91         buttonN = msg.button(QMessageBox.No)
92         buttonN.setText('Non')
93         msg.exec_()
94         if(msg.clickedButton() == buttonY):
95             c=conn.cursor()
96             c.execute(Query1, (token,))
97             c.execute(query,(token,))
98             rowIndex=self.listWidget_Raci.currentRow()
99             self.listWidget_Raci.takeItem(rowIndex)
100
101         self.listWidget_Raci.setCurrentRow(0)
102     else:
103         self.label_Message.setText('il faut selectionner un radical!')
104
105     def DeleteSuffixe(self):
106         if(self.listWidget_Suff.currentItem() != None):
107             if(self.listWidget_Suff.currentItem().text()==''):
108                 self.label_Message.setText('il faut Selectionner un suffixe valide!')
109             else:
110                 token = self.listWidget_Suff.currentItem().text()
111                 db_filename = 'BDD.db'
112                 #schema_filename = 'BDD.sql'
113                 query='delete from Suffixe where ecrit_suff = ?'
114                 Query1='delete from Mot where ecrit_suff =?'
115                 with sqlite3.connect(db_filename) as conn:
116                     msg=QMessageBox()
117                     msg.setIcon(QMessageBox.Question)
118                     msg.setText('En cas de suppression,tous les mots construits avec ce suffixe seront supprimés')
119                     msg.setStandardButtons(QMessageBox.Yes|QMessageBox.No)
120                     msg.setDefaultButton(QMessageBox.No)
121                     buttonY = msg.button(QMessageBox.Yes)
122                     buttonY.setText('Oui')
123                     buttonN = msg.button(QMessageBox.No)
124                     buttonN.setText('Non')
125                     msg.exec_()
126                     if(msg.clickedButton() == buttonY):
127                         c=conn.cursor()
128                         c.execute(Query1, (token,))
129                         c.execute(query,(token,))
130                         rowIndex=self.listWidget_Suff.currentRow()
131                         self.listWidget_Suff.takeItem(rowIndex)
132
133                     self.listWidget_Suff.setCurrentRow(0)
134
135         else:
136             self.label_Message.setText('il faut selectionner un suffixe!')
137
138
139
140
141     def setupUi(self, Form_Mot):
142         Form_Mot.setObjectName("Form_Mot")
143         Form_Mot.resize(644, 547)
144         self.Btn_Ajouter_Pref = QtWidgets.QPushButton(Form_Mot)
```

Chapitre III : Formalisation linguistique

```
145 self.Btn_Ajouter_Pref.setGeometry(QRect(40, 20, 111, 31))
146 self.Btn_Ajouter_Pref.setObjectName("Btn_Ajouter_Pref")
147 self.Btn_Ajouter_Racine = QtWidgets.QPushButton(Form_Mot)
148 self.Btn_Ajouter_Racine.setGeometry(QRect(240, 20, 111, 31))
149 self.Btn_Ajouter_Racine.setObjectName("Btn_Ajouter_Racine")
150 self.Btn_Ajouter_Suff = QtWidgets.QPushButton(Form_Mot)
151 self.Btn_Ajouter_Suff.setGeometry(QRect(450, 20, 111, 31))
152 self.Btn_Ajouter_Suff.setObjectName("Btn_Ajouter_Suff")
153 self.label = QtWidgets.QLabel(Form_Mot)
154 self.label.setGeometry(QRect(40, 330, 101, 21))
155 self.label.setObjectName("label")
156 self.label_2 = QtWidgets.QLabel(Form_Mot)
157 self.label_2.setGeometry(QRect(250, 330, 101, 21))
158 self.label_2.setObjectName("label_2")
159 self.label_3 = QtWidgets.QLabel(Form_Mot)
160 self.label_3.setGeometry(QRect(460, 330, 101, 21))
161 self.label_3.setObjectName("label_3")
162 self.listWidget_Pref = QtWidgets.QListWidget(Form_Mot)
163 self.listWidget_Pref.setGeometry(QRect(20, 120, 151, 192))
164 self.listWidget_Pref.setObjectName("listWidget_Pref")
165 self.listWidget_Raci = QtWidgets.QListWidget(Form_Mot)
166 self.listWidget_Raci.setGeometry(QRect(220, 120, 151, 192))
167 self.listWidget_Raci.setObjectName("listWidget_Raci")
168 self.listWidget_Suff = QtWidgets.QListWidget(Form_Mot)
169 self.listWidget_Suff.setGeometry(QRect(430, 120, 151, 192))
170 self.listWidget_Suff.setObjectName("listWidget_Suff")
171 self.Btn_Ajouter_Mot = QtWidgets.QPushButton(Form_Mot)
172 self.Btn_Ajouter_Mot.setGeometry(QRect(100, 430, 161, 41))
173 self.Btn_Ajouter_Mot.setObjectName("Btn_Ajouter_Mot")
174 self.Btn_Supprimer_Racine = QtWidgets.QPushButton(Form_Mot)
175 self.Btn_Supprimer_Racine.setGeometry(QRect(240, 70, 111, 31))
176 self.Btn_Supprimer_Racine.setObjectName("Btn_Supprimer_Racine")
177 self.Btn_Supprimer_Pref = QtWidgets.QPushButton(Form_Mot)
178 self.Btn_Supprimer_Pref.setGeometry(QRect(40, 70, 111, 31))
179 self.Btn_Supprimer_Pref.setObjectName("Btn_Supprimer_Pref")
180 self.Btn_Supprimer_Suff = QtWidgets.QPushButton(Form_Mot)
181 self.Btn_Supprimer_Suff.setGeometry(QRect(450, 70, 111, 31))
182 self.Btn_Supprimer_Suff.setObjectName("Btn_Supprimer_Suff")
183 self.label_4 = QtWidgets.QLabel(Form_Mot)
184 self.label_4.setGeometry(QRect(10, 380, 91, 21))
185 self.label_4.setObjectName("label_4")
186 self.lineEdit = QtWidgets.QLineEdit(Form_Mot)
187 self.lineEdit.setGeometry(QRect(100, 380, 161, 31))
188 self.lineEdit.setObjectName("lineEdit")
189 self.textEdit_DefMot = QtWidgets.QTextEdit(Form_Mot)
190 self.textEdit_DefMot.setGeometry(QRect(390, 380, 211, 91))
191 self.textEdit_DefMot.setObjectName("textEdit_DefMot")
192 self.label_5 = QtWidgets.QLabel(Form_Mot)
193 self.label_5.setGeometry(QRect(310, 380, 51, 16))
194 self.label_5.setObjectName("label_5")
195 self.label_Message = QtWidgets.QLabel(Form_Mot)
196 self.label_Message.setGeometry(QRect(190, 500, 301, 20))
197 self.label_Message.setText("")
198 self.label_Message.setObjectName("label_Message")
199
200 self.retranslateUi(Form_Mot)
```

Chapitre III : Formalisation linguistique

```
201 QtCore.QMetaObject.connectSlotsByName(Form_Mot)
202 #connection widgets with actions
203 self.Btn_Ajouter_Mot.clicked.connect(self.Ajouter_Mot)
204 self.Btn_Ajouter_Pref.clicked.connect(self.OpenPrefixe)
205 self.Btn_Ajouter_Racine.clicked.connect(self.OpenRacine)
206 self.Btn_Ajouter_Suff.clicked.connect(self.OpenSuffixe)
207 self.Btn_Supprimer_Pref.clicked.connect(self.DeletePrefixe)
208 self.Btn_Supprimer_Racine.clicked.connect(self.DeleteRacine)
209 self.Btn_Supprimer_Suff.clicked.connect(self.DeleteSuffixe)
210
211 # Execute query to get data from tables
212 self.remplir()
213 def remplir(self):
214     db_filename = 'BDD.db'
215     schema_filename = 'BDD.sql'
216
217     db_is_new = not os.path.exists(db_filename)
218
219     with sqlite3.connect(db_filename) as conn:
220         if db_is_new:
221             print('Creating schema')
222             with open(schema_filename, 'rt') as f:
223                 schema = f.read()
224                 conn.executescript(schema)
225
226             print('Inserting initial data')
227
228         else:
229             print('Database exists!!')
