

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



**Faculté des Lettres et des Langues**  
**Département de français**

**Mémoire de master**

**Option : Sciences du langage**

**Vers un prototype de catégorisation automatique des textes**  
**écrits par domaine de spécialité**

Présenté par :

BOUND AOUI Fazia  
BOUSSAID Nouha

Encadré par :

M. BEDDAR Mohand

Année universitaire  
**2019/2020**

## **Remerciements**

Nous tenons à adresser nos vifs remerciements :

Tout d'abord à notre encadrant Beddar Mohand pour nous avoir orientées, conseillées et soutenues lors de la réalisation de ce travail.

A l'entreprise de GENIE LOGICIEL pour ses conseils et son aide notamment dans la réalisation de toute la partie informatique.

Aux membres du jury d'avoir accepté d'examiner notre travail de recherche.

A notre ami particulier : Boudaha Yaakoub qui nous a toujours encouragé dans la voie de la recherche et nous a aidé en permanence jusqu'à la fin de ce travail.

Enfin, à tous nos proches, amis et à toutes les personnes qui ont participé à la réalisation de ce travail, par leurs soutiens et leurs encouragements, de près ou de loin.

## **Dédicaces**

Je dédie ce modeste travail à :

Mes chers parents qui ont été mon appui tout au long de mes études, toujours à mes cotés pour m'encourager et me donner de la force pour aller au bout de mes objectifs. J'espère qu'ils trouveront dans ce travail toute ma reconnaissance ainsi que mon amour intégral.

A mes deux chères sœurs Cylia et Alicia et à mon petit frère Said.

A mes amis et à tous ceux qui ont contribué de près ou de loin, par leurs encouragements et leurs conseils à la réalisation de ce travail.

Boundaoui Fazia

Je dédie ce modeste travail accompagné d'un profond amour :

A l'homme de ma vie, mon précieux offre de dieu, mon exemple éternel, celui qui s'est toujours sacrifié pour moi afin d'atteindre mes rêves, à toi mon cher papa « Boussaid abd el krim ».

A la lumière de ma vie, mon réconfort et ma certitude, mon guide dans l'existence, je ne saurai jamais t'exprimer à quel point tes encouragements m'ont un trésor, à la plus belle perle du monde, à toi ma tendre maman.

A mon grand-papa « Boussaid Nassir » tant aimé, tu as toujours été à mes cotés pour me soutenir et m'encourager, que ce travail traduit mon gratitude et mon affection.

A mes quatre sœurs, mes jumelles et ma raison de vivre Hassiba, Tinhinane, Nour ElHouda et Rania . Je leurs souhaite une vie pleine de succès et d'amour.

A Mes chers et adorable frère Juba, Fouad, Ilyan et Danny que j'aime profondément.

Je dédie ce travail à la très chère personne après mes parents, ma source de joie et de bonheur, mon fiancé et mon tout Massi, pour l'aide et l'encouragement qu'il m'a toujours accordé.

A ma chère binôme :Boundaoui Fazia qui est pour moi une amie et une sœur.

A toute mes amies qui m'ont toujours encouragé durant cette année d'étude et à qui je souhaite plus de succès.

A tous ceux qui me sont chers.

A toute personne qui m'a aidé à franchir un horizon dans ma vie.

Boussaid Nouha

## Sommaire

1-Introduction générale .....	9
-------------------------------	---

### Chapitre 1 Les langues et les textes de spécialité

1- Introduction .....	14
2- La langue .....	14
3- Les langues de spécialité .....	14
4. Langue générale vs langue de spécialité .....	17
5. Les caractéristiques de langue de spécialité .....	19
6. Le lexique .....	20
7. Le terme.....	21
8. La terminologie .....	21
9. La fiche terminologique .....	22
10. L'unité terminologique .....	22
11. Le texte de spécialité.....	23
12. Les caractéristiques des textes de spécialité .....	23
13. Conclusion .....	24

### Chapitre 2 Le traitement automatique des langues

1- Introduction .....	26
2- Bref historique sur le traitement automatique des langues .....	26
3- Définition du traitement automatique des langues .....	27
4- Les applications du TAL .....	28
5- Les niveaux de traitement automatique des langues .....	30
6- Conclusion .....	31

### Chapitre 3 Analyse des textes de spécialité et présentation de l'application

1-Introduction .....	34
2-Analyse des textes de spécialité .....	34
3-Présentation de l'application.....	44
4-Description de l'interface utilisateur .....	52
5-conclusion.....	58

Conclusion générale .....	59
---------------------------	----

Annexes.....	62
--------------	----

Références bibliographiques.....	92
----------------------------------	----

## Les abréviations

TAL : Traitement Automatique des Langues.

FOS : Français pour Objectif Spécifique.

Ex : exemple.

LS : Langue de Spécialité.

TS : Texte de Spécialité.

MIT : Massachusset Institut of Technology.

LADL : Laboratoire de Documentaire et linguistique dirigé.

TA : Traduction automatique.

GAT : Génération Automatique de Texte.

RI : Recherche d'Information.

ETC : et cetera (etcétera).

Ut : unité terminologique.

CCIS : carcinomes canaux in situ.

Hba : hémoglobine A.

CIVD : Coagulation IntraVasculaire Disséminée.

ARPA: Advanced Research Project Agency.

OS: Operating System.

RT: Recette Totale.

P: Prix du bien.

Q: Quantités vendues.

OMS: Organisation Mondiale de Santé.

HO: hydroxide.

WWW: World Wide Web.

GDT: Grand Dictionnaire Terminologique.

Med : médecine.

Phys : physique.

Psych : psychologie.

Chim : chimie.

Biol : biologie.

Éco : Économie.

Math : mathématique.

Dr : droit.

Élect : électricité.

Infr : informatique.

Agric : agriculture.

Socio : sociologie.

Spr : sport.

Philo : philosophie.

Ling : linguistique.

# **Introduction**

### 1- Introduction partielle

La langue est l'outil de communication et d'expression le plus fiable pour l'individu, elle nous sert à incarner nos idées, nos pensées, nos volontés, les choses abstraites, à affirmer nos convictions, à recevoir confirmation ou infirmation de nos énonciations et ainsi transmettre des informations. Elle est acquise le plus souvent de façon naturelle en étant en contact et au contact avec ses proches et sa communauté.

Nous utilisons une langue générale ou langue de communication pour les conversations de tous les jours et pour les échanges de la vie quotidienne, contrairement à une langue de spécialité, propre à une situation de communication circonscrite à un domaine particulier. En effet, cette dernière correspond à des situations d'emploi professionnel et la technicité de formulation est susceptible de changer et de varier selon les besoins de communication.

Les langues de spécialité sont des langues simples mais spécialement destinées pour un usage déterminé par les limites permettant d'établir une bonne compréhension entre les personnes actives dans un même domaine et rendre compte de connaissances particulières. Comme le confirme Goffman « *un ensemble complet des phénomènes linguistiques que se produisent dans une sphère précise de communication limitée par des sujets, des termes, des intentions et des conditions spécifiques* »<sup>1</sup>. Elles sont composées de textes spécialisés qui contiennent des mots de vocabulaire usuel et de vocabulaire spécialisé. La présente recherche porte sur l'identification automatique de domaine des textes de spécialité. Elle s'inscrit dans le domaine du traitement automatique des langues (TAL) qui est défini par Marcel Cori comme suit : « Le traitement automatique des langues (TAL) est constitué des méthodes et des programmes qui prennent pour données des productions langagières, quand ces méthodes et programmes tiennent compte des spécificités des langues humaines. »<sup>2</sup>. C'est un domaine pluridisciplinaire qui comporte à la fois la linguistique, l'informatique et la formalisation. Il englobe un ensemble de méthodes et de programmes qui vise à créer des logiciels traitant les données de la langue de façon automatique.

---

<sup>1</sup> GOFFMAN, CARBRE Maria Teresa, *La terminologie : théorie, méthode et application*, Ottawa, Canada, 1998, p.118

<sup>2</sup> Marcel, C ; *Traitement automatique des langues et formalisation en linguistique*, disponible sur <http://www.tal.univ-paris3.fr/plurita/cours3-2004/slides-cours3-2004-1.pdf> (page consulté le 14 janvier 2020)

### **2- Motivations et objectifs**

Nous avons choisi ce thème car il s'inscrit dans le domaine du traitement automatique des langues auquel nous nous intéressons beaucoup et le fait de l'avoir choisi comme champ de recherche relève de notre volonté d'approfondir les connaissances que nous avons acquises dans ce domaine.

Notre travail de recherche est innovant au niveau de notre département, nous allons ainsi joindre deux domaines différents : l'informatique et la linguistique. Ce sera également dans le but de développer les recherches dans ce domaine récent en Algérie.

De plus, l'application que nous allons réaliser porte sur une méthode automatique qui identifie facilement les domaines des textes de spécialité dans des domaines variés (médecine, économie, informatique, chimie et physique), mais également valable pour les autres domaines. Elle permet d'aider les enseignants des langues de spécialité ou du FOS d'avantage dans leur élaboration des cours.

Notre objectif principal dans ce présent travail de recherche consiste à concevoir et à réaliser une application informatique capable de catégoriser des textes écrits selon leurs domaines de spécialités de façon autonome, il indiquera la spécialité de chacun des textes de spécialité (médecine, économie, informatique, physique, chimie) à l'aide du dictionnaire terminologique qui contient différentes étiquettes, et qui peut marcher ultérieurement avec d'autres textes et d'autres domaines.

### **3- Problématique**

De nos jours, beaucoup de logiciels d'identification des textes ont fait leur apparition, ils permettent de faciliter la tâche et de rendre le travail plus rapide contrairement à une méthode manuelle classique. C'est d'ailleurs cela qui nous anime dans ce travail et nous a amener à poser la problématique suivante :

- Est-il possible de s'appuyer sur un outil informatique pour catégoriser des textes en fonction de leurs domaines de spécialité ?

Il nous paraît intéressant également de savoir à partir de la problématique principale :

- Sur quel matériel linguistique pourrons-nous nous baser pour catégoriser des textes de spécialité ?
- Existe-t-il une terminologie commune à plusieurs domaines de spécialité ?

### **4- Hypothèses :**

Afin d'élucider cette problématique, nous proposons des réponses provisoires qui seront vérifiées, affirmées ou infirmées par l'analyse du corpus. Nous avons ainsi formulé les hypothèses suivantes :

- ✓ Nous supposons qu'il serait possible de créer une application qui permet d'identifier les domaines des textes de spécialité.
- ✓ La catégorisation d'un texte de spécialité reposerait sur quelques critères linguistiques : lexique spécifique, le recours à d'autres systèmes de représentation (schéma, des illustrations...), la nominalisation et des formules impersonnelles.
- ✓ Chaque domaine de spécialité posséderait sa propre terminologie.

### **5- Méthodologie et corpus**

Dans notre travail de recherche, l'étude s'effectuera à partir d'une analyse d'un corpus composé de 53 textes appartenant à divers domaines de spécialité (médecine, économie, informatique, physique, chimie) issus de sites internet variés.

En se basant sur *le grand dictionnaire terminologique*<sup>3</sup>, nous identifions le domaine de 1591 termes et concepts présents dans ces textes de spécialité afin de mettre en place une formalisation en élaborant un dictionnaire terminologique étiqueté. Ensuite à partir de cela nous concevrons une application qui permet de catégoriser automatiquement ses textes par leur domaine de spécialité

### **6- Plan du travail**

Notre travail se divise en trois chapitres, dans lesquels nous allons vérifier les hypothèses posées précédemment. Ils sont précédés par une introduction générale dans laquelle nous avons

---

<sup>3</sup> *Le grand dictionnaire terminologique (GDT)* disponible en ligne sur <http://www.granddictionnaire.com>

## Introduction générale

---

présenté notre sujet de recherche, nos motivations et les objectifs visés, notre problématique, nos différentes hypothèses ainsi que notre méthodologie de travail.

Le premier chapitre sera constitué de la partie théorique, nous allons aborder les aspects définitoires des langues de spécialité et des langues communes tout en faisant la distinction entre ces dernières. Ensuite, nous présentons des notions et concepts fondamentaux en rapport avec les langues de spécialité (ses caractéristiques, terme, lexique terminologie, fiche terminologique...) ainsi que les disciplines et les informations clés qui ont une relation avec notre étude.

Dans le second chapitre, nous présenterons notre domaine d'étude qui est le traitement automatique des langues ainsi que ses domaines et champs de recherche associés.

Le troisième et dernier chapitre expose la pratique, il se compose de trois sous-parties dont la première sera consacrée à l'analyse des textes de spécialité et à l'identification de leurs caractéristiques ; dans la deuxième sous-partie, nous expliquerons le fonctionnement de l'application que nous avons conçue, et qui permet d'identifier le domaine auquel se rapporte chaque texte de spécialité (médecine, économie, informatique, physique, chimie) qui a été inséré. Ainsi, nous décrirons chaque étape de cette application pour pouvoir donner le résultat final. Et la troisième sous-partie, décrira l'interface de notre application et présentera l'utilité de chaque onglet de cette dernière.

# **Chapitre 1**

**Les langues et les textes de spécialité**

## 1- Introduction partielle

Dans ce premier chapitre, nous allons d'abord aborder les aspects définitoires des langues de spécialité et des langues communes tout en faisant la distinction entre ces dernières. Ensuite, nous présentons les notions et concepts fondamentaux en rapport avec les langues de spécialité (ses caractéristiques, terme, lexicologie terminologique, fiche terminologique...).

## 2- La langue

La langue est une faculté dont a besoin le locuteur pour exprimer et transmettre ses pensées et ses intentions, elle représente son appartenance à une communauté, en la distinguant d'autre. Dans ce sens, J. DUBOIS, définit la langue comme étant «*un instrument de communication, un système de signes vocaux spécifique aux membres d'une communauté.*» (1999:260). C'est en effet grâce à cette langue que se transmette d'une génération à l'autre les éléments fondateurs de l'identité culturelle.

## 3- Les langues de spécialité

Le terme générique de langue de spécialité est utilisé depuis les années 1960, pour distinguer la langue des domaines de spécialité de la langue générale (usuelle). Malgré de nombreuses études menées sur cette dernière, les chercheurs ne sont pas toujours d'accord sur l'appellation de cette notion, on peut donc trouver plusieurs appellations pour désigner la LS : *langue spécialisée, langue scientifique, langue technique.*

De manière très générale, Galisson et Coste définissent la LS comme une : « *Expression générique pour désigner les langues utilisées dans des situations de communication (orales ou écrites) qui impliquent la transmission d'une information relevant d'un champ d'expérience particulier* »<sup>4</sup>. D'après cette définition, la langue de spécialité réunit les deux situations de

---

<sup>4</sup> R.GALISSON, D.COSTE, *Dictionnaire de didactique des langues étrangères*, p511, Didier, 1976

communication, orale et écrite qui vise à transmettre des connaissances particulières dans un domaine donné.

M.T.Cabré précise que les langues de spécialité « *sont des codes de type linguistique, différenciés de la langue générale, constitués de règles et d'unités spécifiques* »<sup>5</sup>. Dans ce même sens, elle ajoute que :

« (...) *toute langue de spécialités est une simple variante de la langue générale. On peut donc dire que la langue de la physique et celle de l'informatique sont aussi différentes l'une de l'autre que la langue des maçons et celle des menuisiers. On arrive à définir les langues de spécialité comme de simples variantes lexicales de la langue générale [...] selon une autre façon de les définir, les langues de spécialités seraient des sous-ensembles, fondamentalement pragmatiques, de la langue dans son sens global* »<sup>6</sup>.

De là, on comprend que la LS est un sous-ensemble, un fragment de la langue générale qui sert à la transmission du savoir relevant d'un champ d'expérience particulier et qui se caractérise par des règles spécifiques et un lexique déterminé.

Jean Dubois nous donne une autre définition selon laquelle la langue de spécialité est définie comme: « *un sous-système linguistique tel qu'il rassemble les spécificités linguistiques d'un domaine particulier. En fait, la terminologie, à l'origine de ce concept, se satisfait très généralement de relever les notions et les termes considérés comme propres à ce domaine. Sous cet angle, il y a donc abus à parler de langue de spécialité, et vocabulaire spécialisé convient mieux* » (Dubois, 1994 : p. 440). De ce fait, la LS est une variante de langue qui se caractérise par une terminologie, une syntaxe réduite et surtout un vocabulaire spécialisé qui est propre à des situations de communication dans un domaine particulier.

Selon RONDEAU<sup>7</sup>; l'ensemble des langues de spécialité est réparti en trois zones :

- 1. La zone mitoyenne :** zone qu'il considère "*la plus rapprochée de la langue commune*". Il considère cette zone comme étant la plus proche de la langue commune car c'est sur la langue commune que le vocabulaire spécialisé se construit.

<sup>5</sup> Marea Teresa Cabré, *la terminologie* : traduit du catalan et adapté par Monique C. Cormier, et John C. Humberly,

*Théorie, méthode et applications*, Les presses de l'université d'Ottawa, Armand Colin, 1998.p.119

<sup>6</sup> Ibid. P.119

<sup>7</sup> RONDEAU, Guy. « *Introduction a la terminologie* », Québec, Gaétan Morin, 1991, p24.

2. **La zone centrale** : zone où se trouvent des termes communs à plusieurs domaines.
3. **La zone des ensembles ultra spécialisés** : "*zone des techniques de la recherche d'avant-garde*" (Ex. chimie, mathématique). C'est une zone où le vocabulaire employé est très opaque et hermétique car il est créé, définit et compris par un groupe socioprofessionnel restreint.

### 3-1- Les langues de spécialité : hétérogènes ou homogènes ?

Les LS naissent du besoin des spécialistes qui travaillent dans un même domaine de s'exprimer et communiquer entre eux de façon claire et concise. Elles transmettent des informations propres à un domaine particulier.

Les LS ne sont pas homogènes mais plutôt hétérogènes « *On peut donc dire que la langue de la physique et celle de l'informatique sont aussi différentes l'une de l'autre que la langue des maçons et celle des menuisiers (...)* »<sup>8</sup>. Chaque domaine a sa façon de formuler ses connaissances et ses informations par l'outil de la langue qui varie pour les véhiculer. Son style particulier, ses termes techniques reflètent les usages du domaine et son degré de spécialisation.

Chaque LS a donc son mode de fonctionnement spécifique qui la distingue de l'autre. Autrement dit, toute LS a ses propres caractéristiques : sémantique, syntaxe, lexique et stylistique. Par exemple, le lexique de la langue de la chimie n'est pas le même avec celui de la langue médicale.

### 3-2- Spécialisée ou de spécialité

Plusieurs appellations ont été données pour les LS : langues spéciales (F.Saussure, 1975), langue spéciale et technique (J.Vendryés, 1968), langue de groupe particuliers (A.Meillet, 1975), langue des sciences (M. Fuchs, 1966), langue professionnelle (W.Wartburg, 1963) ou encore langues techniques et scientifiques (B. Quemada, 1955).<sup>9</sup> Pour certains, langue spécialisée, pour d'autres enfin la langue de spécialité.

<sup>8</sup> Marea.T. Cabre, *la terminologie* : traduit du catalan et adapte par Monique C. Cormier, et John C. Humberly, théorie, méthode et applications, Les presses de l'université d'Ottawa, Armand Colin, 1998, p.119

<sup>9</sup> BEDDAR.M.2013. « *Vers un prototype de traduction automatique contrôlée français/arabe appliquée aux domaines à sécurité critique* » thèse de doctorat, université DE FRANCHE-COMTE

En effet, la langue spécialisée se définit par rapport à la langue naturelle. Elle ne fonctionne pas comme une langue autonome avec des caractéristiques spécifiques, mais comme sous-ensemble de la langue commune, « aucune théorie linguistique, quelle qu'elle soit, n'a jamais isolé le fonctionnement des langues spécialisées de celui des langues naturelles en général » (Lerat 1997 : 2)

Pour notre part et pour donner une suite cohérente à notre recherche, nous avons opté pour la dénomination langue de spécialité pour de nombreuses raisons. « Les langues de spécialité sont des codes de type linguistique, différenciés de la langue commune, constitués de règles et d'unités spécifiques » (Cabré 1998 :118-121). C'est une langue différente ayant des caractéristiques et des structures spécifiques. Par ailleurs, ce qui marque la spécificité de LS est « l'usage »<sup>10</sup>. Elle se caractérise sur le point de vue sémantique par la monosémie et l'univocité des termes, contrairement aux langues communes où on trouve des cas de synonymie très élevés, et d'autre part par « une syntaxe réduite et un vocabulaire spécialisé »<sup>11</sup>. De même, elle est difficilement compréhensible par un non-spécialiste surtout pour les domaines ultra-spécialisés. Ainsi, pour comprendre une LS ou un texte de spécialité, il est utile voire nécessaire d'être spécialiste du domaine. Cela n'est pas le cas pour les langues communes où tous les textes et discours peuvent être compris aisément par les locuteurs, c'est ce qui fait d'ailleurs d'elle une langue à part.

#### 4- Langue générale vs langue de spécialité

Plusieurs spécialistes (chercheurs, terminologues, lexicographes, linguistes et traducteurs) ont mis l'accent sur les différences entre la langue générale (ou commune) et les langues de spécialité. Ils distinguent les traits distinctifs entre ces deux ensembles ainsi que les caractéristiques de chaque langue.

