

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

CENTRE UNIVERSITAIRE DE RELIZANE
INSTITUT DES LETTRES ET DES LANGUES
DEPARTEMENT DE FRANÇAIS



Mini - projet de fin d'étude pour l'obtention
du diplôme de licence de français (LMD)

Option: Didactique en français langue étrangère
et ingénierie pédagogique

Étude Lexicométrique D'un Manuel Scolaire

Le manuel de la 2^{ème} AS de FLE en Algérie

L'encadreur

M . Mohamed BOUDAUD

le co - encadreur

M. Abd Ouahid TIOUDIOUINE

Mini - projet réalisé par

Sabrina BENCHENNI

Année universitaire: 2013 / 2014

Dédicace

Je dédie ce modeste travail à:

A mes parents, aucun hommage ne pourrait être à la hauteur de l'amour dont ils ne cessent de me combler. Que dieu leur procure bonne santé et longue vie.

A toute ma famille, et mes amis, Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible, je vous dis merci.

A mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu te garde dans son vaste paradis, à toi mon père.

A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur; maman que j'adore.

Aux personnes dont j'ai bien aimé la présence dans ce jour, à tous mes frères et mes sœurs, je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour leurs conseils, aides, et encouragements.

Aux personnes qui m'ont toujours aidés et encouragés, à mes encadreurs qui m'ont accompagnés durant mon chemin d'études supérieures, mes aimables professeurs « **Mr TIOUIDIOUINE et Mr BOUDAUD**».

A toute la promotion de 3^{ème} année didactique (2014).

Remerciement

On remercie Allah le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mini-projet.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour, sans l'aide et l'encadrement de **M. TIOUDIOUINE** et **M. BOUDAUD**, je les remercie pour la qualité de leurs encadrements exceptionnels, pour la patience, la rigueur et la disponibilité de chacun d'eux durant notre préparation de ce mini-projet.

J'exprime ma profonde gratitude à **M. BOUDAUD** pour son aide pratique et son soutien moral et ses encouragements.

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenu de près ou de loin, et plus particulièrement à **M. TIOUDIOUINE Abdouahid**.

Nous sommes conscients de l'honneur que nous a fait Mme Nafissa ZAROUALI en étant présidente du jury et M .Mohamed SAHI et Mme Saliha HADJARI d'avoir accepté d'examiner ce travail.

Notre remerciement s'adresse également à tous nos professeurs pour leur générosité et la grande patience dont ils ont su faire preuve, malgré leurs charges académiques et professionnelles.

Introduction générale:

L'idée de recourir à l'informatique pour analyser le lexique d'un texte dans le cadre d'un travail de fin d'étude, provient d'une proposition dans une réunion organisée par mes deux encadreurs qui avaient présenté auparavant une communication sur ce thème à Strasbourg; en France, en novembre 2013.

Il faut mettre en valeur que si le lexique constitue pour les apprenants comme pour les enseignants un accès privilégié à la langue étrangère, sa place va connaître une grande importance dans l'apprentissage de cette langue.

C'est pourquoi, nous proposons comme question principale : qu'attendons-nous d'un logiciel lexicométrique qui s'appuie sur la statistique linguistique dans l'analyse d'un texte?

Et si nous assistons actuellement à un regain d'intérêt pour l'étude du lexique, elle passe par des démarches privilégiant à la fois l'analyse des relations intrinsèques aux mots et des relations extrinsèques liées aux connotations sociales et culturelles.

Nous pouvons citer que la disponibilité gratuite de logiciels lexicométriques sur l'ordinateur, qui permettent une analyse quantitative de texte est la première raison qui a influencé notre choix.

Notre travail comporte des démarches qui montrent le fonctionnement de logiciels lexicométriques comme: **LEXICO3** et **CORDIAL**, ayant pour but d'avoir une idée générale sur la nature des textes existantes dans le manuel scolaire de 2^{ème} année secondaire.

1. Problématique:

Notre problématique se résume en une question principale :

Quel est le contenu lexicographique du manuel scolaire de la 2^{ème} AS et quels sont les champs lexicaux et les termes dominants, en utilisant l'analyse lexicométrique?

2. Hypothèse:

L'analyse lexicométrique du manuel de la 2^{ème} AS part sans présupposés et sans idées reçues. Ce n'est qu'à la fin de notre analyse que nous pouvons émettre un avis.

3. justification du choix:

Les raisons ci-dessous sont le principal moteur qui nous a poussés à travailler sur l'analyse lexicométrique:

-C'est un thème inédit, autrement dit, un sujet récent et peu traité par les étudiants.

-La réalisation d'un accès pénétrant aux statistiques à l'aide des nouveaux logiciels; c'est un désir qu'on va réaliser.

-Trouver le sens et le vocabulaire employés dans les textes du manuel scolaire que nous avons déjà étudié lors que nous étions lycéenne; c'est comme faire un retour aux sources et redécouvrir à un échelon supérieur des textes que nous lisions, sans un réel approfondissement.

-Le lexique nous apparaît négligé dans les apprentissages en langues, spécifiquement en français langue étrangère, car dans la plupart des manuels de langue, le lexique n'a guère de place, sauf rarement pour résoudre d'autres problèmes, notamment orthographiques.

4. La méthodologie:

La méthode suivie s'appuiera sur l'emploi des logiciels lexicométriques **LEXICO3** et **CORDIAL** qui permettent une exploitation complète et rapide du corpus textuel : le manuel de la 2^{ème} AS.

Cette analyse traitera les mots signifiants, les mots outils, les signes de ponctuation, les adjectifs et des statistiques sur la sémantique et la stylistique.

- A l'aide des logiciels lexicométriques, nous analyserons les quatre projets du manuel scolaire de 2^{ème} année secondaire.

5. Plan:

Notre mémoire contient deux parties :

Dans la partie théorique nous traiterons la définition de quelques concepts sur la lexicométrie et un aperçu sur les logiciels lexicométriques.

Dans la partie pratique, nous entamerons notre travail par une description du manuel scolaire de 2^{ème} AS, puis nous décrivons les quatre projets de ce manuel, suivi par l'exploitation des logiciels lexicométriques sur notre corpus.

Enfin, nous terminerons notre travail de recherche par une conclusion générale ou nous exposerons notre résultat final.

L'APPORT THEORIQUE:

PARTIE 01:

DEFINITIONS DES CONCEPTS

PARTIE 02:

UN APERÇU SUR DES LOGICIELS LEXICOMETRIQUES

PARTIE 01: DEFINITION DES CONCEPTS

Introduction:

Nous introduirons cette partie par un bref historique sur la lexicométrie, puis nous définissons la lexicométrie avant de citer les notions clés et les différentes fonctionnalités de la lexicométrie.