230             self.listWidget_Pref.clear()
231             self.listWidget_Raci.clear()
232             self.listWidget_Suff.clear()
233             Racine_Query='Select ecrit_raci from Racine'
234             Prefixe_Query='Select ecrit_pref from Prefixe'
235             Suffixe_Query='Select ecrit_suff from Suffixe'
236
237
238             cur = conn.cursor()
239             cur.execute(Racine_Query)
240             rows=cur.fetchall()
241             self.listWidget_Raci.addItem('')
242             for row in rows:
243                 self.listWidget_Raci.addItem(row[0])
244             cur.execute(Prefixe_Query)
245             rows = cur.fetchall()
246             self.listWidget_Pref.addItem('')
247             for row in rows:
248                 self.listWidget_Pref.addItem(row[0])
249             cur.execute(Suffixe_Query)
250             rows = cur.fetchall()
251             self.listWidget_Suff.addItem('')
252             for row in rows:
253                 self.listWidget_Suff.addItem(row[0])
254
```

Chapitre III : Formalisation linguistique

```
256 ##### Define functions
257 def Ajouter_Mot(self):
258     db_filename = 'BDD.db'
259     schema_filename = 'BDD.sql'
260
261     db_is_new = not os.path.exists(db_filename)
262
263     with sqlite3.connect(db_filename) as conn:
264         if db_is_new:
265             print('Creating schema')
266             with open(schema_filename, 'rt') as f:
267                 schema = f.read()
268                 conn.executescript(schema)
269
270             print('Inserting initial data')
271
272
273
274     if(self.lineEdit.text()=="or self.textEdit_DefMot.toPlainText()=="'):
275         self.label_Message.setText('If faut remplir tous les champs!!')
276
277     else:
278
279         if(self.listWidget_Pref.currentItem()== None):
280             self.label_Message.setText("il faut selectionner un préfixe")
281
282         else :
283             if(self.listWidget_Raci.currentItem()==None):
284                 self.label_Message.setText("il faut selectionner un Radical")
285
286             else:
287                 if(self.listWidget_Suff.currentItem()==None):
288                     self.label_Message.clear()
289                     self.label_Message.setText("il faut selectionner une Suffixe")
290
291                 else:
292                     if(self.listWidget_Suff.currentItem().text()==""): Suffixe=None
293                     else: Suffixe=self.listWidget_Suff.currentItem().text()
294
295                     if (self.listWidget_Raci.currentItem().text() == ""): Racine = None
296                     else: Racine = self.listWidget_Raci.currentItem().text()
297
298                     if (self.listWidget_Pref.currentItem().text() == ""): Prefixe = None
299                     else: Prefixe = self.listWidget_Pref.currentItem().text()
300
301                     if(Suffixe== None and Prefixe==None and Racine==None):
302                         self.label_Message.setText('selectionnez au moins un composant du mot')
303                     else:
304                         c = conn.cursor()
305                         Mot = self.lineEdit.text().capitalize()
306                         Definition = self.textEdit_DefMot.toPlainText().capitalize()
307                         Tup = (Mot, Racine, Definition, Prefixe, Suffixe)
308                         query1 = "Select ecrit_raci from Mot where ecrit_mot=?"
309                         c.execute(query1, (Mot,))
310                         row = c.fetchone()
311                         if (row == None):
```

Chapitre III : Formalisation linguistique

```
312         Query = 'insert into Mot values(?,?,?,?)'
313         c.execute(Query, Tup)
314         self.lineEdit.clear()
315         self.textEdit_DefMot.clear()
316         self.label_Message.setText('Mot ajouté avec succès')
317     else:
318         self.label_Message.setText('Ce mot existe déjà dans la base de données')
319
320     self.listWidget_Pref.setCurrentRow(0)
321     self.listWidget_Raci.setCurrentRow(0)
322     self.listWidget_Suff.setCurrentRow(0)
323
324
325     def Supprimer_Pref(self):
326         db_filename = 'BDD.db'
327         schema_filename = 'BDD.sql'
328
329         db_is_new = not os.path.exists(db_filename)
330
331         with sqlite3.connect(db_filename) as conn:
332             if db_is_new:
333                 print('Creating schema')
334                 with open(schema_filename, 'rt') as f:
335                     schema = f.read()
336                     conn.executescript(schema)
337
338                 print('La base de données est vide !!')
339
340
341
342     else:
343         print('')
344
345
346     def retranslateUi(self, Form_Mot):
347         _translate = QtCore.QCoreApplication.translate
348         Form_Mot.setWindowTitle(_translate("Form_Mot", "Ajouter Mot"))
349         self.Btn_Ajouter_Pref.setText(_translate("Form_Mot", "Ajouter Préfixe"))
350         self.Btn_Ajouter_Racine.setText(_translate("Form_Mot", "Ajouter Radical"))
351         self.Btn_Ajouter_Suff.setText(_translate("Form_Mot", "Ajouter Suffixe"))
352         self.label.setText(_translate("Form_Mot", "Selectionner Préfixe"))
353         self.label_2.setText(_translate("Form_Mot", "Selectionner Radical"))
354         self.label_3.setText(_translate("Form_Mot", "Selectionner Suffixe"))
355         self.Btn_Ajouter_Mot.setText(_translate("Form_Mot", "Ajouter Mot"))
356         self.Btn_Supprimer_Racine.setText(_translate("Form_Mot", "Supprimer Radical"))
357         self.Btn_Supprimer_Pref.setText(_translate("Form_Mot", "Supprimer Préfixe"))
358         self.Btn_Supprimer_Suff.setText(_translate("Form_Mot", "Supprimer Suffixe"))
359         self.label_4.setText(_translate("Form_Mot", "Nouveau Mot"))
360         self.label_5.setText(_translate("Form_Mot", "Définition"))
361
362
363     if __name__ == "__main__":
364         import sys
365         app = QtWidgets.QApplication(sys.argv)
366         Form_Mot = QtWidgets.QWidget()
367         ui = Ui_Form_Mot()
368         ui.setupUi(Form_Mot)
369         Form_Mot.show()
370         sys.exit(app.exec_())
371
```

Figure 6 : Création de la fenêtre d'ajout de mots.

Chapitre III : Formalisation linguistique

Les lignes 1 à 117 représentent la fenêtre d'ajout des préfixes et leurs définitions.

```
Interface.py x Main_Window.py x Popup_mot.py x popup_Prefixe.py x popup_Racine.py x popup_Suffixe.py x
1  ...