Maria Teresa Cabré décrit la langue générale (commune) comme suit :

---

<sup>10</sup> « (...) La spécificité des langues de spécialité se manifeste surtout au niveau de l'usage » Cabré (1998 : 112)

<sup>11</sup> « Les langues de spécialité(s) se caractériseraient par une syntaxe réduite et un vocabulaire spécialisé. Il s'agirait d'un sous-système moins complet de la langue générale qui privilégie certaines tournures ou constructions, certains modèles (...) » (Binon & Verlande, 1999 :40).

« 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 ses besoins d'expression et selon les caractéristiques de chaque situation de communication. Cependant, au-delà de cette diversité foisonnante, toute langue possède un ensemble d'unités et de règles que tous ses locuteurs connaissent. Cet ensemble de règles, d'unités et de restrictions qui font partie des connaissances de la majorité des locuteurs d'une langue constitue ce qu'on appelle la langue commune ou générale [...] » (Cabré, 1998 : p. 115).

Rondeau la définit aussi comme « l'ensemble des mots et expressions qui, dans le contexte où ils sont employés, ne se réfèrent pas à une activité spécialisée »<sup>12</sup>

La langue générale (commune) est donc la langue de communication quotidienne, elle est employée dans les échanges de tous les jours, de la vie ordinaire et dans les situations courantes ; c'est la langue commune à tous les locuteurs. Elle utilise un vocabulaire simple, usuel dans un contexte normal qui ne relève pas du discours, de vocabulaire et de termes spécialisés.

Par contre, pour Lerat, la langue de spécialité n'est pas à dissocier de la langue générale mais elle est une partie intégrante de cette dernière. Elles utilisent les mêmes règles syntaxiques et ce n'est qu'au niveau sémantique qu'il y a une différence ; autrement dit, son vocabulaire et son lexique sont attachés à un domaine donné. « Une langue spécialisée ne se réduit pas à une terminologie : elle utilise des dénominations spécialisées (les termes), y compris des symboles non linguistiques, dans des énoncés mobilisant les ressources ordinaires d'une langue donnée. On peut donc la définir comme l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées » (Lerat, 1995 : p. 21).

Pour conclure, la langue de spécialité et la langue générale ne se différencient pas clairement l'une de l'autre. Il y a bien une différence mais qui consiste plutôt en degré qu'en nature. La spécificité des langues de spécialités se manifeste surtout au niveau des usages comme le confirme CABRE « Les unités de la langue commune sont employées dans les situations que

---

<sup>12</sup> RONDEAU, Guy. « Introduction à la terminologie ». Boucherville, Gaétan Morin, 1991, p. 26

*l'on peut qualifier de non-marquées, [...] les situations dans lesquelles on emploie les langues de spécialité peuvent être considérées comme marquées »<sup>13</sup>*

## 5- Les caractéristiques de langue de spécialité

Comme nous l'avons mentionné auparavant, chaque LS (donc chaque domaine) a son propre système linguistique et ses propres caractéristiques, distincts du système de la langue dite ordinaire ou usuelle.

Les langues de spécialité se caractériseraient par « *une syntaxe réduite et un vocabulaire spécialisé. Il s'agirait d'un sous-système moins complet de la langue générale qui privilégie certaines tournures ou constructions, certains modèles : ex. la nominalisation dans la presse, les procédés de distanciation (ex. passivation, locutions impersonnelles) dans le discours scientifique, juridique, etc.* » (Binon & Verlande, 1999 :40). De ce fait, la syntaxe des LS possède des règles grammaticales, des expressions et des tournures syntaxiques qui existent déjà dans la langue générale. Elle se caractérise par l'emploi des pronoms « nous » et « on » de modestie, le pronom invariable « il », les expressions impersonnelles, la nominalisation et les tournures passives qui marquent le caractère impersonnel des phrases.

En plus de toutes ces caractéristiques, il convient d'ajouter la prédominance du présent de l'indicatif comme temps pour les verbes et une prévalence des phrases de types déclaratives. Notons également la diversité d'emploi des coordonnants et des autres unités de liaisons conjonctives et adverbiales ainsi que la pronominalisation pour remplacer les expressions nominales. Les phrases des LS se révèlent structurellement complexe. En effet, la longueur des phrases et l'emploi des constructions réalisées avec des participes et infinitif contribuent à sa complexification.

D'autre part, les LS possèdent un vocabulaire et un lexique spécifique lié à un domaine d'activité précis. Elles utilisent des termes spécifiques qui peuvent être partagés avec d'autres spécialités et bien sûr avec la langue générale. De même, pour communiquer entre eux les spécialistes utilisent une terminologie propre à leur domaine, ils utilisent des outils linguistiques et parfois non linguistiques (symboles, formules) caractéristiques du domaine concerné.

---

<sup>13</sup>CABRÉ, Maria Térésa. *Terminologie : théorie, méthode et applications*. Paris : Armand Colin, 1998, p.115

En outre, la sémantique des langues de spécialité « *est plus simple que celle des langues en général* » Lerat (1995 : 83). Elle se caractérise par la monosémie et l'univocité des termes (mono référentiel) sans connotation, plus particulièrement, dans les domaines hautement spécialisés, ce qui réduit la présence de polysémie et synonymie. Contrairement aux domaines moins spécialisés, la monosémie est moins respectée.

## 6- Le lexique

Le grand dictionnaire Larousse définit le lexique comme un « *ensemble des mots formant la langue d'une communauté et considéré abstraitement comme l'un des éléments constituant le code de cette langue* ». <sup>14</sup>

Dans la même idée, Jean-Pierre Cuq le définit ainsi, le lexique : « *désigne l'ensemble des unités constituant le vocabulaire d'une langue, d'une communauté linguistique, d'un groupe sociale, ou d'un individu..* ». <sup>15</sup>

A partir de ces deux définitions, le lexique est l'ensemble des mots disponibles dans une langue propre à une communauté et à un groupe social. C'est un élément très important qui forme le code d'une langue. Il contient tous les mots des différents domaines et spécialités.

Le lexique englobe tous les mots de la langue, il comporte également certains termes des différents domaines, techniques ou professionnels. C'est pour cela, que l'on distingue deux sous-ensembles dans le lexique :

- **Le lexique général** : commun à tous les locuteurs, c'est celui de la vie quotidienne au contexte de normalité avec des mots simples, usuels, neutres. Il n'y a pas de termes recherchés ou spécialisés.
- **Le lexique spécifique** : il fait référence à une terminologie propre à un domaine particulier et déterminé, il est utilisé par des locuteurs d'une communauté linguistique spéciale avec l'usage d'un fond lexical que l'on trouve dans la langue usuelle.

---

<sup>14</sup> Le grand dictionnaire la Rousse, 2000 P.907.

<sup>15</sup> Jean-Pierre Cuq. *Dictionnaire de didactique du français langue étrangère et seconde*. CLE international, Paris, 2003, P.155.

## 7- Le terme

*Le terme est un élément lexical dont sa signification dépend d'un domaine de spécialité. Kocourek établit que « le terme est une unité lexicale définie dans les textes de spécialité. L'ensemble des termes s'appelle la terminologie. [...] Les termes sont [...] des unités lexicales dont le sens est défini par les spécialistes dans les textes de spécialité ». (Kocourek 1982, p. 77). Il symbolise un concept ou une notion. Il peut être simple (ex : hépatique) ou complexe (ex : hépatite virale).*

## 8- La terminologie

Le terme « terminologie » est polysémique. Dans le premier sens, la terminologie, désigne une discipline linguistique qui étudie la dénomination des notions utilisés dans un domaine d'activité particulier, comme le confirme Dubois : « *L'étude systématique de la dénomination des notions (ou concepts) spécifiques de domaines spécialisés des connaissances ou des techniques* » (J. Dubois, et al. 2005: 481).

Dans sa deuxième acception, la terminologie désigne l'«*ensemble de mots techniques appartenant à une science, un art, un auteur ou un groupe social* »<sup>16</sup>. C'est donc l'ensemble des termes spécialisés propres à une science ou à un domaine d'activité. Autrement dit, un vocabulaire spécifique à un domaine particulier. Exemple : la terminologie du sport, de la médecine...

Dans le troisième et dernier sens, terminologie désigne une pratique ou une activité qui consiste à traiter les problèmes attachés aux termes. En effet, dans ce sens on pourrait parler plus exactement de terminographie, elle englobe les activités et pratiques dont le but est de décrire les termes dans des dictionnaires spécialisés. Pour cela, on la rapproche de la terminologie qui étudie l'aspect théorique des dénominations tandis que la terminographie étudie l'aspect pratique.

La terminologie garantit une bonne communication et la transmission des connaissances entre les spécialistes au sein d'une langue.

---

<sup>16</sup> Silvia Pavel et Diane Nolet, *Précis de terminologie*. Bureau de la traduction, 2001 [http://www.termsciences.fr/sites/termsciences/IMG/pdf/precis\\_de\\_terminologie\\_Pavel](http://www.termsciences.fr/sites/termsciences/IMG/pdf/precis_de_terminologie_Pavel) (page consultée le 09/01/2020)

## 9- La fiche terminologique

La fiche terminologique est un outil de présentation des données qui englobe toutes les informations caractérisant les termes d'un domaine de spécialité comme : terme, domaine, sous-domaine, catégorie grammaticale, définition, exemple, ... Dans ce cas, Pavel et Nolet l'affirment c'est un «*modèle de présentation des données qui regroupe en divers champs tous les renseignements disponibles relatifs à un concept spécialisé (termes et marques d'usage, justifications textuelles, domaines, langues, etc.)*»<sup>17</sup> (Pavel et Nolet 2002.p.108).

La fiche terminologique donne des informations sur les concepts et les termes du domaine de spécialité. Ils peuvent être repartis dans des champs différents, chacun correspond à «*autant de catégories de données ou d'éléments apportant, pour chaque terme vedette, un type d'information spécifique*»<sup>18</sup> (Gouadec 1990 : 38).

## 10- L'unité terminologique

L'unité terminologique représente l'étiquette donnée à un concept dans un ensemble de concepts pour distinguer un terme d'un autre dans une langue générale, elle désigne non seulement sa relation avec ce concept de spécialité, mais aussi sa lexicalisation dans le texte qui traite ce dernier. Elle peut être un mot, un sigle, un symbole...

Dans notre cas, il s'agit d'étiquette de domaine qu'on donne pour chaque terme qui figurent dans les textes de spécialité de notre corpus pour les distinguer des autres termes dans le dictionnaire terminologique.

---

<sup>17</sup>[https://www.lli.ulaval.ca/fileadmin/llt/fichiers/recherche/revue\\_LL/vol35/A.Francoeur.pdf](https://www.lli.ulaval.ca/fileadmin/llt/fichiers/recherche/revue_LL/vol35/A.Francoeur.pdf), (consulté en ligne le 12 janvier 2020)

<sup>18</sup> Ibid

## 11- Le texte de spécialité

Un texte de spécialité est défini comme un ensemble cohérent permettant la transmission et/ou la vulgarisation des connaissances par un langage dit spécifique. Il est destiné à des spécialistes d'un domaine « *Il est souvent considéré a priori comme un discours vrai* » (Tukia1983, 34) car son contenu est généralement le résultat d'une recherche. Son but est d'informer les spécialistes d'un même domaine ou de vulgariser un savoir à un large public.

Pour Cabré, les textes spécialisés se définissent par trois conditions :<sup>19</sup>

- Conditions discursives : les propriétés de la situation spécialisée de ce type de communication
- Conditions cognitives : le thème qui est traité et la façon dont il est traité.
- Conditions linguistiques : les conditions textuelles générales (précision, concision et systématisme, les deux dernières avec des degrés différents suivant les conditions discursives), la forme macro et micro textuelle, et surtout les unités lexicales propres au domaine dont il est question dans le texte.

## 12- Les caractéristiques des textes de spécialité

Le texte de spécialité est généralement le reflet d'une situation de communication particulière ayant pour objectif de transmettre des connaissances bien déterminées dans un domaine donné.

Pour atteindre cet objectif, le TS doit respecter certaines règles et une méthodologie touchant aussi bien la forme que le contenu afin de bien structurer la pensée du spécialiste.

Le TS se révèle par plusieurs caractéristiques qui le différencient des autres types de textes dont certains éléments : lexical, grammatical, énonciation, thématique et textuel.

Sur le plan lexical :

- Présence d'une terminologie spécifique.
- Absence de polysémie et de synonymie.

---

<sup>19</sup> Marea Teresa Cabre. « *Constituer un corpus de textes de spécialité* ». Institut universitari de lingüística aplicada. Universitat Pompeu Fabra (barcelone). Cahier de ciel .2007/2008.p.38

- Emploi de certains symboles, formules...
- Emploi des abréviations, siglaisons, acronymes et néologismes.

Sur le plan grammatical et selon Kocourek (1991), on distingue quatre groupes<sup>20</sup> :

- sélection des catégories grammaticales.
- complexité de la structure.
- condensation syntaxique.
- impersonnalité de la phrase.

Sur le plan d'énonciation :

- objectivité du discours
- absence de la deuxième personne du singulier et du pluriel (*tu, vous*)
- emploi de la troisième personne du singulier.
- emploi des formes passives, du pronom de modestie (*nous*), des articles indéfinis.

Sur le plan textuel et thématique :

- thème propre à un domaine donné (*médecine, économique, chimie ...*)
- absence total de genre littéraire.
- suite logique dans la progression du texte : entre le thème et les rhèmes (ce qu'on dit du thème).

### 13- Conclusion partielle

Les langues de spécialité se distinguent des autres langues par des particularités à plusieurs niveaux : syntaxiques, lexical, sémantique. Elles sont hétérogènes, ont un mode de fonctionnement spécifique et indépendant des langues générales. Elles ont également une terminologie spécifique en effet, dans chaque domaine précis on utilise des termes bien particuliers, qui renvoient à des situations bien définies dans ces domaines, et qu'on ne peut pas trouver dans des autres situations utilisés dans d'autres domaines.

---

<sup>20</sup>Marea Teresa Cabre. « Constituer un corpus de textes de spécialité ». Institut universitari de lingüística aplicada. Universita Pompeu Fabra (barcelone). Cahier de ciel .2007/2008.p.47

# **Chapitre 2**

**Le traitement automatique des langues**

## 1- Introduction partielle

Le traitement automatique des langues (TAL) est un domaine récent apparu dans les années cinquante, dans un contexte scientifique, initié par Z. HARRIS et N. CHOMSKEY, joint deux spécialités : l'informatique et la linguistique, son objectif est de concevoir des logiciels et applications qui peuvent traiter automatiquement des données linguistique, comme : dictionnaires électroniques, traduction automatique, classification automatique des textes (qui est notre objectif dans cette recherche). Notre recherche s'inscrit donc bien dans ce domaine.

Dans ce chapitre, nous présenterons ce domaine de manière concise, nous définirons le traitement automatique des langues, ferons un bref historique, présenterons quelques une de ses applications ainsi que quelques méthodes de traitement.

## 2- Bref historique de traitement automatique des langues

Les premiers travaux en TAL ont commencé au début des années cinquante aux Etats-Unis et ont porté sur la traduction automatique (TA) avec la création du premier traducteur qui permettait la traduction des phrases russe vers l'anglais pendant la période de la guerre froide.

En 1950, Alan Turing publie son article « *Computing machinery and intelligence* » portant sur le thème de l'intelligence artificielle où il aborde une méthode d'évaluation qui s'appelle « *le test de turing* »<sup>21</sup>. Il permet de tester l'intelligence artificielle d'une machine pour imiter la conversation humaine.

En 1952, Yehoshua Bar-Hillel organise la première conférence sur la traduction automatique, au MIT (Massachusetts Institute of Technology).

En 1953, Bar-Hillel émet l'hypothèse qu'une machine peut être conçu et délimiter les structures des phrases d'une langue à condition que la syntaxe de cette langue soit présentée sous forme des données compréhensible pour la machine.

Dans les années 1960-1980, le TAL se libère des théories de l'information pour entrer dans le domaine de l'intelligence artificielle. T.Winograd a réalisé un programme au nom de

---

<sup>21</sup> <https://sites.google.com/site/int3llig3nc3artifici3ll3/le-test-de-turing> consulté le 23/03/2020 à 20h20

SHRDLU<sup>22</sup> qui permet à l'utilisateur de discuter avec un ordinateur à base de terme en anglais pour déplacer divers objets tels que des cubes, des cônes, et des sphères. Un système commercial pour la traduction automatique a été créé en 1968 au nom de SYSTRAN<sup>23</sup> pour améliorer la traduction du russe vers anglais.

A partir des années 1990 et jusqu'à nos jours, le TAL est influencé par l'intelligence artificielle, beaucoup de modèles ont été développés par Microsoft et Alibaba<sup>24</sup> et ils ont même pu récemment battre les humains dans la lecture et la compréhension grâce aux modèles d'apprentissage automatique de traiter de grandes quantités d'informations avant de fournir des réponses précises aux questions qui leur sont posées.

Aujourd'hui, le TAL est plus développé, beaucoup d'applications comme des traducteurs automatiques, de la recherche documentaire, des classificateurs des textes... sont là pour témoigner de l'importance et de l'utilité de ce domaine pour les utilisateurs.

### 3- Définition du traitement automatique des langues

Le Traitement Automatique des Langues (TAL) est un domaine pluridisciplinaire à la croisée de la linguistique, de l'informatique et de l'intelligence artificielle. Il étudie les méthodes qui permettent à la machine de représenter le langage humain. Il lie étroitement linguistes et informaticiens.

Le TAL se préoccupe de développer des outils informatiques (logiciels, programmes, applications...) capables de traiter des données linguistiques de manière automatique, c'est-à-dire sans intervention humaine.

Ce champ de recherche est initialement étudié dans des laboratoires de recherche telle que le LADL (Laboratoire De Documentaire et Linguistique) dirigé par M.GROSS.<sup>25</sup>

---

<sup>22</sup><https://www.tomshardware.fr/lhistoire-de-lintelligence-artificielle-en-images/18/> consulté le 24/03/2020 à 15:36

<sup>23</sup> <https://fr.wikipedia.org/wiki/Systran> consulté le 24/03/2020 à 15:46

<sup>24</sup> <https://fr.wikipedia.org/wiki/Alibaba> consulté le 24/03/2020 à 18:00

<sup>25</sup> Djaidar, O & Fortas, D. 2018. *Vers un prototype de reconnaissance et d'extraction des constructions à verbe support*, mémoire de master, université de Bejaia.

## **4- Les applications du TAL**

Le langage qu'il soit écrit ou oral peut donner lieu à des produits ou service du TAL. En effet, le TAL permet de développer des produits ou des services déjà existants. Ses applications sont multiples et variées, elles peuvent traiter, produire, lier, rechercher, classer, analyser et traduire de manière intelligente les informations disponibles sous forme textuelle, parmi lesquelles nous citerons :

### **4-1- Traduction automatique**

Parmi les premières recherches du domaine du TAL, on trouve la traduction automatique (TA), elle utilise un logiciel informatique pour traduire un texte d'une langue source vers une langue-cible, nous distinguons deux types de traduction la première est la traduction assistée par ordinateur où l'homme traduit avec un soutien informatique lui facilitant la tâche. La deuxième est la traduction complètement automatique qui se réalise sans intervention humaine.

### **4-2- Correcteur orthographique**

Il s'agit ici d'un logiciel qui permet d'analyser un texte afin de détecter et corriger les fautes d'orthographe et les erreurs qu'il contient. Cette fonction de correction est souvent incluse dans des logiciels où l'utilisateur doit saisir une importante quantité de texte (comme par exemple les logiciels de traitement de texte).

Ce correcteur compare les mots présents dans le texte aux mots du dictionnaire, s'ils sont présents ils sont acceptés sinon des propositions proches de ces mots seront mises en avant par le correcteur orthographique.

### **4-3- Résumé automatique**

Le résumé automatique consiste en une compression d'un document textuel avec une technique informatique. Le résumé automatique de texte peut se faire selon trois approches : l'abstraction, l'extraction et enfin la compression. L'abstraction se base sur la rédaction des phrases pas forcément existantes dans la version originale du texte mais en restant bien sûr dans le même sujet. La seconde se limite à extraire les phrases les plus pertinentes du texte et les enchaîner de façon à produire un extrait. La dernière, consiste à faire un résumé par compression de phrases après leur extraction.

#### 4-4- Génération automatique de textes

La génération automatique de texte (GAT) consiste à produire des textes dans une langue via un logiciel de façon automatique, syntaxiquement, sémantiquement et pragmatiquement corrects et compréhensibles par l'utilisateur, à partir des représentations informatisées.

La génération, généralement se partage en deux étapes :

- le " quoi dire ? " qui consiste à déterminer l'ensemble des idées d'un texte à produire.
- le " comment le dire ? " qui traite son expression en langue naturelle.

#### 4-5- Synthèse vocale

La synthèse vocale est une technique informatique qui permet à une machine de transformer un texte écrit en une série de son. Le but est de « *créer des paroles artificielles à partir d'un texte quelconque* » (Baloul, 2003).

