

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



Faculté des Lettres et des Langues
Département de français

Mémoire de master

Option : Sciences du langage

**Etude lexicométrique comparative des manuels scolaires de français du
public et du privé : cas du manuel de 1^{ère} année moyenne**

Présenté par :

M^{elle} Boudraham Amel

M^{elle} Bouilles Linda

Le jury :

M. Seghir Atmane, président

M. Beddar Mohand, directeur

Mme. Bounouni Ouidad, examinatrice

- Année universitaire 2022/2023-

Remerciements

Avant tout et à tout moment, nous tenons à remercier «ALLAH» le tout Puissant et le Miséricordieux de nous avoir illuminé et ouvert les portes du savoir, et nous avoir donné le courage, la force, la volonté afin de pouvoir accomplir ce travail de recherche.

Tout d'abord nous tenons à exprimer notre profonde gratitude à notre Directeur de recherche M. Beddar Mohand d'avoir dirigé notre travail ainsi que pour sa disponibilité et ses précieux conseils qui ont contribué à alimenter notre réflexion et à l'élaboration de ce modeste travail.

Nous ne manquerons pas de remercier les membres du jury qui ont bien voulu examiner et évaluer notre mémoire.

Nos vifs remerciements à nos chères familles, à nos amis, pour leur immense soutien, leurs encouragements et leur confiance en nous.

Nos sincères remerciements à toutes les personnes qui nous ont aidées de près ou de loin.

Dédicaces

Louange à Allah, Seigneur de l'univers et le *possesseur de toute la grâce*

Avec l'expression de ma reconnaissance, je dédie ce modeste travail:

- ◆ A la mémoire de mon cher père **Hamid** qui nous a quitté trop tôt, qui aurait souhaité voir sa fille ce jour là, je pris Dieu le tout puissant de lui accorder sa sainte miséricorde et l'accueillir dans son vaste paradis.

- ◆ A ma source de tendresse, de douceur et de bienveillance. Ma chère mère **Houria** que nulle dédicace ne puisse exprimer mes sincères sentiments, pour sa patience, son aide et ses sacrifices, qu'Allah la garde pour moi.

- ◆ A mes très chers frères : **Kamel, Smail et Hassen** pour leurs confiance et encouragements durant toutes mes études.

- ◆ A mes chères sœurs : **Louiza , Farida, Nassima , Zahra et Lamia** (et leurs maris) en particulier **Sofiane** qui était toujours là pour moi .

- ◆ A mes nièces et neveux.

- ◆ A tous les membres de ma famille ; cousins, cousines, oncles, tantes, petits et grands .

- ◆ A mon binôme et très chère amie **Amel**.

- ◆ A mes copines, A mes meilleurs amis et A toutes les personnes qui m'ont impressionnée et inspirée de près et de loin.

Linda

Dédicaces

Louange à Allah, Seigneur de l'univers et le *possesseur de toute la grâce*.

Je dédie ce modeste travail que j'ai accompli avec tant de passions à toutes les personnes qui m'ont soutenu durant toute cette période.

- ❖ Un grand merci à l'ensemble de ma famille et plus particulièrement à mes parents pour leur amour, leur confiance, leurs conseils ainsi que leur soutien inconditionnel qui m'a permis de réaliser les études pour lesquelles je me destine et par conséquent ce mémoire, que dieu leur apporte santé et longue vie Inshallah.
- ❖ A mes chers frères pour leur appui et leur encouragement tout au long de mon parcours universitaire.
- ❖ A mes chères sœurs, **Souhila** et **Nadia**.
- ❖ A une personne spéciale qui a été toujours là pour moi.
- ❖ A ma binôme et chère amie **Linda**.
- ❖ A mes chères amies, **Amina** et **Nesrine**.

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infaillible. Merci d'être là pour moi.

Amel

Table des matières

Introduction générale	8
Chapitr 01 :	11
Introduction partielle	12
I/Autour de la lexicologie.....	12
1- Le mot	13
a/ Mot plein et mot vide	14
b/ Mot simple et mot composé.....	14
c/ Lexie et lexème	14
d/ Le sème	15
2- Le lexique	15
3/ Le champ lexical et champ sémantique	16
II/ Autour de l'analyse du discours	16
a/ Discours et texte	17
b/ Analyse textuelle	17
c/ Etude comparative	18
III/ Autour de la lexicométrie	18
a/ Histoire et évolution.....	19
b/ Définition.....	19
IV/ Le logiciel Iramuteq.....	20
a/ Occurrence et cooccurrence.....	20
b/ Le segment.....	21
V/ Le système éducatif en Algérie	Error! Bookmark not defined.
VI/ Ecole publique et école privée	Error! Bookmark not defined.
VII/ Ecole « Les Iris »	Error! Bookmark not defined.
VIII/ Le statut de la langue française en Algérie	Error! Bookmark not defined.
a/ La place de la langue française à Bejaia	Error! Bookmark not defined.
IX/ Le manuel scolaire	24
a/ L'importance du manuel scolaire.....	25
Chapitre 02 :	27
Introduction	28
I/ Présentation de corpus	28
1- Le manuel de l'école publique (figure 01)	28

2- Le manuel de l'école privée (figure 02)	30
II/ Cadre méthodologique.....	32
Le logiciel « Iramuteq »	32
1- Les différentes analyses du logiciel.....	33
a/ Statistiques textuelles	33
b/ Spécificités et AFC	35
c/ Classification méthode Reinert	35
e/ Nuages de mots	37
Analyse de données	Error! Bookmark not defined.
1- Corpus	Error! Bookmark not defined.
2- Caractéristiques des corpus	38
3- Analyse des corpus	41
1- Statistiques textuelles	41
a- Manuel public	42
b- Manuel privé.....	43
Formes actives	44
c-1/ Manuel public	44
d/1 Manuel privé	50
2/ Classification méthode Reinert.....	55
b/1 Manuel public	55
b/2 Manuel privé	56
3/Analyse similitude.....	59
c/1 Manuel public	58
c/2 Manuel privé	59
4/ Nuage de mot.....	62
d/1 Manuel public	60
d/2 Manuel privé	61
Conclusion partielle.....	62
Synthèse.....	65
Conclusion générale	63
Références et bibliographiques	66
Table des tableaux.....	
Annexes.....	
Résumé.....	

INTRODUCTION GÉNÉRALE

La langue fait référence aux moyens de communication, au système d'expression oral ou écrit utilisé dans les communautés linguistiques. Toute langue nécessite un apprentissage et une maîtrise, elle forme donc un ensemble de mots et un ensemble de règles de fonctionnement. « *Elle est la partie sociale du langage extérieure à l'individu, qui à lui seul ne peut ni la créer ni la modifier, elle n'existe qu'en vertu d'une sorte de contrat passé entre les membres de la communauté* » (SAUSSURE, 1916)

La langue comme système comprend plusieurs autres disciplines, comme la lexicographie, qui vise à étudier les unités lexicales qui composent le lexique.

La linguistique est une discipline qui a pour objet l'étude scientifique de la langue, elle est apparue au début du 19^{ème} siècle. Cette science est marquée par plusieurs chercheurs dont les préoccupations sont relatives au langage. Parmi eux, nous citerons FERDINAND de SAUSSURE considéré comme le père de la linguistique moderne, il la définit comme « *une science qui a pour objet la langue en elle-même et pour elle-même* » (1916).

Les sciences du langage s'inscrivent dans le champ plus large des sciences humaines et sociales, elles se situent au carrefour des autres branches : la psychologie, les sciences cognitives, la philosophie ainsi que la sociologie, cette discipline ne porte pas sur ce qu'il faut dire, mais sur ce qui est réellement dit.

En effet, notre travail de recherche s'inscrit dans le champ vaste de la lexicologie/analyse de discours en général et dans l'analyse lexicométrique en particulier. La lexicométrie est une discipline qui prend en charge l'analyse informatisée du discours et du lexique, elle permet d'effectuer de nombreuses analyses sur des corpus étendus.

Problématique

Notre recherche vise principalement à mettre en évidence les caractéristiques textuelles des programmes du français de l'enseignement public et privé. C'est pour cette raison, en effet que nous avançons la problématique suivante : « Quelles sont les caractéristiques lexicologiques et lexico-sémantiques des programmes pédagogiques de l'enseignement privé et public en 1^{ère} année moyenne (1^{ère} AM) ? »

En filigrane de cette problématique principale, découlent plusieurs questions subsidiaires :

1. Quels sont les vocabulaires les plus prédominants dans les deux manuels ?
2. Quel est le lexique employé dans chaque manuel scolaire ?

Hypothèse(s)

Afin de répondre à ces questions, nous avons émettre les hypothèses suivantes

1- Il nous semble que le vocabulaire du programme privé serait plus étendu ; l'enseignant utiliserait des termes spécifiques, un vocabulaire propre à son domaine, en se basant généralement sur des textes littéraires, poétiques,... par opposition au vocabulaire des programmes publics, qui s'appuie spécifiquement sur des textes scientifiques.

2- Nous supposons que le lexique du manuel privé est plus riche et variés, notamment dans les activités faites en classe (la compréhension de l'écrit et de l'oral,...etc.) vu qu'ils sont d'avantage orientés vers des éléments issus de la culture française.

Motivation du choix du sujet

Le principal moteur qui nous a poussées à travailler sur l'analyse lexicométrique se justifie par le fait que c'est un thème inédit à l'Université de Bejaia, c'est-à-dire un sujet récent et peu traité dans notre domaine, qui nous permettra sans aucun doute d'avoir une profonde connaissance des statistiques textuelles à l'aide d'un nouveau logiciel « Iramuteq ».

Notre recherche a pour objectif d'analyser le sens du vocabulaire employé dans les deux manuels scolaires de l'école publique et privée.

Corpus et méthodologie

Notre corpus est composé de deux supports pédagogiques textuels utilisés dans les deux écoles sous forme de texte (version Pdf).

Notre étude sera analytique car la méthode suivie s'appuiera sur l'emploi du logiciel lexicométrique « Iramuteq » qui permet de faire des analyses complètes et rapides du corpus textuel.

Plan de la recherche

Notre travail se subdivise en deux chapitres : théorique et pratique. Dans la première partie, nous présenterons d'abord notre sujet en donnant quelques définitions des concepts de base tel que : la lexicologie, la lexicométrie, le logiciel,... pour en tracer l'itinéraire et connaître le rôle du manuel dans l'établissement public ainsi que le privé.

Dans la partie pratique nous opterons pour une étude lexicométrique comparative. Notre choix a été porté sur deux manuels de français relatifs à la première année du cycle moyen d'une école privée «Les Iris» et l'école publique en utilisant le logiciel IRaMuTeQ.

CHAPITRE 01 :
CONSIDÉRATIONS THÉORIQUES

Introduction partielle

Nous introduirons ce chapitre par définir le concept de la lexicologie en générale, et nous aborderons brièvement l'analyse du discours, ensuite on va entamer le concept de la lexicométrie et ses multiples fonctionnalités ; pour en présenter le logiciel Iramuteq avant de citer les caractéristiques du manuel scolaire des deux établissements « Ecole publique et école privé ».

I/Autour de la lexicologie

La lexicologie est une branche récente de la linguistique qui étudie le fonctionnement des composantes lexicales de la langue, ce terme a été inventé dans les années 1970, elle s'intéresse à l'apparition de nouvelles unités lexicales et l'évolution du sens des mots. D'après A.POLOGUERE « *Il serait donc légitime de considérer la lexicologie, la discipline qui étudie les phénomènes lexicaux, comme étant la branche maitresse de la linguistique* » (2002).

En effet, la lexicologie a pour objet l'étude du vocabulaire et du lexique d'une langue dans ses relations avec d'autres composantes de la langue. Selon SIOUFI et VAN RAEMDOKCK « *la lexicologie est définie comme discipline descriptive à deux objets : le lexique et le mot* » (2012).

De plus, elle ne s'attache pas seulement à l'étude du mot et du lexique, mais elle décrit aussi l'organisation de ce lexique à partir de la notion de « champ », en distinguant deux types de champs : « champ lexical » et « champ sémantique ».

La lexicologie n'est pas seulement conçue comme une branche de la sémantique structurale, c'est-à-dire liée à la sémantique visant à étudier le sens des termes lexicaux, mais aussi la forme des unités lexicales et même aux relations existantes entre le lexique et la syntaxe. D'après MEL'CUK et al « *la lexicologie est une discipline théorique qui a pour objet l'étude générale du lexique* » (2002).

La lexicologie est historique et descriptive et peut être orienté vers une ou plusieurs langues. La lexicologie historique est l'étude diachronique de l'évolution du vocabulaire depuis l'origine d'une langue jusqu'à nos jours.

La lexicologie descriptive s'intéresse au vocabulaire d'une langue dans le cadre d'une période déterminée, étude synchronique.

Une approche diachronique permet d'expliquer l'état actuel du vocabulaire, tandis qu'une approche synchronique permet de révéler les facteurs qui conditionnent son mouvement progressif. Par conséquent, la synchronie est étroitement liée à la diachronie.

Un système linguistique représente un ensemble d'unités hiérarchiques, distinguant les phonèmes, les morphèmes, les mots et les propositions, des unités les plus simples aux plus complexes. Chacun de ces types d'unités forme un niveau dit structurel. Il s'agit respectivement des niveaux phonologique, morphologique, lexical et syntaxique.

La lexicologie étudie les unités de niveau de lexique : Un mot et son équivalent fonctionnel. Cette branche est étroitement liée à la morphologie et à la syntaxe. Bien qu'elle s'intéresse au rôle des mots dans l'enrichissement du vocabulaire, la morphologie observe les spécificités des parties du discours et unifie en fait la phonétique (phonologie).

1- Le mot

Dans le langage courant, le mot est un ensemble de lettres correspondant à un sens dans le langage, séparé à gauche et à droite par un blanc des autres éléments du texte, il marque plusieurs définitions « *Le mot est en grammaire traditionnelle l'unité linguistique minimale dotée de signification. Groupe de phonèmes perçu comme autonome, il est à l'oral isolable par les pauses et les accents et à l'écrit repéré par des blanc typographiques* » (C.Détrie, P Siblot, B.Verine,).

ANTOINE MILLET conçoit le mot comme « *un résultat de l'association d'un sens donné à un ensemble de sens donné susceptible d'un emploi grammatical donné* »

Le mot lexical (lexème), est celui qui fournit l'entrée du dictionnaire. Il apparaît comme une forme basique du mot graphique, dépourvu de toutes les variations formelles possibles : l'adjectif est signalé au masculin singulier, le verbe à l'infinitif... c'est ce mot lexical qui sert de base au classement des parties du discours. Le mot graphique correspond à une suite de lettres entre deux blancs, en effet un même mot graphique peut renvoyer à plusieurs mots grammaticalement différents. Or, le mot phonétique renvoie à une suite de sons entre deux pauses. Le problème que pose une telle définition est que les mots phonétiques ne correspondent pas nécessairement aux mots graphiques.