8
9
10 from PyQt5 import QtCore,QtWidgets
11 from PyQt5.QtWidgets import QMessageBox
12
13
14 class Ui_Form(object):
15
16
17     def __init__(self,parent=None):
18         self.Parent=parent
19         super().__init__()
20
21
22     def setupUi(self, Form):
23         Form.setObjectName("Form")
24         Form.resize(315, 294)
25         self.Ln_Pref = QtWidgets.QLineEdit(Form)
26         self.Ln_Pref.setGeometry(QtCore.QRect(130, 40, 161, 31))
27         self.Ln_Pref.setObjectName("Ln_Pref")
28         self.TextEdit_Def = QtWidgets.QTextEdit(Form)
29         self.TextEdit_Def.setGeometry(QtCore.QRect(130, 90, 161, 141))
30         self.TextEdit_Def.setObjectName("TextEdit_Def")
31         self.Btn_Ajouter_Pref = QtWidgets.QPushButton(Form)
32         self.Btn_Ajouter_Pref.setGeometry(QtCore.QRect(90, 250, 131, 31))
33         self.Btn_Ajouter_Pref.setCheckable(False)
34         self.Btn_Ajouter_Pref.setObjectName("Btn_Ajouter_Pref")
35         self.label = QtWidgets.QLabel(Form)
36         self.label.setGeometry(QtCore.QRect(30, 40, 71, 31))
37         self.label.setObjectName("label")
38         self.label_2 = QtWidgets.QLabel(Form)
39         self.label_2.setGeometry(QtCore.QRect(30, 140, 81, 31))
40         self.label_2.setObjectName("label_2")
41
42         self.retranslateUi(Form)
43         QtCore.QMetaObject.connectSlotsByName(Form)
44         self.Btn_Ajouter_Pref.clicked.connect(self.AjouterPrefixe)
45         #self.Btn_Ajouter_Pref.clicked.connect(self.on_submit)
46
47     def retranslateUi(self, Form):
48         _translate = QtCore.QCoreApplication.translate
49         Form.setWindowTitle(_translate("Form", "Ajouter Préfixe"))
50         self.Btn_Ajouter_Pref.setText(_translate("Form", "Ajouter"))
51         self.label.setText(_translate("Form", "Préfixe"))
52         self.label_2.setText(_translate("Form", "Définition"))
53
54
55
56     def AjouterPrefixe(self):
57         Pref=self.Ln_Pref.text().capitalize()
58         Def=self.TextEdit_Def.toPlainText().capitalize()
59         import os
60         import sqlite3
61
62         db_filename = 'BDD.db'
```

Chapitre III : Formalisation linguistique

```
63 schema_filename = 'BDD.sql'
64
65 db_is_new = not os.path.exists(db_filename)
66
67 with sqlite3.connect(db_filename) as conn:
68     if db_is_new:
69         print('Creating schema')
70         with open(schema_filename, 'rt') as f:
71             schema = f.read()
72             conn.executescript(schema)
73
74         print('Inserting initial data')
75
76
77     else:
78         print('Database exists!!')
79
80     c = conn.cursor()
81     msgBox = QMessageBox()
82     if(Pref!='' and Def!=''):
83         tuple=(Pref,Def)
84         Prefixe_Query = 'Select * from Prefixe where ecrit_pref= ?'
85         cur = conn.cursor()
86         cur.execute(Prefixe_Query,(Pref,))
87         row = cur.fetchone()
88
89         if(row == None):
90             c.execute('insert into Prefixe values(?,?)', tuple)
91             self.Ln_Pref.setText('')
92             self.TextEdit_Def.setText('')
93             self.Parent.listWidget_Pref.addItem(Pref)
94             msgBox.setIcon(QMessageBox.Information)
95             msgBox.setText('Préfixe ajouté avec succès')
96             msgBox.exec()
97         else:
98             msgBox.setIcon(QMessageBox.Critical)
99             msgBox.setText("le préfixe existe déjà dans la base de données!")
100            msgBox.exec()
101
102     else:
103         msgBox.setIcon(QMessageBox.Warning)
104         msgBox.setText('Il faut remplir tous les champs!')
105         msgBox.exec()
106
107
108
109
110 if __name__ == "__main__":
111     import sys
112     app = QtWidgets.QApplication(sys.argv)
113     Form_Mot = QtWidgets.QWidget()
114     ui = Ui_Form()
115     ui.setupUi(Form_Mot)
116     Form_Mot.show()
117     sys.exit(app.exec_())
```

Figure 7 : Création de la fenêtre d'ajout de préfixes.

Chapitre III : Formalisation linguistique

Les lignes 1 à 108 représentent la fenêtre d'ajout des radicaux.

```
Interface.py x Main_Window.py x Popup_mot.py x popup_Prefixe.py x popup_Racine.py x popup_Suffixe.py x
1 |...
8
9
10 from PyQt5 import QtCore, QtGui, QtWidgets
11 from PyQt5.QtWidgets import QMessageBox
12
13
14 class Ui_Form3(object):
15
16     def __init__(self, parent=None):
17         self.Parent=parent
18         super().__init__()
19     def setupUi(self, Form3):
20         Form3.setObjectName("Form3")
21         Form3.resize(345, 300)
22         self.Btn_Ajouter_Racine = QtWidgets.QPushButton(Form3)
23         self.Btn_Ajouter_Racine.setGeometry(QtCore.QRect(100, 240, 131, 31))
24         self.Btn_Ajouter_Racine.setCheckable(False)
25         self.Btn_Ajouter_Racine.setObjectName("Btn_Ajouter_Racine")
26         self.label = QtWidgets.QLabel(Form3)
27         self.label.setGeometry(QtCore.QRect(40, 30, 71, 31))
28         self.label.setObjectName("label")
29         self.Ln_Raci = QtWidgets.QLineEdit(Form3)
30         self.Ln_Raci.setGeometry(QtCore.QRect(140, 30, 161, 31))
31         self.Ln_Raci.setObjectName("Ln_Raci")
32         self.label_2 = QtWidgets.QLabel(Form3)
33         self.label_2.setGeometry(QtCore.QRect(40, 130, 81, 31))
34         self.label_2.setObjectName("label_2")
35         self.TextEdit_Def = QtWidgets.QTextEdit(Form3)
36         self.TextEdit_Def.setGeometry(QtCore.QRect(140, 80, 161, 141))
37         self.TextEdit_Def.setObjectName("TextEdit_Def")
38
39         self.retranslateUi(Form3)
40         QtCore.QMetaObject.connectSlotsByName(Form3)
41         self.Btn_Ajouter_Racine.clicked.connect(self.AjouterRacine)
42     def retranslateUi(self, Form3):
43         _translate = QtCore.QCoreApplication.translate
44         Form3.setWindowTitle(_translate("Form3", "Ajouter Radical"))
45         self.Btn_Ajouter_Racine.setText(_translate("Form3", "Ajouter"))
46         self.label.setText(_translate("Form3", "Racine"))
47         self.label_2.setText(_translate("Form3", "Définition"))
48
49     def AjouterRacine(self):
50         Raci=self.Ln_Raci.text().capitalize()
51         Def=self.TextEdit_Def.toPlainText().capitalize()
52         import os
53         import sqlite3
54
55         db_filename = 'BDD.db'
56         schema_filename = 'BDD.sql'
57
58         db_is_new = not os.path.exists(db_filename)
59
60         with sqlite3.connect(db_filename) as conn:
61             if db_is_new:
62                 print('Creating schema')
```

Chapitre III : Formalisation linguistique

```
63     with open(schema_filename, 'rt') as f:
64         schema = f.read()
65     conn.executescript(schema)
66
67     print('Inserting initial data')
68
69
70     else:
71         print('Database exists!!')
72
73     c = conn.cursor()
74     msgBox = QMessageBox()
75     if(Raci!='' and Def!=''):
76         tuple=(Raci,Def)
77         Prefixe_Query = 'Select * from Racine where ecrit_raci= ?'
78         cur = conn.cursor()
79         cur.execute(Prefixe_Query, (Raci,))
80         row = cur.fetchone()
81
82         if (row == None):
83             c.execute('insert into Racine values(?,?)', tuple)
84             self.Ln_Raci.setText('')
85             self.TextEdit_Def.setText('')
86             self.Parent.listWidget_Raci.addItem(Raci)
87             msgBox.setIcon(QMessageBox.Information)
88             msgBox.setText('Radical ajouté avec succès')
89             msgBox.exec()
90
91         else:
92             msgBox.setIcon(QMessageBox.Critical)
93             msgBox.setText("le radical existe déjà dans la base de données")
94             msgBox.exec()
95
96         else:
97             msgBox.setIcon(QMessageBox.Warning)
98             msgBox.setText('Il faut remplir tous les champs!')
99             msgBox.exec()
100     if __name__ == "__main__":
101         import sys
102         app = QtWidgets.QApplication(sys.argv)
103         Form3 = QtWidgets.QWidget()
104         ui = Ui_Form3()
105         ui.setupUi(Form3)
106         Form3.show()
107         sys.exit(app.exec_())
```

Figure 8 : Création de la fenêtre d'ajout des radicaux.

Chapitre III : Formalisation linguistique

Les lignes 1 à 105 représentent la fenêtre d'ajout des suffixes et leurs définitions.