1. Bref historique sur la lexicométrie :

-La lexicométrie a été développée dans les années 60 par des linguistes influencés par deux courants intellectuels majeurs:

-Le structuralisme et le quantitavisme¹

-Le renouveau de la lexicométrie a été réalisé par l'opprobre jeté sur la lexicométrie qui n'a pas rebuté Tous les historiens.

-Certains comme André SALEM, Alain guerre an ou jean PHILLIPE GEUET ont continué à s'y intéresser, et se sont servi des critiques qui leur avaient été adressées pour améliorer les logiciels de lexicométrie.

2. Définition de la lexicométrie:

La lexicométrie consiste à aborder un texte avec un prisme quantitatif; en comptant l'usage des mots, leurs emplacements, leurs proportions et l'occurrence de celui-ci.

L'analyse lexicale connue en France en tant que **lexicométrie** repose sur l'exploration statistique du lexique d'un corpus de textes.

3. Notions clés de la lexicométrie:

Un **texte** est formé par l'ensemble des caractères des éléments constituant un discours ou l'échange entre deux locuteurs ou l'expression graphique d'un auteur.

Nous approximerons le texte comme étant un ensemble écrit et non un ensemble oral.

La réalisation d'un seul texte ou d'un ensemble de textes déterminé constitue un **corpus**, il est utilisable comme unité ou comme un ensemble de fragments.

D'après **André SALEM**², la lexicométrie regroupe un ensemble de méthodes qui permettent de décrire quantitativement les séquences textuelles constitutives d'un **corpus**.³

¹Idées exprimées successivement par Rousseau et Loche. Jean jacques Rousseau, les confessions. Paris, Gallimard, 2009 ,858 p et John loche, Essai sur l'entendement humain, Paris, Vrin, 2002, 640p

²André SALEM, professeur à l'université Paris3, Sorbonne Nouvelle, <http://www.tal.univ-paris3.fr/asalem>

³<http://halshs-archives.fr> consulté le 29/03/2014 à 18 :20.

Le **lexique** est formé de toutes les occurrences possibles et toutes les formes articulées que constituent l'ensemble des occurrences(ou formes lexicales) potentielles ou réalisées d'une langue.

Toute suite de lettres ou de caractères non séparées par un blanc et délimitée par deux blancs sera appelée une *forme* ou un *vocable*.

La **fréquence absolue** d'une forme dans un texte sera le nombre de ses occurrences dans celui-ci. Sa **fréquence relative** sera le rapport de sa fréquence absolue par la taille du texte qui la contient.

L'occurrence est la réalisation d'un groupe ou d'un ensemble de caractères graphiques d'un alphabet séparés par le délimiteur et formant une unité dissociable dans le flot des caractères particulièrement par ce qu'il peut être voisé en représentant une unité de parole (sensée) pour un locuteur déterminé d'une langue.

Nous désignerons par l'appellation *segment répété*, toute suite de formes qui se répète dans un texte donné.

Certaines formes peuvent apparaître avec une fréquence de réalisation plus ou moins importante dans un texte déterminé ou pas de tout. Certaines autres formes ne sont jamais été attestées celles qui n'ont attestées qu'une seule fois sont dite **hapax**.

Les index et les concordanciers peuvent à leur tour donner lieu à des calculs spastiques et à des représentations graphiques, ils peuvent être globaux (portant sur tout le texte) ou sélectifs (portant sur un seul personnage ou une seule forme).

Avec les index, nous pouvons aussi nous intéresser à l'étude des temps et des personnes verbales⁴.

⁴ <http://halshs-archives-ouvertes.fr> consulté le 29/03/2014 à 18 :20

4. Les différentes fonctionnalités de la lexicométrie:

L'analyse lexicométrique permet d'abord un décompte des mots les plus fréquemment utilisés dans un corpus, afin de repérer, au delà des formes généralement très banales qui occupent le haut de la liste et les termes particuliers les plus récurrents dans un entretien.

Elle permet aussi l'étude des spécificités langagières associées à un sujet et /ou à un type de locuteur. L'analyse lexicométrique prend tout son intérêt lorsqu'elle ne porte plus sur l'intégralité du corpus ; mais sur une sous- partie, que vous aurez préalablement isolée.

En ajoutant, l'étude des catégories de répondants ou les sujets dans lesquels une forme est la plus fréquente. La lexicométrie compare les textes entre eux pour mieux établir les contrastes et les cohérences.

La comparaison des œuvres d'un même auteur ou de plusieurs auteurs, les essais d'attribution ou de rapprochement d'œuvres d'un auteur inconnu à un ensemble d'auteurs connus et bien d'autres application comme la linguistique ; il est assez courant de situer.

La lexicométrie est donc une discipline qui fournit des données exactes sur des textes, elle se développe plus particulièrement dans le monde de la recherche universitaire⁶. Nous pouvons citer aussi l'aide à la constitution de dictionnaires, les études d'occurrences avant ou après lemmatisation⁷, l'analyse des discours non directifs.

⁶ [http:// www. Sonal-info- com. /](http://www.Sonal-info-com/)consulté le 29/03/2014 à 17 : 20

⁷ la lemmatisation est la seconde étape de la plupart des analyses lexicométriques, elle consiste à trouver les différentes formes revêtues par les mots du texte et à les simplifier en une seule forme

PARTIE 02:

APERCU SUR LES LOGICIELS LEXICOMETRIQUES

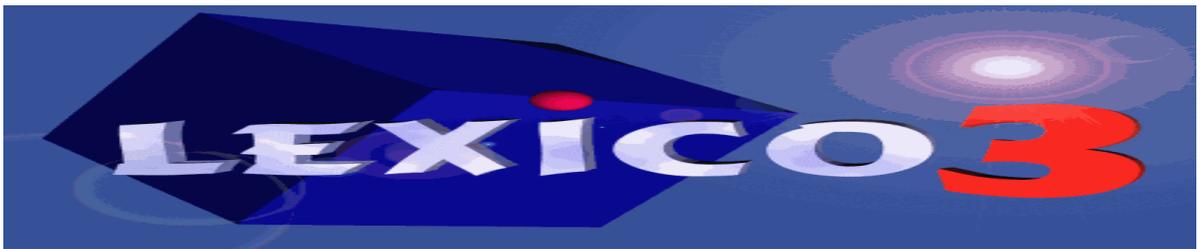
Partie 2 : Un aperçu sur les logiciels lexicométrique

Introduction:

Cette deuxième partie s'appuie sur des logiciels lexicométriques et son fonctionnement sur l'ordinateur.

1. Lexico 03:

Figure 1.1: le logo de lexico3



Le logiciel lexico 03 est développé par André SALEM, Serge Fleury, Cédric Lamelle et William Martinez; il permet de mener des analyses contractives et chronologiques et offre les fonctionnalités suivantes :

Inventaire des segments répétés, concordances, décomptes portant sur les formes graphiques, spécificités par partie, spécificités chronologiques, analyse factorielles portant sur les formes, recherche de groupes de formes, graphiques de ventilation en fréquences absolues relatives ou histogrammes de spécificités⁸.