Elle s'appuie sur des techniques de traitement des langues pour transformer le texte orthographique en version phonétique, et des techniques de *traitement du signal* pour au final transformer cette version phonétique en son.

#### 4-6- Recherche d'information

La Recherche d'Information (RI) est un type de système qui permet de trouver, dans une grande collection de documents, l'information recherchée par rapport à la requête <sup>26</sup> d'un utilisateur.

La recherche de l'information s'articule autour de deux étapes essentielles :<sup>27</sup>

- La phase d'indexation transforme les documents afin de créer une représentation de leur contenu textuel qui soit utilisable par le système de recherche de l'information.
- Enfin la phase de recherche qui se base sur un formalisme précis défini par un modèle de système de recherche de l'information dont l'objectif est de présenter les

---

<sup>26</sup> Exprime le besoin d'information de l'utilisateur

<sup>27</sup> Soulimani. M. 2014. *Réalisation d'un analyseur morphologique du texte arabe « AMTAR »*, mémoire de master, université de Bejaia.

documents répondant à la requête de l'utilisateur. Les documents présentés sont considérés comme les plus pertinents. (Boulaknade1, 2008)

#### **4-7- Classification et catégorisation de documents**

La classification et la catégorisation de documents a pour objectif de trier automatiquement des textes. Elle peut prendre plusieurs formes, on peut citer quelques-unes : classification par thème, par genre, par opinion, ou encore par type (comme dans ce travail de recherche). Elle est réalisée par des algorithmes propres à des systèmes de traitement de l'information.

L'intérêt de cette dernière est d'organiser des connaissances de manière à rechercher ou extraire une information efficacement.

### **5- Les niveaux de traitement automatique des langues**

Tous les niveaux d'analyse linguistique sont pertinents pour l'analyse du TAL comme le niveau morphologique, le niveau syntaxique, le niveau sémantique, et enfin le niveau pragmatique.

#### **5-1- Niveau morphologique**

Elle étudie la forme, la structure interne et les variations des mots. C'est à dire la formation de ceux-ci à partir des morphèmes sans tenir compte des règles syntaxiques.

##### **5-1-1. Racinisation**

La racinisation est un procédé de suppression de terminaisons des mots en ne laissant que le radical ou la racine.

##### **5-1-2. Lemmatisation**

Le but de lemmatisation est de regrouper tous les mots de la même famille dans un texte pour retrouver leurs formes canoniques, qui est le lemme. Certaines conjugaisons peuvent rendre cette tâche complexe pour des ordinateurs. Exemple : retrouver la forme canonique du verbe « aller » depuis des formes très modifiées « va, vont... » Car ils ne présentent plus le même radicale.

### **5-1-3. Segmentation**

La segmentation consiste à découper chaque mot pour faire apparaître les segments<sup>28</sup> qui le constituent.

### **5-2- Niveau syntaxique**

L'analyse syntaxique étudie les règles d'enchaînement des mots dans le but d'obtenir des phrases cohérentes et correctes, par l'intermédiaire d'un arbre syntaxique<sup>29</sup>.

Dans ce cadre, elle associe à chaque fragment de phrase sa structure syntaxique en les segmentant en syntagmes<sup>30</sup> : syntagme nominal, syntagme verbal, syntagme adjectival, ...

### **5-3- Niveau sémantique**

L'analyse sémantique vise à chercher la signification et le sens des mots figurants dans les phrases utilisées dans une langue donnée, en restant dans le même sens et dans la même logique que les éléments du texte.

### **5-4- Niveau pragmatique**

L'analyse pragmatique traite la signification de la phrase par rapport aux conditions situationnelles et contextuelles, donc elle est utilisée pour exprimer le sens le plus exact des mots. Autrement dit, elle prend en compte le contexte et la situation réelle pour poser le sens.

## **6- Conclusion partielle**

Nous concluons que le TAL est une discipline qui appartient au domaine de l'Intelligence artificielle (I.A.). Il est au cœur des nouvelles applications technologiques et informatiques qui utilisent la langue. Il vise à créer de systèmes informatiques intelligents capables de reconnaître, de comprendre, d'interpréter et de reproduire le langage humain sous ses différentes formes. Ses applications sont aujourd'hui essentiellement et semblent fournir les

---

<sup>28</sup> Unité dégagée par segmentation.

<sup>29</sup> Représente la structure syntaxique d'une phrase.

<sup>30</sup> Forment une unité à l'intérieur de la phrase

perspectives les plus prometteuses après de nombreuses évolutions tels que : la traduction automatique, correcteur orthographique, résumé automatique, synthèse vocal...

# **Chapitre 3**

**Analyse des textes de spécialité et  
présentation de l'application**

## 1- Introduction partielle

Après avoir abordé les aspects définitoires des notions et concepts fondamentaux en rapport avec les langues de spécialité, ainsi que les informations clés qui ont une relation avec cette recherche. Nous allons, dans ce chapitre, identifier et analyser les caractéristiques des textes de spécialités de notre corpus. Ensuite, nous allons expliquer le fonctionnement de l'application que nous avons conçue et décrirons chaque étape du processus permettant d'atteindre le résultat final.

## 2- Analyse des textes de spécialité

Dans cette première partie, nous avons choisi d'analyser linguistiquement les textes de diverse spécialité (médecine, économie, informatique, chimie et physique.) choisis pour notre corpus ainsi que certains des termes qui les composent. Le but principal de cette analyse est de percer les caractéristiques linguistiques qui les distinguent des autres textes.

### 2-1- Caractéristiques morphosyntaxiques

Généralement dans les TS et plus précisément dans les textes du domaine de la médecine, l'économie, l'informatique, la chimie et la physique, les spécialistes utilisent les mêmes règles syntaxiques que celles que l'on trouve dans les textes de la langue générale. Toutefois, ces TS se caractérisent par des structures spécifiques à ces domaines dans l'objectif principal est d'assurer l'objectivité du discours.

#### 2-1-1. Les pronoms personnels

Nous avons remarqué lors de l'analyse des textes de spécialité, l'usage des pronoms personnels : il, on, nous.

- **1<sup>er</sup> exemple tiré du corpus texte N° 19** : « Elle est la réunion du matériel et du système d'exploitation (que nous noterons OS par la suite pour Operating System) »
- **2<sup>ème</sup> exemple tiré du corpus texte N° 38** : « Nous admettons que l'équilibre thermodynamique est atteint dans ces conditions »

- **3<sup>ème</sup> exemple tiré du corpus texte N° 46** : « nous pouvons résumer la fonction de production ainsi :  $Y=F(K, L)$ . (...), En représentant cette hausse par la lettre  $z$ , nous obtenons alors la quantité produite suivante :  $zY= F(zK, zL)$  »
- **4<sup>ème</sup> exemple tiré du corpus texte N° 49** : «  $p$  le prix d'une firme dont nous étudions le comportement. (...), et  $Q_i$  les ventes de la firme à laquelle nous nous intéressons »

En effet, le " nous " ici est utilisé pour prendre un peu de distance avec l'égo (je) et insérer le récepteur dans le sujet, aussi pour rappeler que l'auteur n'est pas parti de rien et que le contenu du TS est le résultat d'une longue recherche ; ou bien encore tout simplement par une modestie de la part de l'auteur.

Le pronom " on " joue le même rôle que le pronom " nous ", il permet d'inclure le destinataire, et de rejeter une partie de la responsabilité à l'égard des informations abordées. Comme dans les exemples suivants :

- **1<sup>er</sup> exemple tiré du corpus texte N° 04** : « On retrouve ce minéral dans de très nombreux fruits et légumes »
- **2<sup>ème</sup> exemple tiré du corpus texte N° 07** : « On lui associe souvent de l'aspirine à faible dose en continu, afin de limiter le risque thrombotique »
- **3<sup>ème</sup> exemple tiré du corpus texte N° 17** : « on utilise un navigateur (ou browser en anglais) »
- **4<sup>ème</sup> exemple tiré du corpus texte N° 22** : « On peut trier les données en utilisant un ou plusieurs critères (appelés aussi « clés »). (...) On peut effectuer un tri en fonction des valeurs. (...) on peut placer les cellules en haut ou en bas »
- **5<sup>ème</sup> exemple tiré du corpus texte N° 26** : « il doit être interprété par un autre programme, qu'on appelle interpréteur »
- **6<sup>ème</sup> exemple tiré du corpus texte N°29** : « Mais on doit utiliser des compilateurs différents »

Le pronom personnel " il " est employé avec les verbes impersonnels comme « falloir » :

- **Exemple tiré du corpus texte N° 27** : « Il faut alors ajouter l'option -classpath suivie des répertoires »

- **Exemple tiré du corpus texte N° 42** : « il faut retenir le cas où la circulation monétaire demeure insuffisante par suite d'une modification des habitudes de paiement »
- **Exemple tiré du corpus texte N° 47** : « il faut calculer la productivité globale des facteurs »

Ou dans d'autres formes à l'instar de la langue générale. Comme : « *il est nécessaire* », « *il est recommandé* », « *il suffit de* » ; « *il s'agit* » ... :

- **Exemple tiré du corpus texte N° 02** : « il est également nécessaire de retirer un ou plusieurs ganglions lymphatiques axillaires »
- **Exemple tiré du corpus texte N° 09** : « il est recommandé de limiter sa consommation à moins de 200 grammes 24 heures avant l'examen »
- **Exemple tiré du corpus texte N° 19** : « Il est donc nécessaire que quelque chose dans l'ordinateur permette la communication entre l'homme et la machine »
- **Exemple tiré du corpus texte N° 17** : « Il suffit de cliquer sur un lien pour être dirigé vers une nouvelle page »
- **Exemple tiré du corpus texte N°43** « Il s'agit d'un phénomène persistant qui fait monter l'ensemble des prix »

Il est utilisé également pour remplacer des sujets dans des phrases, comme cela se fait dans la langue générale. Voici quelques exemples :

- **Exemple tiré du corpus texte N°03** : « Le calcium lié aux protéines est une forme de réserve de calcium pour les cellules, mais il n'a pas de fonction active dans l'organisme »
- **Exemple tiré du corpus texte N°04** : « (...) apport en potassium provient de l'alimentation humaine puisqu'il n'est absolument pas synthétisé par l'organisme »
- **Exemple tiré du corpus texte N°05** : « le parasite passe dans le système sanguin de la personne, où il se multiplie et peut entraîner la maladie »

Par contre, nous avons remarqué une absence totale des pronoms personnels " tu ", " vous "). En bref, c'est l'usage de ces pronoms qui marque le caractère impersonnel des phrases dans ces textes de spécialité et leur l'objectivité.

### 2-1-2. Les temps verbaux

Nous avons constaté la prédominance du présent de l'indicatif : le présent de vérité générale vu que ce temps est largement prédominant dans les textes scientifiques. Exemples :

- Texte N° 01 : « La plupart des personnes présentent un ictère (jaunissement de la peau et des yeux) se sentent fatiguées et faibles »
- Texte N° 21 : « une petite barre verticale permet, par cliqué-glissé, d'afficher plus ou moins »
- Texte N° 30 : « comme toutes les réactions chimiques, se font sans variation de masse de chacun des éléments ».
- Texte N° 31 : « L'acide éthanoïque réagit avec l'eau pour donner l'ion acétate »
- Texte N° 42 : « Cette définition, issue des théories quantitativistes, ne suffit pas à rendre compte de l'inflation »

Le passé composé s'impose quand à lui dans les formes passives. Exemples :

- Texte N° 02 : « Elle est toujours complétée d'une radiothérapie »
- Texte N° 05 : « La malaria est traitée grâce à L'Artémisine »
- Texte N°35 : «  $K_a$  est appelée constante d'acidité du couple A/B et la quantité  $-\log(K_a)$  est appelée  $pK_a$  du couple »
- Texte N°42 : « L'inflation a été souvent identifiée avec un accroissement de la circulation monétaire »
- Texte N°46 : « L'offre est déterminée par la fonction de production »
- Texte N°49 : « Le modèle de concurrence monopolistique a été introduit par l'économiste Chamberlin »

### 2-1-3. Les types de phrases

Le type de phrases qui prédomine est les phrases affirmatives et déclaratives. Exemples :

- Texte N°02 : « La chirurgie non conservatrice (ou mastectomie) consiste à retirer la totalité du sein y compris l'aréole et le mamelon. »
- Texte N°19 : « Un ordinateur est constitué de matériel (hardware) et de logiciel (software). »

- **Texte N°32** : « Le voltamètre d'Hoffmann est le dispositif expérimental permettant d'étudier l'électrolyse de l'eau. »
- **Texte N°47** : « Les productivités du travail et du capital, apparentes ou effectives sont des productivités partielles. »

Pour transmettre des connaissances telles qu'elles sont, sans marquer la présence de l'auteur pour assurer l'objectivité.

### 2-1-4. La condensation syntaxique

Les phrases utilisées dans les textes des domaines étudiés dans ce travail (médecine, économie, chimie, informatique et la physique) sont tellement chargées d'information qu'elles ont un caractère de complexité. C'est pour cette raison que l'auteur ici a recours à des procédés d'économie du langage pour apporter beaucoup de connaissances en peu de mots, c'est ce que l'on appelle la condensation syntaxique et lexicale. Parmi ces procédés, nous avons :

- **La nominalisation** est utilisée de manière très fréquente dans nos TS, c'est un procédé qui consiste à transformer un verbe ou un adjectif en substantif ou une phrase en syntagme nominal. Ex : « *insuffisance hépatique* » ; « *traitement médical* » ; « *augmentation de la production* » ; « *réactions de combustion* » ; « *fonction de demande* »
- **La voix passive** où le sujet subit l'action, le sujet de la phrase active devient complément d'agent de la phrase passive et le complément direct de la phrase active devient sujet de la phrase passive. Cette forme est très présente dans nos textes. Ex :
  - **Texte N°17** : « *Internet est issu du réseau Arpanet, qui a été conçu en 1969* » ;
  - **Texte N°25** : « *Si le lien a été créé sur un texte* » ;
  - **Texte N°42** : « *L'inflation a été souvent identifiée avec un accroissement de la circulation monétaire* » ;
  - **Texte N°38** : « *La réaction est catalysée par de l'acide phosphorique* ».
- **La pronominalisation** est une transformation qui remplace un syntagme nominal par un pronom pour supprimer répétitions et lourdeurs de phrase, afin d'avoir un écrit clair et concis. Ex :
  - **Texte N°02** : « La chirurgie non conservatrice (ou mastectomie) consiste à retirer la totalité du sein y compris l'aréole et le mamelon » ;

- **Texte N°09** : « Le niveau sanguin de créatinine étant lié à l'activité musculaire, il est recommandé de cesser toute activité physique » ;
- **Texte N°57** : « Une entreprise peut obtenir des capitaux sur le marché financier de deux manières. Elle peut émettre des titres de dette à revenu fixe ».
- **Les propositions subordonnées participiales**, il s'agit d'une proposition **sans subordonnant** dont le noyau verbal est un **participe ayant son propre sujet**. Comme dans les exemples suivants :
  - **Texte N°43** : « la consommation des biens et services, mesurée à quantité et qualité égales »
  - **Texte N°38** : « Toutes ces réactions étant à l'équilibre »
  - **Texte N°43** : « La monnaie étant l'étalon des valeurs »

## 2-2- Caractéristiques morphologiques et lexicale des termes

Le lexique des LS de médecine, économie, chimie, informatique et la physique est purement spécialisé car ce sont des domaines ultra-spécialisés. En effet, nous avons remarqué une présence abondante de terminologie spécialisée. Ex :

- **la terminologie de médecine** : *La chirurgie, une radiothérapie, symptômes, les nodules thyroïdiens, maladies, médicaments ...;*
- **la terminologie d'économie** : *crise financière, investissement, marchés, production, entreprise, coût, consommateur... ;*
- **la terminologie informatique** : *Logiciel ; Excel, diapositives, lien hypertexte, Diaporama, Java, Python, Windows....*

Nous avons remarqué également la répétition de mêmes unités pour un concept car il n'existe ni synonymie ni polysémie dans un même texte. Exemples :

- **Texte N°01** : « L'insuffisance hépatique est une dégradation sévère de la fonction hépatique, elle est due à un trouble ou à une substance qui endommage le foie (...). Le médecin pose, en général, le diagnostic d'insuffisance hépatique en se basant sur les symptômes et les résultats de l'examen clinique et des analyses de sang (...). L'insuffisance hépatique peut être la conséquence de tout type de maladie hépatique, telle qu'une hépatite virale (le plus souvent l'hépatite B ou C) (...) »
- **Texte N°34** : « Les réactions de combustion, comme toutes les réactions chimiques, se font sans variation de masse de chacun des éléments. Les équations chimiques qui

décrivent ces réactions doivent donc satisfaire cette contrainte. À titre d'exemple, les réactions de combustion du carbone et du méthane sont respectivement :  $C + O_2 \rightarrow CO_2$  et  $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$ . On distingue dans une telle réaction, les réactifs, qui subissent la réaction, que l'on sépare en combustible »

- **Texte N°4 :** « L'inflation est la perte du pouvoir d'achat de la monnaie qui se traduit par une augmentation générale et durable des prix (...). L'inflation doit donc être distinguée de l'augmentation du coût de la vie car elle ne prend pas en compte la variation des quantités achetées en réponse à l'évolution des prix (...). L'inflation est un phénomène à propos duquel les controverses entre économistes sont nombreuses (...) »

Nous allons identifier des catégories lexicales des termes présentes dans nos TS. Ensuite nous passerons aux procédés de la création lexicale (analyse morphologique des termes).

### 2-2-1. Les catégories lexicales

Selon la grammaire traditionnelle, on classe les mots en neuf catégories: adjectif, adverbe, article, conjonction, interjection, nom, préposition, pronom et verbe. En revanche, on trouve parmi ces dernières quelques unes de manière abondante dans nos textes de spécialité :

- La catégorie nominale : on trouve les substantifs ; ex « *Le médecin* » ; « *La chirurgie* » ; « *Le courrier* » ; « *Un tableur* » ; « *la quantité* » ; « *les électrons* » ; « *le marché* » ; « *Une entreprise* ». Et les syntagmes nominaux ; ex « *Une nouvelle technologie* » ; « *données numériques* » ; « *Les options thérapeutiques* » ; « *un trouble héréditaire* » ; « *des bases faibles* » ; « *une crise économique* ».
- La catégorie adjectivale : elle regroupe en effet, les adjectifs ; ex « *Sévère* » ; « *élevé* » ; « *active* » ; « *simples* » ; « *faible* » ; « *répulsives* » ; « *insuffisante* » ; « *durable* », les adjectifs verbaux et les participes adjectivés ; ex « *une fièvre élevée* » ; « *enfants affectant* ».
- La catégorie verbale : englobe tous les types de verbes : les verbes d'action ; ex « *Réduire* » ; « *produire* » ; « *transformer* » ; « *utiliser* », les verbes impersonnels comme : « *il semble* » ; « *il faut* » ; et les verbes d'état ; ex « *Etre* » ; « *paraître* » ; « *devenir* » ; « *reste* » etc.
- La catégorie adverbiale: contient des adverbes ; ex : « *Beaucoup* » ; « *encore* » ; « *très* » ; « *avant* » ; « *à gauche* » ; « *bien* » etc.

- locutions adverbiales ; ex : « *en général* » ; « *en haut* » ; « *en bas* » ; « *de suite* » ; « *de là* » ; « *de moins en moins* » etc.
- La catégorie prépositive : contient des prépositions « *De ; pour ; à ; selon ; chez ; parmi ; en ; sur ; dans ; par ; avec* » ; et des locutions prépositives ; comme : « *Afin de ; à moins de ; à partir de ; d'après ; en cas de ; par rapport à* »
- La catégorie conjonctive : regroupe des conjonctions de coordination « *Mais ; donc ; et ; ou ; car ; ni* » ; et des locutions conjonctives « *Bien que ; sans que ; de même que ; ainsi que* »

En effet, kocourek (1991 :31) classe ces catégories en deux classes : la classe lexicale faible qui englobe les prépositions, les adverbes et les conjonctions. Et La classe lexicale forte regroupe la classe verbale, nominale et adjectivale. Elle a un rôle très important dans les LS car elle inclut toute la terminologie et les procédures propres à un domaine de spécialité.

### 2-2-2. Les procédés de la création lexicale des termes

#### a) Dérivation

La dérivation est un procédé de formation qui consiste à créer des nouvelles unités lexicales par l'adjonction d'affixes à un radical de la langue cible. L'affixe peut être préfixe ou suffixe. Elle est présente avec tous ses types dans nos TS de (médecine, économie, informatique, chimie, physique)

- **La dérivation affixale** : Elle se fait par l'ajout d'un préfixe à une base ou d'un suffixe. En effet, c'est le procédé le plus utilisé dans les TS de (médecine, économie, informatique, chimie, physique). Exemples : « *saponification* ; *réplication* ; *codage* ; *regulation* ; *Anticoagulation* ; *Antigrippal* ; *immobiliser* ; *indispensable* »
- **La dérivation parasynthétique** : Il consiste à réunir en parallèle la préfixation et la suffixation. Donc il s'agit de l'adjonction d'un préfixe et d'un suffixe à la même base. L'emploi de ce mode de dérivation est rare dans notre corpus. Prenant quelques exemples : « *Intervention* ; *emplacement* ; *déshydratation* ; *indéfinitement* »
- **La dérivation régressive** : elle se fait par la suppression d'un suffixe ou à partir d'une base. Exemples : *la Fonction* → de verbe « fonctionner » ; *le médecin* → de nom « médecine » ; *la demande* → de verbe « demander » ; *la forme* → de verbe

« former »; *la charge* → de verbe « charger »; *une saignée* → de verbe « saigner »; *le calcul* → de verbe « calculer »; *l'analyse* → de verbe « analyser »; *la structure* → de verbe « structurer ».