```
Interface.py x Main_Window.py x Popup_mot.py x popup_Prefixe.py x popup_Racine.py x popup_Suffixe.py x
1 |...
8
9
10 from PyQt5 import QtCore, QtGui, QtWidgets
11 from PyQt5.QtWidgets import QMessageBox
12
13
14 class Ui_Form2(object):
15     def __init__(self, parent=None):
16         self.Parent=parent
17         super().__init__()
18
19     def setupUi(self, Form2):
20         Form2.setObjectName("Form2")
21         Form2.resize(316, 300)
22         self.Btn_Ajouter_Suff = QtWidgets.QPushButton(Form2)
23         self.Btn_Ajouter_Suff.setGeometry(QtCore.QRect(90, 240, 131, 31))
24         self.Btn_Ajouter_Suff.setCheckable(False)
25         self.Btn_Ajouter_Suff.setObjectName("Btn_Ajouter_Suff")
26         self.label_2 = QtWidgets.QLabel(Form2)
27         self.label_2.setGeometry(QtCore.QRect(30, 130, 81, 31))
28         self.label_2.setObjectName("label_2")
29         self.label = QtWidgets.QLabel(Form2)
30         self.label.setGeometry(QtCore.QRect(30, 30, 71, 31))
31         self.label.setObjectName("label")
32         self.Ln_Suff = QtWidgets.QLineEdit(Form2)
33         self.Ln_Suff.setGeometry(QtCore.QRect(130, 30, 161, 31))
34         self.Ln_Suff.setObjectName("Ln_Suff")
35
36         self.TextEdit_Def = QtWidgets.QTextEdit(Form2)
37         self.TextEdit_Def.setGeometry(QtCore.QRect(130, 80, 161, 141))
38         self.TextEdit_Def.setObjectName("TextEdit_Def")
39
40         self.retranslateUi(Form2)
41         QtCore.QMetaObject.connectSlotsByName(Form2)
42         self.Btn_Ajouter_Suff.clicked.connect(self.AjouterSuffixe)
43     def retranslateUi(self, Form2):
44         _translate = QtCore.QCoreApplication.translate
45         Form2.setWindowTitle(_translate("Form2", "Ajouter Suffixe"))
46         self.Btn_Ajouter_Suff.setText(_translate("Form2", "Ajouter"))
47         self.label_2.setText(_translate("Form2", "Définition"))
48         self.label.setText(_translate("Form2", "Suffixe"))
49     def AjouterSuffixe(self):
50         Suff=self.Ln_Suff.text().capitalize()
51         Def=self.TextEdit_Def.toPlainText().capitalize()
52         import os
53         import sqlite3
54
55         db_filename = 'BDD.db'
56         schema_filename = 'BDD.sql'
57
58         db_is_new = not os.path.exists(db_filename)
59
60         with sqlite3.connect(db_filename) as conn:
61             if db_is_new:
62                 print('Creating schema')
63                 with open(schema_filename, 'rt') as f:
```

Chapitre III : Formalisation linguistique

```
62     with open(schema_filename, 'rt') as f:
63         schema = f.read()
64         conn.executescript(schema)
65
66     print('Inserting initial data')
67
68
69     else:
70         print('Database exists!!')
71
72     c = conn.cursor()
73     msgBox = QMessageBox()
74     if(Suff!='' and Def!=''):
75         tuple=(Suff,Def)
76         Prefixe_Query = 'Select * from Suffixe where ecrit_suff= ?'
77         cur = conn.cursor()
78         cur.execute(Prefixe_Query, (Suff,))
79         row = cur.fetchone()
80
81         if (row == None):
82             c.execute('insert into Suffixe values(?,?)', tuple)
83             self.Ln_Suff.setText('')
84             self.TextEdit_Def.setText('')
85             self.Parent.listWidget_Suff.addItem(Suff)
86             msgBox.setIcon(QMessageBox.Information)
87             msgBox.setText('Suffixe ajouté avec succès!')
88             msgBox.exec()
89
90         else:
91             msgBox.setIcon(QMessageBox.Critical)
92             msgBox.setText("le suffixe existe déjà dans la base de données!")
93             msgBox.exec()
94     else:
95         msgBox.setIcon(QMessageBox.Warning)
96         msgBox.setText('Il faut remplir tous les champs!')
97         msgBox.exec()
98     if __name__ == "__main__":
99         import sys
100         app = QtWidgets.QApplication(sys.argv)
101         Form2 = QtWidgets.QWidget()
102         ui = Ui_Form2()
103         ui.setupUi(Form2)
104         Form2.show()
105         sys.exit(app.exec_())
```

Figure 9 : Création de la fenêtre d'ajout de suffixes.

Chapitre III : Formalisation linguistique

8. Description de l'interface :

8.1. Fenêtre principale

Après avoir exécuté le programme, la fenêtre principale s'affiche comme présenté ci-dessous. Elle est composée d'une zone de texte pour chercher le mot voulu, deux boutons d'ajout et de suppression des mots, une colonne des mots avec le même radical, et enfin des zones de texte dans lesquels les morphèmes s'afficheront.

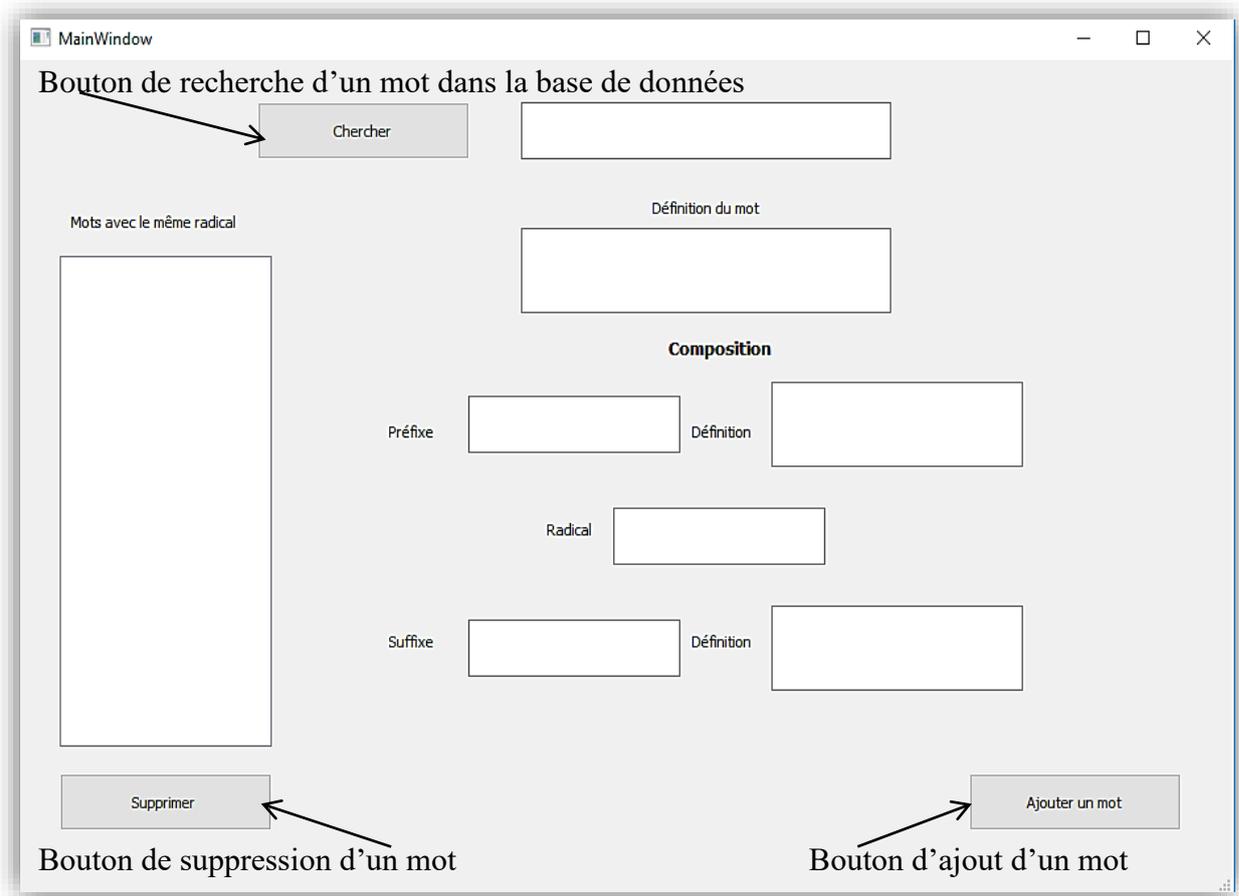


Figure 10 :Fenêtre principale.