Le centre de lexicométrie et d'analyse automatique des textes a été créé en 1997 à Paris 3, il a mis lexico comme un logiciel conçu pour le traitement lexicométrique de textes comportant plusieurs centaines de milliers d'occurrences.

Pour pouvoir utiliser ce logiciel correctement, il faut mettre en place un corpus qui sera au format txt, ce fichier txt doit être réalisé selon les normes définies par lexico 39

⁸-<http://www://cavi-univ-paris3.fr>, consulté le 08/04/2014 à 09:45.

⁹-<http://tal.univ-paris3.fr> / consulté le 08/04/2014 à 15 : 30.

La version que nous utilisons est 3,6 pour Windows, il permet d'effectuer des analyses lexicométriques sur les concordances, les formes graphiques, la segmentation répétée et les formes.

2. Hyperbase :

Figure 1.2 : l'interface d'Hyperbase



Le logiciel Hyperbase est conçu et développé par Etienne Brunet¹⁰, il est considéré comme un outil d'automatisation pour compléter l'offre de collaboration de bain de Hyper office.

Hyperbase est l'un des plus anciens parmi les logiciels hypertextuels et/ou statistiques disponibles sur le marché français : LEXICO d'A.SALEM, SPHINX d'Y.Baulac, ALCESTE de Max Reinert, CORDIAL de la société Synapse, WEBLEX de S.Heiden, SPAD-T de lebart, INTEX de Maurice Gross, TROPES de la société Acetic.

En douze années de développement le logiciel a beaucoup évolué, il repose sur des lois statistiques et les techniques¹¹ lexicométriques.

Ce logiciel propose des fonctions nouvelles, notamment thématiques, et permet une exploitation plus souple et plus fine du corpus ou nous pouvons l'isoler ou réunir librement les textes et les sous-ensembles.

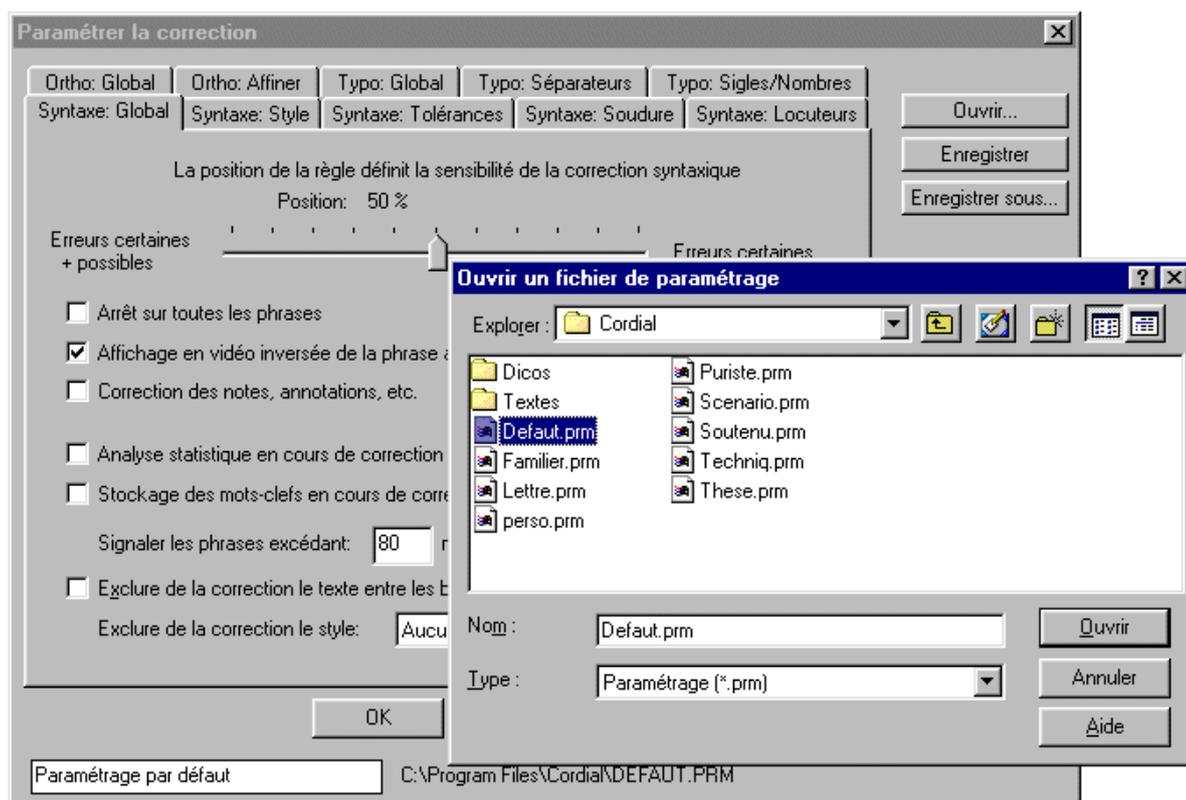
¹⁰ Etienne BRUNET : créateur du logiciel HYPERBASE email : brunet @ univers.f

¹¹ http://www.unce.fr/bc/rubrique_38 consulté le 10/04/2014

A travers le logiciel Hyperbase, nous pouvons garantir l'autorisation d'un ensemble de traitements sur des corpus de textes prédéfinis ou saisis par l'utilisateur.

Hyperbase nous permet aussi de construire des bases de données personnalisées, comme MS Access, mais est beaucoup plus facile à utiliser et accessible en ligne¹².

3-cordial :



Le logiciel CORDIAL est né d'un correcteur orthographique intégré en 1988. Après quatre ans ce logiciel mua en correcteur grammatical.

L'acquisition d'un correcteur externe ne se justifie que si la qualité de ce dernier est nettement supérieure à celle des correcteurs déjà présents.

Comme l'ont démontré de nombreux comparatifs, cordial est effectivement très supérieur en correction et, année après année, l'écart avec la concurrence se creuse.

¹² <http://translate.google.com/user-content.com/> consulté le 11/04/2014.

Cordial est partout, il est le seul correcteur à pouvoir remplacer le correcteur intégré dans la suite office, permettant ainsi à l'utilisateur de ne pas changer ses habitudes. CORDIAL s'intègre également dans les navigateurs, les messageries¹³...

Il est compatible avec les dernières nouveautés telles que Windows8

Le correcteur typographique, orthographique et grammatical de référence du français associe aux outils bureautiques pour plus de simplicité.