Certains mots ont une signification émotionnelle et aident à traduire les émotions humaines, les attitudes émotionnelles face à la réalité. Ce sont des mots à fonction expressive. La plupart des mots autonomes ont la capacité d'exprimer des idées et de nommer des objets et leurs indices.

a/ Mot plein et mot vide

En linguistique, on parle de mot vide en s'attachant à son sens. En général, il s'agit des mots-outils d'une langue, ses mots grammaticaux (déterminants, pronoms, prépositions, conjonctions, etc.), en opposition aux mots pleins, les mots lexicaux (noms, adjectifs, verbes, adverbes), ceux qui donnent du sens, qui décrivent le monde. Les mots vides ne servent qu'à l'articulation dans le texte, et parfois ils indiquent le genre, le nombre, l'intensité, la manière, etc.

b/ Mot simple et mot composé

On dit qu'un mot dérivé est un mot simple et un mot composé est un mot construit, or un mot composé est formé à partir de deux mots simples. Les mots français sont divisés en mots simples et en mots composés. Les linguistes sont habitués à conserver les unités lexicales qui ne peuvent pas être décomposées en unités plus petites significatives comme de simples mots. En revanche, les mots construits sont composés de deux ou plusieurs morphèmes ou mots. Certains termes sont formés par dérivation, d'autres par composition. Vous pouvez également rencontrer des phrases lexicalisées et des constructions qui corrigent les structures syntaxiques.

c/ Lexie et lexème

La lexie représente des unités fonctionnelles significatives du langage et peuvent être des mots simples (morphèmes lexicaux ou grammaticaux) ou des mots composés. Il contient plusieurs mots (composés) car il peut s'agir d'un vocabulaire complexe qui implique un ordre fixe (formule). Contrairement au lexème qui est une unité abstraite appartenant à la langue, un morphème qui porte une signification lexicale. Il peut être un composant de la lexie, comme il peut se trouver à l'état libre.

d/ Le sème

En linguistique, le sème est la plus petite unité de sens non susceptible, une caractéristique sémantique pertinente lors de l'analyse du sens des mots. Ce dernier peut contenir un ou plusieurs sèmes. Contrairement avec les morphèmes, qui représentent des signes linguistiques autonomes. Les Sèmes ne se réalisent pas seulement au sein de constructions sémantiques, c'est-à-dire des faisceaux de sèmes qui caractérisent la signification des morphèmes, appelés sémèmes. Selon la définition de Pottier le sème est un « *trait distinctif de la substance du signifié d'un signe (au niveau du morphème), et relativement à un ensemble donné de signes* » (1974 p. 330), quant au sémème est « *l'ensemble des sèmes d'un signe, au niveau du morphème dont c'est la substance du signifié* » (1974 p. 331).

2- Le lexique

Le lexique est divisé en classes fonctionnelles appelées "parties du discours". Ils portent sur la nature du langage par opposition à sa fonction. Un lexique représente un ensemble d'unités qui constituent son matériau de base. Les chercheurs n'ont aucun moyen valable de répertorier le lexique d'une langue à un moment donné. On estime que le dictionnaire d'une langue civilisée comme le français contient plus de 200 000 mots, sans compter les noms propres.

Le lexique est constitué de toutes les occurrences possibles et de toutes les formes segmentaires qui composent l'ensemble des occurrences (ou formes lexicales) potentielles ou réalisées de la langue. Un caractère ou une série de caractères non séparés par des espaces et séparés par deux espaces sont appelés formes ou termes.

Le lexique implique la connaissance des règles de formation des mots, c'est-à-dire des règles morphologiques liées à la grammaire

D'après Gracia-Debanc, Masseron, ronveaux dans l'ouvrage intitulé "Enseigner le lexique"
« *Le lexique (en langue) comme système de relations entre unités peut difficilement se dispenser d'un enseignement explicite, progressif et doté d'une terminologie spécifique qui excède les seuls emprunts à la rhétorique et à ses figures de mots.*» (2014)

En fait, le lexique est bien plus qu'une simple liste alphabétique. Il est structuré à la fois sur le plan sémantique et formel. La sémantique lexicale examine la construction sémantique du vocabulaire, analyser le sens des mots et les relations significatives entre les mots. La

morphologie lexicale examine l'organisation formelle du lexique. Il analyse la structure des mots et les relations formelles et sémantiques entre les mots. Ensemble, ces deux aspects forment une structure complexe qui est incorporée dans l'ensemble du système linguistique.

Le lexique est lié à la syntaxe car les mots sont utilisés dans des phrases. Le sens d'un énoncé ne découle pas uniquement du sens des mots qui le composent. Cela dépend également du contexte de la syntaxe de la phrase et de la prononciation. La sémantique lexicale fait donc partie de la sémantique, c'est-à-dire de la théorie sémantique.

3- Le champ lexical et champ sémantique

En lexicologie, le terme "champ" est utilisé pour désigner la structure d'un domaine particulier de la langue.

Les deux termes « champ sémantique » et « champ lexical » sont souvent confondus. En général, la notion "champ sémantique" est réservée à la caractérisation de traits spécifiques des termes lexicaux, et le champ lexical a pour but de décrire les relations entre plusieurs éléments lexicaux.

Un champ lexical est défini comme un ensemble de mots appartenant au même sujet. Par conséquent, il combine une sémantique commune, c'est-à-dire des noms, des adjectifs et des verbes liés par le même sens. Il ne doit pas être confondu avec les mots de la même famille. Il peut également avoir des significations implicites ou explicites selon le contexte dans lequel il est présenté. En d'autres termes, le champ lexical résume tous les mots directement ou indirectement liés au sujet.

A l'inverse, le champ sémantique résume tous les sens donnés à un mot particulier. Il s'agit de polysémie, cela signifie que les mots prennent différentes significations dans une phrase en fonction du contexte. Comme il peut être défini comme l'ensemble des sens qu'un mot ou un groupe de mots peut prendre dans différents contextes.

II/ Autour de l'analyse du discours

L'analyse du discours est un domaine des sciences du langage qui traite des unités textuelles dans leurs rapports à leurs conditions de production. Ce concept est une expression qui a été employée pour la première fois par Z.S. Harris en 1952 pour désigner une étude

distributionnelle des unités transphrastiques. P. Charaudeau et D. Maingueneau rapportent l'analyse du discours à la relation entre texte et contexte et excluent les travaux de pragmatique qui portent sur des énoncés décontextualisés (en dehors de tout contexte). Vers le milieu des années soixante, on voit émerger les courants qui vont constituer l'actuel champ de l'analyse du discours comme : l'analyse conversationnelle, l'ethnographie de la communication, la pragmatique, les théories de l'énonciation et la linguistique textuelle.

a/ Discours et texte

Le concept de discours est très ancien, existant depuis l'époque de la philosophie classique et a été proposé et mis en avant par G. Guillaume. Il est défini par opposition à d'autres termes de la linguistique classique.

Pour Benveniste, la notion de discours est proche de celle d'énonciation : « *c'est la langue en tant qu'assumée par l'homme qui parle, et dans la condition d'intersubjective qui seule rend possible la communication linguistique* ». J.M. Adam définit le texte comme étant le discours dépourvu de son contexte. Selon D. Maingueneau, on a tendance à parler de texte pour « *des productions verbales orales ou écrites et qui sont structurées de manière à durer, à être répétées, à circuler loin de leur contexte originel* ».

En termes académiques, un texte est essentiellement tout ce qui a du sens pour ceux qui l'étudient. Il présente les statistiques.

b/ Analyse textuelle

L'analyse de données textuelles, inspirée de la linguistique structurale et de l'analyse du discours, l'analyse des données textuelles est à la fois qualitative et quantitative. L'objectif est de qualifier des éléments textuels à l'aide de catégories et de les quantifier en analysant leurs distributions statistiques. Cette approche a été fortement influencée par les travaux de Jean-Paul Bendskrill et a été appliquée aux corpus de textes littéraires et politiques depuis les années 1960 .

L'analyse de données textuelles (ADT) se situe au carrefour de plusieurs disciplines, notamment : la linguistique, l'analyse du discours, les statistiques, l'informatique, le traitement de la recherche et la psychosociologie.

L'analyse textuelle est un terme utilisé pour étudier et comprendre des textes. Cela comprend l'exploration du langage, des symboles, des motifs et des images dans le texte, en d'autres termes, l'analyse textuelle est l'analyse statistique d'un corpus, un ensemble de documents en langage naturel.

L'analyse de données textuelles est défini principalement par son objet, ou "données textuelles". Par conséquent, la spécification ADT inclut simplement toutes les méthodes qui traitent le texte comme des données analytiques.

La raison en est que l'analyse textuelle est en effet un domaine très dynamique et en même temps un domaine très fragmenté et plus spécifiquement segmenté.

c/ Etude comparative

Une étude comparative analyse et résume les similitudes, les différences et les tendances entre deux ou plusieurs cas qui partagent un intérêt ou un objectif commun.

En d'autres termes, une étude comparative est une étude qui compare deux éléments ou plus pour déterminer leurs similitudes et leurs différences. Il peut être appliqué dans tous les domaines.

L'objectif général de cette recherche est de comprendre comment et pourquoi des concepts sont similaires ou dissemblables. Ceci nous permet de tirer des conclusions de cette comparaison méthodologique.

III/ Autour de la lexicométrie

a/ Histoire et évolution

Le développement de l'analyse automatique est passé par plusieurs étapes et a dû surmonter des obstacles épistémologiques et matériels pour arriver à son état actuel. Les premiers projets d'analyse automatisée du discours remontent aux Italiens Roberto Busa et Georges Kingsley Zipf.

La lexicométrie s'est développée en France depuis les années 1970, et les travaux de statisticiens comme Étienne Brunet, Charles Muller, Pierre Guiraud, Maurice Tournier ou Jean-Paul Benzcrill sont à l'origine de l'analyse factorielle des correspondances (Merchant

2013). Cependant, l'essor de cette méthode est lié à l'essor des outils numériques et au développement de la puissance de calcul des ordinateurs. Les ordinateurs facilitent la vulgarisation des recherches dans les dictionnaires. Deux auteurs ont contribué à la description et à la conception.

Ludovic Lebart et André Salem ont combiné deux livres de base sur les méthodes utilisées.

En 1988, "l'analyse statistique des données textuelles" a été produite, suivie en 1994 par "l'analyse statistique du texte". En parallèle, des logiciels libres sont développés pour mettre en œuvre ces procédures statistiques. B.Iramtek, 2007. Cependant, l'analyse lexicographique a révolutionné l'étude de l'analyse du discours. En effet, le développement d'outils informatiques de plus en plus performants permettant l'analyse automatisée de textes a permis de traiter des corpus écrits et parlés plus importants, ou macro corpus.

b/ Définition

La lexicométrie est une discipline qui prend en charge l'analyse informatisée du discours et du lexique. Cette jeune discipline est appelée également « analyse du discours assistée par ordinateur », ou encore « traitement automatisé du discours » G. Mounin, la définit comme le « *Domaine de la lexicologie dans lequel les procédures de la statistique sont utilisées pour l'étude quantitative du lexique* » (1974 : 203), Pour Leimdorfer et Salem, on regroupe sous le terme de lexicométrie « *toute une série de méthodes qui permettent d'opérer, à partir d'une segmentation, des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire* » (1995 : 133). Enfin, pour Maingueneau, il s'agit d'une « *Discipline auxiliaire de l'analyse du discours qui vise à caractériser un ensemble discursif (souvent un positionnement) par rapport à d'autres appartenant au même espace grâce à l'élaboration informatique de réseaux quantifiés de relations significatives entre ses unités. Il s'agit par conséquent d'une démarche essentiellement comparative* » (2009 : 81).

La lexicométrie s'intéresse au discours à travers l'analyse du lexique. Pour ce faire, elle utilise des méthodes et des outils mathématiques et informatiques.

Il existe de nombreux logiciels de statistiques lexicales, mais les plus connus dans le monde francophone sont Hyperbase d'Etienne Brunet et Lexico d'André Salem. Citons aussi : Cordial, Sphinx, Alceste, Tropes, Sato, Weblex, Prospero, Astarte et plus encore. L'analyse lexicographique a été divisée en deux divisions principales depuis sa création. La première a

été fondée par Charles Müller à l'Université de Strasbourg et est appelée en raison de son intérêt pour le traitement statistique des textes littéraires. La revue *Lexicometrica* est aujourd'hui le principal support de diffusion de ses recherches. La lexicométrie littéraire accorde une attention particulière au style de l'auteur, à la richesse et au développement du vocabulaire. La seconde, la lexicologie politique se développe vers la fin des années 1960, grâce aux travaux de Maurice Tournier au Département de lexicographie et documents politiques de l'ENS de Fontenay-Saint-Cloud.

La lexicométrie est chargée « *d'examiner, à partir de corpus de textes soumis à comparaison, comment les termes échangés dans l'espace public autour des enjeux de pouvoir rendent compte des luttes d'appropriation ou de dépossession symboliques qui se jouent dans le lieu même de l'échange* » (Bonnafous et Tournier, 1995 : 69).

IV/ Le logiciel Iramuteq

Iramuteq est un logiciel libre distribué sous licence GNU GPL (v2), développé par Pierre Ratinaud. Il s'agit donc d'une interface R qui fournit plusieurs outils d'analyse visant à analyser des corpus et des matrices.

Le logiciel IRaMuTeQ permet d'effectuer des analyses statistiques sur des corpus textuels et des tableaux basés sur la méthode de Max Reinert (classification hiérarchique descendante des tableaux à travers les formes pleines et les segments de texte).

En 2015, Cécile Alduy et Stéphane Wahnich ont utilisé IRaMuTeQ pour créer le livre "Marine Le Pen prix aux mots". "Décoder le nouveau discours de première ligne".

IRaMuTeQ s'appuie sur les langages libres R et Python développés à l'Institut des sciences sociales et de la recherche appliquée (LÉRASS, Université de Toulouse) et soutenus par le LabEx "Structurés des mondes sociaux".

a/ Occurrence et cooccurrence

L'occurrence est l'apparition d'une unité linguistique dans le corpus, selon Gorge Mounin « *l'occurrence d'une unité linguistique qui, étant réalisée dans le discours, et observable dans ses particularités dénombrables dans un corpus* » (Dictionnaire de la linguistique ,2003).

L'occurrence peut être définie comme 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.

Quant à la cooccurrence est notamment l'apparition simultanée de deux ou plusieurs éléments ou classes d'éléments dans le même discours.

b/ Le segment

Le segment est l'unité dégagée par la segmentation, « *Le segment peut être un phonème, un morphème, un syntagme, et même une phrase. La phrase est alors considérée comme segment de niveau supérieur et le phonème le segment de niveau inférieur. C'est pour désigner un morphème que le terme segment est le plus employé* » (D. D. L. 1976).

La segmentation est une opération qui consiste à découper un énoncé ou une séquence afin de faire apparaître les éléments qui le constituent, « *La segmentation est à la base de l'analyse structurale. (...) une phrase peut être segmentée en syntagmes, un syntagme en morphèmes, un morphème en phonèmes. Limites: la segmentation ne rend pas compte de l'amalgame (au ne peut être segmenté)* » (D.D.L, 1976).