Chapitre III : Formalisation linguistique

8.2. Fenêtre d'ajout de mots

Cette fenêtre est composée de sept (7) fonctionnalités : l'ajout ou la suppression de l'un des morphèmes des mots, et d'un bouton de confirmation d'ajout du mot voulu. Une zone de texte est mise en place pour la définition du mot.

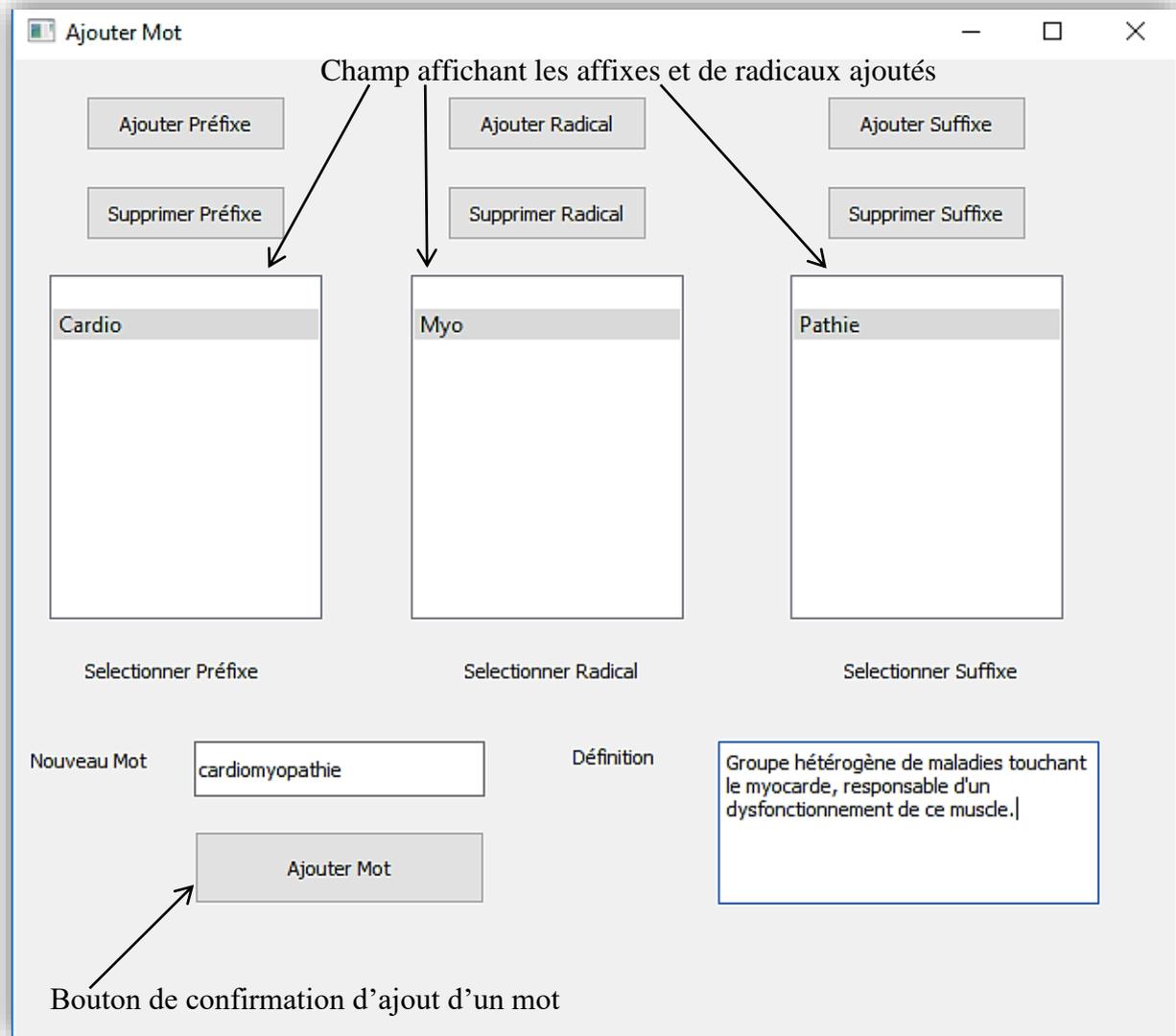


Figure 11 : Fenêtre d'ajout de mots.

Chapitre III : Formalisation linguistique

8.3. Fenêtres d'ajout de morphèmes et leurs définitions

Ces deux fenêtres permettent l'ajout des affixes et leurs définitions en écrivant dans les zones de textes dédiés à cela.

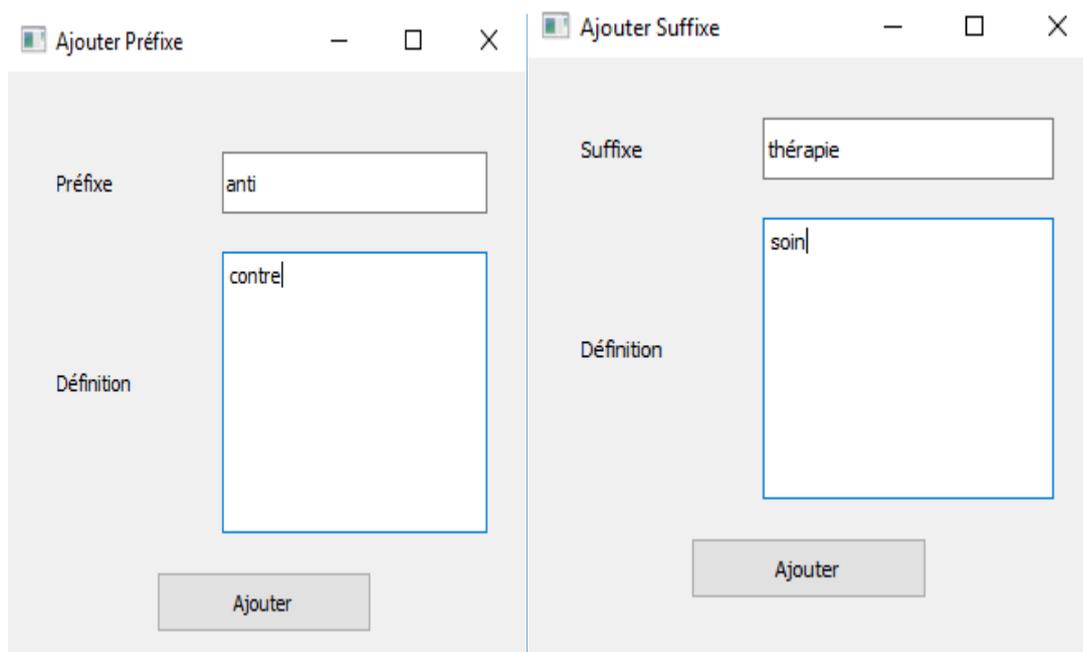


Figure 12 : Fenêtre d'ajout de préfixe.

Figure 13 : Fenêtre d'ajout de suffixe.