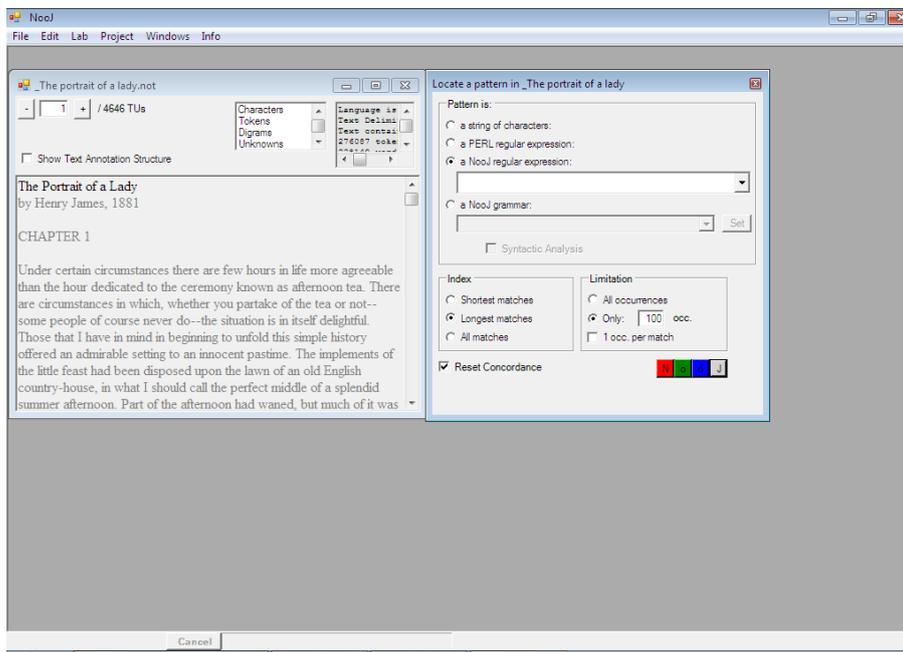
Il corrige les textes, il fonctionne avec des logiciels. Il aide l'utilisateur à rédiger, il analyse les documents, il nous offre aussi une mise à jour permanente, c'est à dire nous pouvons bénéficier des mises à jour permanente, sur le modèle des mises à jour d'individus. Donc, le logiciel cordial est un outil pour accompagner et perfectionner la maîtrise de la langue française pour écrire sans risque de commettre de fautes¹⁴.

¹³[http://www. Amazon.fr/](http://www.Amazon.fr/)consulté le 09/04/2014 à 17 :59

¹⁴[http://www. Je telecharge.com/Bureautique/2277. php/](http://www.Je_telecharge.com/Bureautique/2277.php/)consulté le 09/04/2014 à 18 :10

Nooj:

Figure 1.4 : nooj ouvert sur notre corpus



Nooj est un système de traitement de corpus, reprenant et améliorant les fonctionnalités D'INDEX, conçu pour l'enseignement des langues et de la linguistique.

Nooj présente des fonctionnalités de TAL¹⁵ qui paraissent prometteuses pour l'enseignement des langues et de la linguistique.

Son principale avantage est sa simplicité d'utilisation : il permet à la fois à l'enseignant de constituer des ressources linguistiques et de les paramétrer afin de constituer des projets pédagogiques destinés aux apprenants¹⁶.

Par ailleurs, à moyen terme, d'autres fonctionnalités plus avancées de Nooj, en particulier les transformations syntaxiques, pouvaient aussi être exploitées comme démonstrateurs d'opérations linguistiques.

L'installation de Nooj peut être réalisée par un téléchargement gratuit à partir du site web Nooj : <http://www.nooj4nlp.net>.

¹⁵ TAL : c'est l'abréviation de traitement automatique des langues

¹⁶ <http://alsic.revues.org/> consulté le 12 /04/2014 à 20:26

Nooj tout comme INTEX, est un environnement de développement linguistique permettant de construire, de tester et de gérer des descriptions formalisées à large couverture des langues naturelles, sous formes de dictionnaires et de grammaires électroniques.

La nouvelle mouture du logiciel INTEX, appelée Nooj (INooj05), [Silberztein04] a été réécrite en particulier pour répondre aux besoins des utilisateurs pédagogiques.

Conclusion partielle :

Dans cette partie théorique, nous avons donné dans la première partie des définitions des notions clés en lexicométrie, après un bref historique et une définition de cette science, enfin nous avons cité les fonctionnalités de la lexicométrie.

Dans la seconde partie, nous avons donné un aperçu sur les quatre logiciels lexicométriques les plus utilisés et les plus disponibles sur le net, à savoir, LEXICO 3, HYPERBASE, CORDIAL, NOOJ.

L'APPORT PRATIQUE

PARTIE 01:

LA DESCRIPTION DU MANUEL SCOLAIRE DE 2^{ème} AS

PARTIE 02:

L'ANALYSE LEXICOMETRIQUE DE MANUEL SCOLAIRE

DE 2^{ème} AS

PARTIE 01 : LA DESCRIPTION DU MANUEL SCOLAIRE DE 2^{ème} ANNEE SECONDAIRE

INTRODUCTION :

Nous débuterons cette première partie par les caractéristiques du manuel secondaire puis nous décrirons les projets de celui-ci en général.

1. Les caractéristiques du manuel scolaire de la 2^{ème} AS:

Le manuel scolaire de français destiné aux élèves de 2^{ème} AS est un ensemble didactique ouvert, où le professeur et l'élève pourront puiser les matériaux nécessaires qui leur serviront à réaliser le programme officiel. Il comporte une série d'activités qui se déroulent dans les séquences d'apprentissage intégrées et dans des projets didactiques, ces derniers sont au nombre de quatre dans ce manuel ; chaque projet est intitulé différemment par rapport aux autres, comme il a ses objets d'étude, ses séquences et ses intentions communicatives.

2. La description des projets de manuel scolaire de 2^{ème} AS:

Comme nous l'avons déjà signalé ; il y a quatre projets constituant le manuel scolaire de la 2^{ème} AS.

Le projet (01) vise à concevoir et réaliser un dossier documentaire pour présenter les grandes réalisations scientifiques et techniques de notre époque.

Exposer pour démontrer ou pour présenter un fait, une notion, un phénomène et commenter des représentations graphiques et/ ou iconiques sont des faits à réaliser dans les pratiques discursives et intentions communicatives de ce projet.

La démarche de projet est privilégiée pour installer les compétences visées par le programme, ce projet vise comme objet d'étude le discours objectif.

Le projet (02) consacré à mettre en scène un procès pour défendre des valeurs humanistes. Via trois séquences l'élève va :

*Plaider une cause

*Dénoncer une opinion, un fait, un point de vue

*Débattre d'un sujet d'activité

Le projet(03) s'appuie sur la présentation de lycée, le village, la ville ou le monde de vos rêves, pour faire partager vos idées, vos inspirations. Il a pour objet d'étude la réalisation de reportage touristique et le récit de voyage dans une première phase et dans une deuxième phase la nouvelle d'anticipation, c'est-à-dire l'élève va relater pour informer et agir sur le destinataire, la ou il doit rédiger un récit de voyage et produire un texte touristique à partir d'un reportage qui est réalisé dans la première étape.