## b) Composition

La composition consiste à construire des mots nouveaux, à partir d'une base préexistante, qui se réalise par l'adjonction de deux notions nouvelles d'une unité lexicale soit par l'adjonction d'un préfixe à un radical, soit par l'association de deux lexèmes. Dans notre corpus, nous avons :

- **Composition savante :** C'est la juxtaposition des éléments qui peuvent être grecs et latins. elle utilise des sources différentes de base lexicales pour former des mots composés. Ce sont des mots dits « savants », elle est très fréquente dans nos TS vu qu'ils sont des textes scientifiques surtout dans les textes de médecine. Par exemple : *hyperthyroïdie* (hyper du grec+ dérivé de verbe thyroïde+ suffixe ie); *polyglobulie* (poly du grec + globule du latin + ie); *Plasmodium* (du latin plasma+ du suffixe latin odium); *Myalgie* (Composé de préfixe myo- et de algie, tiré du grec); *Homogène* (mot grec formé du préfixe homo+la racine genos); *hypertexte* (préfixe grec hyper+ texte du latin textum) ; *rachitisme* (rachis du grec+ suffixe -isme du latin)

## c) Les emprunts

Il apparaît clairement dans les domaines étudiés dans notre corpus qu'il y a des créations de beaucoup de termes basés sur l'emprunt. Il s'agit d'adopter des traits linguistiques d'une langue source à une langue cible. A noter que les termes relatifs à (la médecine, l'économie, la chimie, l'informatique et la physique) ont été, presque tous, transférés au français à partir d'anglais. Notre corpus contient une grande quantité d'unités terminologiques (UT) empruntées de cette langue citée dernièrement, car les spécialistes recourent toujours à la terminologie d'autres langues au détriment de leurs synonymes en français. En effet, aucun système linguistique (langue de spécialité) ne peut survivre en complète autonomie et faire face aux progrès de l'humanité et de la science. Prenons quelques exemples :

«*Calcium*» ; «*web*»; «*wifi*»; «*electron*»; «*ion*»; «*hormone*»; «*windows*»; «*cathode*»; «*firme*»

### d) L'abréviation

Les LS de (médecine, économie, informatique, chimie, physique) se constituent de plusieurs abréviations. Il s'agit de réduire un mot ou un ensemble de mots à un certains de ses caractères. L'abréviation recourent à plusieurs procédés, nous citons les plus fréquents dans notre corpus :

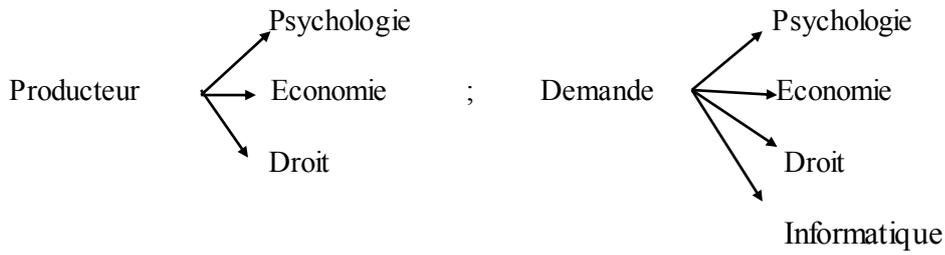
- **Siglaion** c'est-à-dire la réduction d'un mot ou d'une expression à ses lettres initiales qui forment un mot unique. Ex : *(CCIS)* → « Carcinomes canaux in situ » ; *(HbA)* → « hémoglobine A » ; *(CIVD)* → « coagulation intravasculaire disséminée » ; *(ARPA)* → « Advanced Research Project Agency » ; *(OS)* → « Operating System » ; *(RT)* → « recette totale » ; *(p)* → « prix du bien » ; *(Q)* → « quantités vendues » ; *(OMS)* → « organisation mondiale de santé » ; *(HO)* → « hydroxyde » ; *(WWW)* → « World Wide Web »
- **Symbole** est un signe spécifique à un domaine précis, conventionnel, qui peut notamment consister en une ou plusieurs lettres, correspondant à une réalité, à un élément, à une unité de mesure, à une opération. Dans notre corpus, nous avons noté l'usage des symboles et des formules en intensité surtout dans les textes de la chimie et la physique. Ex :  $(\Leftrightarrow)$  → Équivalence logique ;  $(\beta)$  → radioactivité bêta

En effet, le but de l'usage de ces derniers est d'assurer l'économie du discours et de l'information tout en gardant l'objectivité et la vérité scientifique.

### 3-3- Caractéristiques sémantiques

D'après l'analyse sémantique des termes présents dans nos textes, nous avons constaté qu'il existe deux qualités de sens pour les termes, ils peuvent être :

- **Monosémique** : lorsqu'un terme n'a qu'un seul sens et qu'il appartient à un seul domaine de spécialité. Généralement, ce genre de termes appartient au vocabulaire scientifique et technique comme dans notre cas pour la médecine, la chimie, la physique. Ex : *Tumorectomie* → médecine ; *Sulfurique* → chimie ; *Electron* → physique
- **polysémique** : lorsqu'un même terme a plusieurs sens et/ou qu'il appartient à plusieurs domaines en même temps. C'est-à-dire qu'un seul terme peut avoir plusieurs domaines de spécialité. Comme c'est le cas des termes issus de textes d'économie. Ex



#### 4- Présentation de l'application

La formalisation de notre corpus sous forme d'une application a permis la mise en place d'un prototype dont les différentes étapes, que nous développerons ensuite, sont illustrées dans la figure suivante :

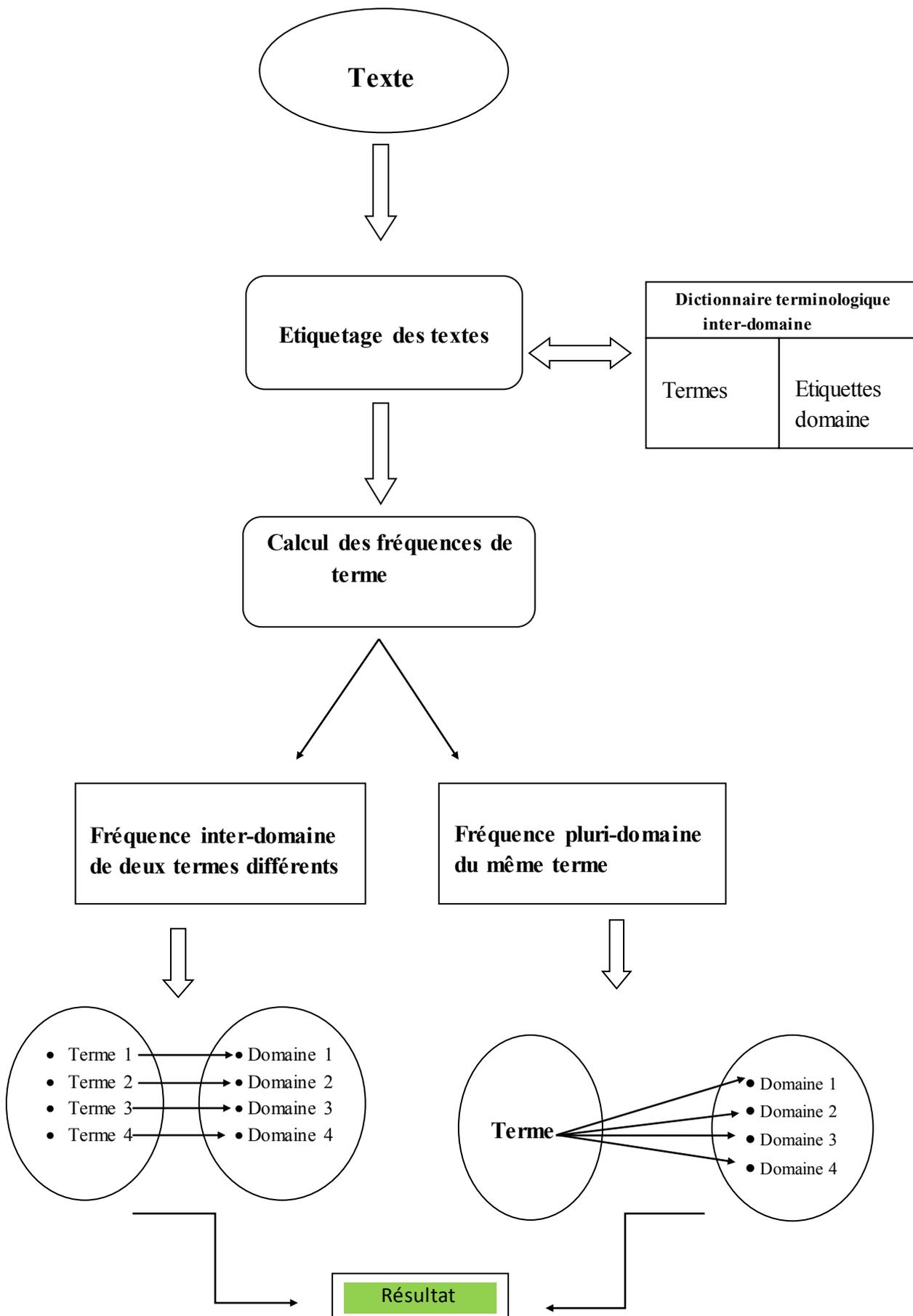


Figure 01 : Processus de catégoriseur automatique des textes écrits par domaine de spécialité.

- **Dictionnaire terminologique inter-domaine**

Pour que l'application puisse détecter les domaines de spécialité de chaque texte, nous avons élaboré un dictionnaire terminologique dans lequel nous mettons les termes des textes de spécialité de notre corpus avec la forme canonique <sup>31</sup> puis nous leur donnons les étiquettes domaines en se basant sur *le grand dictionnaire terminologique (GDT)*, pour que le système les reconnaisse en consultant notre dictionnaire.

Terme	Etiquette domaine
‘‘Bilan’’	Biol
‘‘Bilan’’	Phys
‘‘Ester’’	Biol
‘‘Ester’’	Chim
‘‘Butanol’’	Chim
‘‘Solvant’’	Med
‘‘Solvant’’	Chim
‘‘Estérification directe’’	Chim
‘‘Acide carboxylique’’	Chim
‘‘Alcool’’	Chim
‘‘Catalyseur’’	Biol
‘‘Catalyseur’’	Chim
‘‘Distillation’’	Chim
‘‘Distillation’’	Med
‘‘Volatilité’’	Chim
‘‘Volatilité’’	Infr
‘‘Ébullition’’	Phys
‘‘Ébullition’’	Med
‘‘Acide acétique’’	Chim
‘‘Acétate’’	Chim
‘‘Butyle’’	Chim
‘‘Réduire’’	Infr
‘‘Réduire’’	Dr
‘‘Réduire’’	Math
‘‘Réduire’’	Psych
‘‘Consommation’’	Éco
‘‘Consommation’’	Élect
‘‘Consommation’’	Infr
‘‘Énergétique’’	Élect
‘‘Énergétique’’	Phys
‘‘Séparation’’	Phys
‘‘Séparation’’	Spr
‘‘Séparation’’	Élect
‘‘Mélange’’	Med
‘‘Mélange’’	Chim

<sup>31</sup> La forme canonique est la forme d'un mot la plus simple utilisée comme entrée dans les dictionnaires

"Technique"	Med
"Technique"	Eco
"Technique"	Med
"Alternative"	Infr
"Position"	Phys
"Ethanol"	Chim
"Addition"	Med
"Addition"	Math
"Addition"	Chim
"Catalysé"	Chim
"Acide phosphorique"	Chim
"Acide phosphorique"	Agric
"MPa"	Phys
"Température"	Med
"Température"	Phys
"Cinétique"	Med
"Cinétique"	Phys
"Cinétique"	Biol
"Thermodynamique"	Phys
"Thermodynamique"	Med
"Réacteur"	Psych
"Réacteur"	Chim
"Effluent"	Med
"Enthalpie libre"	Med
"Atome"	Med
"Atome"	Chim
"Atome"	Infr
"Proton"	Phys
"Proton"	Med
"Proton"	Chim
"Ampholyte"	Chim
"Hydroxyde"	Chim
"Ho-"	Chim
"Acide chlorhydrique"	Chim
"Acide sulfurique"	Chim
"H2SO4"	Chim
"Ion sulfate"	Chim
"Hydroxyde de sodium"	Chim
"Organomagnésien"	Chim
"Cétone"	Chim
"Aldéhyde"	Chim
"Aldéhyde"	Med
"Alcène"	Chim
"Éthylène "	Chim
"Éthylène"	Med
"Éthanol"	Chim
"-OH"	Chim
"Estérification"	Chim

“Estérification”	Biol
“Méthanol”	Chim
“Déshydratation”	Med
“Déshydratation”	Chim
“Déshydratation”	Phys
“Déshydratation”	Biol
“Oxydation”	Chim
“Aldéhyde”	Chim
“Aldéhyde”	Med
“Acide carboxylique”	Chim
“Donneur”	Dr
“Donneur”	Med
“Donneur”	Biol
“Récepteur”	Med
“Récepteur”	Biol
“Récepteur”	Psych
“Récepteur”	Élect
“Concentration”	Eco
“Concentration”	Agric
“Concentration”	Chim

**Tableau 01 : Dictionnaire terminologique inter-domaine.**

- **Etiquetage des textes**

Dans cette étape, le système consulte automatiquement le dictionnaire terminologique qui permet de repérer les termes issus des textes de spécialité de notre corpus et de détecter les domaines auxquels ils appartiennent afin de faciliter leur recensement et le calcul de leurs fréquences dans chaque texte, ce qui constitue la prochaine étape.

Prenons quelques exemples :

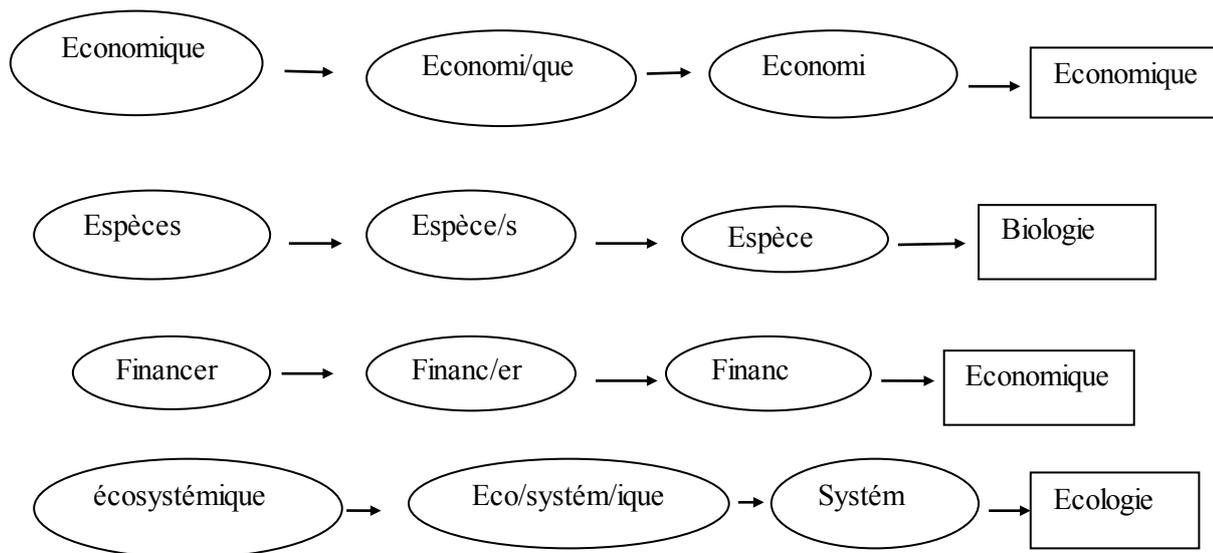
“*Concentration*” → philosophie

“*Reins*” → médecine

“*Électron*” → physique

“*Gestion*” → économique

Le système, dans certains cas, ne prend que la forme canonique des mots, qui est la forme de base sans les désinences. Il les segmente automatiquement en excluant les affixes (suffixes et préfixes) des radicaux. Comme dans les exemples ci-dessous:



**Figure 02 : segmentation automatique des affixes**

- **Calcul des fréquences de termes**

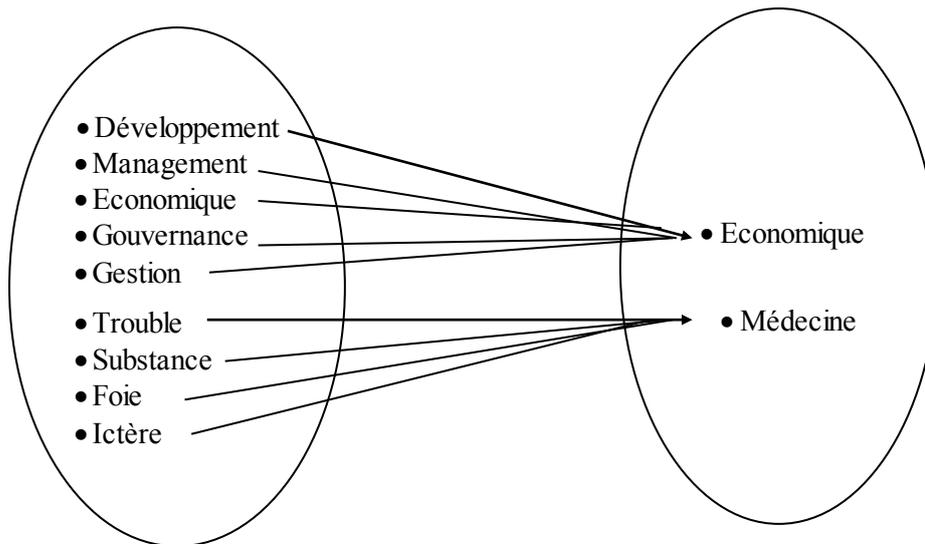
Le système, à cette étape du processus, calcule les fréquences des termes dans les textes (1591 termes) après les avoir préalablement étiqueter des textes. Il recense combien de termes appartiennent au domaine 1, au domaine 2, au domaine 3... pour pouvoir donner le résultat final qui aboutit à la définition du domaine du texte.

Dans notre cas, deux cas peuvent apparaitre, nous les développons ci-après :

**- Fréquence inter-domaine de deux termes différents:**

Dans la plupart des situations, le système trouve (recense) les termes appartenant au même domaine, il les dénombre pour en faire des statistiques.

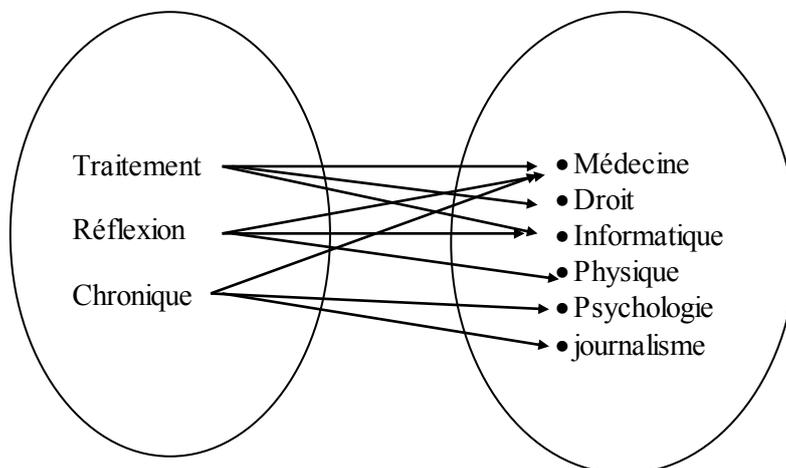
Cette relation est illustrée dans la figure ci-dessous :



**Figure 03 : Relation termes/ inter-domaine**

**- Fréquence pluri-domaine de même terme:**

Il existe des cas où le système peut trouver un même terme correspond à plusieurs domaines en même temps. Cette relation est illustrée dans la figure ci-dessous :



**Figure 04 : Relation terme/ pluri-domaine.**

- **Produit cartésien**

Nous utilisons la notion du produit cartésien car elle participe à la mise en place de notre dictionnaire terminologique.

Le produit cartésien de deux ensembles A et B est l'ensemble de tous les couples dont la première composante appartient à l'ensemble A et dont l'extrémité est un élément de l'ensemble B.

Dans notre cas, l'ensemble A est constitué des termes issus des textes de spécialité présents dans notre corpus, et B est l'ensemble des étiquettes domaines auxquels les termes appartiennent.