V/ Le système éducatif en Algérie

Le système éducatif algérien est divisé en plusieurs niveaux : l'enseignement préparatoire, l'enseignement fondamental (primaire et collège), l'enseignement secondaire, l'enseignement professionnel et enfin l'enseignement supérieur.

C'est l'ordonnance du 16 avril 1976 qui organise l'éducation et la formation en Algérie. A partir de 1972, date du début de l'arabisation de l'enseignement, le français a reçu le statut de langue étrangère ou seconde. .

L'éducation de base est une étape d'enseignement obligatoire de neuf ans. À partir de l'année scolaire 2003/2004, le système comprend cinq années d'enseignement primaire et quatre années d'enseignement secondaire. L'éducation est un droit fondamental, un puissant outil de développement et l'un des meilleurs moyens d'élever les normes et de promouvoir la paix et la stabilité.

L'école est une institution chargée de dispenser une formation générale collective aux enfants d'âge scolaire et préscolaire, d'instruire les élèves, de leur transmettre des connaissances et de

leur apprendre tout ce dont ils ont besoin pour grandir et se développer, s'épanouir : lire, écrire, compter, comprendre comment le monde est né, comprendre l'histoire humaine, maîtriser la technologie.... C'est un milieu de vie où les apprenants interagissent et vivent des situations variées, offrant de nombreuses occasions d'apprendre à vivre en société.

L'école favorise la cohésion sociale par la diffusion d'une culture et d'une langue communes, elle contribue à l'accession à un statut professionnel et permet à chacun d'obtenir des qualifications qui assurent sa place dans la société par son travail.

VI/ Ecole publique et école privée

L'enseignement en Algérie est dispensée différemment dans les deux institutions. Dans les écoles publiques, l'État couvre tous les frais nécessaires. En revanche, dans les écoles privées les études sont payantes par les parents d'élèves, c'est eux qui prennent en charge tous les frais scolaires de leurs enfants.

Les programmes diffusés dans les deux établissements ne sont pas les mêmes. Chaque établissement d'enseignement a ses propres méthodes d'enseignement et ses propres principes. Les écoles privées suivent le programme des pays européens comme la France. A l'opposé des écoles publiques qui suivent un programme national incitant les parents à choisir entre deux établissements d'enseignement, en effet la langue d'enseignement est le français, une langue qui leurs facilitera la communication orale et l'ouverture sur un monde meilleur.

Il n'y a pas si longtemps, les écoles privées ont été officiellement créées par le gouvernement algérien.

Selon le ministère de l'Éducation nationale, le nombre d'établissements d'enseignement est de 28 839, dont 20 272 écoles primaires, 5 909 universités et 2 658 écoles secondaires. Quant aux écoles privées, il y a 310 écoles, 190 écoles dans la Wilaya d'Alger et 7 écoles au niveau de la Wilaya de Bejaia, parmi ces écoles nous avons choisi école « Les Iris».

En 1990, le développement du secteur privé a commencé au sein du système éducatif algérien et en 2005, l'État a décidé de formaliser les établissements privés selon les conditions de l'État. L'enseignement privé est considéré comme complémentaire de l'enseignement public. Selon l'article de M. SAADOUNE, la mise en garde lancée par BOUTEFLIKA en avril 2005 « il est tout à fait clair que toute institution privée qui ne tient pas compte du fait que l'arabe

est la langue nationale et officielle, et qui ne lui accorde pas une priorité absolue, est appelée à disparaître.» (Le quotidien d'Oran, 26 février 2006).

Les écoles privées sont des écoles payantes qui ne sont pas contrôlées par les gouvernements locaux, étatiques ou centraux et conservent le droit de choisir leurs élèves. Ces derniers sont souvent issus de milieux privilégiés et sont souvent bien préparés à la réussite scolaire. Contrairement aux écoles publiques relevant du ministère de l'Éducation nationale, son objectif est de fournir aux enfants de citoyens algériens une éducation de base universellement accessible et une éducation gratuite et financée par l'État.

VII/ Ecole « Les Iris »

L'établissement d'éducation et d'enseignement privé « les Iris » au centre de la wilaya de Bejaia « Bd Krim Belkacem, Bejaia 06000 » Exerce sa mission d'enseignement et d'éducation dans le cadre d'un réseau d'associations depuis juillet 2005.

Il compte parmi les premiers établissements agréés et autorisés à ouvrir par le ministère de l'éducation nationale, à compter de septembre 2005 juste après la promulgation du décret exécutif numéro 04-90 du 24 mars 2004 fixant les conditions de création d'ouverture et de contrôle des établissements privés, sous l'autorité de Monsieur AZZOUG Ahmed, Gérant de la SARL les IRIS. et sous la direction de l'épouse d'AZZOUG, LEHCANI Sihem. Cette école est autorisée à accepter des élèves des niveaux suivants : Préscolaire, primaire et collège dès l'obtention de son agrément. L'école exploite un manuel de français appelé "Fleur d'Encre".

VIII/ Le statut de la langue française en Algérie

L'Algérie est un pays francophone depuis la colonisation (1830 - 1962) jusqu'à présent, d'après Grandguillaume .G « la langue française, la langue officielle durant la colonisation, a conservé une place importante : celle-ci s'est même accrue après l'Indépendance, du fait de l'extension considérable de la scolarisation, et la place qu'elle a toujours conservée dans l'enseignement, selon des proportions variables »

En effet, la langue française est considérée comme la première langue étrangère dans la société algérienne, elle tient une place supérieure dans plusieurs secteurs tels que l'enseignement, elle est enseignée dès la troisième année primaire. Selon R.Sbaa « la réalité

empirique indique que la langue française occupe en Algérie une situation sans conteste, unique au monde. Sans être la langue officielle, elle véhicule l'officialité, sans être la langue d'enseignement, elle reste une langue de transmission du savoir, sans être la langue d'identité, elle continue à façonner de différentes manières et par plusieurs canaux, l'imaginaire collectif » (Culture et plurilinguisme en Algérie).

a/ La place de la langue française à Bejaia

Après l'indépendance, la ville de Bejaia est connue par sa civilisation ; elle se caractérise par un multilingue (berbère, arabe, français,...) dont le français occupe une place importante chez les locuteurs bougeottes, il est pratiqué sans aucune difficulté dans leur vie quotidienne et professionnelle.

Aujourd'hui, l'usage de cette langue est concentré dans divers enseignement et disciplines, en particulier dans le secteur de l'enseignement public et privé. Ce dernier considère le français comme la langue d'instruction et d'éducation. Par conséquent, le niveau de cette langue s'améliore constamment, tant à l'oral qu'à l'écrit.

IX/ Le manuel scolaire

Le manuel scolaire est un élément central de la pratique de l'enseignement, il est considéré comme l'un des facteurs les plus efficaces pour améliorer la qualité de l'éducation, en particulier dans les Etats dont les systèmes éducatifs manquent de ressources. En effet des recherches ont montré que l'amélioration de la disponibilité des manuels scolaires a un impact positif sur les performances scolaires (Seguin 1989), les manuels sont donc un outil de réussite scolaire. Pour Gérard et Roegiers, ils sont *«intentionnellement structuré pour s'inscrire dans un processus d'apprentissage en vue d'en améliorer l'efficacité»* (1993, 35), Cela se fait en fournissant aux apprenants un corpus de connaissances pour découvrir, apprendre et comprendre de nouvelles choses. Cela permet aux enseignants d'aider les élèves à suivre une leçon ou une série d'exercices, tout en permettant aux parents de soutenir et de surveiller l'apprentissage de leurs enfants.

Cependant, cet outil pédagogique de base n'est pas un moyen facile de diffuser les connaissances car son contenu n'est pas seulement consacré à l'enseignement mais également à l'éducation« *par transmission, de manière plus ou moins explicite, de modèles de*

comportement sociaux, de normes et de valeurs.» (UNESCO, 2008, 14). C'est-à-dire qu'il a à la fois un côté cognitif et un côté intellectuel, c'est là que les manuels ont un pouvoir considérable. « A travers les connaissances ; mais à travers aussi les opinions sur l'organisation sociale et politique d'un pays, le manuel comporte des appels qui s'adressent à l'élève, lui suggèrent ce qu'il faut aimer et respecter et ce qu'il faut haïr et mépriser. Le manuel participe ainsi à la construction du système de valeurs de l'élève. Il lui transmet des modèles d'identification, lui trace des idéaux et l'oriente, contribuant ainsi au renforcement de son Surmoi» (Ansart 1984, cité par Cromer et Hassani - idrissi, 2011, 2).

a/ L'importance du manuel scolaire

L'usage d'un manuel scolaire joue un rôle très important dans l'enseignement des langues. C'est un lien entre les enseignants et les élèves, il est considéré aussi comme étant la base et le noyau de cet enseignement/apprentissage (Jean-Pierre Cuq 2003).

En d'autres termes, le manuel procure une représentation claire et nette du programme et aide les enseignants et les apprenants à se positionner par rapport au processus d'enseignement dans la mesure les enseignants savent quoi et comment enseigner, et l'élève pourra se référer plus facilement à tout ce qu'il a appris. De plus cet outil constitue une riche bibliothèque de documentation d'exemple, d'exercices et d'activités qui développe l'interaction entre l'enseignant et les élèves en classe ainsi que l'interaction entre les élèves eux-mêmes. En fait il est important d'indiquer que l'emploi du manuel scolaire ne désigne pas forcément le respecter entièrement parce que le manuel scolaire n'est qu'un indicateur et un déclencheur des activités et des situations didactique et pédagogiques (J.P Robert, 2002)

Conclusion partielle

Dans ce premier chapitre, nous avons mis l'accent sur quelques notions de base, tels que la lexicologie, la lexicométrie et l'analyse du discours. En premier lieu, nous avons pu définir la lexicologie, son objet d'étude et tout ce qui est en relation avec cette discipline, suivie d'une partie consacré à la lexicométrie où nous avons présenté le logiciel iramuteq et ses différentes analyses.

En deuxième lieu nous avons expliqué le système éducatif, l'école publique et l'école privée, ainsi que le statut de la langue française en Algérie en général et à Bejaia en particulier.

Enfin, nous avons tenu à exposer le manuel scolaire et son importance dans l'enseignement.

CHAPITRE 02:
ANALYSE DE DONNÉES

Introduction partielle

Après avoir terminé le premier chapitre qui est consacré à la théorie, nous allons passer au second chapitre qui est destiné à la pratique, dans cette partie nous abordons la description des manuels de l'école privée et l'école publique de la première année moyenne, en effet nous allons effectuer une analyse lexicométrique sur ce corpus textuel par un logiciel des statistiques textuelles « Iramuteq ».

Pour cela, nous les comparons d'abord par rapport à la richesse, à la variété de ces supports et enfin à leur contenu.

I/ Présentation de corpus

Notre corpus s'agit de deux manuels de l'école publique et l'école privé (version PDF), que nous allons présenter comme suit :

1- Le manuel de l'école publique (figure 01)

Le nouveau manuel de 1AM moyenne édité par ENAG (Entreprise Nationale des Arts Graphique) est organisé par quatre auteurs ANISSA MADAGH Inspectrice de l'éducation et de l'enseignement moyen, CHAFIK MERAGA professeur de français cycle moyen HALIM BOUZELBOUDJEN professeur de français cycle moyen.

Ce manuel constitue 179 pages, il est en concordance total avec les programmes et organise donc les apprentissages selon la pédagogie du projet à l'exemple de tous les manuels algériens, il englobe trois projet, chaque projet se déclinant en séquence

- Le projet 1 s'intitule « afin de célébrer les journées mondiales de la propreté »

Ce projet est subdivisé en trois séquences :

- La séquence 1 s'intitule : « J'explique l'importance de se laver correctement.
 - La séquence 2 s'intitule : « J'explique l'importance de manger convenablement.
 - La séquence 3 s'intitule : « J'explique l'importance de bouger régulièrement.
- Le projet 2 s'intitule « Je réalise avec mes camarades un dossier documentaire. »

Ce projet est subdivisé en trois séquences :

- La séquence 1 s'intitule : « J'explique les progrès de la science. »
- La séquence 2 s'intitule : « J'explique les différents pollutions. »
- La séquence 3 s'intitule : « J'explique le dérèglement du climat. »
- Le projet 3 s'intitule : « Sous le slogan pour une vie meilleur, je réalise avec mes camarades. » ce projet est subdivisé sur deux séquences
- La séquence 1 s'intitule : « J'incite à l'utilisation des énergies renouvelables. »
- La séquence 2 s'intitule : « J'agis pour un comportement éco-citoyen. »



-Figure 01-

2- Le manuel de l'école privée (figure 02)

Le manuel utilisé dans l'école privée « Les Iris » s'intitule « Fleurs d'encre » emporté de France, édité par deux auteurs ; Chental Bertagna, professeur en collège et à l'IUFM, et Françoise Carrier-Nayrolles, professeur en collège et coordinatrice académique de lettres.

Ce livre contient 383 pages et composé de quatre parties, et chaque partie se subdivise en séquences.

- La première partie « Récits d'aventures et de voyages », repartie en trois séquences :
 - L'appel de l'aventure.
 - Aventuriers : de la réalité au mythe.
 - Sur les traces de Marco Polo.
- La deuxième partie « Preux chevaliers et gentes dames » avec deux séquences :
 - Les romans de chrétien de Troyes.
 - Le mythe de Tistan et Iseult.
- La troisième partie « Le rire au moyen Age », cette partie englobe deux séquences :
 - Bourgeois, prêtres et vilains dans les fabliaux.
 - Le Roman de Renart.
- La quatrième partie « Divertissements littéraires », divisé e, trois séquences :
 - Maitres, valets et servantes.
 - Au fil de temps et des saisons.
 - Les artistes à la cour.



-Figure 02-

Description

Selon ces deux brèves présentations des deux manuels, nous constatons que le nombre de textes, de projets ainsi que la thématique sont différents pour chaque programme, notant que la plus part des textes du manuel public sont scientifique tout en gardant le même genre contrairement à celui du privé où les éléments culturels sont plus riches dont chaque projet traite un genre différent, à titre d'exemple : le poème, le théâtre, le mythe,...

C'est pour cela que le programme de l'école privée est le plus riche, grâce notamment à leurs supports qui sont variés par rapport au celui du public.

II/ Cadre méthodologique

La lexicométrie regroupe « toute une série de méthodes qui permettent d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire à partir d'une segmentation » (Salem, 1986).

Le principe lexicométrique de base consiste à examiner un corpus indexé et partitionné à partir d'un nombre limité de variables, analyser les caractéristiques de différentes partitions et vérifier sa pertinence statistique.

La démarche lexicométrique contrastive permet de comparer le lexique de plusieurs textes, de plusieurs parties d'un même texte et de mesurer ce lexique sur la base d'un corpus cohérent ; c'est-à-dire la taille des différentes parties déterminée par le nombre des mots de ces dernières. En effet, nous avons choisi une méthode lexicométrique comparative de deux manuels scolaires par un logiciel de statistique « Iramuteq ».