Cette fenêtre permet d'ajouter les radicaux en les écrivant dans la zone de texte illustrés dans

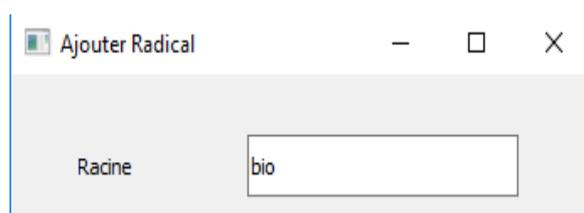


Figure 14 : Fenêtre d'ajout des radicaux.

Chapitre III : Formalisation linguistique

8.4. Fenêtre principale affichant les résultats

Après avoir effectué la recherche d'un mot, nous le sélectionnons dans la colonne affichant les mots ayant le même radical, afin d'afficher les morphèmes qui le composent accompagné de leurs définitions.

Exemple d'un mot segmenté avec cette application est : myocardopathie, les résultats sont affichés comme suit :

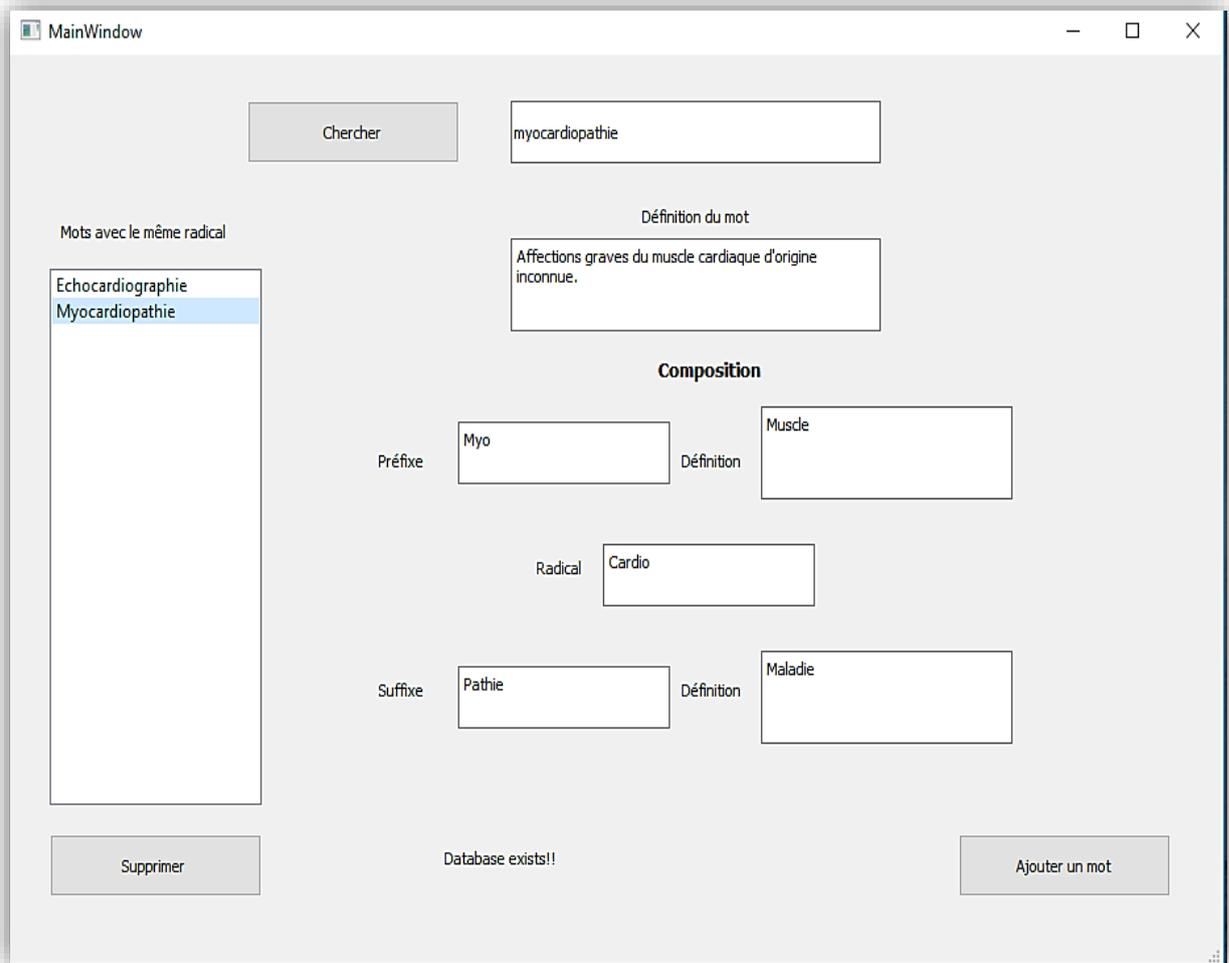


Figure 15 : Fenêtre principale affichant les résultats.

Conclusion

Dans ce chapitre, nous avons mis en place quatre dictionnaires : celui des affixes et leurs définition ainsi qu'un dictionnaire des mots ajoutés dans l'application et leurs définition.

Pour la création de notre application, nous nous sommes basées sur une approche purement linguistique.

Grâce au langage Python, nous sommes parvenues à la création d'une application permettant de segmenter des termes médicaux à composition savante et hybride.

Chapitre III : Formalisation linguistique

Le développement informatique a permis à la linguistique de développer le domaine du traitement automatique des langues. C'est un travail complémentaire ; l'informatique met à disposition des outils permettant d'informatiser le langage, et la linguistique fournit à son tour la description du fonctionnement du langage humain.

Conclusion

générale

Notre projet de fin d'étude s'inscrit dans le domaine du traitement automatique des langues. Ce projet a pour but de concevoir un prototype capable de segmenter les termes médicaux à composition savante et hybride.

Dans un premier temps, nous avons tracé et discuté la différence entre la langue générale et la langue de spécialité. Ensuite nous avons invoqué la notion de la terminologie qui est une discipline qui prend en charge le vocabulaire de spécialité ; dans notre cas, nous avons cité la terminologie médicale, dans laquelle la notion du terme nous intéresse.

Dans un deuxième temps, nous avons développé le domaine de la lexicologie, dans lequel les procédés de formation du lexique français sont étudiés. Ces procédés nous les avons cités et définis, tout en mettant l'accent sur la composition savante et hybride. Enfin, nous avons consacré une partie pour l'étude morphologique des termes médicaux à composition savante et hybride.

Dans un troisième temps, nous avons consacré un chapitre dédié au domaine du traitement automatique des langues. Ce domaine relativement récent se trouve à l'intersection de la linguistique et l'informatique, grâce auquel le langage naturel est informatisé. Ceci permet la création de diverses applications, que nous avons citées.

Arrivant à l'objet de recherche de notre mémoire, nous nous sommes intéressées à l'élaboration d'une application pouvant segmenter les termes médicaux à composition savante et hybride compte tenu de leur complexité morphologique, mais aussi afin de faciliter la compréhension du terme étant décomposé en morphèmes. C'est pour cette raison que nous avons mis en place trois dictionnaires lexicaux comportant les préfixes, radicaux et suffixes de notre corpus, complétés par un autre dictionnaire des termes avec leurs significations qui sont ajoutés à la base de données de notre prototype. En cherchant un mot donné, l'application affiche à la fois sa définition, les mots ayant le même radical, et enfin les morphèmes qui le composent suivi de leurs significations.

D'autres mots peuvent s'ajouter au fur et à mesure, soit en utilisant la même base de données, soit en travaillant sur des mots nouveaux.

Pour conclure, nous avons pu réaliser une application assez performante vis-à-vis de nos attentes et besoins, cependant une amélioration est envisageable vu que nous avons essayé de simplifier la réalisation du prototype compte tenu de la crise sanitaire actuelle.

Le lemmatiseur FrenchStemmer n'est pas assez développé par rapport à ses modules, son développement pourrait ouvrir de nouvelles perspectives concernant la réalisation d'applications semblables à celle que nous venons de réaliser. Ce point peut être pris en considération lors des prochains projets fins d'étude en traitement automatique des langues, afin de promouvoir ce domaine.