Prenons les exemples des deux cas apparus dans notre analyse : **Cas 01 :**

Ensemble A = Développement, Management, Economique, Gouvernance, Gestion, Trouble, Substance, Foie, Ictère → **(termes)**

Ensemble B = Economique, Médecine → **(domaines)**

Alors le produit cartésien sera :

$A*B =$  (Développement, Economique), (Management, Economique), (Economique, Economique), (Gouvernance, Economique), (Gestion, Economique)

, (Trouble, Médecine), (Substance, Médecine), (Foie, Médecine), (Ictère, Médecine).

**Cas 02 :**

Ensemble A = Traitement, Réflexion, Chronique → **(termes)**

Ensemble B = Médecine, Droit, Informatique, Physique, Psychologie, Journalisme → **(domaines)**

$A*B =$  (Traitement, Médecine), (Traitement, Droit), (Traitement, Informatique)

, (Réflexion, Médecine), (Réflexion, Informatique), (Réflexion, Physique)

, (Chronique, Médecine), (Chronique, Journalisme), (Chronique, Psychologie).

- **Résultat**

Cette étape est la dernière du processus, elle correspond à la phase de catégorisation des textes. Après avoir fait l'étiquetage des textes en consultant le dictionnaire terminologique pour calculer les fréquences selon le domaine des termes. En effet, c'est le domaine qui a le pourcentage le plus élevé auquel le système va attribuer le texte.

## 5- Description de l'interface utilisateur

Notre application permet d'identifier automatiquement les domaines des textes écrits, elle se compose de différentes parties. Nous allons maintenant présenter l'interface de l'application et la logique sur laquelle est basée.

### 4-1- Interface d'authentification

Pour que l'utilisateur puisse utiliser l'application, il doit d'abord s'authentifier en introduisant le mot de passe.



Figure 05 : interface d'authentification

## 4-2- Fenêtre principale

Après avoir introduit son mot de passe, l'utilisateur aura l'accès à l'application et la fenêtre principale nommée « catégoriseur des textes écrits par domaine de spécialité » s'affiche. Sur cette dernière figurent : le nom de notre université, notre faculté, notre département, nos noms de toutes les deux qui avons élaboré ce mémoire, le nom de notre encadrant, l'intitulé de notre thème de recherche, la date et l'heure et ainsi que trois onglets qui servent à insérer trois données à l'application : le premier pour la création de liste des spécialités, le deuxième pour la création du dictionnaire terminologique et le troisième pour la création des textes.



Figure 06: fenêtre principale

- a. **Onglet « création » de liste des spécialités** : il permet de créer et d'y avoir accès à l'ensemble des spécialités, afin d'étiqueter les termes, avec la possibilité

d'ajouter, modifier, supprimer et afficher une spécialité. Il suffit d'écrire le nom de spécialité sur la case de « spécialité » et en cliquant sur le bouton « ajouter », la spécialité sera ajoutée à l'application.



Figure 07: fenêtre de liste des spécialités

**b. Onglet « création » du dictionnaire terminologique :** il permet de créer le dictionnaire terminologique composé des termes et spécialités auxquels ils appartiennent, avec la possibilité d'ajouter, modifier et supprimer.

Il suffit d'écrire le nom de terme sur la case de « terme » et le nom de spécialité au quelle fait parti et en cliquant sur le bouton « ajouter » le terme sera intégré dans l'application.

Pour la suppression ou la modification d'un terme, il suffit de suivre les mêmes étapes mais en cliquant sur le bouton « modifier » ou « supprimer » au lieu « ajouter ».

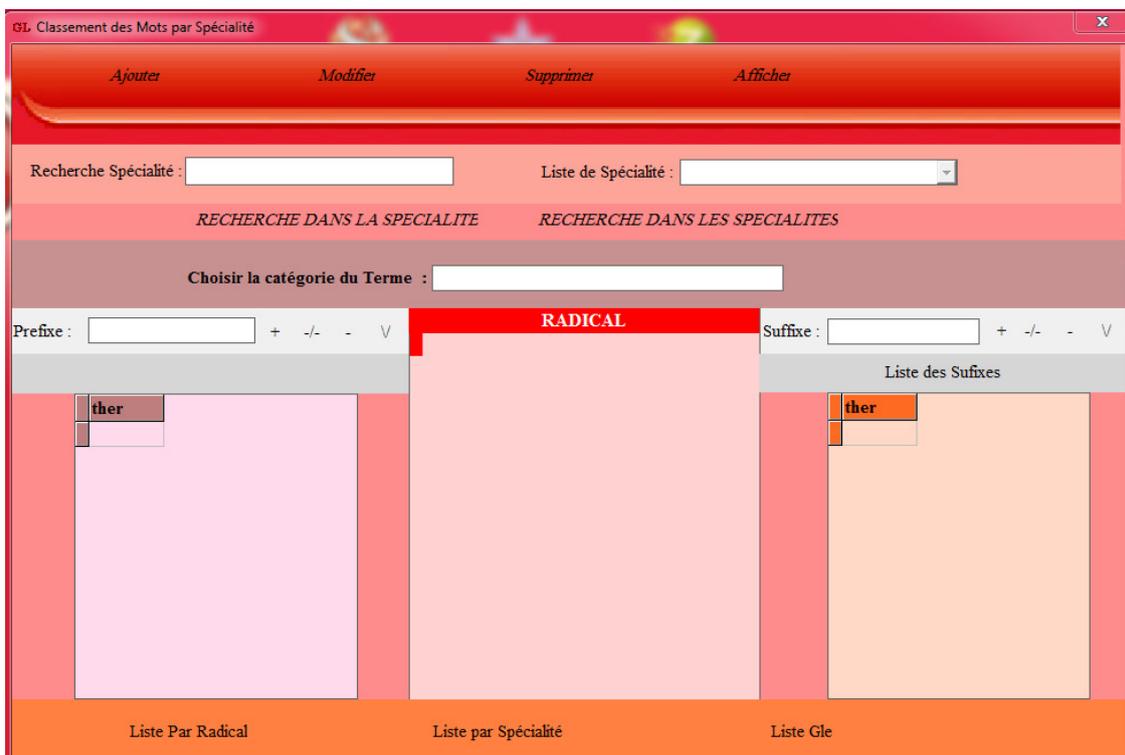
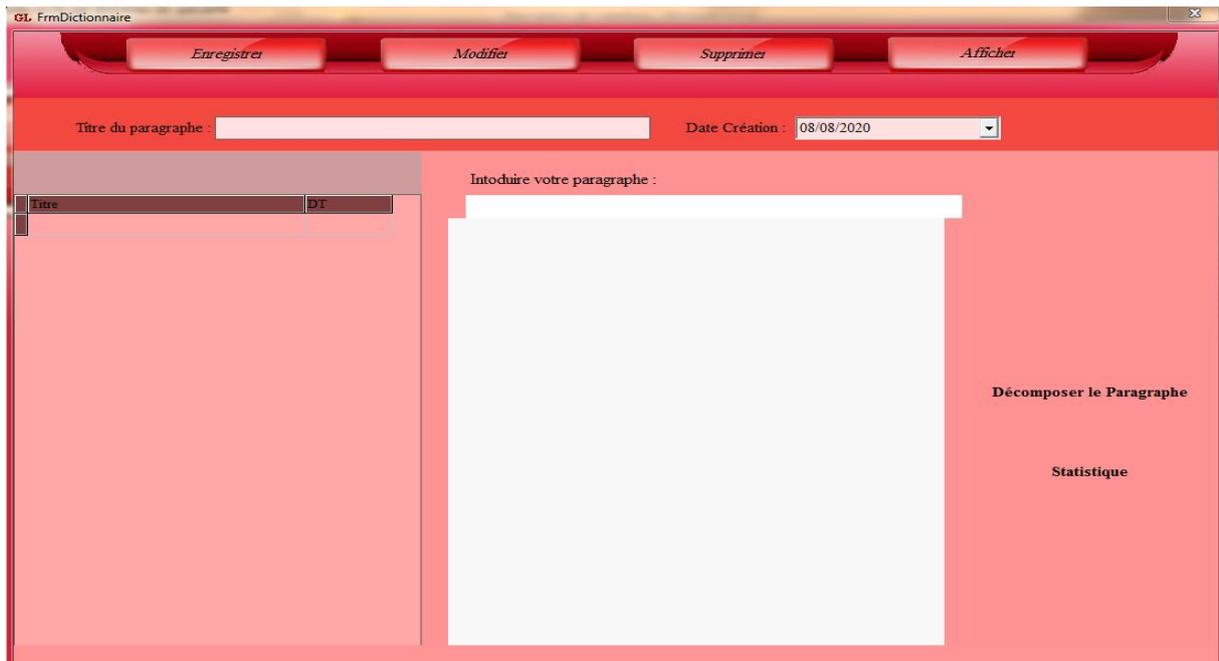


Figure 08: fenêtre de dictionnaire terminologique

- c. Onglet « création » des textes :** cet onglet permet de créer et saisir les textes. Il nous propose une fenêtre qui est composée d'une zone de texte où peut être saisi notre texte, une case pour numéroté les textes et deux boutons à droite, l'un pour décomposer le texte et l'autre pour voir le résultat « le domaine du texte ». Ainsi, quatre onglets tout en haut de la fenêtre : «enregistrer» pour valider le texte saisi, «supprimer» pour la suppression d'un texte, «modifier» pour la modification d'un texte et «afficher» pour l'affichage des textes.



**Figure 09: fenêtre de la saisie du texte**

#### **4-3- Fenêtre de résultat**

Après avoir validé le texte et cliquer sur le bouton « domaine du texte », le résultat de domaine de texte est affiché dans la fenêtre de résultat qui est composée de liste des spécialités avec le pourcentage à côté. En effet, l'application va attribuer le texte à la spécialité qui a le pourcentage le plus élevé.

The screenshot shows a window titled 'Fiche des Statistique' with a red header containing the word 'STATISTIQUE'. Below the header is a table with three columns: 'Specialite', 'Pourcentage', and 'Um'. The table lists various specialties and their corresponding percentages.

Specialite	Pourcentage	Um
▶ médecine	30,00	%
biologie	13,33	%
informatique	13,33	%
économie	11,67	%
mathématique	6,67	%
psychologie	6,67	%
physique	6,67	%
sport	5,00	%
électricité	3,33	%
agriculture	1,67	%
chimie	1,67	%

Figure 10: fenêtre du résultat

Exemple d'identification du domaine d'un texte avec cette application :

The screenshot shows the 'FrmDictionnaire' application interface. At the top, there are buttons for 'Enregistrer', 'Modifier', 'Supprimer', and 'Afficher'. Below these, there are input fields for 'Titre du paragraphe' (containing 'texte 02') and 'Date Création' (containing '08/08/2020'). The main area is divided into two sections: a list of entries on the left and a text input area on the right. The list has columns for 'Titre' and 'DT'. The text input area contains a paragraph of text about antiviral medications. To the right of the text input area, there are two buttons: 'Décomposer le Paragraphe' and 'Statistique'. An arrow points from the 'Statistique' button towards the text input area.

Titre	DT
texte 01	22/07/2020
▶ texte 02	25/07/2020
texte 02	29/07/2020
texte 02	22/07/2020
texte 02	24/07/2020
texte 02	22/07/2020
texte 02	24/07/2020
texte 02	24/07/2020
texte 02	29/07/2020
texte 02	25/07/2020
texte 02	29/07/2020
texte 02	25/07/2020
texte 02	24/07/2020
texte 02	29/07/2020
texte 02	02/08/2020

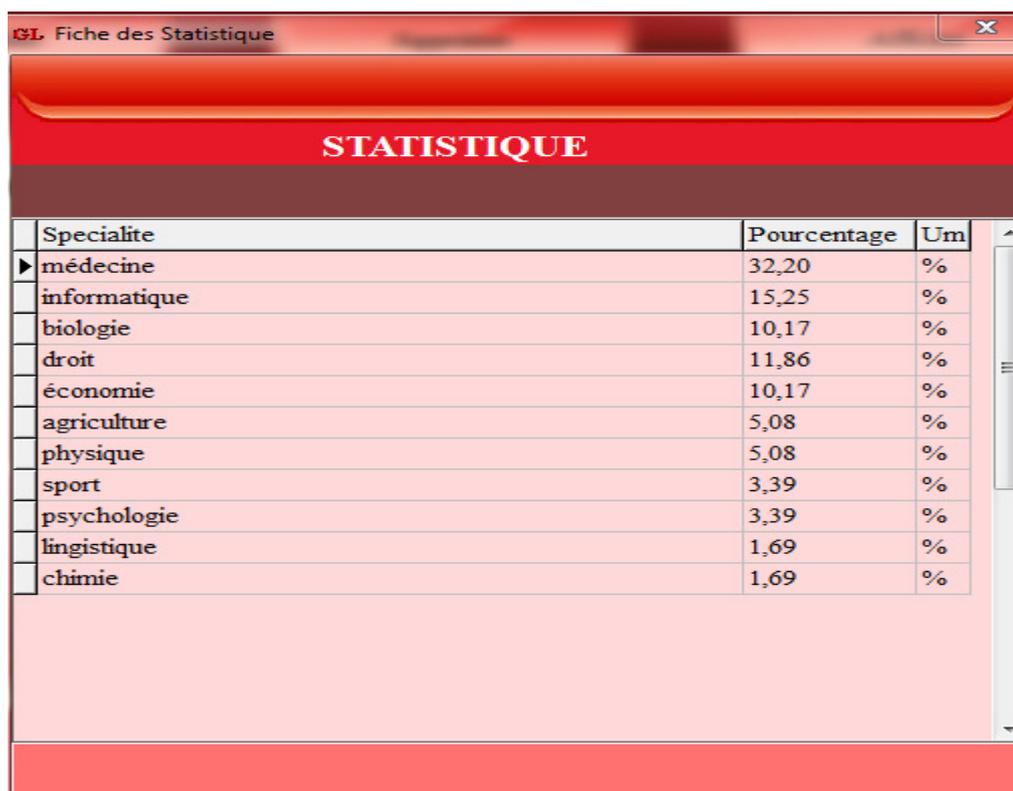
Introduire votre paragraphe :

Les médicaments antiviraux, administrés 1 à 2 jours après le début des symptômes, entraînent une diminution de la fièvre, de la sévérité des symptômes et du retour à une activité normale. Le traitement par des antiviraux est en général recommandé chez le patient à haut risque (dont tous les patients hospitalisés) qui développe des symptômes de type grippaux ; cette recommandation est basée sur les données suggérant que le traitement précoce peut éviter les complications chez ces patients. Les médicaments de la grippe comprennent les suivants : Oseltamivir, zanamivir et péramivir (inhibiteur de la neuraminidase) - Baloxavir (inhibiteur de l'endonucléase, nouveau). Les inhibiteurs de la neuraminidase perturbent la libération du virus influenza par les cellules infectées et bloquent ainsi la propagation de l'infection. L'inhibiteur de l'endonucléase baloxavir interfère avec la réplication virale en bloquant la transcription de l'ARN viral. Il est actif contre la grippe A et B et peut être une nouvelle option de traitement importante en cas de résistance aux inhibiteurs de la neuraminidase.

Décomposer le Paragraphe

Statistique

Figure 11 : saisie du texte à identifier



Specialite	Pourcentage	Um
▶ médecine	32,20	%
informatique	15,25	%
biologie	10,17	%
droit	11,86	%
économie	10,17	%
agriculture	5,08	%
physique	5,08	%
sport	3,39	%
psychologie	3,39	%
linguistique	1,69	%
chimie	1,69	%

Figure 12: résultat d'identification du domaine

## 5- Conclusion partielle

Dans ce chapitre, nous avons analysé les caractéristiques des textes de spécialités linguistiquement (caractéristiques morphosyntaxiques, caractéristiques morphologiques, caractéristiques sémantiques) et nous avons mis en place, à l'aide du langage de programmation Java, une application capable d'identifier le domaine auquel se rapporte chaque texte de spécialité qui lui a été inséré (médecine, économie, informatique, chimie et physique) nous avons mis également en place un dictionnaire terminologique inter-domaine dans lequel nous mettons les termes des textes de spécialité à côté les étiquettes domaines en se basant sur le grand dictionnaire terminologique(GDT).

# **Conclusion générale**

### Conclusion générale

Les langues de spécialité correspondent à une situation de communication dans un domaine d'activité précis. Elles sont composées des textes spécialisés qui contiennent des termes et un vocabulaire spécifiques propre à un domaine particulier.

L'objet de ce mémoire de recherche a porté sur la conception d'un catégoriseur automatique des textes écrits par domaine de spécialité. Notre réflexion s'est centrée d'emblée sur ces questions suivantes : est-il possible de s'appuyer sur un outil informatique pour catégoriser des textes en fonction de leurs domaines de spécialité ? Sur quel matériel linguistique pourrions-nous nous baser pour catégoriser des textes de spécialité ? Y-a-t-il une terminologie commune à plusieurs domaines de spécialité ?

Nous avons choisi de construire notre corpus à partir de (53) textes appartenant à plusieurs domaines de spécialité (médecine, économie, informatique, physique, chimie) et tirés de diverses sites internet et les analyser ensuite linguistiquement.

Cette analyse montre que les textes de spécialité sont le reflet d'une situation de communication particulière qui transmet des connaissances bien déterminées dans un domaine précis et définis et qui ont des caractéristiques linguistiques particulières qui les différencient des autres types de textes ou des textes générales. En effet, les TS se révèlent grâce à un thème propre à un domaine donné (médecine, économique, chimie...), avec une terminologie spécifique, l'emploi de formules, des symboles et d'abréviations, la condensation syntaxique, l'impersonnalité des phrases et leur complexité de structure.

Pour identifier les domaines des textes de spécialité, nous avons réussi à mettre en place un système catégoriseur automatique des textes écrits selon leurs domaines de spécialité. Nous avons choisi le logiciel java pour développer cette application basée sur des données linguistiques.

Au cœur de ces données linguistiques, nous avons élaboré un dictionnaire terminologique qui contient des termes issus des textes de spécialité de notre corpus avec la forme canonique, en ajoutant les étiquettes domaines à côté.

## Conclusion générale

---

Cette application détecte les domaines de spécialité des textes qui lui ont été présentés. Elle repose sur trois étapes essentielles pour donner le résultat :

- la première étape est l'étiquetage des textes, où le système consulte automatiquement le dictionnaire terminologique qui permet de repérer les termes des textes de spécialité qui sont insérés et détecter leurs domaines auxquels ils appartiennent.
- Le calcul des fréquences de termes est la deuxième l'étape, Le système ici a calculé les fréquences d'apparition des termes dans les textes après avoir fait l'étiquetage des textes, il a recensé combien de termes appartiennent au domaine 1, domaine 2, domaine 3...
- La troisième et dernière étape est le résultat final, l'application donne le domaine auquel appartiennent les textes insérés. Le domaine qui a le pourcentage le plus élevé est alors détecté par le système comme domaine de texte.

Cette recherche novatrice que nous venons de réaliser dans le domaine des langues de spécialité et du traitement automatique des langues ouvre des perspectives dans d'autres domaines tel que la didactique. En outre, le catégoriseur que nous avons conçu peut faire l'objet d'outil et de support dans l'enseignement des LS dans la classe du FLE; il peut être, par ailleurs, un facilitateur pour l'enseignant et en même temps une aide pour l'apprenant.

# **Annexes**

## **Texte 01**

L'insuffisance hépatique est une dégradation sévère de la fonction hépatique, elle est due à un trouble ou à une substance qui endommage le foie. La plupart des personnes présentent un ictère (jaunissement de la peau et des yeux) se sentent fatiguées et faibles, et perdent l'appétit. Les autres symptômes sont une accumulation de liquide dans la cavité abdominale (ascite) et une tendance aux ecchymoses et aux saignements.

Le médecin pose, en général, le diagnostic d'insuffisance hépatique en se basant sur les symptômes et les résultats de l'examen clinique et des analyses de sang. Son traitement consiste généralement à contrôler la consommation de protéines, limiter le sodium dans l'alimentation, éviter complètement l'alcool et traiter la cause ; néanmoins, une transplantation hépatique est parfois nécessaire.

L'insuffisance hépatique peut être la conséquence de tout type de maladie hépatique, telle qu'une hépatite virale (le plus souvent l'hépatite B ou C), une cirrhose et une lésion hépatique due à l'alcool ou à des médicaments comme le paracétamol.

Cet état ne se produit que lorsqu'une grande partie du foie est endommagée. L'insuffisance hépatique peut se développer rapidement en quelques jours ou semaines (aiguë) ou de façon progressive sur plusieurs mois ou années (chronique).

**[www.msmanuals.com](http://www.msmanuals.com)**

## **Texte 02**

La chirurgie est le plus souvent réalisée en premier et peut être suivie d'une chimiothérapie et/ou d'une radiothérapie et/ou d'une hormonothérapie. Elle a pour objectif d'enlever les tissus atteints par les cellules cancéreuses.

Elle est parfois précédée d'un traitement médical, dit alors « néoadjuvant » (une chimiothérapie ou une hormonothérapie) qui peut permettre de réduire la taille de la tumeur afin de faciliter l'intervention.