1- Le logiciel « Iramuteq »

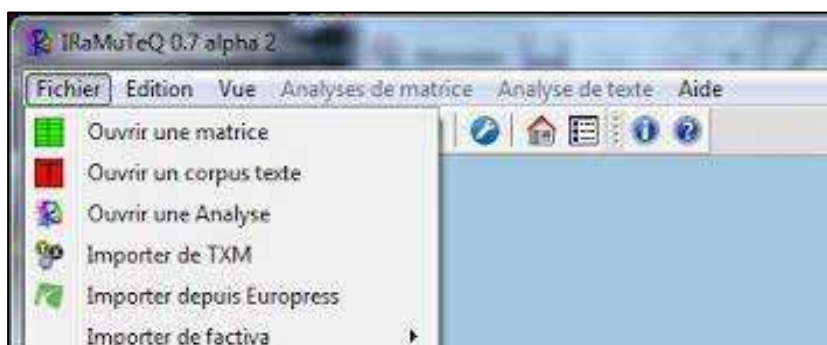


Le centre de la lexicométrie et de l'analyse automatique des textes a été créé à 1997 à Paris 3 ; Iramuteq est un logiciel libre développé par P.Ratinaud, conçu pour le traitement lexicométrique de textes.

Iramuteq permet de faire des analyses statistiques de corpus textuels, basé sur une segmentation du texte à analyser, le logiciel donne le moyen d'avoir une visibilité globale des termes contenus par les textes qui lui est soumis à l'aide des représentations graphiques.

Ainsi, ouvrir un corpus texte permet de charger un fichier texte qui respecte la mise en forme, comportant des lignes étoilées entre chaque réponse.

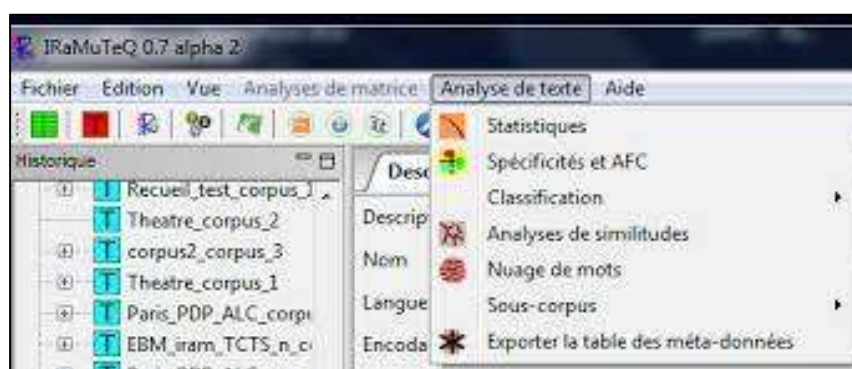
Le texte à analyser ne doit pas comporter d'étoiles, car le caractère * est réservé aux caractéristiques sur les textes dans la ligne étoilée dédiée. Une fois le nom du fichier renseigné, IRaMuteQ propose de paramétrer la transformation du corpus pour effectuer l'analyse de texte.



-Figure03-

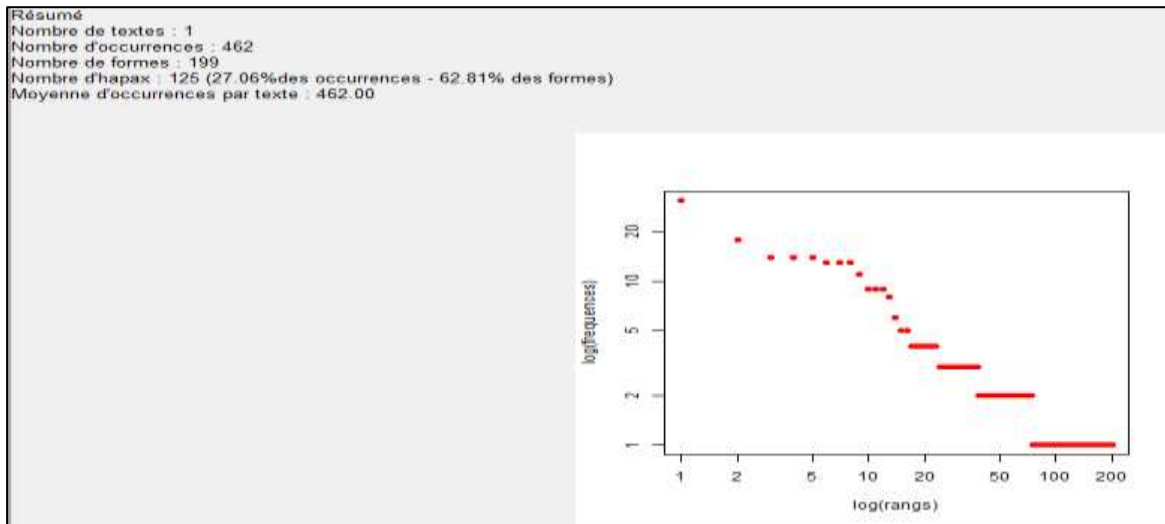
2- Les différentes analyses du logiciel

Ce logiciel propose différents types d'analyses basées sur : statistiques textuelles, spécificités et analyses factorielles des correspondances (AFC), classification méthode Reinert, analyse de similitude et nuage des mots, comme le montre la figure ci-dessous (Figure 04). Notre recherche s'intéresse notamment à quatre analyses essentielles : statistiques textuelles, classification méthode Reinert, analyse de similitude et nuage des mots.



-Figure 04-

2.1/ Statistiques textuelles



-Figure05-

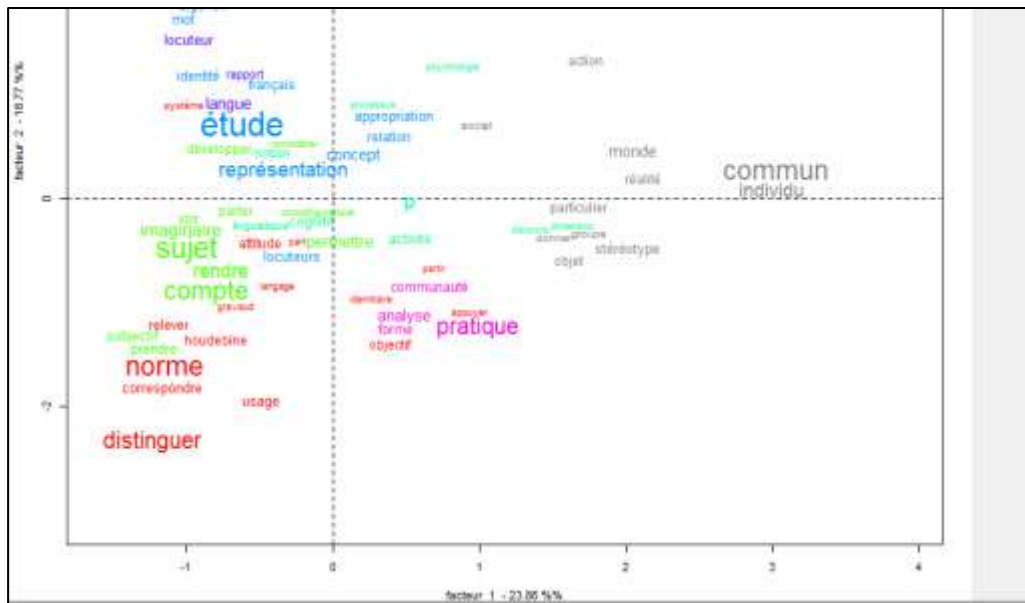
Cette analyse propose des statistiques simples sur les corpus texte : effectif des formes actives, supplémentaires et liste des hapax, sous forme d'un logarithme.

- ❖ **Les formes actives** : ou mots pleins, sont les formes qui constituent des noms et des verbes.
- ❖ **Les formes supplémentaires** : ou mots outils, présentent des mots de liaison (de, et, en, etc.) ou des déterminants (la, le, etc.).
- ❖ **Les hapax** : sont les formes qui n'apparaissent qu'une seule fois dans le corpus les, c'est-à-dire les formes dont la fréquence est 1.

Dans cette analyse, IRaMuTeQ affiche tout le lexique du corpus, il faut d'abord choisir de lemmatiser ou non les formes/mots et paramétrer les catégories de mots à prendre en compte dans les calculs.

La lemmatisation selon Mounin est une « opération consistant à regrouper les formes occurrentes d'un texte ou d'une liste sous des adresses lexicales » (1974), les occurrences sont réduites à leur racine ; les verbes sont ramenés à l'infinitif, les noms au singulier et les adjectifs au masculin singulier.

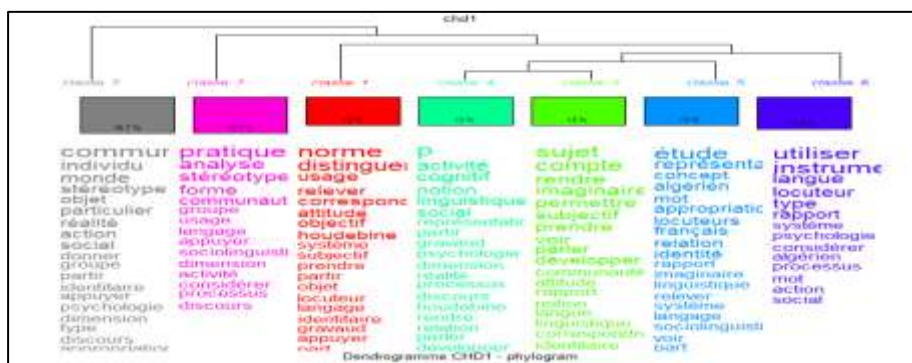
2.2/ Spécificités et AFC



-Figure06-

L'analyse factorielle des correspondances est une méthode statistique d'analyse des données qui permet d'analyser et de hiérarchiser les informations contenues dans un tableau rectangulaire de données, l'AFC sert à analyser le lien entre deux variables qualitatives, or elle produit une analyse factorielle des correspondances sur un tableau de contingence qui croise formes actives et les variables.

2.3/ Classification méthode Reinert

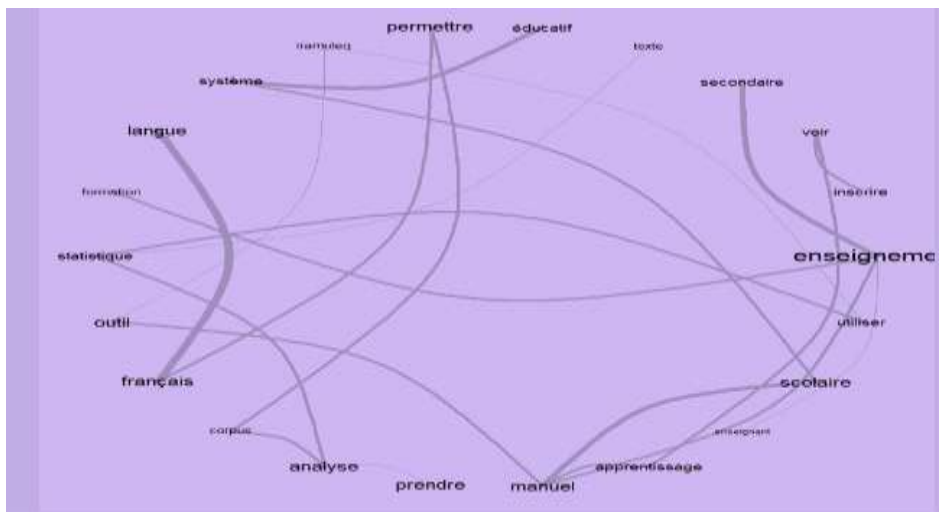


-Figure07-

Cette analyse décrite par Reinert (1983) propose une classification hiérarchique descendante, cette classification est proposée selon trois modalités : classification simple sur texte, classification simple sur segment de texte et classification double sur RST, elle consiste à décrire les lois de distribution du vocabulaire dans les textes.

La classification descendante hiérarchique (CDH) est une procédure croisant plusieurs techniques. Elle est principalement élaborée en relation avec une pratique d'analyse de données en psychosociologie, avec en particulier l'analyse des réponses libres à des questions ouvertes, Selon Max Reinert (1983, 1987, 1990) : « *La procédure descendante de la classification s'explique par le souhait de posséder des classes de cardinaux assez élevés, bien différenciées les unes des autres afin de les décrire correctement* »

2.4/ Analyses de similitudes



-Figure08-

L'analyse de similitudes est une méthode d'analyse des données s'inscrivant dans une approche issue de la théorie des graphes pour l'étude des relations entre les parties d'un ensemble sous forme d'arbres maximum.

Cette analyse produit des graphes à partir de la librairie « igraph » de R, la matrice de similitude est calculée à partir de l'un des indices proposés, la plupart de ces indices sont issus de la librairie « proxy » de R.

2.5/ Nuages de mots



-Figure09-

Le nuage de mots est une représentation graphique des termes du corpus rassemblé sous forme de nuage, il fait apparaître les mots essentiels qui caractérisent une recherche.

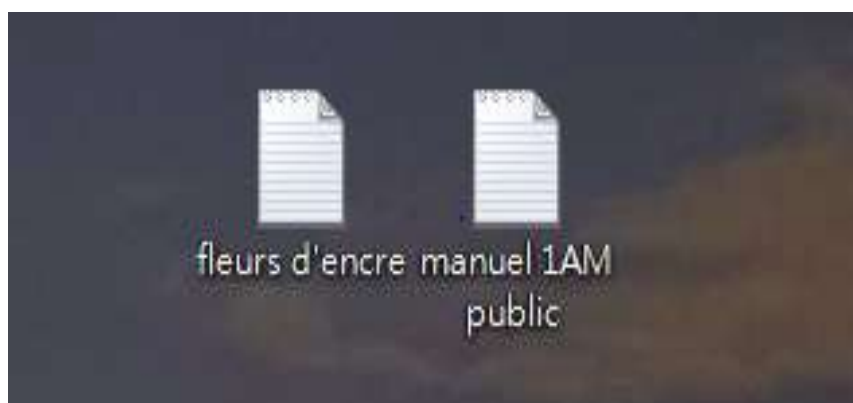
Les termes qui ont le plus grand nombre d'occurrences occupent des positions de choix dans ce nuage, plus un terme revient, plus il sera mis en valeur, ainsi la taille de chaque terme dans le nuage (police), détermine son occurrence.

III/ Analyse de données

1- Corpus

Après avoir rassemblé notre corpus composé de deux manuels scolaires privé et public en version PDF, l'ordinateur ne peut pas faire la différence entre les parties de ces corpus, il faut donc les lui indiquer par des balises installées par le logiciel R.

Nous avons donc été obligé de les convertir d'abord en fichier .doc, c'est-à-dire fichier Word, ensuite les transformer en fichier .Txt ; car les logiciels lexicométriques ne peuvent traiter que ce genre de fichiers.