En outre, il va relater pour se présenter un monde futur ou il doit rédiger un texte d'anticipation, imaginer et présenter le monde de demain.

Le Projet (4) destiné pour la classe de lettres, il a pour objet d'étude le discours théâtral, dans ce projet l'élève doit mettre en scène un spectacle pour émouvoir ou donner à réfléchir.

A travers trois séquences à l'élève devient capable à dialoguer pour raconter.

PARTIE 02:

L'analyse lexicométrique du manuel scolaire de 2ème AS

INTRODUCTION

Cette deuxième partie sera consacrée à l'analyse lexicométrique du manuel scolaire de la 2^{ème} AS. Nous présentons d'abord les étapes préparatoires à l'analyse lexicométrique, puis nous passerons à l'exploration et l'application des deux logiciels lexicométriques (**lexico3**) et (**cordial**) sur le manuel scolaire de 2^{ème} AS.

1. Les étapes préparatoires à l'analyse lexicométrique

1.1 Préparation du manuel scolaire :

L'ordinateur ne peut pas faire la différence entre les parties d'un corpus, il faut donc les lui indiquer par des balises. Celles-ci se présentent de la façon suivante, pour le logiciel **lexico3** : <partie=partie1>

Alors, nous avons été obligé de scanner le manuel scolaire, puis nous avons converti ce fichier **.doc**, c'est -a- dire en un fichier **Word**. Enfin nous l'avons modifié en fichier **.txt**, car les logiciels lexicométrique ne peuvent examiner que ce genre de fichiers.

2. Analyse lexicométrique du corpus

2.1 Principales caractéristiques du corpus

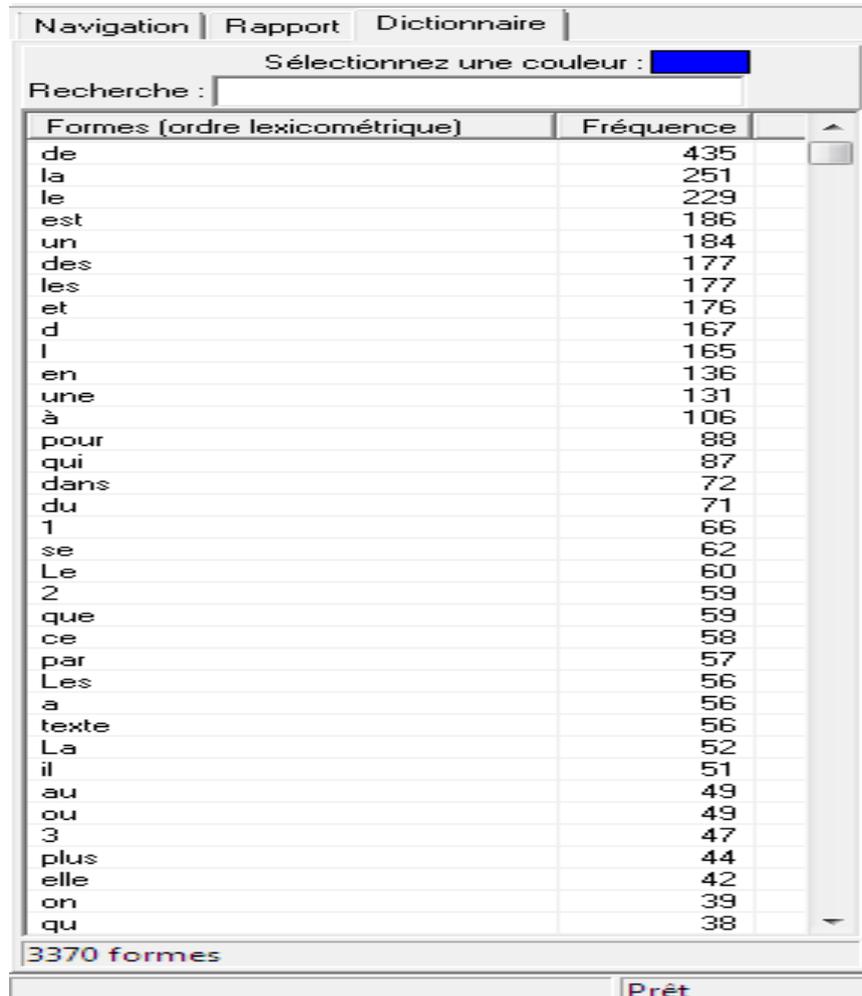
Tableau 2.1 : principales caractéristiques du corpus

Nombre d'occurrences:		10874	Nombre de formes:		3356		
Nombre d'hapax:		2248	Fréquence maximale:		435		
	Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
✓	1	manuel	10874	3356	2248	435	de

Comme le tableau le montre, notre corpus est composé de 3473 formes qui génèrent un texte de 10874 occurrences et 2248 hapax. La fréquence maximale est 435, elle revient à la forme « de ».

Voici maintenant une figure qui affiche les vocables les plus fréquents de notre corpus.

Figure2.1 : les mots outils



Formes (ordre lexicométrique)	Fréquence
de	435
la	251
le	229
est	186
un	184
des	177
les	177
et	176
d	167
l	165
en	136
une	131
à	106
pour	88
qui	87
dans	72
du	71
1	66
se	62
Le	60
2	59
que	59
ce	58
par	57
Les	56
a	56
texte	56
La	52
il	51
au	49
ou	49
3	47
plus	44
elle	42
on	39
qu	38

3370 formes

Prêt

2.1.1 Les mots outils :

Nous constatons que presque tous les vocables sont des mots outils (ou grammaticaux). C'est un phénomène toujours vérifié en lexicométrie. Les mots-outils qui composent une petite partie de la langue française (5%), couvrent presque 50% de la taille d'un corpus textuel.

Ce phénomène n'est pas étonnant car les mots outils¹ sont employés souvent pour

¹ les unités linguistiques sont classées en deux catégories : les mots pleins (nom, verbe, adjectif qualificatif,...) et les mots outils (pronom, article, adjectif non qualificatif, conjonction,...).

communiquer oralement et même par écrit. Nous avons pu faire la même remarque dans les travaux que nous avons consultés².

En ce qui concerne les mots pleins, le premier vocable qui apparaît dans le dictionnaire est le vocable « texte » à la 27ème place avec une fréquence égale à 56.

2.1.2 Les mots signifiants :

Voici un tableau approvisionné par les logiciels lexicométriques concernant les 10 premiers mots signifiants pour but d'entreprendre dans l'analyse lexicométrique de notre corpus.