Deux types d'interventions chirurgicales peuvent être pratiqués : une chirurgie mammaire conservatrice, appelée tumorectomie ou segmentectomie ou une chirurgie mammaire non conservatrice, appelée mastectomie.

La chirurgie conservatrice (ou tumorectomie ou segmentectomie) consiste à retirer la tumeur et une petite quantité des tissus qui l'entourent de façon à conserver la plus grande partie de votre sein. Elle est privilégiée aussi souvent que possible, en concertation avec vous. Elle est toujours complétée d'une radiothérapie.

La chirurgie non conservatrice (ou mastectomie) consiste à retirer la totalité du sein y compris l'aréole et le mamelon. Dans ce cas, différentes techniques de reconstruction du sein peuvent vous être proposées. Dans certains cas de carcinomes canauxaires in situ (CCIS), l'ablation de quelques ganglions est nécessaires (ganglion sentinelle).

En cas de tumeur infiltrante, il est également nécessaire de retirer un ou plusieurs ganglions lymphatiques axillaires (exérèse du ganglion sentinelle ou curage ganglionnaire). Dans ce cas, l'objectif est de préciser si la tumeur s'étend au-delà du sein, de déterminer si un traitement complémentaire par chimiothérapie ou hormonothérapie est nécessaire ou si une radiothérapie des ganglions est justifiée. Le curage ganglionnaire, quand il est nécessaire, permet de réduire le risque de récurrence.

**[www.e-cancer.fr](http://www.e-cancer.fr)**

### **Texte 03**

Le calcium est l'un des électrolytes de l'organisme, qui sont des minéraux portant une charge électrique lorsqu'ils sont dissous dans les liquides corporels tels que le sang (mais la majorité du calcium de l'organisme est sous forme non chargée). Bien que la majeure partie du calcium de l'organisme soit stockée dans les os, une partie circule dans le sang. Environ 40 % du calcium présent dans le sang est fixé (lié) à des protéines, principalement l'albumine. Le calcium lié aux protéines est une forme de réserve de calcium pour les cellules, mais il n'a pas de fonction active dans l'organisme. Seul le calcium non lié a un effet sur les fonctions de l'organisme. Ainsi, l'hypocalcémie n'entraîne des problèmes que lorsque le taux de calcium non lié est bas. Le calcium non lié a une charge électrique (ionique), on l'appelle donc aussi

calcium ionisé. Le taux de calcium dans le sang peut être modérément bas sans que cela ne produise de symptômes. Si les taux de calcium sont bas pendant une période prolongée, cela peut induire une peau sèche et squameuse, des ongles cassants et des cheveux ternes. Les crampes des muscles du dos et des jambes sont fréquentes. Avec le temps, l'hypocalcémie peut affecter le cerveau et induire des symptômes neurologiques ou psychologiques, tels qu'une confusion, des pertes de mémoire, un délirium, une dépression et des hallucinations. Ces symptômes disparaissent si le taux de calcium est rétabli.

Un taux de calcium extrêmement bas peut provoquer des fourmillements (souvent au niveau des lèvres, de la langue, des doigts et des pieds), des douleurs musculaires, des spasmes musculaires de la gorge (susceptibles d'induire des difficultés respiratoires), une raideur et des spasmes musculaires (tétanie), des crises et des troubles du rythme cardiaque.

[www.merckmanuals.com](http://www.merckmanuals.com)

## Texte 04

Le potassium est principalement concentré dans l'eau des cellules, à l'inverse du sodium, qui lui, est plus largement répandu au niveau de l'eau extracellulaire au sein de laquelle se dissolvent les nutriments et où se produisent de multiples réactions chimiques. C'est la raison pour laquelle le sodium et le potassium sont intrinsèquement liés puisque ce sont les échanges entre ces deux minéraux qui régissent l'hydratation cellulaire, même si chacun possède indépendamment du message nerveux, dans la sécrétion de l'acide gastrique, ou encore dans la régulation rénale des ions sodium et chlore.

Le corps humain renferme entre 100 et 50 g de potassium, dont une large majorité est contenue dans les muscles, l'apport en potassium provient de l'alimentation humaine puisqu'il n'est absolument pas synthétisé par l'organisme. On retrouve ce minéral dans de très nombreux fruits et légumes : avocats, bananes, agrumes, pommes de terre, pois, salades, tomates, brocolis, épinards. En outre, les produits laitiers constituent également une source de potassium importante, tout comme les céréales complètes, les poissons gras ou encore les noix.

Les déséquilibres alimentaires peuvent être à l'origine d'une déficience en potassium. Dans certains cas, cette carence doit faire l'objet d'une supplémentation pour prévenir l'apparition de

troubles apparentés qui peuvent être plus au moins graves : troubles digestifs : vomissements, nausées... ; trouble de la motricité : crampes, courbatures, hypertension artérielle ; asthénie, dépression...

Sur le long terme, une telle déficience peut engendrer des lésions nerveuses et cardiaques pouvant conduire au décès.

**[www.femmeactuelle.fr](http://www.femmeactuelle.fr)**

## **Texte 05**

La malaria, appelée aussi paludisme, fait partie des maladies les plus meurtrières dans le monde et touche chaque année entre 400 et 500 millions de personnes. Elle survient dans la plupart des régions tropicales et subtropicales de la planète. L'Afrique est le continent le plus touché par la malaria, mais de nombreuses autres régions sont concernées. Au total, la malaria est endémique dans une certaine de pays et elle concerne environ 40 pour cent de la population mondiale.

Cette maladie est un facteur important d'anémie chez les enfants affectant gravement leur développement. Elle a, par ailleurs, de graves effets économiques en Afrique car elle ralentit la croissance économique et perpétue le cercle vicieux de la pauvreté.

La malaria existe sous 4 formes, dont le plasmodium Falciparum qui est à l'origine des cas sévères et mortels. La transmission survient par la pique d'un moustique appelé l'anophèle présent dans certaines parties du globe. Seule la femelle pique entre le coucher du soleil et l'aube, afin de nourrir ses œufs avec les protéines du sang humain.

Les premiers signes d'alerte apparaissent après une semaine voire plusieurs mois. En se multipliant dans l'organisme, le parasite provoque une fièvre élevée, des céphalées et des douleurs articulaires, comme lors d'une forte grippe. Sans traitement efficace, l'infection évolue souvent vers une forme sévère de malaria qui peut mener au coma et la mort.

Lorsqu'un anophèle infecté par le parasite pique un être humain, le parasite passe dans le système sanguin de la personne, où il se multiplie et peut entraîner la maladie, voire la mort. Inversement, le parasite passe de l'homme infecté à l'insecte sain et le cycle continue.

La malaria est traitée grâce à l'Artémisine ; c'est une substance extraite d'une plante chinoise, l'Armoise, considérée par l'OMS comme étant très efficace.

La prévention du paludisme consiste en une préservation contre les piqûres de moustiques (moustiquaire, port de vêtements longs, insecticides, répulsifs...) et un assainissement des zones humides.

[www.centre.chl.lu](http://www.centre.chl.lu)

## Texte 06

Les spécialistes distinguent plusieurs types de nodules thyroïdiens

- Le nodule thyroïdien froid ou inactif est bénin dans 95 % des cas et ne sécrète pas d'hormones thyroïdiennes. La fonction thyroïdienne reste alors normale.
- Le nodule thyroïdien chaud ou actif sécrète de grandes quantités d'hormones thyroïdiennes à l'origine d'une hyperthyroïdie. Il est bénin.
- Le kyste est une sorte de poche remplie de liquide ou de sang (kyste hémorragique). Il est bénin et ne produit pas d'hormones thyroïdiennes. Mais son volume peut rapidement augmenter et entraîner des douleurs.
- Les nodules thyroïdiens accompagnant certaines maladies de la thyroïde, telles que :
  - La thyroïdite subaiguë (augmentation du volume de la thyroïde avec douleurs et fièvre) ;
  - Les thyroïdites lymphocytaires (maladies auto-immunes), comme la thyroïdite d'Hashimoto, qui peuvent entraîner une hypothyroïdie.

Les nodules thyroïdiens peuvent être isolés ou non. La présence simultanée de plusieurs nodules est désignée par le terme de goitre multinodulaire.

[www.sante-sur-le-net.com](http://www.sante-sur-le-net.com)

## Texte 07

Lorsque la polyglobulie est importante, c'est-à-dire avec un hémocrite très élevé, le premier traitement consiste à pratiquer des saignées. Une saignée consiste à prélever entre 300 et 450 ml de sang (selon le poids et la taille). Les saignées permettent de diminuer rapidement l'hémocrite et de réduire le risque de complications. Parallèlement, le traitement au long cours de la maladie de Vaquez repose sur la prise d'un traitement oral dit « myélofreinateur » qui réduit la production des globules rouges et des plaquettes par la moelle osseuse. Il existe

deux médicaments principaux, cytostatiques, inhibiteurs de la synthèse de l'ADN, dont l'un est un alkylant. L'un ou l'autre est prescrit à doses variables selon chaque patient. En général, le traitement est pris « à vie ». On lui associe souvent de l'aspirine à faible dose en continu, afin de limiter le risque thrombotique.

[www.hemato.chu-limoges.fr](http://www.hemato.chu-limoges.fr)

## **Texte 08**

Les hormones thyroïdiennes de synthèses agissent comme les hormones naturelles, via des récepteurs nucléaires. La lévothyroxine (T4) est la plus utilisée en traitement substitutif. Tout comme l'hormone naturelle, elle sera transformée en T3 (forme active) par les désiodases. La liothyronine (T3) peut être utilisée si un effet rapide ou transitoire est nécessaire. Bien que l'association T3 + T4 puisse paraître plus proche de la sécrétion endogène d'hormones thyroïdiennes, le ratio T3/T4 de la préparation commerciale disponible est environ 2 fois supérieur au ratio physiologique et est peu utilisé en pratique. Les hormones thyroïdiennes utilisées en thérapeutique ont les mêmes effets métaboliques, cardiovasculaires et sur le développement que les hormones naturelles et permettent de corriger les symptômes de l'hypothyroïdie. Elles peuvent également être utiles en cas de tumeur de la thyroïde, en freinant la TSH qui stimule les cellules thyroïdiennes.

[www.Pharmacomedicale.org](http://www.Pharmacomedicale.org)

## **Texte 09**

L'examen sanguin de la créatinine consiste en une prise de sang veineux, en général au niveau du pli du coude. Le niveau sanguin de créatinine étant lié à l'activité musculaire, il est recommandé de cesser toute activité physique intense au cours des 48 heures précédant l'examen. De plus, parce que la viande rouge est riche en créatine, il est recommandé de limiter sa consommation à moins de 200 grammes 24 heures avant l'examen, voire de ne pas en consommer du tout. La consommation de boissons ou de produits diurétiques est à proscrire la veille et le jour de l'examen. La clairance urinaire de créatinine, qui permet de mesurer la quantité de créatinine retirée du sang par les reins, est estimée à partir de la totalité du volume urinaire produit en 24 heures. La concentration normale de créatinine dans le sang est comprise entre 6 et 12 mg/l (milligrammes par litre) chez l'homme et entre 4 et 10 mg/l chez la femme. Ces valeurs peuvent varier selon les sources. Un calcul normal de la clairance urinaire de

créatinine est compris entre 107 et 139 ml/min (millilitres par minutes) chez l'homme de moins de 40 ans, et entre 87 et 107 ml/min chez la femme de moins de 40 ans.

[www.passeportsante.net](http://www.passeportsante.net)

## Texte 10

La drépanocytose (parfois appelée anémie falciforme) est un trouble héréditaire qui entraîne la production d'un type anormal d'hémoglobine appelé hémoglobine S (HbS ou HgbS). Les tests de dépistage de la drépanocytose déterminent la présence et la quantité relative d'hémoglobine S dans un échantillon de sang ou bien ils détecteront les mutations dans les gènes qui produisent l'hémoglobine, pour aider à diagnostiquer la drépanocytose et/ou identifier les porteurs du trait drépanocytaire. L'hémoglobine est la protéine dans les globules rouges qui se lie à l'oxygène dans les poumons pour le transporter dans les tissus partout dans le corps. Règle générale, l'hémoglobine A (HbA – l'hémoglobine *adulte*) forme la majeure partie de l'hémoglobine présente dans les globules rouges normaux des adultes, les petites quantités restantes étant constituées d'hémoglobine A2 et d'hémoglobine F. Avant la naissance, les bébés produisent, en temps normal, de grandes quantités d'hémoglobine F (HbF – hémoglobine *fœtale*), qui est ensuite remplacée par l'hémoglobine A, à titre d'hémoglobine prédominante, peu de temps après la naissance.

[www.msmanuals.cn](http://www.msmanuals.cn)

## Texte 11

L'augmentation de la production des urates peut être provoquée par un turnover augmenté des nucléoprotéines dans les pathologies hématologiques (p. ex., lymphome, leucémie, anémie hémolytique) et dans des pathologies avec augmentation du taux de prolifération cellulaire et de mort cellulaire (p. ex., psoriasis sévère, traitements cytotoxiques du cancer, radiothérapie). L'augmentation de la production d'acide urique peut également être secondaire à une anomalie héréditaire primitive et à l'obésité car la production d'acide urique est corrélée avec la surface corporelle. Dans la plupart des cas, la cause de la production excessive d'urate en est inconnue mais certains cas sont liés à des anomalies enzymatiques; le déficit en hypoxanthine-guanine phosphoribosyl-transférase (la déficience complète est le syndrome de Lesch-Nyhan) est une cause possible, de même que l'hyperactivité de la phosphoribosyl-pyrophosphate synthétase.

[www.msmanuals.com](http://www.msmanuals.com)

## Texte 12

Les options thérapeutiques de la thrombose de la veine rénale comprennent l'anticoagulation par l'héparine, une thrombolyse et thrombectomie percutanée ou chirurgicale. L'anticoagulation à long terme avec une héparine de faible poids moléculaire ou de la warfarine orale doit être débutée immédiatement si aucune intervention invasive n'est prévue. L'anticoagulation minimise le risque de nouvelles thromboses, favorise la recanalisation des vaisseaux avec les caillots existants et améliore la fonction rénale. Le traitement anticoagulant doit être poursuivi pendant au moins 6 à 12 mois ou, si un trouble d'hypercoagulabilité (p. ex., syndrome néphrotique persistant) est présent, indéfiniment. L'utilisation d'un cathéter percutané pour thrombectomie ou thrombolyse est actuellement recommandée. La thrombectomie chirurgicale est rarement pratiquée mais elle doit être envisagée chez les patients qui ont une thrombose veineuse aiguë rénale bilatérale et des lésions rénales aiguës qui ne peuvent pas être traitées par une thrombectomie et/ou une thrombolyse par cathéter percutanée. La néphrectomie est effectuée uniquement si l'infarctus est total (dans certains cas) ou si la pathologie causale la rend nécessaire.

**Www.msmanuals.com**

## Texte 13

Les médicaments antiviraux, administrés 1 à 2 jours après le début des symptômes, entraînent une diminution de la fièvre, de la sévérité des symptômes et du retour à une activité normale. Le traitement par des antiviraux est en général recommandé chez le patient à haut risque (dont tous les patients hospitalisés) qui développe des symptômes de type grippaux; cette recommandation est basée sur les données suggérant que le traitement précoce peut éviter les complications chez ces patients. Les médicaments de la grippe comprennent les suivants:

- Oseltamivir, zanamivir et péramivir (inhibiteur de la neuraminidase)
- Baloxavir (inhibiteur de l'endonucléase, nouveau)

Les inhibiteurs de la neuraminidase perturbent la libération du virus influenza par les cellules infectées et bloquent ainsi la propagation de l'infection. L'inhibiteur de l'endonucléase baloxavir interfère avec la réplication virale en bloquant la transcription de l'ARN viral. Il est actif contre la grippe A et B et peut être une nouvelle option de traitement importante en cas de résistance aux inhibiteurs de la neuraminidase.

**Www.msmanuals.com**

## Texte 14

L'examen des frottis périphériques est important en cas de thrombopénie car les numérations plaquettaires automatisées montrent parfois une pseudothrombopénie due à l'agrégation plaquettaire provoquée par une réaction avec le réactif EDTA présent dans certains tubes de prélèvement sanguin. En outre, des schistocytes peuvent être observés, ce qui peut être en faveur d'un purpura thrombotique thrombopénique, d'un syndrome hémolytique urémique ou d'une coagulation intravasculaire disséminée (CIVD, Caractéristiques des éléments sanguins dans les troubles thrombopéniques). Une ponction de moelle osseuse (myélogramme) est indiquée si le frottis montre des anomalies autres que la thrombopénie, telles que des globules rouges nucléés ou des globules blancs anormaux ou immatures. Le myélogramme révèle le nombre et l'aspect des mégacaryocytes et représente le test diagnostique de certitude pour nombre de maladies causes d'insuffisance médullaire. Cependant, un nombre et un aspect normal des mégacaryocytes ne signifient pas toujours que la production de plaquettes est normale. Par exemple, en cas de thrombopénie immunitaire, la production plaquettaire peut être diminuée en dépit de l'aspect normal et de l'augmentation du nombre de mégacaryocytes.

[www.msmanuals.com](http://www.msmanuals.com)

## Texte 15

Le traitement du rachitisme hypophosphatémique repose sur le phosphore neutre en solution ou comprimés. La dose initiale chez l'enfant est de 10 mg/kg (basée en phosphore élément) po qid. La supplémentation en phosphates abaisse les concentrations du Ca ionisé et inhibe en outre la conversion du calcitriol, induisant une hyperparathyroïdie secondaire et une exacerbation des pertes urinaires de phosphates. Ainsi, la vitamine D est administrée sous forme de calcitriol, initialement de 5 à 10 ng/kg po bid. Cependant, ce n'est pas le cas dans le rachitisme hypophosphatémique héréditaire avec hypercalciurie, où les taux de 1,25-dihydroxyvitamine D3 sont élevés et où l'utilisation du calcitriol peut être néfaste. La dose de phosphate doit parfois être augmentée pour assurer la croissance osseuse ou soulager les douleurs osseuses. Une diarrhée peut limiter la dose orale de phosphate. Une augmentation de la phosphatémie, une diminution de la phosphatase alcaline, la guérison des lésions rachitiques et l'amélioration de la vitesse de croissance se produisent. Une hypercalcémie, une hypercalciurie et une

néphrocalcinose avec diminution de la fonction rénale peuvent compliquer le traitement. Les patients suivant ce traitement doivent bénéficier d'évaluations de contrôle fréquentes.

**www.msmanuals.com**

## **Texte 16**

Pour le vaccin anti-grippal inactivé, les effets indésirables se limitent habituellement à une légère douleur au point d'injection. La fièvre, les myalgies et d'autres effets systémiques sont relativement rares; cependant, les sujets qui ont été vaccinés peuvent penser à tort que le vaccin provoque la grippe. Ces réactions ne constituent pas une contre-indication aux vaccinations futures, lesquelles doivent être encouragées. Les flacons multidoses contiennent du thimérosal, un agent de conservation à base de mercure. Les préoccupations du public au sujet d'un possible lien entre le thimérosal et l'autisme se sont révélées infondées; cependant des flacons à dose unique, exempts de thimérosal, sont disponibles. Les effets indésirables associés au vaccin antigrippal vivant atténué sont modérés; une rhinorrhée est le plus fréquent et un léger wheezing peut survenir.

**www.msmanuals.com**

## **Texte 17**

Internet est un réseau informatique qui relie des ordinateurs du monde entier entre eux et qui leur permet d'échanger des informations. Les données sont transmises par l'intermédiaire de lignes téléphoniques, de câbles ou de satellites. Pour communiquer entre eux, les ordinateurs connectés à internet utilisent un langage commun nommé protocole et sont équipés de logiciels(ou programmes) permettant l'échange de données.

Internet est issu du réseau Arpanet, qui a été conçu en 1969 par l'Agence pour les projets de recherches avancée (ARPA, Advanced Research Project Agency) pour le département américain de la Défense. Réservé à l'origine aux militaires, le réseau arpanet s'est étendu progressivement aux universités et aux administrations américaines.

L'internet relie des millions de personnes à travers le monde. Internet n'appartient à personne et personne ne le contrôle .Les utilisateurs d'Internet (appelés internautes) ont accès à de nombreux services, dont World Wide et le courrier électronique .Le terme «World Wide Web» (souvent abrégé en WWW ou Web) signifie «toile d'araignée mondiale» en français. C'est un gigantesque ensemble de pages électroniques dites pages Web, reliées entre elles par des liens hypertextes. Il suffit de cliquer sur un lien pour être dirigé vers une nouvelle page.

Pour accéder à des pages Web, on utilise un navigateur (ou browser en anglais). Un navigateur est un logiciel qui permet notamment de consulter des moteurs de recherche. Ces moteurs sont très utiles pour trouver une information. Le Web ne sert pas seulement à trouver des informations. Il permet entre autres de récupérer (ou télécharger) des fichiers électroniques, d'acheter ou de vendre des objets... Par ailleurs, les amateurs de jeux vidéo peuvent, grâce à internet, affronter de nombreux autres joueurs dans le monde entier.