-Figure 01- (fichier Txt)

2- Caractéristiques de corpus

Notre recherche s'inscrit dans le champ de la lexicométrie et englobe les différentes analyses mentionnées ci-dessus., Nous intéresserons donc à quatre analyses à savoir : statistiques textuelles, classification méthode Reinert, analyse de similitude et nuage de mots.

Après avoir analysé notre corpus dans le logiciel, nous constatons en premier lieu deux descriptions, ces dernières nous donnent des informations principales de ces manuels tel que le nom, la langue et l'encodage,...etc.

- Première description « manuel public »

The screenshot shows a window titled "Description manuel 1AM public_corpus_1" with the following data:

Description du corpus	
Nom	manuel 1AM public_corpus_1
Langue	french
Encodage	cp1252
originalpath	C:\Users\LENOVO\Desktop\manuel 1AM public .txt
pathout	C:\Users\LENOVO\Desktop\manuel 1AM public_corpus_1
date	Tue Mar 14 19:54:48 2023
time	0h 0m 6s
Paramètres	
ucemethod	1
ucesize	40
keep_caract	^a-zA-Z0-9àÀáÂãÄåÁäÅæÉëÊëËèÈìíîïóÔõÖöØøùÚúÛüÜçÇœƒ'ñÑ.:;!?'_-
expressions	1
Statistiques	
Nombre de textes	1
Nombre de segments de texte	823
occurrences	28951
Nombre de formes	4751
Nombre d'hapax	2486 - 52.33 % des formes - 8.59 % des occurrences

-Figure 02-

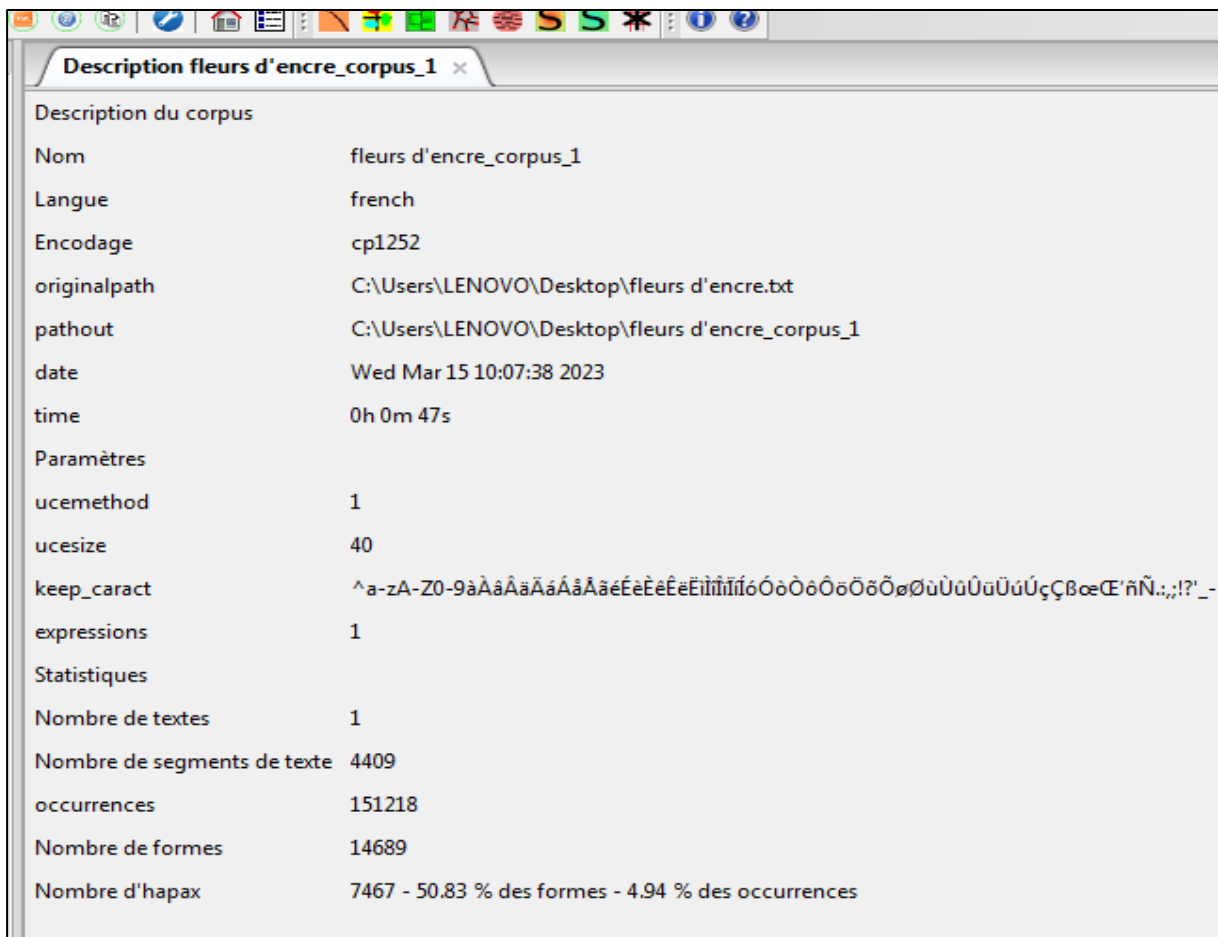
Nombre de textes	01
Nombre de segments de texte	823
Occurrences	28951
Nombre de formes	4751
Nombre d'hapax	2486

Tableau 01

Description

Le tableau 01, nous donne des informations sur notre premier corpus (le manuel public). Nous constatons que le nombre d'occurrences présentes (28951) suivi par le nombre de formes (4751), nombre d'hapax (2486) et enfin le nombre de segments de texte (823) sur un seul texte analysé.

- Deuxième description « manuel privé »



The screenshot shows a window titled "Description fleurs d'encre_corpus_1" with a list of corpus parameters and statistics. The window has a standard Windows taskbar at the top with various application icons.

Description du corpus	
Nom	fleurs d'encre_corpus_1
Langue	french
Encodage	cp1252
originalpath	C:\Users\LENOVO\Desktop\fleurs d'encre.txt
pathout	C:\Users\LENOVO\Desktop\fleurs d'encre_corpus_1
date	Wed Mar 15 10:07:38 2023
time	0h 0m 47s
Paramètres	
ucemethod	1
ucesize	40
keep_caract	^a-zA-Z0-9àÁâÃäÅæÉèÊëËìíîïïóÔõÖøÙúÛüÝÇßœƒ'ñÑ.,;:!?_-'
expressions	1
Statistiques	
Nombre de textes	1
Nombre de segments de texte	4409
occurrences	151218
Nombre de formes	14689
Nombre d'hapax	7467 - 50.83 % des formes - 4.94 % des occurrences

-Figure 03-

Nombre de textes	01
Nombre de segments de texte	4409
Occurrences	151218
Nombre de formes	14689
Nombre d'hapax	7467

Tableau 02

Description

Comme nous pouvons le constater à travers le tableau 02, le nombre le plus fréquent est toujours celui des occurrences (151218) suivi par le nombre de formes (14689), nombre d'hapax (7467) et enfin le nombre de segments de texte (4409) sur un seul texte analysé.

Après les deux descriptions obtenues par Iramuteq, nous remarquons une différence de nombre de segments de textes, d'occurrences, de nombre de formes ainsi que le nombre d'hapax dans les deux corpus.

Nous constatons que les données du manuel privé sont plus élevées que celles du public.

3- Analyse de corpus

3-1/ Statistiques textuelles

Cette analyse contient cinq onglets :

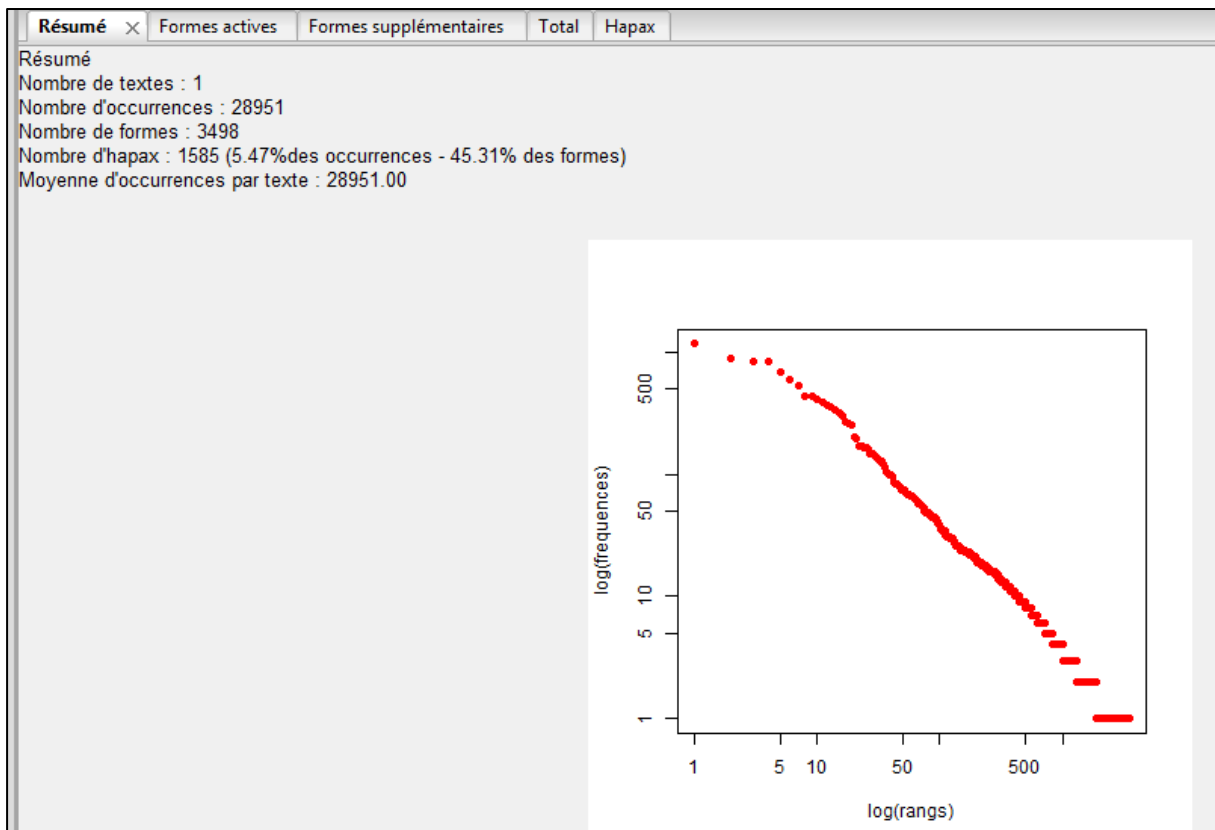
- 1^{er} onglet : Résumé = description générale du corpus (nombre de textes, d'occurrences, de formes...)
- 2^{ème} onglet : formes actives = liste des formes / mots actifs (avec leur catégorie grammaticale) par fréquences décroissantes.

- 3^{ème} onglet : formes supplémentaires = listes des formes / mots supplémentaires par fréquences décroissantes.
- 4^{ème} onglet : total = ensemble des mots par fréquences décroissantes.
- 5^{ème} onglet : hapax = mots du corpus présents une seule fois.

Sur chaque forme / mot, en cliquant droit, nous trouvons les formes associées qui permettent de visualiser les regroupements (lemmatisation) et concordancier qui affiche le contexte d'utilisation du mot dans le corpus.

a- Manuel public

Tout comme les informations précédentes (voir la tableau 1), la figure 04 résume le résultat obtenu illustré par un logarithme qui classe les fréquences dans le sens vertical et les lignes dans le sens horizontal.



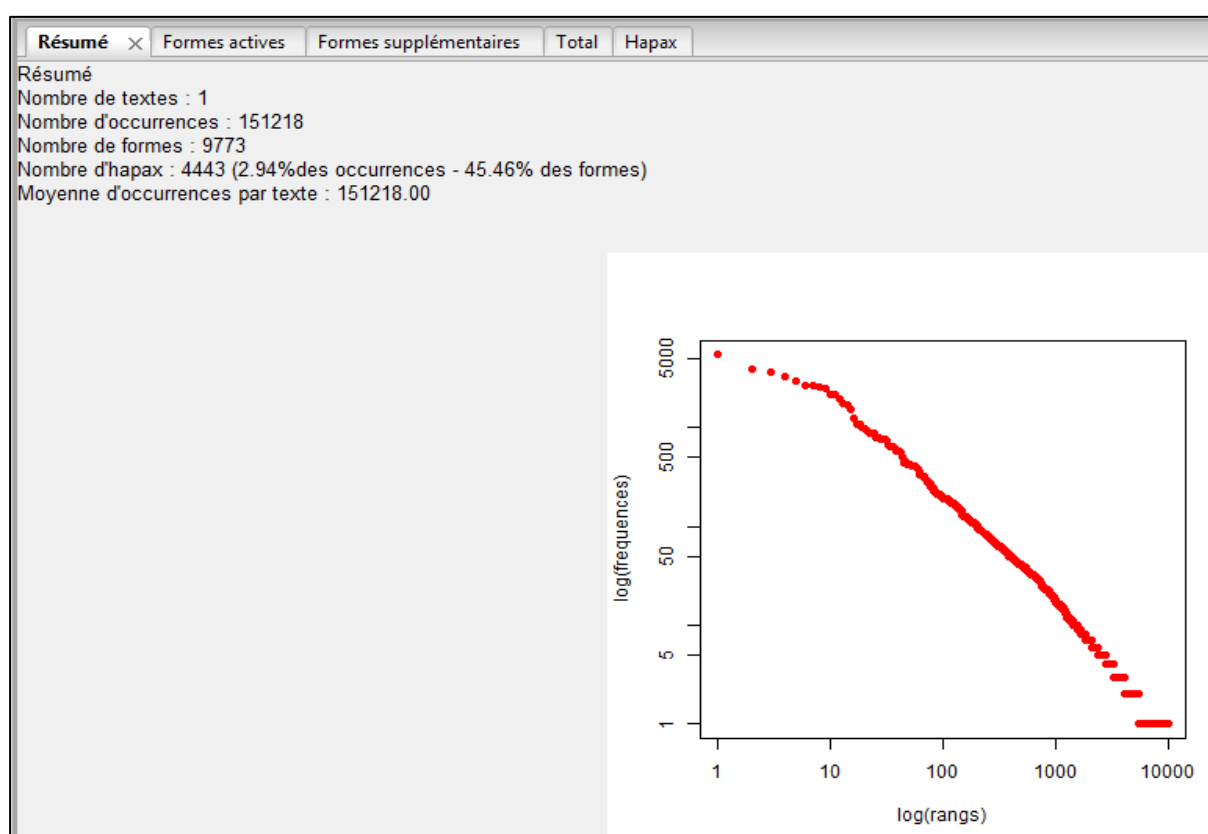
-Figure 04-

Description

Le tableau 1 ci-haut représente les caractéristiques du manuel public en général, c'est-à-dire sans lemmatisation, par contre la figure 04 montre ainsi l'ensemble de données de même manuel avec lemmatisation ; cette dernière donne la racine de chaque forme.

b-Manuel privé

Comme mentionné dans le tableau précédent (tableau 02), la figure 05 résume le résultat obtenu illustré par un logarithme en donnant le nombre de texte, d'occurrences, de formes, d'hapax et enfin la moyenne d'occurrences par texte.