Tableau 2.2 : mots signifiants

Rang	forme	fréquence
27	texte	56
48	explication	25
51	mots	24
74	thème	17
82	paris	14
83	séquence	14
85	information	14
86	phrase	14
87	point	13
88	sens	14

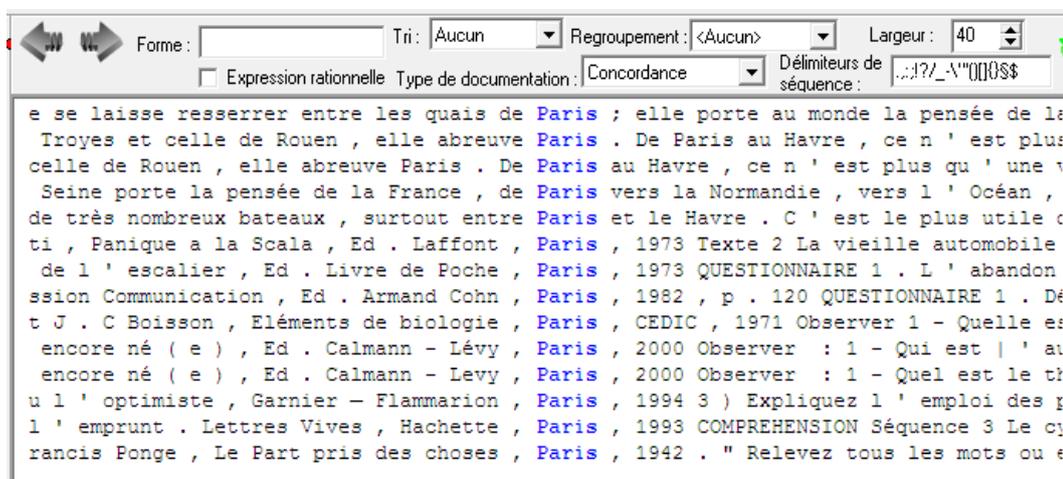
²Aïcha ALIANE, étude lexicométrique du roman « Topographie idéale pour une agression caractérisée de Rachid BOUDJEDRA » Mémoire de licence, centre universitaire de relizane, 2013.

Aoued CHEMLLEL, et BELKHOJA Nour el Houda, étude lexicométrique de la chronique tranche de vie du journal le quotidien d'Oran. Mémoire de licence, centre universitaire de relizane, 2013.

Nous remarquons à partir de ce tableau 2.2 que toutes les formes sauf « Paris » sont des mots techniques, qui appartiennent au métalangage.

Revenant maintenant à l'étude du mot « Paris » :

Figure 2.2 : Concordance de la forme *Paris*



Comme nous pouvons le voir, un retour au corpus par l'ensemble des contextes linguistiques du vocable « Paris », en bleu dans notre figure, montre qu'il est souvent utilisé pour désigner des maisons d'édition qui existent à Paris.

Après ces premiers mots signifiants, nous avons relevé une autre liste rangée dans le tableau suivant :

Tableau 2.3 : Champ lexical de science

Fréquence	Forme	Fréquence	Forme
16	vie	5	fleurs
16	terre	5	soleil
13	monde	5	volcan
8	science	4	phénomène
6	plantes	3	séisme
5	biologie	3	climat
5	compagne		papillon

Nous voyons dans ce tableau qui montre les mots signifiants classés par ordre de fréquence décroissante que ces vocables peuvent être rangés dans deux champs lexicaux : celui de la terre et de la science.

Nous citons maintenant quelques exemples sur ces deux termes :

*« A cette fascination s'ajoute l'admiration pour le jeu décrit par les **chimistes** des interactions entre **molécules** capable de produire la danse de la lumière. »* (p20)

*« C'est cela la **science** : un effort jamais achevé pour naître au monde en faisant naître en notre esprit un modèle de monde. »* (p20)

*« Le président de la conférence a repris la proposition des ONG d'exclure le **nucléaire** pour les pays en développement mais, de manière contradictoire, il a admis que les investissements nucléaires dans les ex-pays de l'Europe de l'Est pouvaient entrer dans le grand marchandage du **CO²**. »* (P52)

Ces résultats nous poussent à chercher encore d'autres vocables pour déterminer la nature des textes de notre corpus. Voici un autre tableau qui montre les mots signifiants classés par ordre de fréquence décroissante

Tableau2.4 : Formes liées au champ lexical des animaux

Fréquence	Forme
11	papillon
6	animaux
5	abeilles
4	oiseau
3	coccinelles
3	insectes

Nous observons que tous ces formes relèvent le champ lexical des animaux en général, des insectes, en particulier.

Deux exemples de mot« papillon» relevé de notre corpus :

« Lorsque le sucre élaboré dans les tiges surgit au fond des fleurs, comme des tasses mal lavées - un grand effort se produit par terre d'où les **papillons** tout à coup prennent leur vol. » (p 11)

« Bès lors le **papillon** erratique ne se pose plus qu'au hasard de sa course, ou tout comme. » (p11)

A partir de ces deux exemples et d'autres, il nous semble que ce langage scientifique puisé dans des livres scientifiques en grande partie de biologie présente beaucoup de difficultés quant à sa compréhension par des élèves de lycée (des apprenants). Comment un élève qui apprend la langue française pourrait-il le faire à partir de textes aussi compliqués que les suivants :

1) « Lorsque le sucre élaboré dans les tiges surgit au fond des fleurs, comme des tasses mal lavées - un grand effort se produit par terre d'où les papillons tout à coup prennent leur vol.

Mais comme chaque chenille eut la tête aveuglée et laissée noire, et le torse amaigri par la véritable explosion d'o~ les ailes symétriques flambèrent.

Dès lors le papillon erratique ne se pose plus qu'au hasard de sa course, ou tout comme.

Allumette volante, sa flamme n'est plus contagieuse. Et d'ailleurs, ii arrive trop tard et ne peut que constater les fleurs écloses. N'importe : se conduisant en lampiste, ii vérifie la provision d'huile de chacune. Il pose au sommet des fleurs la guenille atrophiée qu'il emporte et venge ainsi sa longue humiliation amorphe de chenille au pied des tiges.

Minuscule voilier des airs maltraité par le vent en pétale superfétatoire, ii vagabonde au jardin. »

2) Les relations dans un écosystème

Lorsque les biologistes étudient un milieu, ils utilisent le terme d'écosystème qui désigne d'une part toutes les espèces animales et végétales qui habitent un même lieu, d'autre part les relations qu'elles établissent entre elles et avec le milieu.