Le courrier «électronique permet d'envoyer un message électronique (aussi appelé e-mail) à un ou plusieurs internautes .pour cela, il suffit de connaître l'adresse électronique (ou adresse e-mail) de celui à qui on envoie le message et d'être équipé d'un logiciel de messagerie.

Une nouvelle technologie, appelée wifi (contraction de wireless-fidelity, signifiant «qualité sans fil»), permet également d'accéder à Internet .toutes les données sont transmises par ondes radio d'un ordinateur à l'autre, sans modem ni ligne téléphonique .ce type de connexion est très rapide, mais ne fonctionne que dans certaines zones.

**[www.thomasta.com](http://www.thomasta.com)**

## **Texte 18**

Une mémoire ECC (mémoire à code correcteur d'erreur) est une mémoire contenant des bits supplémentaires servant à détecter et à corriger une éventuelle erreur ou altération de l'information qu'elle contient (par exemple lors d'un transfert). La technique la plus simple est celle du bit de parité (Parity check code), selon cette technique l'information est codée sur n bits et la mémoire contient un n+1 ème bit qui indique si le nombre de bits codant l'information contenue dans les n bits est pair (bit=0) ou impair (bit=1). C'est un code détecteur d'erreur. Une altération de deux bits (ou d'un nombre pair de bits) ne modifiant pas la parité du décompte ne

sera donc pas décelée par ce code : supposons que l'information 10010 soit altérée en 01100 (le bit de parité ne change pas car le nombre de 1 de l'information altérée est toujours pair, il y en a toujours 2 !). Ce code est simple et peu coûteux, il est en fait un cas particulier simple de codage linéaire systématique inventés par les spécialistes du domaine.

[www.rmdiscala.developpez.com](http://www.rmdiscala.developpez.com)

## **Texte 19**

Un ordinateur est constitué de matériel (hardware) et de logiciel (software). Cet ensemble est à la disposition de un ou plusieurs utilisateurs. Il est donc nécessaire que quelque chose dans l'ordinateur permette la communication entre l'homme et la machine. Cette entité doit assurer une grande souplesse dans l'interface et doit permettre d'accéder à toutes les fonctionnalités de la machine. Cette entité douée d'une certaine intelligence de communication se dénomme « la machine virtuelle ». Elle est la réunion du matériel et du système d'exploitation (que nous noterons OS par la suite pour Operating System). Le système d'exploitation d'un ordinateur est chargé d'assurer les fonctionnalités de communication et d'interface avec l'utilisateur. Un OS est un logiciel dont le grand domaine d'intervention est la gestion de toutes les ressources de l'ordinateur : mémoires, fichiers, périphériques, entrée-sortie, interruptions, synchronisation...

[Www.rmdiscala.developpez.com](http://www.rmdiscala.developpez.com)

## **Texte 20**

Un tableur est un logiciel permettant de manipuler des données numériques et d'effectuer des calculs de manière automatique sur des nombres stockés dans des tableaux. Il permet également de créer facilement des représentations graphiques à partir des données saisies (diagrammes en barres, courbes, diagrammes circulaires, etc.). Microsoft Excel est le tableur de Microsoft Office. Un document Excel est appelé classeur. Il est composé d'une ou plusieurs feuilles de calcul. À l'ouverture d'un classeur, Excel présente une feuille de calcul, ainsi que des onglets en bas à gauche de l'écran permettant l'accès à d'autres feuilles de calcul.

[Https://elearning.univ-annaba.dz](https://elearning.univ-annaba.dz)

## Texte 21

Un classeur est constitué de feuilles, ayant chacune un onglet. Cliquer sur un onglet permet d'afficher la feuille correspondante. L'onglet sur fond blanc désigne la feuille active. Par défaut, le classeur comporte 3 feuilles, donc 3 onglets. On peut en rajouter, le nombre maximal étant 255 feuilles. A gauche des onglets, des boutons fléchés de défilement permettent d'accéder à l'onglet souhaité, dès lors qu'ils sont nombreux et que certains sont masqués. Pour afficher la liste de toutes les feuilles du classeur : clic droit sur l'un de ces boutons. Cliquer ensuite sur une feuille permet de l'afficher. Suite aux onglets correspondant aux feuilles existantes, un onglet d'info-bulle « Insérer une feuille de calcul » permet d'insérer une nouvelle feuille. A droite des onglets, une petite barre verticale permet, par cliqué-glissé, d'afficher plus ou moins d'onglets (le pointeur se transforme en double-fl)

[www.coursbardon-microsoftoffice.fr](http://www.coursbardon-microsoftoffice.fr)

## Texte 22

On peut trier les données en utilisant un ou plusieurs critères (appelés aussi « clés »), chacun correspondant à un en-tête de colonne. Après application d'un tri, une flèche apparaît dans la case du menu déroulant de la colonne concernée. Dirigée vers le haut, elle indique un tri croissant ; vers le bas, un tri décroissant. On peut effectuer un tri en fonction des valeurs, également en fonction de la couleur de cellule ou de police, ou de l'icône de la cellule. En fonction des valeurs, un tri croissant peut être : de A à Z (textes), du plus petit au plus grand (nombres), du plus ancien au plus récent (dates). Quel que soit l'ordre (croissant ou décroissant), les cellules vides sont placées en dernier. En fonction des couleurs ou des icônes, on peut placer les cellules en haut ou en bas.

[www.coursbardon-microsoftoffice.fr](http://www.coursbardon-microsoftoffice.fr)

## Texte 23

Pour créer un tableau croisé dynamique, sélectionnez d'abord une cellule quelconque de la plage des colonnes de données. Puis affichez la fenêtre « Créer un tableau croisé dynamique » : sous l'onglet « Insertion », dans le groupe « Tableaux », activez le bouton « Tableau croisé dynamique »

Dans cette fenêtre, indiquez :

- L'emplacement des données à analyser : vérifiez la plage de données, modifiez-la si nécessaire (cliquez dans la zone, puis sélectionnez la plage des données à analyser).
- L'emplacement où sera créé le tableau croisé dynamique : cliquez dans la zone, puis sélectionnez la première cellule de l'emplacement prévu pour le tableau. Validez. Les « Outils de tableau croisé dynamique » se répartissent sur les deux onglets « Options » et « Création ». Sur la feuille, apparaissent un espace réservé au tableau croisé dynamique, ainsi que le volet « Liste de champs de tableau croisé dynamique » sur le côté droit.

**[www.coursbardon-microsoftoffice.fr](http://www.coursbardon-microsoftoffice.fr)**

## **Texte 24**

Le masque des diapositives contient des informations relatives au modèle : les styles de police, les tailles et les emplacements des espaces réservés, la conception de l'arrière plan et les jeux de couleurs. Il permet les modifications globales, comme le remplacement du style de police, et l'application des modifications dans toutes les diapositives de votre présentation. Le masque des diapositives convient particulièrement pour les opérations suivantes ::

- Modifier la police ou les puces
- Insérer un élément dessiné comme un logo que vous souhaitez reproduire sur plusieurs diapositives
- Modifier les emplacements, la taille et la mise en forme des espaces réservés. En règle générale, le modèle contient également un masque de titre, dans lequel vous pouvez apporter des modifications s'appliquant aux diapositives ayant une mise en page Diapositive de titre.

**<https://it-it.facebook.com>**

## **Texte 25**

Un lien hypertexte permet, lors de la présentation en Diaporama de vos diapositives, d'accéder, par un simple clic sur du texte ou un objet, à une autre diapositive dans la même présentation ou dans une autre présentation, à une adresse électronique, à une page Web ou à un fichier. Si le lien a été créé sur un texte, le texte apparaît souligné et d'une couleur spécifique : il devient hypertexte. Lorsque vous amenez le pointeur sur un lien hypertexte (texte ou objet), une

infobulle affiche un texte que vous pouvez définir en cliquant sur le bouton [Info-bulle] dans le dialogue de création du lien hypertexte.

**[www.microfer28.fr](http://www.microfer28.fr)**

## **Texte 26**

Java est un langage interprété, ce qui signifie qu'un programme compilé n'est pas directement exécutable par le système d'exploitation mais il doit être interprété par un autre programme, qu'on appelle interpréteur. Un programmeur Java écrit son code source, sous la forme de classes, dans des fichiers dont l'extension est .java. Ce code source est alors compilé par le compilateur javac en un langage appelé bytecode et enregistre le résultat dans un fichier dont l'extension est .class. Le bytecode ainsi obtenu n'est pas directement utilisable. Il doit être interprété par la machine virtuelle de Java qui transforme alors le code compilé en code machine compréhensible par le système d'exploitation. C'est la raison pour laquelle Java est un langage portable : le bytecode reste le même quelque soit l'environnement d'exécution.

**[Https://www.esen.tn](https://www.esen.tn)**

## **Texte 27**

La compilation s'effectue par la commande javac suivie d'un ou plusieurs noms de fichiers contenant le code source de classes Java. Par exemple, javac MyProg.java compile la classe MyProg dont le code source est situé dans le fichier MyProg.java. La compilation nécessite souvent la précision de certains paramètres pour s'effectuer correctement, notamment lorsque le code source fait référence à certaines classes situées dans d'autres répertoires que celui du code compilé. Il faut alors ajouter l'option -classpath suivie des répertoires (séparés par un sous Windows et sous Unix) des classes référencées.

Par exemple javac -classpath /prog/exos1:/cours MyProg.java compilera le fichier MyProg.java si celui-ci fait référence à d'autres classes situées dans les répertoires /prog/exos1 et /cours. Le résultat de cette compilation est un fichier nommé MyProg.class contenant le bytecode correspondant à la source compilée. Ce fichier est créé par défaut dans le répertoire où la compilation s'est produite. Il est cependant fortement souhaitable de ne pas mélanger les fichiers contenant le code source et ceux contenant le bytecode. Un répertoire de destination où

sera créé le fichier MyProg.class peut être précisé par l'option -d, par exemple : javac -d /prog/exos1 -classpath /cours MyProg.java

<https://www.esen.tn>

## Texte 28

Dans un programme Java, une classe de composant graphique est une classe qui hérite de java.awt.Component. Il est indispensable d'hériter de cette classe car elle implémente de nombreuses méthodes nécessaires à l'affichage du composant, en renvoyant la dimension du composant, sa police de caractères, sa couleur de fond, etc. La seconde classe essentielle pour l'écriture d'interfaces graphiques est la classe Container.

Un container est un composant (qui hérite donc de Component) mais avec la particularité de pouvoir contenir d'autres composants. Ainsi, une interface graphique est un objet d'une classe de type java.awt.Container (par exemple une fenêtre ou une boîte de dialogue) regroupant des simples composants et d'autres containers qui, eux aussi, contiennent d'autres simples composants et containers, et ainsi de suite.

[www.emse.fr](http://www.emse.fr)

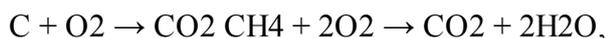
## Texte 29

Les avantages d'un langage interprété sont la simplicité (on ne passe pas par une étape de compilation avant d'exécuter son programme) et la portabilité (un langage tel que Python est censé fonctionner aussi bien sous Windows que sous Linux ou Mac OS, et on ne devrait avoir à effectuer aucun changement dans le code pour le passer d'un système à l'autre). Cela ne veut pas dire que les langages compilés ne sont pas portables, loin de là ! Mais on doit utiliser des compilateurs différents et, d'un système à l'autre, certaines instructions ne sont pas compatibles, voire se comportent différemment. En contrepartie, un langage compilé se révélera bien plus rapide qu'un langage interprété (la traduction à la volée de votre programme ralentit l'exécution), bien que cette différence tende à se faire de moins en moins sentir au fil des améliorations.

[www.pythone-monsite.com](http://www.pythone-monsite.com)

## Texte 30

Les réactions de combustion, comme toutes les réactions chimiques, se font sans variation de masse de chacun des éléments. Les équations chimiques qui décrivent ces réactions doivent donc satisfaire cette contrainte. À titre d'exemple, les réactions de combustion du carbone et du méthane sont respectivement :



On distingue dans une telle réaction, les réactifs, qui subissent la réaction, que l'on sépare en combustible, qui est l'agent réducteur, tel que ceux mentionnés à la section précédente, et comburant, qui est le corps contenant l'agent oxydant, le plus fréquemment l'air. Les produits de la combustion d'hydrocarbures comprennent du dioxyde de carbone ( $\text{CO}_2$ ) et de l'eau, qui selon les conditions de pression et de température, pourra être sous forme liquide ou vapeur.

[www.cours-examens.org](http://www.cours-examens.org)

## Texte 31

Toutes les molécules  $\text{CH}_3\text{COOH}$  de l'acide éthanoïque ne réagissent pas avec l'eau : on dit que l'acide éthanoïque est un acide faible. Par rapport à un acide fort, de même concentration, la solution d'acide éthanoïque est moins acide et présente donc un pH plus élevé. De même, il existe des bases qui ne réagissent que de façon limitée dans l'eau (comme l'ammoniac  $\text{NH}_3$  ou l'ion acétate  $\text{CH}_3\text{COO}^-$ ) : ce sont des bases faibles.

L'acide éthanoïque réagit avec l'eau pour donner l'ion acétate (ou éthanoate) ; de même l'ion acétate (ou éthanoate) réagit avec l'eau pour donner l'acide éthanoïque. Acide éthanoïque et ion acétate sont donc associés par la réaction acido-basique. On dit qu'ils constituent un couple acido-basique. On dit que A et B constituent un couple acide - base si les espèces A et B sont conjuguées, c'est-à-dire liées par la relation :  $\text{A} \leftrightarrow \text{B} + \text{H}^+$ . Dans l'eau, on observe l'équilibre :



[www.keepschool.com](http://www.keepschool.com)

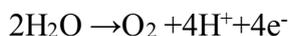
## Texte 32

Le voltamètre d'Hoffmann est le dispositif expérimental permettant d'étudier l'électrolyse de l'eau. On parle d'électrolyse de l'eau, mais, pour assurer une bonne conductivité électrique, l'eau pure ne suffit pas et on doit ajouter un électrolyte, comme de l'acide sulfurique  $\text{H}_2\text{SO}_4$ .

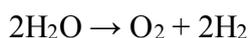
Sous l'action d'une différence de potentiel de quelques volts, une réaction électrochimique apparaît au niveau des électrodes :

La cathode délivre des électrons qui décomposent l'eau selon la réaction de réduction :  $2\text{H}_2\text{O} + 2\text{e}^- \rightarrow \text{H}_2 + 2\text{OH}^-$

Côté anode, les molécules d'eau subissent la réaction d'oxydation suivante :



Le bilan donne l'équation de décomposition suivante:



La réaction produit deux fois plus de dihydrogène que de dioxygène. La loi d'Avogadro justifie que le volume de gaz dans le tube de gauche ( $\text{H}_2$ ) est deux fois plus important que le volume de gaz récupéré dans le tube de droite ( $\text{O}_2$ ).

[www.edumedia-sciences.com](http://www.edumedia-sciences.com)

## Texte 33

La saponification désigne une réaction chimique qui permet la fabrication du savon. Elle consiste à hydrolyser, sous l'effet d'une base forte, un ester en un ion carboxylate et un alcool. La réaction de saponification est l'inverse de l'estérification. La saponification est une réaction lente mais totale. Elle dégage également une importante quantité de chaleur : elle est fortement exothermique. Dans le cas de la fabrication du savon, l'hydrolyse de corps gras se fait en milieu alcalin par une base (soude ou potasse) et produit du glycérol et des carboxylates, une espèce détergente qui fait que le savon lave. Les carboxylates de sodium donnent des savons durs et les carboxylates de potassium des savons plutôt mous sinon liquides. Le savon de Marseille, par exemple, est issu de la saponification par la soude d'huiles végétales, essentiellement.

[Www.futura-sciences.com](http://www.futura-sciences.com)

## Texte 34

Électrolyse de l'eau salée.

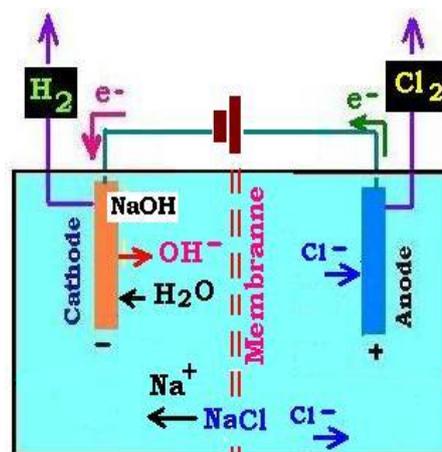
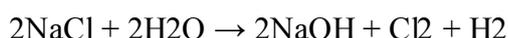
On utilise une membrane séparant les deux parties de l'électrolyseur. Les espèces chimiques présentes en solution sont l'eau  $H_2O$  et les ions  $Na^+$  et  $Cl^-$  provenant de la dissociation du chlorure de sodium  $NaCl$  :

$$NaCl \rightarrow Na^+ + Cl^-$$

Sur l'anode, d'où partent des électrons, il se produit une oxydation: le chlore  $Cl^-$  est oxydé en dichlore en fournissant des électrons :



Sur la cathode, l'eau  $H_2O$  passe à la cathode où il est réduit en captant les électrons en provenance de la pile débitrice pour former des molécules de dihydrogène  $H_2$  et des ions hydroxide  $OH^-$ . C'est la réaction de réduction:  $2H_2O + 2e^- \rightarrow H_2 + 2OH^-$  ( $Na^+ + 1.4.2$ ) + ( $Na^+ + 1.4.3$ ) donnent, avec  $Na^+ + Cl^- = NaCl$ , l'équation bilan ::



[www.scientificsentence.net](http://www.scientificsentence.net)

## Texte 35

Les esters de butanol sont largement utilisés industriellement comme solvant. Ces esters de faible poids moléculaire sont généralement synthétisés par estérification directe d'un acide carboxylique avec un alcool en présence d'un catalyseur acide. Pour déplacer l'équilibre de la réaction, et obtenir des produits purs, la distillation est utilisée industriellement. Néanmoins, dans le cas particulier des esters de butanol, la distillation est énergivore de part les faibles volatilités relatives entre les acides carboxyliques et les esters correspondants. A titre d'illustration, les températures d'ébullition de l'acide acétique ( $118^\circ C$ ) et de l'acétate de butyle ( $126^\circ C$ ) sont relativement proches. Pour réduire la consommation énergétique associée à la

séparation des constituants du mélange, des techniques de séparation alternatives à la distillation peuvent être envisagées.

[www.researchgate.net](http://www.researchgate.net)

### Texte 36

L'éthanol est produit industriellement par addition d'eau sur l'éthylène, en phase gaz :  $C_2H_4 + H_2O \rightleftharpoons CH_3CH_2OH$  La réaction est catalysée par de l'acide phosphorique adsorbé sur un support solide. Elle a lieu sous pression, (6 à 8 MPa), et à température assez élevée (260 à 300°C) à cause de contraintes cinétiques. Nous admettons que l'équilibre thermodynamique est atteint dans ces conditions, pour un temps de séjour dans le réacteur de l'ordre de la minute. Comme sous produit de la réaction, on trouve surtout le diéthyl-éther, qui peut être formé par déshydratation de l'alcool :  $2 C_2H_5OH \rightleftharpoons C_2H_5-O-C_2H_5 + H_2O$  Toutes ces réactions étant à l'équilibre, on utilise le modèle RGIBBS d'Aspen pour décrire le réacteur : il suffit de spécifier la température, la pression et la liste des effluents possibles du réacteur. Le modèle minimise alors l'enthalpie libre du mélange sous contrainte de bilans sur les atomes constitutifs des molécules.

[www.perso.imt-mines-albi.fr](http://www.perso.imt-mines-albi.fr)

### Texte 37

Un acide est une entité chimique capable de perdre un proton H pour donner une base. À l'inverse, une base est une espèce capable de capter un proton H pour donner un acide. Un couple acide/base est généralement noté  $AH/A^-$ , ou  $AH/A$  suivant le couple concerné. On dit que  $A^-$  est la base conjuguée de l'acide AH. Chaque acide possédant une base conjuguée, ces deux entités forment un couple acide/base. L'eau est un cas particulier car elle peut aussi bien jouer le rôle d'un acide que d'une base, c'est une espèce ampholyte. Elle est en effet capable de perdre un proton pour donner un ion hydroxyde  $HO^-$  mais elle peut aussi gagner un proton pour former l'ion oxonium  $H_3O^+$ . L'eau est ainsi un acide dans le couple  $H_2O/HO^-$ ; elle est une base dans le couple  $H_3O^+/H_2O$ . Parmi les acides les plus connus, l'acide chlorhydrique (HCl) est une espèce gazeuse appartenant au couple  $HCl/Cl^-$  et l'acide sulfurique ( $H_2SO_4$ ) est un liquide qui a pour base conjuguée l'ion sulfate  $SO_4^{2-}$ . Une des bases les plus employées est l'hydroxyde de sodium NaOH, solide blanc également appelé soude.

## Texte 38

Les alcools peuvent être obtenus par hydrolyse d'un halogénoalcane, d'un organomagnésien ou d'un ester ; par réduction d'une cétone ou d'un aldéhyde ; ou encore par hydratation d'un alcène (par exemple l'éthylène pour l'éthanol). Cette dernière méthode est largement employée dans l'industrie chimique pour synthétiser les alcools très demandés, comme l'éthanol.