-Figure 05-

Description

L'analyse textuelle effectuée pour le manuel privé nous montre bien que les occurrences présentent le nombre le plus haut avec (151218), par ailleurs le nombre d'hapax (4443) est classé en dernier par un taux total de 48.4 % (2.94 % des occurrences/ 45.46 % des formes).

Après l'application de la lemmatisation dans les deux corpus, nous avons constaté que le nombre de formes diminue.

Le logiciel fait une lemmatisation à l'aide de ses dictionnaires, et peut ainsi regrouper les formes au singulier et au pluriel sous une même forme, les verbes conjugués sous la forme infinitive.

Formes actives

Les formes actives sont les formes que le logiciel Iramuteq prend en compte, elles contiennent trois colonnes (mots, fréquence et type).

1/ Ecole publique

Nous avons pris pour illustrer nos propos un échantillon de 145 formes actives, en effet, pour chaque corpus nous avons choisi de prendre un nombre de cinq images au minimum, (voir les images ci-dessous) :

manuel 1AM public_stat_3 x

Résumé Formes actives x Formes supplémentaires Total Hapax

Forme	Freq. ↓	Types
texte	388	nom
abc	167	nom
utiliser	149	ver
verbe	143	nom
phrase	126	nom
mot	101	nom
expliquer	97	ver
eau	84	nom
nom	77	nom
camarade	68	nom
souligner	67	ver
suivant	67	adj
écrire	67	ver
énergie	65	nom
groupe	63	nom
passer	63	ver
santé	58	nom
présent	54	adj
observer	52	ver
exemple	49	nom
activité	48	nom
adjectif	48	nom
consigne	48	nom
mettre	48	ver
réaliser	45	ver
pollution	44	nom
retenir	44	ver
analyse	43	nom
prendre	43	ver

- Figure 06-

manuel 1AM public_stat_3 x					
Résumé		Formes actives x	Formes supplémentaires	Total	Hapax
Forme	Freq.	↓	Types		
projet	42		nom		
propre	41		adj		
aider	38		ver		
sujet	36		nom		
enfant	35		nom		
laver	35		ver		
procédé	35		nom		
algérien	34		nom		
grand	34		adj		
hygiène	34		nom		
conséquence	33		nom		
importance	32		nom		
air	31		nom		
donner	31		ver		
indicatif	31		nom		
partir	31		ver		
séquence	31		nom		
vie	31		nom		
composer	30		ver		
exprimer	30		ver		
question	30		nom		
algérie	29		nr		
année	29		nom		
cause	27		nom		
rédiger	27		ver		
complet	26		adj		
conjuguer	26		ver		
main	26		nom		
participer	26		ver		

-Figure 07-

manuel 1AM public_stat_3			
Résumé	Formes actives	Formes supplémentaires	Total Hapax
Forme	Freq. ↓	Types	
proposition	26	nom	
2ème	25	nr	
comprendre	25	ver	
meilleur	25	adj	
physique	25	adj	
solaire	25	adj	
aller	24	ver	
expression	24	nom	
manger	24	ver	
mondial	24	adj	
scientifique	24	adj	
servir	24	ver	
sport	24	nom	
but	23	nom	
dictionnaire	23	nom	
définition	23	nom	
environnement	23	nom	
impératif	23	adj	
infinitif	23	nom	
maladie	23	nom	
problème	23	nom	
qualificatif	23	nom	
renouvelables	23	nr	
temps	23	nom	
vérifier	23	ver	
éviter	23	ver	
agir	22	ver	
alimentation	22	nom	
journée	22	nom	

-Figure 08-

manuel 1AM public_stat_3 x			
Résumé			
Formes actives x			
Formes supplémentaires			
Total			
Hapax			
Forme	Freq. ↓	Types	
mer	22	nom	
professeur	22	nom	
remplacer	22	ver	
répondre	22	ver	
école	22	nom	
énumération	22	nom	
citoyen	21	nom	
construire	21	ver	
dent	21	nom	
production	21	nom	
titre	21	nom	
trouver	21	ver	
animal	20	nom	
coffre	20	nom	
forme	20	nom	
science	20	nom	
travers	20	nom	
colonne	19	nom	
compréhension	19	nom	
différent	19	adj	
introduire	19	ver	
mauvais	19	adj	
monde	19	nom	
partie	19	nom	
permettre	19	ver	
propreté	19	nom	
sentir	19	ver	
voiture	19	nom	
case	18	nom	

-Figure 09-

manuel 1AM public_stat_3 x					
Résumé		Formes actives x	Formes supplémentaires	Total	Hapax
Forme	Freq. ↓		Types		
convenir	18		ver		
explicatif	18		adj		
jour	18		nom		
lecture	18		nom		
marcher	18		ver		
obtenir	18		ver		
parler	18		ver		
petit	18		adj		
poisson	18		nom		
polluer	18		ver		
produire	18		ver		
soleil	18		nom		
train	18		nom		
utilisation	18		nom		
voir	18		ver		
3ème	17		nr		
accord	17		nom		
aide	17		nom		
ali	17		nr		
courir	17		ver		
danger	17		nom		
explication	17		nom		
fin	17		nom		
ponctuation	17		nom		
premier	17		adj		
rendre	17		ver		
retrouver	17		ver		
scolaire	17		adj		
suivre	17		ver		

-Figure10-

Description

D'après ces figures, la forme active la plus utilisée pour ce corpus est le mot « texte » qui apparaît 388 fois (figure06), nettement devant les formes actives « verbe » et « utiliser », la dernière colonne fournit la catégorie grammaticale du mot qu'est un nom dans ce cas. Par contre, dans les figures7, 8, 9 et 10, nous remarquons que le nombre de fréquences se réduisent.

Nous remarquons que la majorité des formes sont des mots pleins (nom, verbe, adjectif qualificatif,...).

2/ Ecole privée

Pour ce corpus, nous avons pris le même échantillon de 145 formes, représentées dans cinq figures ; chacune contient une colonne de 29 mots :

fleurs d'encre_stat_3 x				
Résumé	Formes actives x	Formes supplémentaires	Total	Hapax
Forme	Freq. ↓		Types	
verbe	895		nom	
p	888		nr	
phrase	638		nom	
mot	580		nom	
passer	466		ver	
temps	427		nom	
nom	403		nom	
voir	378		ver	
groupe	332		nom	
chevalier	330		nom	
aller	320		ver	
présent	318		adj	
sujet	294		nom	
prendre	282		ver	
texte	282		nom	
renart	272		nr	
lire	271		ver	
roman	246		adj	
indicatif	226		nom	
proposition	226		nom	
suivant	222		adj	
complément	219		nom	
écrire	215		ver	
mettre	210		ver	
simple	210		adj	
sentir	208		ver	
participer	192		ver	
roi	192		nom	
récit	190		nom	

-Figure 11-

fleurs d'encre_stat_3 x			
Résumé			
Formes actives x			
Formes supplémentaires			
Total			
Hapax			
Forme	Freq. ↓	Types	
employer	189	ver	
adjectif	188	nom	
grand	188	adj	
verbal	188	adj	
venir	186	ver	
pronom	179	nom	
langue	176	nom	
observer	175	ver	
aventure	173	nom	
scène	173	nom	
ex	170	nr	
parler	170	ver	
relever	169	ver	
voix	168	nom	
partir	164	ver	
forme	160	nom	
robinson	160	nom	
exprimer	159	ver	
personnage	156	nom	
nominal	155	adj	
laver	154	ver	
tristan	154	nr	
trouver	154	ver	
compléter	152	ver	
donner	152	ver	
gras	150	adj	
rédiger	147	ver	
image	146	nom	
recopier	143	ver	

-Figure 12-

fleurs d'encre_stat_3 x					
Résumé		Formes actives x	Formes supplémentaires	Total	Hapax
Forme	Freq. ↓		Types		
ez	130		nr		
classe	129		nom		
grammatical	128		adj		
siècle	127		nom		
souligner	127		ver		
premier	126		adj		
élément	126		nom		
château	124		nom		
liste	124		nom		
asseoir	121		ver		
île	121		nom		
lancelot	120		nom		
argan	118		nr		
réviser	118		ver		
appeler	114		ver		
valeur	114		nom		
lancer	113		ver		
raconter	113		ver		
pirate	112		nom		
vie	112		nom		
conjugaison	111		nom		
poème	111		nom		
relatif	110		adj		
justi	109		nr		
livre	109		nom		
fonction	108		nom		
futur	108		adj		
demander	107		ver		
homme	107		nom		

-Figure 13-

fleurs d'encre_stat_3 x			
Résumé			
Formes actives x			
Formes supplémentaires			
Total			
Hapax			
Forme	Freq. ↓	Types	
expression	106	nom	
français	106	adj	
leçon	106	nom	
mode	105	nom	
employé	103	nom	
action	102	nom	
passif	102	adj	
terminaison	102	nom	
iseult	100	nr	
circonstancier	97	adj	
comprendre	94	ver	
exercer	94	ver	
imparfait	94	adj	
composer	93	ver	
conjuguer	93	ver	
lecture	93	nom	
lieu	93	nom	
écriture	92	nom	
accorder	91	ver	
cod	91	nr	
histoire	91	nom	
expliquer	90	ver	
subordonné	90	adj	
subjonctif	89	nom	
nitif	88	nr	
étude	87	nom	
maître	86	nom	
utiliser	86	ver	
qn	85	nr	

-Figure 14-

fleurs d'encre_stat_3 x				
Résumé	Formes actives x	Formes supplémentaires	Total	Hapax
Forme	Freq. ↓	Types		
moyen	85	nom		
3e	84	nr		
exercice	84	nom		
jouer	84	ver		
petit	83	adj		
polo	83	nom		
toinette	83	nr		
marco	82	nr		
jour	81	nom		
orthographe	81	nom		
base	80	nom		
indiquer	80	ver		
présenter	80	ver		
suivre	80	ver		
accord	79	nom		
beau	79	adj		
fabliau	79	nom		
latin	79	adj		
retenir	78	ver		
bnf	77	nr		
conditionnel	77	nom		
couleur	77	nom		
médiéval	76	adj		
passage	76	nom		
animal	75	adj		
jeune	75	adj		
servir	75	ver		
arriver	74	ver		
découvrir	74	ver		

-Figure 15-

Description

Le logiciel nous a permis de relever une étendue lexicale équivalente à 895 formes pour le mot « *verbe* » qu'est le vocable le plus exploité dans ce corpus, suivi par « *phrase* » avec une fréquence maximale de 888 et « *mot* » avec 580 apparitions.

Les formes actives sont un phénomène toujours vérifié en lexicométrie dans le logiciel Iramuteq.

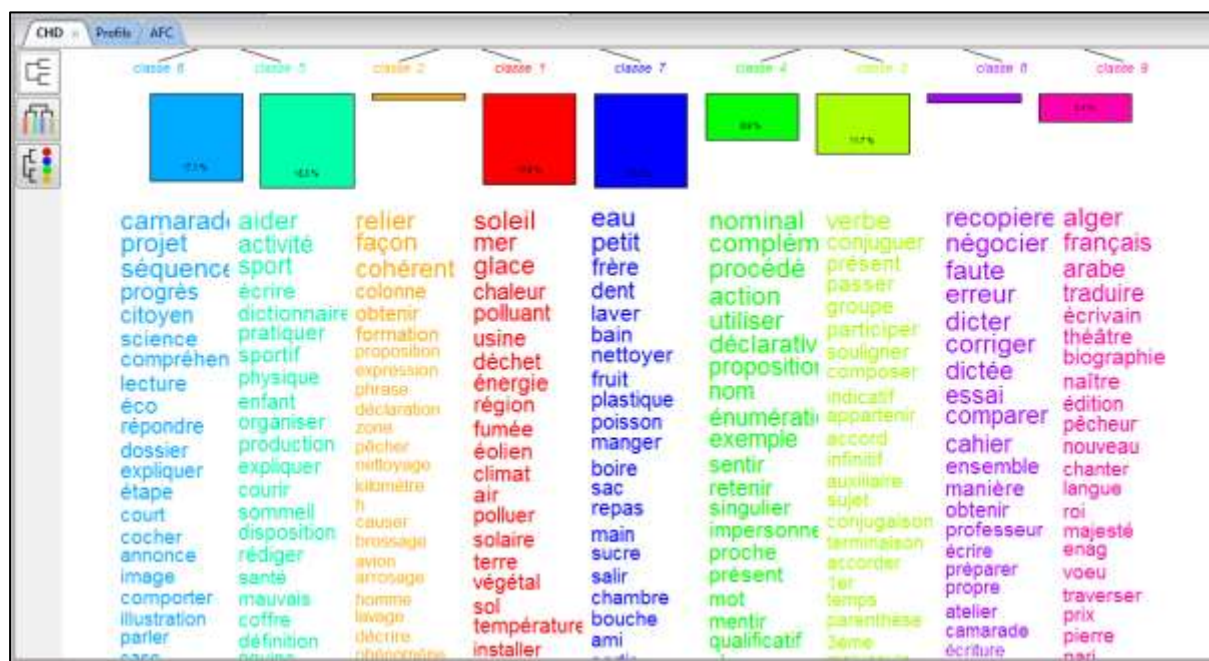
3.2/ classification méthode Reinert

Cette méthode permet de regrouper les grandes thématiques qui se retrouvent dans le corpus étudié sous forme de classes.

Une classe: est un ensemble de mots lexicaux et sémantiques associés à un thème particulier.

1/ Manuel public

Cette figure montre comment cette analyse répartie les termes sous des classes comme ce qui suit :



-Figure 16 -

Description

A partir de cette analyse effectuée sur notre corpus textuel, la méthode Reinert regroupe les termes dans neuf classes (figure16).

Ces classes sont constituées des termes ayant des liens sémantiques et se regroupant autour de la thématique de la classe en question, ainsi l'analyse nous permet de dire que la classe 1 occupe la première position dans le classement avec un pourcentage de (17,6%).

Nous remarquons que la classe 2 et la classe 7 sont en relation avec la classe 1 qui traitent le thème de l'environnement, la nature (soleil, nettoyer, homme, eau, salir,...).

La classe 3 et la classe 4 abordent le thème de structure de la langue (verbe, présent, conjuguer, nom, mot,...)

La classe 5 et la classe 6 occupent une place majeure (troisième thème) qui se focalise sur le programme scolaire et l'enseignement (séquence, projet, camarade, activité, production,...).

Enfin, la classe 8 nous donne des termes qui sont en rapport avec la production écrite (recopier, faute, dictée, écrire,...) et pour le dernier classement (classe 9), elle se focalise quant à elle sur la description du concept de la langue (français, arabe, langue, traduire,...).