Dans un milieu donné, il s'établit un véritable réseau de relations alimentaires, appelé réseau trophique (du grec " trophos" qui signifie nourriture). Dans un réseau, tous les animaux se nourrissent d'autres animaux ou de plantes. Un réseau est constitué de plusieurs chaînes alimentaires I une chaîne alimentaire est une relation simple entre des êtres vivants dans laquelle chaque individu mange celui qui le précède et sert de nourriture à celui qui le suit: cette organisation permet un transfert de matière et d'énergie d'un être à un autre. Le nombre de maillons peut varier d'une chaîne alimentaire o une autre, mais chaque chaîne débute toujours par un végétal chlorophyllien. Ces végétaux sont appelés producteurs parce qu'ils produisent de la matière organique ai partir d'éléments inorganiques. Cette matière est consommée par les autres êtres vivants appelés pour cette raison des consommateurs.

Dans tous les milieux, toute une faune d'animaux de petite taille et une flore de bactéries et de champignons se nourrissent de matière organique morte (cadavres d'animaux, restes de végétaux) dont ils provoquent la décomposition. Ces décomposeurs procèdent à la minéralisation de cette matière organique, assurant ainsi son recyclage en sels minéraux qui servent de nourriture aux plantes chlorophylliennes. »

2.2L'exploitation de logiciel CORDIAL :

Après le téléchargement de ce logiciel, nous avons commencé par son application sur notre corpus, et nous avons obtenu ces résultats :

2.2.1Mots signifiants :

Proportion de substantifs : 54,0%

Proportion d'adjectifs : 14,3%

Proportion de verbes : 20,2%

Proportion d'adverbes : 11,5%

Proportion personnels 1-2^op: 1,8%

Proportion personnels 3^op.: 4,0%

2.2.2Statistiques sur les types lexicaux :

Vocabulaire :

Noms :

% noms communs : 82,1%

% noms propres : 17,9%

% noms composés : 0,4%

% mots très rares : 4,0%

% noms abstraits : 75,7%

% noms communs inconnus: 0,1%

% noms propres inconnus : 0,7%

Nous constatons via ces statistiques que le pourcentage d'emploi des noms communs est très élevé dans notre corpus par rapport aux noms propres et noms composés.

2.2.3 Statistiques sur les types grammaticaux :

Mots outils : 49%

Mots signifiants : 51%

Articles définis : 12,6%

Articles indéfinis : 2,4%

Adjectifs démonstratifs : 0,6%

Adjectifs possessifs : 1,4%

Adjectifs indéfinis : 0,6%

Prépositions : 11,9%

Conj. coordination : 3,1%

Conj. subordination : 1,1%

Pronoms relatifs : 1,3%

Pronoms possessifs : 0,0%

2.2.4 Signes de ponctuation :

Nous ajoutons une autre liste de signes de ponctuation fournie par le logiciel cordial :

Tableau2.5 : les signes de ponctuation

Signe	Taux
. :	33,3%
... :	0,9%
,	50,7%
;	2,0%
:	4,0%
?	1,8%
!	0,5%
() :	4,5%
[] :	0,2%
- :	0,9%

2.2.5 Les adjectifs :

Un autre tableau pour citer certains adjectifs employés dans le manuel scolaire de 2^{ème} AS :

Tableau 2.6 : les adjectifs

Forme	Genre	Fréquence
vivant	AMS	8
Nombreuse	AFS	3
Animal	AMS	3
Chimique	ASIG	3
Chaud	AMS	3
Propre	ASIG	2
Tropical	AMS	2
Océanique	AMS	2
Génétique	ASIG	2
Naturel	AMS	2
Considérable	ASIG	2
Technique	ASIG	2

Nous remarquons que ces adjectifs sont liés au champ lexical de la science. Pour mieux expliquer, nous avons extrait ces exemples de notre manuel scolaire:

« Le moyen le plus employé est la lutte **chimique** (insecticide et fongicide), mais elle présente de **nombreux** inconvénients. » (p 24)

« Cette stratégie consiste à utiliser des ennemis **naturels** des ravageurs des cultures. » (p24)

« En découvrant la molécule d'ADN, les chercheurs n'ont pas seulement élucidé un problème qui semblait toujours **mystérieux**, celui de la « vie » ; ils ont montré que ce qui donne leurs pouvoirs aux êtres vivants repose sur des mécanismes **chimiques** très ordinaires, et par conséquent modifiables. » (P26)

D'autres informations fournis par le logiciel cordial ainsi que:

2.3 Statistiques sur la sémantique :

Le domaine «géologie» est nettement prédominant dans ce texte.

Les deux domaines «biologie» et «bactériologie» sont deux domaines secondaires saillants dans votre texte.

Le domaine «astronomie, astrologie» est un quatrième domaine remarquable.

Dans la thématique de ce texte, la vie est une notion particulière.

D'une façon plus précise, l'analyse des thèmes généraux de ce texte indique une prédominance des thèmes suivants : «sentiments», «comportements» et «politique».

Une analyse plus fine encore de la thématique de ce texte fait apparaître comme thèmes centraux :

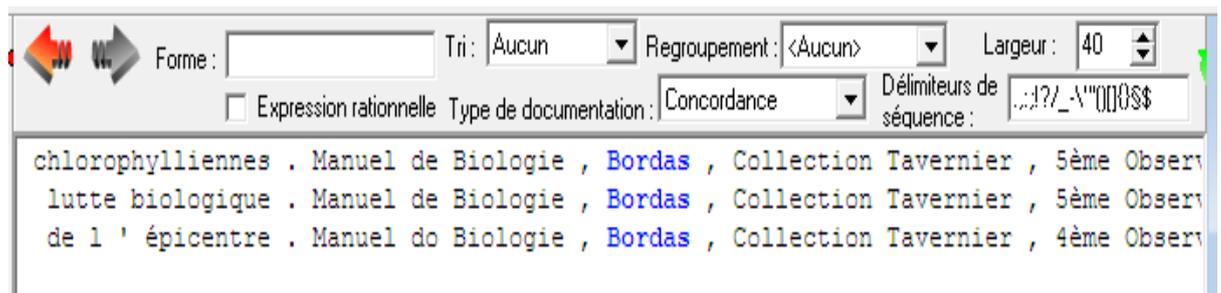
- * tectonique (catégorie : géologie)
- * fonctionnement de l'œil (catégorie : vision)
- * cellule (catégorie : micro-organismes)
- * embryologie (catégorie : reproduction)
- * immunologie (catégorie : immunité, sexualité)
- * chenilles (catégorie : insectes)
- * oiseaux (catégorie : oiseaux)
- * crustacéologie (catégorie : poissons, fruits de mer, pêche)

Une étude lexicologique et sémantique de l'ensemble des phrases indique que les mots forts du texte sont «venir», «ravageur», «puceron», «lithosphère», «coccinelle», «derviche» et «séisme».

Parmi les noms propres, on relève une nette prédominance des mots «Paris» et «Seine».