Nous espérons que ce modeste travail éveillera l'intérêt d'autres chercheurs de continuer et de développer certains points encore plus importants que nous n'avons pas exploités.

Références

Bibliographiq

ues

- APOTHELOZ Denis, 2002, *La construction du lexique français: principes de morphologie dérivationnelle*, Editions OPHRYS. 164p [En ligne] https://books.google.dz/books?id=c9O6vcs6Ru4C&hl=fr&source=gbs_navlinks_s
- BECHARD Hervé, 1992, *phonétique et morphologie du Français moderne et contemporain*, 1ère édition, PUF.
- BEDDAR Mohand, 2013, Thèse de doctorat « *Vers un prototype de traduction automatique contrôlée français/arabe appliquée aux domaines à sécurité critique* », Sciences du langage, Université Franche-Comté.
- BINON Jean & VERLANDE Serge, 1996, *la contribution de la lexicologie pédagogique à l'apprentissage d'une langue étrangère ou seconde, études de la linguistique appliquée*.
- CABRE Maria Térésa, 1998, *La terminologie. Théorie, méthode et applications*, Paris, Armand Colin. 322 p.
- CELLARD Jacques, ALAIN Rey, 1991, *Dictionnaire du français non conventionnel, Titolo Collona*, Paris. Hachette
- CORBIN Danielle, CORBIN Pierre, TEMPLE Martine 2004, *Lexique 16/ La formation des mots : horizons actuels*, Presses universitaires. Septentrion. 271 p. [En ligne] https://books.google.dz/books?id=USFRX-BGglcC&dq=composition+savante&hl=fr&source=gbs_navlinks_s
- DELBECQUE Nicole & M. Jean-Rémi Lapaire, 2006, *Linguistique cognitive: Comprendre comment fonctionne le langage*, De Boeck Supérieur. 404 p. [En ligne] https://books.google.dz/books?id=LwMoKDO2JbcC&dq=%E2%80%A2%09Nicole+Delbecque+%26+M.+Jean-R%3%A9mi+Lapaire,+Linguistique+cognitive:+Comprendre+comment+fonctionne+le+langage,+De+Boeck+Sup%3%A9rieur,+2006.+404+pages&hl=fr&source=gbs_navlinks_s.
- DERROY Louis, 1956, *L'emprunt linguistique*, Chapitre IX. Les degrés de la pénétration, Liège, Presses universitaires de Liège, Les Belles Lettres, Bibliothèque de la faculté de philosophie et lettres de l'université de Liège. p215-234. [En ligne] <https://books.openedition.org/pulg/685?lang=fr>.
- DUBOIS Jean, 1992, *Dictionnaire de linguistique et des sciences du langage*, édition Larousse 1999.
- FARMERIE Etienne, 1998, *Le latin et le grec d'Appien: contribution à l'étude du lexique d'un historien grec de Rome*, Librairie Droz. 459 p. [En ligne] https://books.google.dz/books?id=dR2CfpVju-8C&dq=x%3%A9nisme+linguistique&hl=fr&source=gbs_navlinks_s
- FAURE Pascaline, 2012, *L'anglais médical et le français médical: Analyse linguistico-culturelle comparative et modélisations didactiques*, Archives contemporaines. 188 pages. [En ligne] https://books.google.dz/books?id=IV04fl_uUm8C&dq=l%27encycllop%3%A9die+m

[%C3%A9dicale+DE+MEDICINA+de+l%27hippocrate&hl=fr&source=gbs_navlinks_s](#)

- FRADIN Bernard, 2003, *Nouvelles approches en morphologie*, Presses Universitaires de France. 360 p. [En ligne] https://books.google.dz/books?id=Hd0JCwAAQBAJ&hl=fr&source=gbs_similarbooks.
- FRADIN Bernard, KERLEROUX Françoise, PLENAT Marc, 2009, *Aperçus de morphologie du français*, Presses universitaires de Vincennes. 313 p. [En ligne] https://books.google.dz/books?id=IiodAQAAIAAJ&hl=fr&source=gbs_similarbooks
- GREVISSE Maurice, FAIRON Cédric, SIMON Anne-Catherine, 2018, *Le petit Bon usage de la langue française: La grammaire qui aime les écrivains*, Grevisse langue française. De Boeck Supérieur. 576p.
- JUAN C. Sager, 1990, *A practical Course in Terminology Processing*, John Benjamins, Amsterdam. 246 p
- LAUREN, Christer & PICT Heribert, 2001, « Terminologie aus linguistischer Sicht ». *Terminology Science & Research*. (1-2) p30-40.
- Le Grand Larousse de la langue française. Paris: Larousse, 1978.
- LERAT, 1995, *Les langues spécialisées*, Presses universitaires de France, Paris. P 11.
- M Bénaben, *Manuel de linguistique espagnole*, Editions OPHRYS, 2002. 329 p [En ligne] https://books.google.dz/books?id=o01-HCCJEOEC&dq=%E2%80%A2%09M+B%C3%A9naben,+Manuel+de+linguistique+espagnole,+Editions+OPHRYS,+2002.+329+p&hl=fr&source=gbs_navlinks_s
- MARTIN, R., BERTHET, F. 1998, *Introduction à la lexicologie, sémantique et morphologie*, Dunod.
- MARTINET André, 1960, *Eléments de linguistique générale*, Paris, Armand Colin, p. 20.
- MATORÉ Georges, 1953, *La méthode en lexicologie*, Didier, Paris.
- MILLIARESSI Tatiana & RASIERR François, 2011, *De la linguistique à la traductologie: Interpréter/traduire*, Presses Université. Septentrion, 2011. 324 p. [En ligne] https://books.google.ci/books?id=zZd_wHl7dZ0C
- MORTUREUX Marie Françoise, 1997, *La lexicologie entre langue et discours*, SEDES.
- MOUNIN, G 1974, *Dictionnaire de la linguistique*, Paris, PUF.
- Picht H & Draskau, 1985, *terminology: an introduction*. Surrey: the University of Surrey.
- PICOCHÉ Jacqueline, 1977, *Précis de lexicologie française*, collection « Nathan Université », Paris.
- POLUGUERÉ Alain, 2003, *lexicologie sémantique lexicale : notions fondamentales*. PUM, p42-45
- R. Galisson & D. Coste, 1976, *Dictionnaire de didactique des langues*, Paris, Hachette. 612p.

- ROSTISLAV Kocourek, 1991, *La langue française de la technique et de la science*, Brandstetter, p42.
- SUCHET Myriam, 2009, *Outils pour une traduction postcoloniale: littératures hétérolingues*, Archives contemporaines. p262. [En ligne] https://books.google.dz/books?id=qjxYQ6XV39cC&dq=x%C3%A9nisme+linguistique&hl=fr&source=gbs_navlinks_s
- WUSTER, Eugen, 1931, *Internationale Sprachnormung in der Technik besonders in der Elektrotechnik (Die nationale Sprachnormung und ihre Verallgemeinerung). (Normalisation internationale des langues dans les technologies, en particulier dans Génie électrique (La normalisation en langue nationale et sa généralisation)*, VDI Verlag. p431
- 真華·徐 (Zhen Hua Xu), 2001, *Le néologisme et ses implications sociales*, Harmattan. 174 p. [En ligne] https://books.google.dz/books?id=RKIbAQAIAAJ&q=n%C3%A9ologisme+linguistique&dq=n%C3%A9ologisme+linguistique&hl=fr&sa=X&ved=2ahUKEwj9_u_m38rDqAhWjAmMBHSToAiEQ6AEwBXoECAyQAQg.

Site internet

- <http://www.eurekoi.org/quels-sont-les-laboratoires-interesses-par-le-traitement-automatique-de-la-langue-arabe-dans-le-monde/04/06/2020> 13h24.
- <http://www.france-ioi.org/algo/course.php?idChapter=561&idCourse=2368> 12/06/2020 à 21h28.
- <https://www.culture-informatique.net/cest-quoi-langage-de-programmation/> 05/06/2020 à 17h21.