2/ Manuel privé

Cette représentation est donnée par le logiciel Iramuteq sous forme d'un nombre de classes constituées des termes ayant des liens sémantiques associés à un thème donné, le montre la figure ci-dessous:



-Figure 16-1 -

Description

Comme nous avons précisé auparavant, toujours avec la méthode Reinert ; ce corpus est réparti en dix classes différentes.

Nous avons d'abord les classes 1, 2 et 3 en tête du classement, elles englobent le thème de genres littéraires (comédie, Molière, théâtre, roman, Hugo,...).

Le deuxième thème représenté par les classes 4, 5, 6, et 7 parle de preux chevalier et gente dame (chevalier, reine, angélique, Toinette,...).

La classe 8 et la classe 9 traitent le domaine de la grammaire et la conjugaison (phrase, mot, proposition, simple, indicatif, terminaison,...).

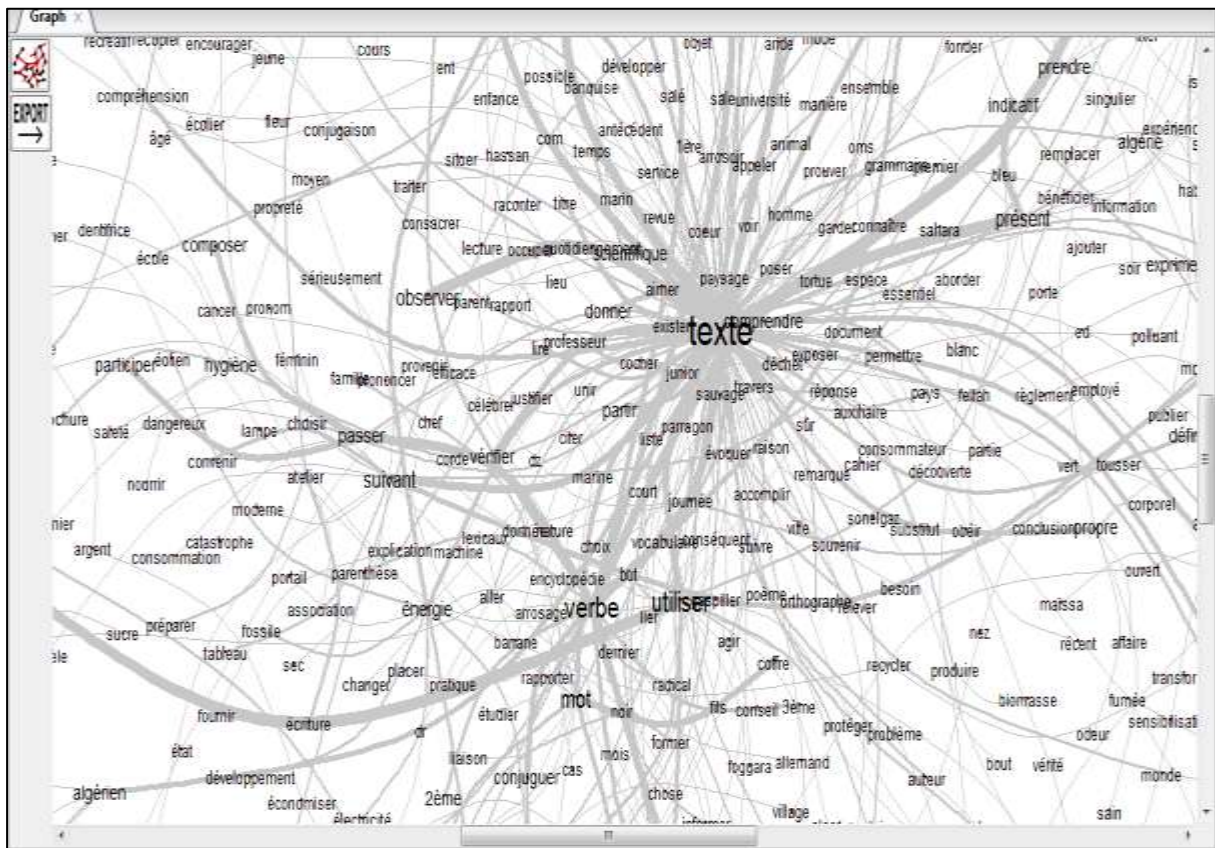
Enfin la dernière classe renvoie à la bibliographie du corpus.

3.3/ Analyse de similitude

Cette analyse de similitude consiste à représenter sur un graphique une arborescence axée sur l'élément principal autour duquel le corpus est construit. Ainsi sur la base des proximités sémantiques, cette analyse crée un arbre autour du terme principal.

1/ Manuel public

Cette figure ci-dessous, nous permet de voir dans une représentation arborescente les différentes ramifications qui existent entre les termes qui composent le corpus :



-Figure 17-

Description

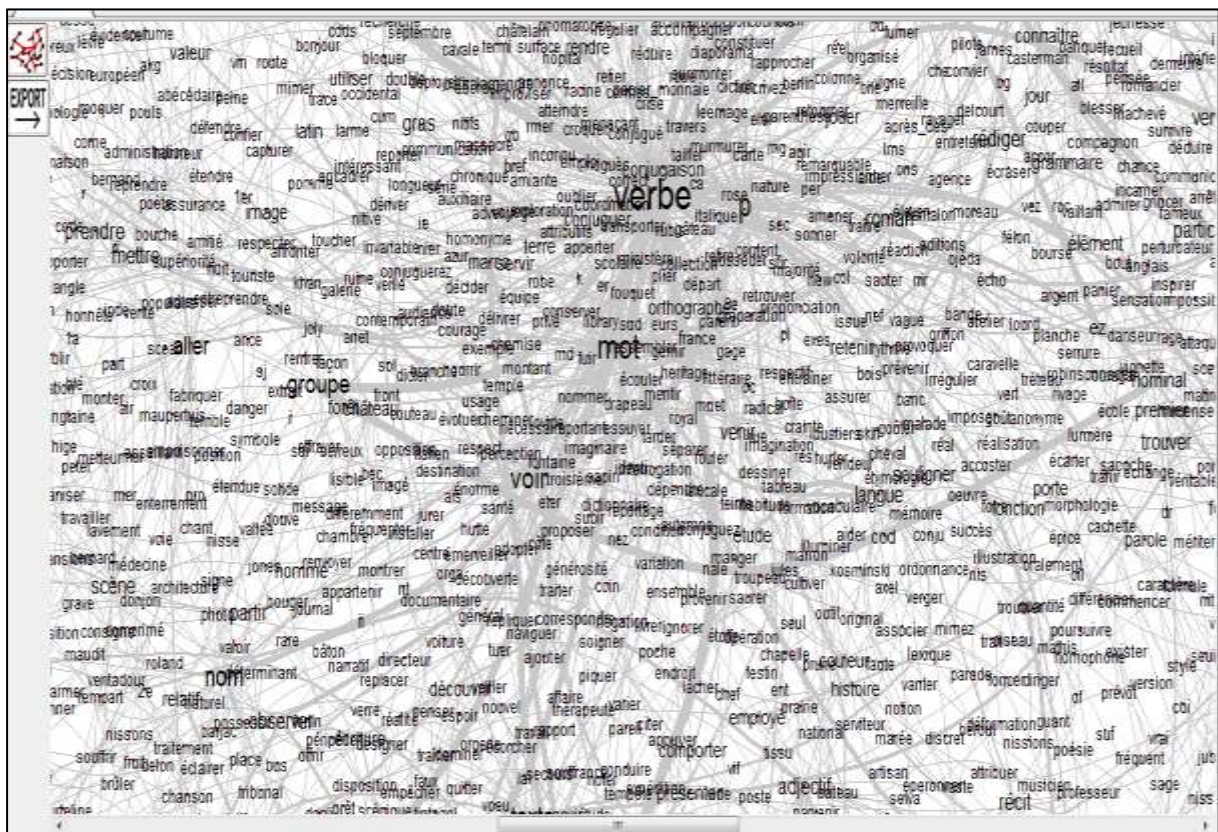
Nous pouvons constater à partir de l'image précédente une communauté lexicale est à la racine de l'arbre, constituée autour de la forme « texte » qui constitue un premier noyau de référence, il a des liens avec les autres parce qu'en réalité c'est à travers lui que chacun des autres termes prennent tout leur sens comme : verbe, utiliser, mot,...

Ainsi nous avons la possibilité d'établir des liens sémantiques entre les termes à travers lesquels nous pourrions analyser les champs conceptuels qui se recoupent via ces termes.

Le logiciel nous donne une vue plus approfondie du texte analysé en faisant des ramifications basées sur la cooccurrence des termes et aussi sur les proximités sémantiques.

2/ Manuel privé

Comme pour le public la figure ci-dessous retranscrit les différentes ramifications liées à une forme noyau.



-Figure 17-1-

Description

Cet arbre décrit un ensemble de ramifications différentes que ces dernières traduisent en quelque sorte les liens logiques entre les différents thèmes qui sont contenus dans ce corpus.

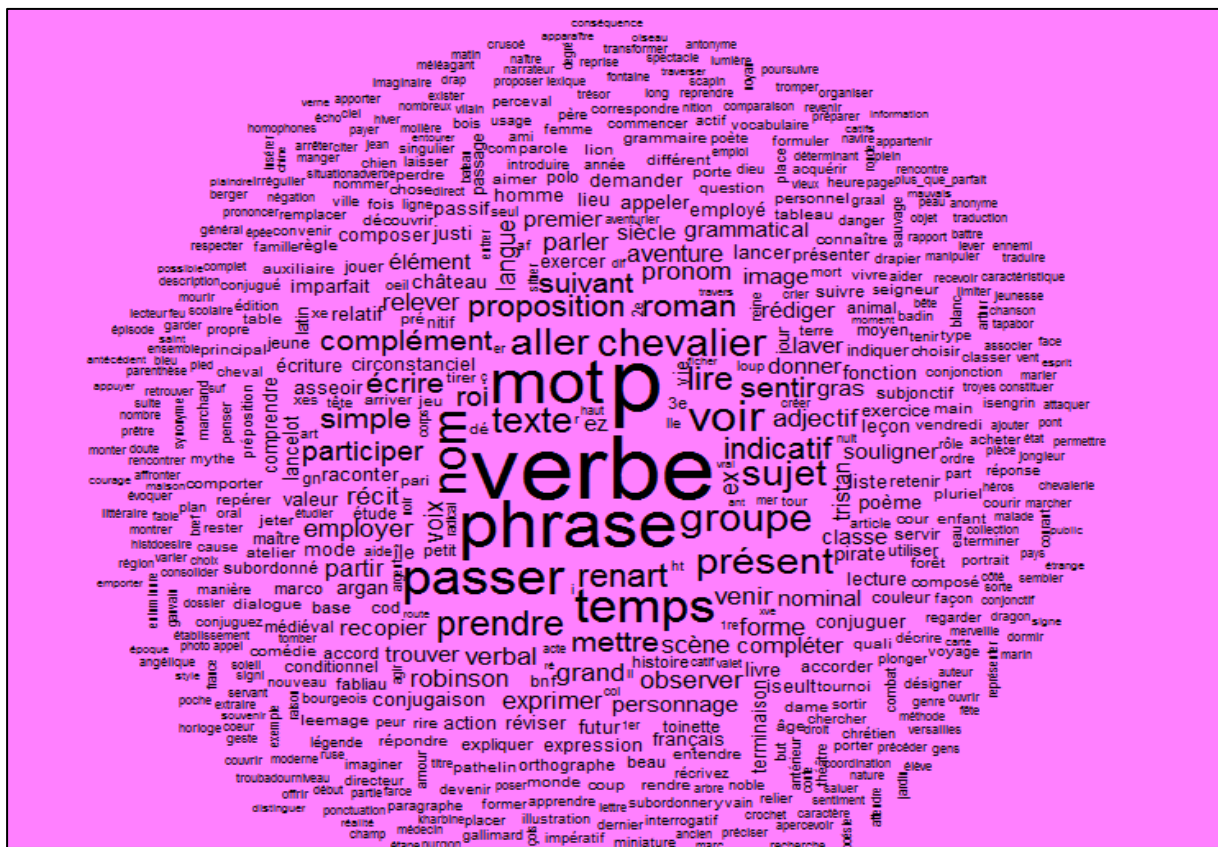
Description

Ce schéma est un outil de lexicométrie qui représente simplement les fréquences des formes en les corrélant à leurs tailles. Elles sont ainsi liées à la description statistique du corpus montrée précédemment, les mots : texte, phrase, verbe, utiliser, mot, expliquer, etc. sont les plus fréquents.

Notons que plus la taille du terme est grande plus ce dernier a une valeur étendue dans le corpus analysé, c'est-à-dire que les mots qui occupent le centre du nuage sont essentiels pour la compréhension de l'analyse.

2/ Manuel privé

De la même manière que celle du premier corpus, nous avons analysé le manuel scolaire du privé avec la méthode nuage de mots (Voir la figure ci-dessous)



-Figure 18-1 -

Description

Dans cette perspective l'intervention du contexte du corpus textuel est très sollicitée dans le but de mettre en lumière certains termes visibles et significatifs.

Le mot « verbe » occupe le cœur de l'image car il présente un nombre plus élevé d'occurrences, non seulement les termes qu'ils l'entourent sont peu visibles mais ils ont des significations sémantiques avec le terme principal : verbe, mot, phrase, nom, texte, etc.

Il est essentiel de noter que l'analyse de similitude et l'analyse en nuages de mots sont complémentaires l'une avec l'autre, car elles représentent presque les mêmes données car elles sont corrélées avec les formes actives obtenues dans l'analyse des statistiques textuelles.

Conclusion partielle

Dans cette partie pratique, nous avons procédé à une brève description des deux manuels scolaires du public et du privé de première année moyenne, puis nous avons réalisé l'analyse lexicométrique de notre corpus à l'aide d'un logiciel de statistique Iramuteq.

Dans ce second chapitre, nous avons réalisé les différentes analyses textuelles proposées par ce logiciel, et nous avons opté pour quatre analyses à savoir : l'analyse des statistiques textuelles, la méthode Reinert, l'analyse de similitude et ainsi que l'analyse en nuage de mots.

A travers ces différentes méthodes d'analyse, nous avons pu relever toutes les différences lexicologiques et lexico-sémantiques dans chaque corpus. Elles sont mises en évidence par le logiciel Iramuteq par un inventaire de formes actives et de formes supplémentaires, par des arbres maximaux ainsi que par des ensembles de mots formant des nuages.

CONCLUSION GÉNÉRALE

Via la lexicométrie, la linguistique est devenue propriétaire des résultats chiffrés. Aujourd'hui cette discipline nous semble indispensable, étant donné la longueur de plus en plus importante des corpus qui peuvent être traités également pour une lecture séparée des résultats.