Nous pouvons signaler que les domaines abordés sont la biologie, la botanique, astronomie, astrologie,...

un exemple qui prouve notre résultat :



Donc, ce que nous avons constaté a été prouvé par le logiciel cordial.

Conclusion partielle :

Dans cette partie pratique, nous avons procédé a une brève description du manuel scolaire de la 2ème AS, puis nous avons réalisé l'analyse lexicométrique de ce manuel, à l'aide des deux logiciels : LEXICO3 et CORDIAL.

Nous avons constaté que le contenu du lexique du manuel est presque totalement scientifique, où prédominent des termes en relation avec la biologie, le nucléaire, la vulcanologie,...

Nous avons aussi constaté par l'intermédiaire de Cordial que les vocables et les phrases sont en général très compliqués. Question : Ces mots et ces structures difficiles ne sont-elles pas un obstacle à l'apprentissage de la langue française ?

Conclusion générale:

Lexico 3 et *cordial* sont les deux logiciels que nous avons exploités pour explorer et analyser le contenu du manuel scolaire de la 2^{ème} année secondaire. Ce dernier revêt une importance particulière dans le processus éducatif et d'enseignement, non seulement d'un point de vue didactique et pédagogique, mais aussi d'un point de vue formatif.

En premier lieu, nous soulignons que nous avons beaucoup appris en employant l'outillage informatique et les logiciels textométriques. Ces outils nous ont permis de constater que l'analyse d'un texte par cette méthode est très efficace et rigoureuse. Nous sortons donc de ce travail enrichi par des compétences informatiques et lexicométriques.

En deuxième lieu, nous sommes arrivées à la conclusion suivante : Les termes et les structures syntagmatiques employés dans notre manuel sont complexes et compliqués.

En outre, l'image que présente ce manuel scolaire à travers les textes, démontre que ceux-ci sont difficilement accessibles. Quantitativement, le manuel est riche, mais la nature des textes abordés est purement éloignée des intérêts des apprenants. Nous signalons que le manuel scolaire de la 2^{ème} AS est un ouvrage plutôt scientifique qu'un manuel destiné à l'apprentissage du FLE.

Pour finir, notons une nouvelle fois que cette analyse nous a permis de découvrir une thématique de textes autour de la science en général, très technique, abstraite, non attirante et qui pourrait constituer un obstacle pour un jeune apprenant auquel est destiné ce manuel.

TABLE DES MATIERES:

INTRODUCTION

GENERALE:.....03

Introduction.....	03
1.Problématique.....	03
2. Hypothèse.....	03
3.Justification de choix.....	04
4.La méthodologie.....	04
5. Le plan.....	04

L'APPORT THEORIQUE:

PARTIE 1 : DEFINITION DES CONCEPTS

Introduction.....	06
1.Bref historique de la lexicométrie.....	06
2. Définition de la lexicométrie.....	06
3. Notion clés de la lexicométrie.....	07
4. Les différentes fonctionnalités de la lexicométrie.....	08

PARTIE 2 : UN APERÇU SUR LES LOGICIELS LEXICOMETRIQUES.....09

Introduction.....	10
1. Lexico3.....	10
2.Hyperbase.....	10
3. Cordial.....	12
4. Nooj.....	14
conclusion.....	15

L'APPORT PRATIQUE :

PARTIE 1 : LA DESCRIPTION DE MANUEL SCOLAIRE DE 2EME ANNEE

SECONDAIRE.....16

Introduction.....17

1. Les caractéristiques de manuel scolaire de 2^{ème} AS.....17

2. La description des projets de manuel scolaire de 2^{ème} AS.....18

PARTIE 2 : ANALYSE LEXICOMETRIQUE DU CORPUS.....19

Introduction.....20

1. Les étapes préparatoires à l'analyse lexicométrique.....20

1.1Préparation de manuel scolaire.....20

2.L'analyse lexicométrique du corpus.....20

2.1Principales caractéristiques du corpus.....20

2.1.1 Les mots outils21

2.1.2 Les mots signifiants.....22

2.2L'exploitation de logiciel cordial.....26

2.2.1 Mots signifiants.....26

2.2.2Statistiques sur les types lexicaux.....27

2.2.3Statistiques sur les types grammaticaux.....27

2.2.4Signes de ponctuation.....28

2.2.5Les adjectifs.....29

2.3 Statistiques sur la sémantique.....30

conclusion partielle :.....31

CONCLUSION GENERALE.....32

BIBLIOGRAPHIE

La bibliographie :

1. Ouvrages théoriques :

Marie-Anne PAVEAU, Francis Grossmann et Gérard Petit, *didactique de lexique, langue, cognition, discours*, Elluq éditions de l',2005.

Francis Grossmann et Sylvie Plane, *les apprentissages lexicaux*, Pu septentrion, 2008.
SALEM André, « Analyse factorielle et lexicométrie synthèses de quelques expériences », Mots, 21982.

DOLPHIN Colette, *Méthodes de la statistique linguistique et vocabulaire fantastique de Malpertuis*, Genève, Editions Slatkine, 1979.

LEBART Ludovic et SALEM André, *Statistique textuelle*, Paris, Dunod, 1994.

2. MEMOIRES :

Aicha ALIANE, étude lexicométrique du roman « Topographie idéale pour une agression caractérisé de Rachid BOUDJEDRA » Mémoire de licence, centre universitaire de relizane, 2013.

Aoued CHEMLLEL,et BELKHOJA Nour el Houda, étude lexicométrique de la chronique tranche de vie du journal le quotidien d'Oran. Mémoire de licence, centre universitaire de relizane, 2013.

3. LIVRES :

Le manuel scolaire de 2^{ème} année secondaire, 2009, office national des publications scolaires.

4.LOGICIELS :

Lexico 3, logiciel d'analyse textuelle conçue par André Salem. Téléchargeable ici : < <http://www.cavi.univ-paris3.fr/ilpga/tal/lexicowww/download.htm> >.

Hyperbase, logiciel d'analyse textuelle conçue par Etienne Brunet. Téléchargeable ici :<ancilla.unice.fr/~brunet/pub/hyperbase.html>.

Cordial, logiciel d'analyse textuelle conçue par la société toulousaine synapse développement.

Nooj, logiciel d'analyse textuelle. Téléchargeable

ici :<<http://www.nooj4nlp.net/pages/download.html>>.

5. SITOGRAPHIE :

<http://www.cavi.-univ-paris3.fr/ilpga/tal/lexicoWWW>(consulté le 14/04-2014)

<http://www.cavi.-univ-paris3.fr/ilpga/> ilpga/

<http://www.unice.fr/bcl/rubrique38>

6.DICTIONNAIRES :

DUBOIS J. et al. *Dictionnaire de linguistique*, Paris, Larousse-Bordas, 2001.