- <https://www.ouest-france.fr/sante/virus/coronavirus/coronavirus-cluster-quatorzaine-super-spreader-les-mots-de-l-epidemie-6783539> 28/05/2020 à 20h13.
- <https://www.home.uni-osnabrueck.de/bschwisc/archives/formation.htm> 23/08/2020 à 21h40.
http://www.clillac-arp.univ-paris-diderot.fr/_media/user/najet_boutmgharine/cours-6-.pdf 23/08/2020 à 23h29
- <https://www.ffn-neurologie.fr/grand-public/quest-ce-que-la-neurologie> 30/07/2020 à 10h44.
- <http://www.culture.fr/Ressources/FranceTerme/Qu-est-ce-que-la-terminologie> 27/06/2020 à 22h38.
- <http://tice-It.info/termscience/mod/book/view.php?id=90> 28/06/2020 à 13h11.
- <https://www.cnrtl.fr/definition/terminologie/substantif> 28/06/2020 à 18h05.

- <http://www.diacronia.ro/ro/indexing/details/A17763/pdf> 18/08/2020 à 23h20.
- <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> 05/06/2020 à 22h36.
- <https://www.cairn.info/revue-langages-2008-3-page-95.htm#> 07/06/2020 à 15h16
- <https://www.erudit.org/fr/revues/ttr/1995-v8-n2-ttr1483/> 04/04/2020 à 01h16.
- <http://www.artificiel.net/test-de-turing> 04/06/2020 à 22h18.
- <https://fracademic.com/dic.nsf/frwiki/70859> 04/06/2020 à 23h31.
- <https://fr.duolingo.com/> 05/06/2020 à 23h47.
- https://www.reverso.net/text_translation.aspx?lang=FR 06/06/2020 à 00h15.
- <http://python-history.blogspot.com/> 12/06/2020 à 23h25
- <https://dictionnaire.lerobert.com/> 27/06/2020 à 14h16.
- <https://www.cnrtl.fr/definition/medecine> 02/07/2020 à 15h54.
- <https://www.cnrtl.fr/definition/Psychiatrie> 03/07/2020 à 10h20.
- <https://fr.wikipedia.org/wiki/Cardiologie> 30/07/2020 à 11h10.
- <https://www.cnrtl.fr/definition/composition/substantif> le 15/08/2020 à 18h15
- <https://cnrtl.fr/definition/lexicologie> 18/08/2020 à 19h29.

Annexe

Corpus

1. **Antibiothérapie** (sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 411)
2. **Aplasia** (cours de Dr. Bordessoule-octobre 2010. P 1)
3. **Apolipoprotéine**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 411)
4. **Arythmie** (CAZAL Julien, 18 octobre 2012, thèse de doctorat « *prévenir la maladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire* », université de Toulouse III-Paul Sabatier. P173).
5. **Athérosclérose**(CAZAL Julien, 18 octobre 2012, thèse de doctorat « *prévenir lamaladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire* », université de Toulouse III-Paul Sabatier. P62).
6. **Bradycardie**. (sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 52)
7. **Cardiomyopathie**(DELOBELLE Julien, 27 octobre 2011, thèse de doctorat « Risque thromboembolique des tachycardies jonctionnelles paroxystiques, université Henri Poincaré Nancy 1. P 53)
8. **Cardiopathie**(CAZAL Julien, 18 octobre 2012, thèse de doctorat « *prévenir lamaladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire* », université de Toulouse III-Paul Sabatier. P295).
9. **Dyslipidémie**(CAZAL Julien, 18 octobre 2012, thèse de doctorat « *prévenir lamaladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire* », université de Toulouse III-Paul Sabatier. P104).
10. **Echocardiographie** (rapport d'évaluation technologique par M. Cédric CARBONNEIL, novembre 2012, « *échographie doppler transthoracique : principales indication et conditions de réalisation* ». p 13)
11. **Electrochoc**(revue médicale liège 2008. P 407)
12. **Endothélium**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 5)
13. **Fibrinogène**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 336)
14. **Hémorragie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 292)
15. **Hépatomégalie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 243)

16. **Hyperlipidémie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 413)
17. **Hyperplasie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 321)
18. **Hypothalamus** (JESUS Pierre, 19 décembre 2014, thèse de doctorat « impact de l'état et de la prise en charge nutritionnels dans les maladies neurodégénératives : approche neuroépidémiologique »universitédeLimoges. P 183.)
19. **Iatrogénie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 409)
20. **Microangiopathie**(DELOBELLE Julien, 27 octobre 2011, thèse de doctorat « Risque thromboembolique des tachycardies jonctionnelles paroxystiques, université Henri Poincaré Nancy 1. P 30)
21. **Néphropathie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 303)
22. **Physiopathologie**(CAZAL Julien, 18 octobre 2012, thèse de doctorat « prévenir la maladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire », université de Toulouse III-Paul Sabatier. P12).
23. **Psychotrope** (CAZAL Julien, 18 octobre 2012, thèse de doctorat « prévenir la maladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire », université de Toulouse III-Paul Sabatier. P335).
24. **Rétinopathie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 302)
25. **Scintigraphie**(CAZAL Julien, 18 octobre 2012, thèse de doctorat « prévenir la maladie cardiovasculaire : socio-ethnologie du risque et de l'incitation sanitaire », université de Toulouse III-Paul Sabatier. P37).
26. **Tachycardie**(sous l'égide du Collège National des Enseignants de Cardiologie et de la société Française, sous la direction de Dominique Lacroix, *cardiologie*, édition Elsevier Masson, 2010. P 42)
27. **Traumatologie**(urgences en traumatologie et orthopédie pédiatrique, 3 novembre 2014, université de Genève. P 3)

Résumé

Plusieurs travaux en TAL tentent à développer des applications performantes. Ces dernières ont pour but de faciliter l'apprentissage, la compréhension ainsi que l'utilisation de la langue de manière générale.

Afin de pouvoir aboutir à ce but, les linguistes fournissent les données linguistiques nécessaires pour pouvoir les informatiser. L'objectif de ce travail de fin d'étude, est justement de proposer un modèle de prototype destiné aux praticiens en médecine.

Ce prototype en question propose une segmentation automatique des termes à composition savante. Cette segmentation leur permettra de pouvoir comprendre un terme donné à partir de ses morphèmes. Une définition est également proposée pour chacun des mots, et chacun des morphèmes.

Mots clés : TAL. Informatique. Médecine. Segmentations. Langues de spécialités. Terminologie médicale.

Summary

Some efficient applications are being developed by many NLP work. The purpose of these applications is to facilitate learning, comprehension as well as the use of language in general.

So as to reach this purpose, linguists provide necessary linguistic data to computerize them. The main purpose of this graduation project is to suggest a prototype assigned to medicine practitioners. This prototype recommends word automatic segmentation from Latin or Greek roots

The segmentation allows the understanding of given words from their morphemes. A definition is provided for each of these words and morphemes.

Keywords: ALP. Computer science. Medicine. Segmentations. Specialty languages. Medical terminology.

ملخص

عدة اعمال في برمجة اللغات تحاول تطوير برامج ذات فعالية. هذه البرامج تسعى لتسهيل التعلم، الفهم واستعمال اي لغة كانت. لتحقيق هذا المبتغى، الأخصائيون في اللسانيات لهم دور كبير في توفير كل قواعد اللغة من اجل برمجتها .

الهدف من مشروع التخرج هذا، هو اقتراح برنامج نموذجي موجه للمختصين في الطب. هذا الاخير يقترح تقسيم اوتوماتيكي للكلمات المشكلة من كلمات ذو أصل اغريقي او لاتيني. عملية التقسيم هذه ستسهل فهم الكلمات المعقدة من خلال اقسامها. خاصية الشرح متوفرة ايضا لكل منهما.

الكلمات المفتاحية: برمجة اللغات. علوم الكمبيوتر. الطب. الانقسامات. اللغات المتخصصة. مصطلح طبي.