L'objectif principal de notre recherche est de mener une étude lexicométrique comparative entre les manuels scolaires du français d'un collège privé et public. Nous avons choisi pour cela les manuels de 1^{ère} année moyenne afin de répondre aux questions de notre recherche sur les différences relatives à la longueur des textes traités, leurs richesses, leur variété ainsi qu'aux contenus proposés.

Iramuteq est le logiciel que nous avons exploité pour examiner et analyser le contenu des deux manuels scolaires, 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 considérables.

En deuxième lieu, Après avoir analysé les deux corpus nous sommes arrivées aux résultats suivants :

- Les manuels privés de français ont un support de texte plus cohérent, dans les manuels privés, par rapport au nombre de textes dans les manuels étatiques, ils sont de plus en plus variés que ceux des manuels publics, et par rapport aux genres proposés dans le projet (ex : théâtre, conte, etc.) sont divers, contrairement avec les manuels publics.
- On a constaté également que les textes traités dans le manuel privé sont des textes essentiellement littéraires, des auteurs français allemand, et anglais (des contes, fables, Romans, théâtres, ...).

Dans le manuel public, par contre, on propose des textes scientifiques comme les extraits d'articles de revue. Par conséquent, le manuel de l'école privée offre à l'apprenant une occasion pour découvrir d'autres cultures en les mettant en contact avec des textes riches en éléments culturels.

Pour conclure, rappelons enfin que cette analyse nous a permis de découvrir des sujets textuels très spécialisés et abstraits rarement abordés dans le domaine de la linguistique qu'est la lexicométrie.

Nous espérons que ce modeste travail permettra d'ouvrir d'autres pistes de recherche y compris une analyse des activités suggérées dans les deux manuels ; trouver d'autres éléments dépendant de la culture qui n'ont pas pu être cités afin de mieux comprendre et d'expliquer la différence entre les deux manuels (privé & public).

RÉFÉRENCES ET BIBLIOGRAPHIQUES

Ouvrages

- AINO NIKLAS-SALIMNEN (1997), *La lexicologie*, Armand Colin, Paris.
- ALISE Lehmann, FRANÇOISE MARTIN-BERTHET (2018), *Avant propos dans Lexicologie*, Armand Colin.
- JEAN-CLAUDE Soulages (2015), *L'analyse du discours Sa place dans les sciences du langage et la communication. Hommage Patrick Charaudeau*, PUR.
- LUDOVIC Lebart, BENEDICTE Pincemin , CELINE Poudat, *Analyse de données textuelles*, Presses de l'Université du Québec.
- MARIE-FRANÇOISE MORTUREUX (2008), *La lexicologie entre langue et discours*, Armand Colin.
- ROLAND ELUERD (2000), *La lexicologie : Que sais-je ?*, Presses Universitaires de France, 2000, 108, Boulevard Saint-Germain, 75006 Paris.

Articles du revue

- CHERIF SOUTI (2015), La lexicométrie ou l'analyse du discours assisté par ordinateur : l'informatique et les mathématiques peuvent apporter à la littérature et la linguistique, Revue EXPRESSIONS, N° : 01, 122-130.
- CHLOE SIX (2019), Analyse lexicale appliquée à une question ouverte à l'aide d'IRaMuTeQ, HAL open science, dumas-02372217.
- CLOTILDE Corone (2020), Outil 51. L'analyse textuelle ; Dans La Boîte à outils de l'analyse de données en entreprise, Revue Cairn-infos ; 168-171.
- GILBERT Grandguillaume (2004), La francophonie en Algérie, Revue Cairn-infos N° : 40, 75-78.

- HADRIA KHELIFI (2016), La lexicométrie : un outil efficient pour l'analyse du discours, HAL open science, hal-01730406, 151-161.
- HUSSAIN BILHAJ (2016), Le manuel scolaire et son rôle dans l'action d'enseignement en classe de F.L.E en Libye, HAL open science, hal-03081361, 47-65.
- JEAN-JACQUES SALONE (2013), Analyse textuelle avec IRaMuTeQ et interprétations référentielles des programmes officiels de mathématiques en quatrième, HAL open science, hal-01671161.
- MARYLIZ RACUNE et STEPHANE ALLAIRE (2022), Contribution à la compréhension du concept de la participation : une analyse lexicométrique, Revue internationale du CRIRE, N° : 01, 1-30.
- NATACHA SOUILLARD (2022), Introduction à la textométrie avec le logiciel IRaMuTeQ, HAL open science, hal-03942366.

Sitographie

- Analyse lexicométrique comparative de deux manuels de 4AM, <https://s13ef6dd46b53ef81.jimcontent.com/download/version/0/module/12992711127/name/Osmani..pdf> (consulté le 19/10/2022).
- Analyse sémique et thématique, https://www.observatoireplurilinguisme.eu/177778500-ct_th%C3%A8se-publications/th%C3%A8se/la-langue/1-analyse-linguistique/11983-la-langue_2_3_analyse-s%C3%A9mique (consulté 17 mars 2023)/
- Analyses statistiques de textes avec le logiciel libre IRaMuTeQ, <https://mate-shs.cnrs.fr/actions/tutomate/tuto47-iramuteq-ratinaud/> (consulté 3 mars 2023).
- Catherine FUCHS « Linguistique-Domains », <http://www.universalis.fr/encyclopedie/linguistique-domains /3-lexicologie-la->

[semantique-et-la-pragmatique/](#) (consulté le 23/03/2023).

- Champ lexical, <https://www.linternaute.fr/dictionnaire/fr/definition/champ-lexical/> (consulté le 14 avril 2023).
- Ecole privée ou école publique : quelles différences ?, <https://www.flexifamily.com/article/ecole-privee-ou-ecole-publique---people-baby/9928359#:~:text=%C3%80%20la%20diff%C3%A9rence%20des%20%C3%A9tablissements,de%20l%C3%A9cole%20publique%20plus%20personnalis%C3%A9es>. (consulté le 11 février 2023).
- Enseignement privé : « Le nombre d'écoles est timide et ne répond aucunement aux besoins » selon Finabi Conseil, <https://www.reporters.dz/enseignement-prive-le-nombre-decoles-est-timide-et-ne-repond-aucunement-aux-besoins-selon-finabi-conseil> (consulté le 12 avril 2023).
- Etude et analyse comparative pour un mémoire, <https://proredaction.com/etude-et-analyse-comparative/> (consulté le 12 avril 2023).
- Initiation à la lexicométrie approche pédagogique à partir de l'étude d'un corpus avec logiciel Iramuteq, https://f.hypotheses.org/wpcontent/blogs.dir/2671/files/2016/04/TD_lexico_Dut_Info_S2_mai2016_v2.pdf (consulté le 26/03/2023).
- Introduction à la lexicologie Conférence 6 La sémantique structurale: l'analyse du sens lexical, <https://slideplayer.fr/slide/10198010/> (consulté le 19 mars 2023).
- Iramuteq (logiciel), [https://fr.wikipedia.org/wiki/Iramuteq_\(logiciel\)](https://fr.wikipedia.org/wiki/Iramuteq_(logiciel)). (consulté février 2023).
- Iramuteq téléchargement et installation, <http://iramuteq.org/telechargement> (consulté le 12/12/2022).

- L'école, <https://www.larousse.fr/dictionnaires/francais/%C3%A9cole/27609>,
<http://www.education.gouv.qc.ca/professionnels/dossiers/education-a-la-sexualite/roles-de-lecole-et-de-la-famille> ,
<https://www.1jourlactu.com/monde/a-quoi-ca-sert-lecole-> (consulté le 12 mars 2023).
- L'objet d'étude de la lexicologie, <https://studfile.net/preview/5115123/> (consulté février 2023).
- La langue française en Algérie : Etat des lieux, <https://arlap.hypotheses.org/7935> (consulté le 20/03/2023).
- La notion et la définition du mot, <https://studfile.net/preview/5115123/page:2/> (consulté février 2023).
- Langue, langage, parole, http://laboiteasaussure.fr/langage_langue_parole.htm (consulté janvier 2023).
- Le manuel scolaire : Au-delà de l'outil pédagogique, l'objet politico-social, <https://journals.openedition.org/edso2014> (consulté le 13/03/2023).
- Le système éducatif algérien, <https://www.asjp.cerist.dz/en/downArticle/50/5/1/155335> (consulté le 19/12/2022).
- Lemme, <https://www.cnrtl.fr/definition/lemme> (consulté mars 2023).
- Les apports d'une analyse statistique des données textuelles pour les recherches en didactique : l'exemple de la méthode Reinert, <https://journals.openedition.org/adsc/458#tocto1n3> (consulté mai 2023).

- Les Iris : établissement privé Bejaia, <https://ecolelesiris.com/> (consulté mars 2023).
- Mot, <https://www.cnrtl.fr/definition/mot> (consulté mars 2023).
- Occurrence et cooccurrence, <https://www.cnrtl.fr/definition/cooccurrence> (consulté mars 2023).
- Pourquoi enseigner la langue française dans nos classes et comment ? <https://lettres-histoire.ac-versailles.fr/spip.php?article1212> (consulté avril 2023).
- Qu'est-ce-que l'approche comparative ? <https://www.soft-concept.com/surveymag/definition-fr/definition-approche-comparative.html> (consulté mars 2023).
- Segment, <https://www.lalanguefrancaise.com/dictionnaire/definition/segment> (consulté mars 2023).
- Statistiques textuelles- Iramuteq, <http://www.iramuteq.org/documentation/html/2-5-1-statistiques-textuelles> (consulté mai 2023).
- Système éducatif algérien, http://www.meric-net.eu/files/fileusers/National%20Report%20template_MERIC-Net_Algeria.pdf (consulté mars 2023).
- Téléchargement R 4.2.2, <https://r-for-windows.updatestar.com/fr> (consulté le 17/12/2022).
- Tutoriel d'installation Iramuteq et R (Windows), <http://iramuteq.org/telechargement/tutoriel-dinstallation-iramuteq-et-r-sur-windows> (consulté le 26/01/2023).

Dictionnaires

- DUCROT Oswald, dictionnaire encyclopédique des sciences du langage, Seuil 1972.
- GEORGES Mounin, le dictionnaire de la linguistique, 1974.

Mémoires et thèses

- GHOUAR Souhila/CHETIOUI Dounia (2020/2021), *Les supports textuels dans les manuels de français : étude comparative entre les écoles publiques et privées (niveau IAM)* Mémoire de Master, université de Bejaia.
- BENSALIM Mohammed (2018/2019) *Étude comparative entre les manuels scolaires de la 1ère génération et ceux de la 2^{ème} génération (Cas du manuel scolaire de la 4ème AP)* mémoire de Master, université Mouhamed Boudiaf, M'sila.
- CHATRI Farida/KOUACHE Ouissam (2016/2017), *L'étude comparative des erreurs en production écrite. Cas de 4ème année moyenne dans l'école publique Manaul-Tekssira Gomez et l'école privée Galilée*, mémoire de Master, université de Bejaia.
- MOUHOUBI Roza/ NABTI Nawal (2016/2017) *Analyse lexico-sémantique et comparative de la presse écrite algérienne d'expression française : Rubrique culturelle des journaux « Liberté » et « El watan »* mémoire de Master, université de Bejaia.
- BENHENNI Sabrina (2013/2014) *Etude lexicométrique d'un manuel scolaire le manuel de la 2^{ème} AS FLE en Algérie*, mini-projet université de Relizane.
- DEHMANE Nadhira (2011/2012) *Analyse lexicométrique de discours politique français en classe de langue*, mémoire de Magister, université El Hadj Lakhdar, Batna.

- MOUASMI Oum el djilali/ MEGHEZI Louiza (2018/2019) *Analyse des erreurs dans la production écrite des apprenants de quatrième année moyenne*, mémoire de Master, université de Ibn-khaldoun, Tiaret.
- HADJOU Drifa/ HALIM Souhila (2020/2021) *Etude psycholinguistique des difficultés langagières chez les enfants souffrants de handicap mental*, mémoire de Master, université de Bejaia.

TABLE DES TABLEAUX

Tableau 01 : description des données du manuel public.....41

Tableau 02 : description du manuel privé.....42

ANNEXES

Manuel de langue française

Première année moyenne

1^{ère}
AM

Par

ANISSA MADAGH

Inspectrice de l'Éducation et de l'enseignement moyen

CHAFIK MERAGA

Professeur de français cycle moyen

HALIM BOUZELBOUDJEN

Professeur de français cycle moyen

Sous la direction de

ANISSA MADAGH

Inspectrice de l'Éducation et de l'enseignement moyen

Conception graphique et couverture:

GRONI MOHAMED ZOHR

Master Édition et diffusion de l'écrit « Les métiers du livre »

fleurs
d'encre



Français 5^e

C. Bertagna & F. Carrier

manuel unique



Existe en version numérique
(voir au dos)

hachette
ÉDUCATION

Image créée à partir de illustrations de © Toulon éditions, 2011-2012



Résumé

Notre travail de recherche s'inscrit dans le cadre de la lexicologie, il est centré sur une étude lexicométrique comparative des supports textuels dans les manuels de français entre le public et le privé : cas du 1AM école « les Iris» dans la ville de Bejaia.

L'intérêt de cette étude à pour objet une analyse statistique textuelle des deux manuels scolaires par le logiciel Iramuteq , pour bien présenter notre sujet nous l'avons organisé en deux chapitres principales , le premier théorique et le deuxième pratique .

De ce fait, nous avons confirmé nos hypothèses, que le lexique de manuel privé est plus riche et variés par rapport au public ainsi que le vocabulaire qui se base sur des textes littéraires, poétiques. Les résultats obtenus montrent la différence totale des programmes, en effet le programme de l'école publique donne l'importance à l'acquisition d'une compétence de communication orale et écrite du moment où le programme de l'école privé accorde une grande importance aux compétences.

Mots-clés: Lexicologie, lexicométrie, iramuteq, analyse, école, privé, public.

ملخص

يندرج عملنا البحثي في إطار علم المعاجم ، ويتركز على دراسة مقارنة معجمية للدعم النصي في الكتب المدرسية الفرنسية بين في مدينة بجاية "les Iris" القطاعين العام والخاص: حالة السنة الأولى متوسط مدرسة

لتقديم موضوعنا جيداً وقد Iramuteq الفائدة من هذه الدراسة هي التحليل الإحصائي النصي للكتابيين المدرسين بواسطة برنامج قمنا بتنظيمه في فصلين رئيسيين ، الأول نظري والثاني عملي.

ونتيجة لذلك ، أكدنا فرضياتنا القائلة بأن معجم الكتب المدرسية الخاصة أكثر ثراءً وتنوعاً مقارنة بالعامية بالإضافة إلى المفردات التي تعتمد على النصوص الأدبية والشعرية. تظهر النتائج التي تم الحصول عليها الاختلاف الكلي بين البرامج ، وبالفعل يعطي برنامج المدرسة العامة أهمية لاكتساب مهارات الاتصال الشفوي والكتابي عندما يعطي برنامج المدرسة الخاصة أهمية كبيرة للمهارات.

الكلمات المفتاحية:،تحليل ، مدرسة ،خاصة ، عامة iramuteq المعجم ، قياس،