Les alcools sont des acides faibles (car ils se dissocient difficilement dans l'eau), susceptibles de libérer un proton (ion  $H^+$ ) issu de la liaison  $-OH$ .

Ils jouent un rôle dans plusieurs réactions, dont certaines sont d'intérêt industriel :

-estérification : alcool + acide donne ester + eau ;

-réduction : ils donnent l'alcane correspondant, ainsi le méthanol est réduit en méthane, l'éthanol en éthane, etc. ;

-déshydratation : par perte d'une molécule d'eau, on obtient un alcène ;

-oxydation ménagée : selon la classe de l'alcool, on obtient un aldéhyde, un acide carboxylique ou une cétone.

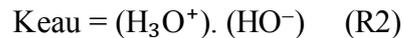
## Texte 39

Dans une solution aqueuse, le produit des concentrations en ions  $H_3O^+$  et  $HO^-$  est constant.

L'eau pure contient une quantité très faible d'ions hydronium et hydroxyde,  $H_3O^+$  (aq) et  $HO^-$  (aq), qui proviennent de l'équilibre :  $H_2O(l) + H_2O(l) \leftrightarrow H_3O^+(aq) + HO^-(aq)$  (R1))

Dans cette réaction, les protons sont transférés d'une molécule d'eau à une autre. On note ainsi que l'eau agit à la fois comme acide (donneur) et comme base (récepteur de proton). La réaction décrite ci-dessus est une réaction d'autoprotolyse (autoprotonation).

La constante d'équilibre correspondante s'écrit :



Cette constante est le produit ionique de l'eau.  $K_{\text{eau}} = 10^{-14} \text{ M}^2$  à  $25^\circ\text{C}$ . La valeur très faible de  $K_{\text{eau}}$  signifie que les concentrations de  $\text{H}_3\text{O}^+$  (aq) et  $\text{HO}^-$  (aq) dans l'eau pure sont très faibles. L'équilibre R-2-1 est fortement déplacé vers la gauche.

D'après la stœchiométrie de l'équation (R1), si on part de 2  $\text{H}_2\text{O}$  (l), on produira  $\text{H}_3\text{O}^+$  (aq) et  $\text{HO}^-$  (aq) en quantités égales. On aura donc, dans l'eau pure,  $(\text{H}_3\text{O}^+) = (\text{HO}^-)$ .

On peut utiliser cette relation dans (R2), et obtenir  $K_{\text{eau}} = (\text{H}_3\text{O}^+)^2$ , ce qui donne  $(\text{H}_3\text{O}^+) = \sqrt{K_{\text{eau}}}$ . Ainsi, dans l'eau pure à  $25^\circ\text{C}$ ,  $(\text{H}_3\text{O}^+) = (10^{-14})^{1/2} = 10^{-7} \text{ mole/litre}$ , et, comme  $(\text{H}_3\text{O}^+) = (\text{HO}^-)$ ,  $(\text{HO}^-) = 10^{-7} \text{ mole/litre}$ .

Quand d'autres substances sont présentes dans l'eau, la relation  $(\text{H}_3\text{O}^+) \cdot (\text{HO}^-) = K_{\text{eau}}$  est encore valable, mais la relation  $(\text{H}_3\text{O}^+) = (\text{HO}^-)$  n'est pas nécessairement vérifiée.

On distinguera alors : les solutions neutres  $(\text{H}_3\text{O}^+) = (\text{HO}^-)$ , les solutions acides  $(\text{H}_3\text{O}^+) > (\text{HO}^-)$ , les solutions basiques  $(\text{HO}^-) > (\text{H}_3\text{O}^+)$ .

[www.emse.fr](http://www.emse.fr)

## Texte 40

L'inflation a été souvent identifiée avec un accroissement de la circulation monétaire, la masse des biens disponibles sur le marché restant invariable. Cette définition, issue des théories quantitativistes, ne suffit pas à rendre compte de l'inflation. De nombreux cas historiques s'inscrivent en faux contre cette notion ; la circulation monétaire a souvent augmenté sans entraîner de changements notables dans l'activité économique. La quantité de monnaie additionnelle peut s'immobiliser dans les encaisses des particuliers, elle peut aussi servir à régler des transactions qui vont en augmentant, ou à financer des investissements nouveaux dont les conséquences sur le niveau de production ne se font pas immédiatement sentir. Enfin, il faut retenir le cas où la circulation monétaire demeure insuffisante par suite d'une modification des habitudes de paiement, d'encaisse, d'épargne, et tend à augmenter pour s'y conformer.

[www.persee.fr](http://www.persee.fr)

## Texte 41

L'inflation est la perte du pouvoir d'achat de la monnaie qui se traduit par une augmentation générale et durable des prix. Il s'agit d'un phénomène persistant qui fait monter l'ensemble des prix, et auquel se superposent des variations sectorielles des prix

La monnaie étant l'étalon des valeurs, la variation de sa propre valeur n'est pas directement mesurable ; on l'évalue à partir des variations des prix à la consommation des biens et services, mesurée à quantité et qualité égales. L'inflation doit donc être distinguée de l'augmentation du coût de la vie car elle ne prend pas en compte la variation des quantités achetées en réponse à l'évolution des prix

L'inflation est un phénomène à propos duquel les controverses entre économistes sont nombreuses : le débat porte sur les conséquences (qui sont parfois considérées comme positives), comme sur les causes. Justifiées par l'incidence concrète de ce phénomène sur l'ensemble de la population, ces controverses sont alimentées par les interrogations posées sur les mesures prises pour la contenir, et sur le degré d'interventionnisme étatique nécessaire pour ce faire.

[www.loretlargent.info](http://www.loretlargent.info)

## Texte 42

Le terme crise financière s'emploie pour désigner un ensemble assez large qui inclut notamment les crises du change, les crises bancaires et les crises boursières, récurrentes dans l'histoire boursière. Mais le terme est également utilisé pour désigner les crises de la dette publique ou des crises qui affectent un marché à terme, voire un marché de produit agricole ; Une crise financière peut concerner seulement quelques pays, ou, initiée dans un pays, peut s'étendre par contagion et devenir internationale et ralentir ainsi l'économie mondiale. Si une crise financière ne concerne dans un premier temps que les marchés financiers, son aggravation conduira à des effets néfastes sur le reste de l'économie, entraînant une crise économique voire une récession. Ces effets sont généralement un resserrement du crédit et donc une baisse de l'investissement, une crise de confiance des ménages.

[www.etudier.com](http://www.etudier.com)

## Texte 43

L'impact d'une hausse du revenu réel (hausse du revenu nominal à prix constants ou baisse des prix à revenu nominal constant) sur les biens consommés s'appelle un effet revenu. Celui-ci peut être positif ou négatif. S'il est positif, par définition, le bien ou le service considéré est un bien normal ; s'il est négatif on dira qu'il s'agit d'un bien inférieur. Le mot inférieur ne veut pas dire qu'il s'agit d'un bien dont on tire une utilité négative, c'est un terme technique : lorsque le revenu augmente, le consommateur substitue au bien considéré d'autres biens. Lorsque le revenu augmente, on substitue un faux filet au hamburger. Lorsque le revenu augmente, on substitue la voiture au bus et au métro, l'avion au train. Très souvent, il s'agit de deux biens rendant un service identique mais dont l'un est d'une variété inférieure à l'autre au sens commun du terme.

[www.lemennicier.bwm-mediasoft.com](http://www.lemennicier.bwm-mediasoft.com)

## Texte 44

L'offre est déterminée par la fonction de production, laquelle représente l'emploi de facteurs de production pour la réalisation de biens et de services. Le modèle d'équilibre général identifie les facteurs de production de façon suivante : K représente le capital, L le travail. Ces facteurs sont supposés être constants en quantité et employés en totalité. Ainsi, si l'on représente la production totale en quantité par la lettre Y, nous pouvons résumer la fonction de production ainsi :  $Y = F(K, L)$ .

Le modèle d'équilibre général retient également l'hypothèse de rendements d'échelle constants, ce qui signifie que la hausse des facteurs de production va entraîner une hausse de la production totale dans les mêmes proportions. En représentant cette hausse par la lettre z, nous obtenons alors la quantité produite suivante :  $zY = F(zK, zL)$ .

La fonction de production inclut également le progrès technologique, autre déterminant de l'offre avec le capital et le travail, les quantités produites pouvant en effet varier avec la même quantité de facteurs de production. Cependant, contrairement à la fonction Cobb-Douglas, le progrès technologique n'est pas identifié en tant que tel dans la fonction de production.

[www.Formation-audit-ecofi.com](http://www.Formation-audit-ecofi.com)

## **Texte 45**

En calculant la productivité d'un facteur, on fait l'hypothèse irréaliste que ce facteur peut être isolé des autres. En réalité le travail et le capital sont combinés pour produire. Les productivités du travail et du capital, apparentes ou effectives sont des productivités partielles. Pour mesurer l'efficacité d'une combinaison productive il faut calculer la productivité globale des facteurs. La production étant obtenue à partir d'une combinaison des deux facteurs, le travail et le capital, la productivité globale des facteurs doit avoir comme dénominateur une expression contenant cette combinaison. Si on veut calculer la productivité globale effective il faudra retenir au dénominateur le flux des services producteurs des facteurs à partir de ceux du travail et du capital en tenant compte de la part respective du travail et du capital dans la production.

**[www.pedagogie.ac-limoges.fr](http://www.pedagogie.ac-limoges.fr)**

## **Texte 46**

Le nombre d'acteurs présents sur le marché ainsi que la nature du produit (homogène ou non) déterminent ensemble la structure du marché ou du secteur. Sur les marchés très concurrentiels, les prix sont tirés vers les coûts marginaux. Cependant seulement un très petit nombre de marchés sont parfaitement concurrentiels. On dit que la concurrence est monopolistique lorsque les produits, même semblables, sont suffisamment différenciés pour bénéficier d'une position de monopole local. Dans ce cas, la concurrence parfaite ne s'applique pas car les biens ne sont pas parfaitement homogènes. Dans une situation de concurrence monopolistique, les producteurs possèdent chacun un pouvoir de marché sur un segment particulier du produit puisque les consommateurs perçoivent les produits comme différenciés. Une entreprise a ainsi une niche de consommateurs qui préfère son bien aux autres biens, même semblables. L'entreprise en question peut ainsi pratiquer un prix supérieur au coût marginal.

**[www.hal.archives-ouvertes.fr](http://www.hal.archives-ouvertes.fr)**

## **Texte 47**

Le modèle de concurrence monopolistique a été introduit par l'économiste Chamberlin. On suppose qu'il existe  $n$  firmes dans un secteur donné de l'économie. On appelle  $p^*$  le prix moyen pratiqué dans ce secteur, pris comme une donnée exogène par chaque firme, et  $p$  le prix

d'une firme dont nous étudions le comportement. Soit  $Q$  la quantité totale de l'offre du secteur, également prise comme une donnée par chaque firme, et  $Q_i$  les ventes de la firme à laquelle nous nous intéressons. La fonction de demande adressée à notre firme s'écrit :  $Q_i = Q(1 - \beta(p - p^*))$  où le paramètre  $\beta$  représente la sensibilité de la demande adressée à notre firme à la déviation moyenne du prix, ou le degré de différenciation de la variété produite par notre firme avec les variétés produites par les firmes concurrentes.

[www.cours-examens.org](http://www.cours-examens.org)

### Texte 48

La production est l'activité économique socialement organisée consistant à créer des biens et services s'échangeant habituellement sur le marché ». On distingue la production marchande et nonmarchande. La production marchande s'échange sur un marché et les revenus qu'elle engendre doivent du moins couvrir les coûts de production. La production non-marchande réalisée par les administrations vise à répondre à des besoins satisfaits hors marché (sécurité, enseignement public). Néanmoins certaines activités comme l'enseignement ou la santé peuvent être à la fois des productions marchandes et non-marchandes.

[www.ekladata.com](http://www.ekladata.com)

### Texte 49

La production consiste à transformer des ressources naturelles et des biens (les consommations intermédiaires qui disparaissent dans le processus productif) en d'autres biens (le produit final), en créant au passage de la valeur. Cette valeur ajoutée est obtenue grâce à la combinaison technique de facteurs de production, travail et outils de production. À l'inverse du travail, ces derniers font eux-mêmes l'objet d'une production, et sont donc accumulables. La croissance résulte de l'augmentation continue des facteurs accumulables. Pour les économistes classiques (Smith, Ricardo et Marx) l'accumulation des biens d'équipement apparaît comme le principal facteur de croissance. Le premier modèle de croissance néo-classique, proposé par Robert Solow en 1956, s'inscrivait également dans cette logique.

[www.universalis.fr](http://www.universalis.fr)

## Texte 50

Les entrepreneurs cherchent à utiliser les facteurs de production et les ressources dans les activités qui promettent les profits les plus élevés. S'il semble plus profitable de produire un bien Q1 plutôt qu'un bien Q2, ils consacrent davantage de facteurs de production à la fabrication de Q1 et moins à celle de Q2. L'objectif de la plupart des entreprises est la maximisation du profit Q.

Une telle entreprise désire donc produire la quantité qui maximise l'écart entre la Recette Totale et le Coût Total. L'entreprise maximise son profit. Le profit économique est défini comme la différence entre les recettes et les coûts de l'entreprise. Profit = Recette Totale – Coût Total,  $Q = RT - CT$ . La Recette Totale dépend des quantités vendues et du prix de vente:  $RT = P \times Q$ . Le Coût Total dépend des prix des facteurs de production et de la quantité produite.

[www.Foad.ugb.sn](http://www.Foad.ugb.sn)

## Texte 51

Les dépenses des consommateurs constituent des recettes pour les producteurs. C'est pourquoi il existe un lien entre l'analyse des dépenses des uns et celles des recettes des autres. La recette totale d'un producteur est égale au produit des ventes par les prix fixés lors de la vente.

$RT = P \times Q$ , Avec RT, recette totale, P, prix du bien et Q quantités vendues. La variation de la recette est donc simultanément soumise aux variations des prix et aux variations des quantités. Si le prix diminue, on peut s'attendre à ce que les quantités augmentent. Il n'est pas sûr que la recette totale du vendeur augmente automatiquement. Tout dépend de l'élasticité de la demande par rapport au prix. Si l'élasticité de la demande par rapport au prix est faible en valeur absolue, cela signifie que l'augmentation du prix n'entraînera pas pour la firme une forte baisse de la demande ; la recette totale pourra augmenter. Au contraire, le fait de baisser son prix entraînera pour le producteur une faible augmentation de la demande, donc une baisse de sa recette totale.

La recette moyenne est la recette par unité vendue, c'est-à-dire le prix.  $RM = RT/Q = PQ/Q = P$  ; La recette marginale Rm est l'accroissement de la recette totale résultant de la vente d'une unité supplémentaire du bien. Lorsque la recette totale est une fonction continue des quantités offertes, La recette marginale est donc la dérivée de la recette totale par rapport à Q. Quand la

recette totale augmente, la recette procurée par les dernières unités est positive. Au contraire, quand la recette totale diminue, cela signifie que les recettes additionnelles procurées par les dernières unités vendues sont négatives. La recette marginale est nulle quand la recette totale est maximale.

**[www.Foad.ugb.sn](http://www.Foad.ugb.sn)**

## **Texte 52**

Le coefficient budgétaire est la part qu'une dépense de consommation (d'un bien ou d'une catégorie de biens) représente dans le total des dépenses de consommation. La loi d'Engel désigne le fait que le coefficient budgétaire des dépenses de premières nécessités (alimentation) diminue lorsque les ressources des ménages progressent. L'effet Giffen constitue une exception à cette « règle » : la hausse du prix d'un bien inférieur peut s'accompagner d'une hausse de sa demande. Si un bien occupe une grande place dans le budget des ménages, la hausse de son prix provoque une baisse du pouvoir d'achat du revenu et conduit le ménage à accroître la demande de ce bien inférieur qui se substitue à d'autres. La structure de la consommation diffère selon le niveau de vie des différentes catégories sociales. Le coefficient budgétaire des dépenses de premières nécessités est plus élevé pour les catégories populaires que pour les catégories moyennes et supérieures. À l'inverse les coefficients budgétaires des biens de moindre nécessité comme les loisirs sont plus élevés dans les catégories supérieures.

**[Www.books.google.fr](http://Www.books.google.fr)**

## **Texte 53**

Une entreprise peut obtenir des capitaux sur le marché financier de deux manières. Elle peut émettre des titres de dette à revenu fixe, il s'agit essentiellement d'obligations. Une obligation représente un engagement contractuel de l'emprunteur de payer au détenteur de l'obligation, des sommes déterminées à des intervalles fixés (intérêt et principal). La date de maturité du titre correspond au dernier versement. Par convention, une dette est dite à court terme si sa maturité est inférieure à une année, à long terme si elle est supérieure à dix ans. Une autre manière d'obtenir des fonds sur le marché financier consiste à émettre des actions. Les montants levés entrent dans les fonds propres de l'entreprise car les actions constituent une partie de son capital. Elles donnent à l'acquéreur un droit au partage des bénéfices (perception

de dividendes). Elles donnent également un droit de participer à l'élection des dirigeants de l'entreprise et aux grandes décisions de la vie de l'entreprise lors de l'assemblée générale.

**[Www.books.google.fr](http://www.books.google.fr)**

# **Références bibliographiques**

## **1- Ouvrage :**

GOFFMAN, CARBRE Maria Teresa, *La terminologie : théorie, méthode et application*, Ottawa, Canada, 1998.

R.GALISSON, D.COSTE, *Dictionnaire de didactique des langues étrangères*, Didier, 1976.

RONDEAU, Guy, *Introduction a la terminologie*, Québec, Gaétan Morin, 1991.

SILVIA Pavel et Diane Nolet, *Précis de terminologie*. Bureau de la traduction, 2001.

MAREA Teresa Cabre, *Constituer un corpus de textes de spécialité*. Cahier de ciel, 2007/2008.

## **2- Dictionnaire :**

Le grand dictionnaire la Rousse, 2000.

Jean-Pierre Cuq. Dictionnaire de didactique du français langue étrangère et seconde. CLE international, Paris, 2003.

## **3- Mémoire :**

Djaider, O &Fortas, D. 2018. Vers un prototype de reconnaissance et d'extraction des constructions à verbe support, mémoire de master, université de Bejaia.

SOULIMANI. M. 2014. « Réalisation d'un analyseur morphologique du texte arabe «AMTAR» », mémoire de master, université de Bejaia.

BEDDAR.M.2013. « Vers un prototype de traduction automatique contrôlée français/arabe appliquée aux domaines à sécurité critique » thèse de doctorat, université De Franche-Comte.

## **4- Site internet :**

[http://www.termosciences.fr/sites/termosciences/IMG/pdf/precis\\_de\\_terminologie\\_Pavel](http://www.termosciences.fr/sites/termosciences/IMG/pdf/precis_de_terminologie_Pavel)

[https://www.lli.ulaval.ca/fileadmin/llt/fichiers/recherche/revue\\_LL/vo135/A.Francoeur.pdf](https://www.lli.ulaval.ca/fileadmin/llt/fichiers/recherche/revue_LL/vo135/A.Francoeur.pdf)

<https://fr.wikipedia.org/wiki/Systran>

<https://fr.wikipedia.org/wiki/Alibaba.com>

## **Résumé**

L'intitulé de ce mémoire, «Vers un prototype de catégorisation automatique des textes écrits par domaine de spécialité.» indique clairement la direction de ce travail centré sur la conception d'une application visant à catégoriser les types de textes à partir d'un dictionnaire de termes issus de diverses spécialités. Deux domaines, la linguistique et le traitement automatique des langues (TAL), que nous avons pris la peine de bien travailler, ont permis de réaliser une application qui répond aux besoins des langues de spécialité, en prenant en compte leurs spécificités.

Dans un souci d'originalité et d'innovation, nous avons choisi un sujet qui est sans doute novateur dans le domaine de la linguistique et qui pourrait se diffuser dans le domaine de la didactique. Notre application permet de catégoriser les types de textes à partir d'un dictionnaire de termes de spécialités. Elle repose sur trois étapes essentielles pour aboutir au résultat. La première étape est l'étiquetage des textes, le système consulte automatiquement un dictionnaire terminologique se rapportant à une ou plusieurs spécialités. Le calcul des fréquences de termes est la seconde étape, le système calcule alors les fréquences d'apparition des termes dans les textes et des domaines s'y rattachant. La troisième et dernière étape est la révélation du résultat final, l'application donne le domaine auquel appartient le texte inséré et traité.

Ce travail nous a permis de répondre à notre problématique de départ et à affirmer les hypothèses que nous avons posées. Les résultats que nous avons obtenus ont ainsi révélés qu'il est possible de créer une application permettant l'identification de textes de spécialité en identifiant le domaine de spécialité en fonction de critères linguistiques ainsi que d'autres systèmes de représentation.

### **Mots clés :**

Catégorisation automatique, domaine de spécialité, application, le traitement automatique des langues (TAL), langues de spécialité, dictionnaire terminologique, termes, identification des textes.