

Université Sidi Mohamed Ben Abdellah
Faculté des Lettres et des Sciences Humaines
Dhar El Mahraz – Fès



Centre d'Etudes Doctorales : Esthétiques et Sciences de l'Homme

Formation doctorale : Langages et Formes Symboliques

Spécialité : Langue et littérature françaises

Statistique lexicale

Le vocabulaire d'Edmond Amran El Maleh

Étude quantitative et sémantique

Thèse pour l'obtention du Doctorat

Préparée par :

Hajar DAHMA

CNE : 0422752843

Sous la direction du Professeur :

Ali HEFIED

Année universitaire : 2019-2020

Dédicace

Je dédie le présent travail à mes parents

Mon mari

Mes frères

Ma très chère Yassmine Riham

Remerciements

M'aventurer dans un champ d'investigation où littérature, linguistique et lexicométrie se rencontre est un privilège inouï. Je ne peux être qu'émue par l'amabilité, le soutien et l'assistance m'ayant été prodigués à volonté par tous ceux qui m'ont épaulé pour relever ce défi.

Je remercie particulièrement mon directeur de recherche M. Ali HEFIED qui m'a permis de rabâcher une étude épineuse. Ses conseils, ses orientations et ses directives émanant bien d'un linguiste et statisticien chevronné m'ont permis, au cours de cette entreprise, de redresser les acquis et de rectifier le tir. Je le remercie également pour sa magnanimité et pour son sens rassis qu'il m'a inculqué.

Ma gratitude ira également à tous les professeurs du Laboratoire de recherche en Langues, Cultures, Dictionnairique et Corporas (LaCuDic), que ces lignes puissent leur exprimer toute ma reconnaissance pour leur écoute et leurs précieux conseils.

Certainement, ce travail n'aurait jamais pu voir le jour sans le soutien assidu de mes parents, je les remercie pour leurs encouragements et leur confiance en moi. Merci d'avoir fait de moi la personne que je suis aujourd'hui.

Mes remerciements s'adressent enfin à mon mari, mes deux frères et ma sœur Bouchra pour leur affection et leur soutien permanent.

À toutes et à tous, merci mille fois.

INTRODUCTION

Depuis 1980, la littérature marocaine d'expression française connaît un nouvel élan. Non seulement les grands auteurs des années 1960 et 1970 comme T. Ben Jelloun, D. Chraïbi, A. Khatibi, etc., continuent d'enrichir la littérature marocaine, mais de nombreux écrivains les rejoignent, parmi eux citons Edmond Amran El Maleh.

C'est vrai que ses idées politiques antisémites ne lui ont pas valu les feux de la rampe, cependant la singularité de son œuvre, la qualité de son écriture, la particularité de ses essais littéraires et même ses essais sur l'art l'érigent comme l'une des figures intellectuelles marocaines les plus illustres de son époque.

Ses textes ne sont pas faciles à assimiler. En effet, par son génie d'écrivain, il nous offre une œuvre délibérément hermétique, où tout semble confus, le temps, les lieux¹, le langage, les propos même du narrateur/auteur et pourtant une logique se poursuit. Son texte se caractérise par l'absence de linéarité chronologique, il se déploie défiant la syntaxe, la grammaire et la ponctuation. Tous ces aspects et bien d'autres ont très largement contribué à accréditer la thèse d'une œuvre labyrinthique. D'ailleurs, El Maleh le confirme lui-même dans un article qu'il a rédigé dans ***Les Cahiers d'Etudes Maghrébines*** :

« Alors ce récit et ceux qui le suivront ne serait qu'un long soupir de nostalgie, (...). Il y a peut-être un auteur, on peut se risquer à penser qu'il existe, il pourrait sans doute nous conduire dans le dédale de ce labyrinthe »².

¹ « Dans l'écriture de mes livres, on ne peut pas dresser une suite de lieux qui par une sorte d'enchaînement suivraient la progression d'un itinéraire. Les lieux deviennent des non-lieux. Je suis séduit par l'idée d'utopie, au sens étymologique. Très rarement, je nomme les villes. Je suis ravi quand le lecteur ne sait pas où il est. » Marie Redonnet, ***Entretiens avec Edmond Amran El Maleh***, Grenoble/Rabat, La pensée sauvage/Fondation Edmond Amran El Maleh, 2005, p.93.

² Edmond Amran EL MALEH, « ***La figure singulière du juif marocain qui écrit en langue française*** », In : ***Les Cahiers d'Etudes Maghrébines***, n°3, juin 1991, p.58.

De même, El Maleh marque un autre écart par rapport à une norme, c'est qu'il refuse d'inscrire ses textes dans un genre bien précis ; il ne veut pas barricader son œuvre dans une catégorie particulière. Ses deux premiers livres, *Parcours Immobile* et *Aïlen ou la nuit du récit*, sont édités sans étiquette du genre avant d'être réédités une deuxième fois avec une indication assez vague « récit ». De même pour *Mille ans, un jour* publié chez la Pensée Sauvage sous l'étiquette de « Roman », avant d'être réédité encore une fois chez André Dimanche, sans indication de genre. Jusqu'à la dernière année de sa vie, il demeurait rétif aux classifications génériques ; quand Le Fennec a publié *Lettres à moi-même* sous l'étiquette de « Journal », il était très en colère.

Cependant, comment aborder une œuvre si singulière telle que fût celle d'Edmond Amran El Maleh ? comment cerner l'écriture d'un écrivain dont l'œuvre est souvent qualifiée d'hermétique ? Comment pouvons-nous approcher une écriture qui ébranle les genres littéraires ? La réponse est peut-être dans la citation suivante d'El Maleh lui-même :

« L'écriture ne serait rien si on ne pose pas la question déterminante du langage, de la langue, et c'est là qu'effectivement réside tout le mystère, mystère en ce qui concerne l'être du langage et ses pouvoirs de création. Oui, c'est là le foyer, le point cardinal de tout ce qui a trait au pouvoir de l'écriture. (...) Chacun sait plus ou moins par expérience qu'on approche d'une réalité pleine de mystère, inépuisable au fur et à mesure qu'on s'y enfonce, que là dans un univers qui se révèle lentement, qui vous tient en sa bordure, à une distance sans cesse grandissante, se révèle une puissance, des pouvoirs qui vous tiennent sous leur empire, que vous n'êtes plus maître, mais esclave et ce sont les mots qui s'expriment alors que vous aviez la croyance naïve que vous vous en serviez pour exprimer vos sentiments, vos pensées. Expérience unique!³ »

³ Edmond Amran El Maleh, *Le Café bleu, Zrirek*, Paris, La Pensée sauvage, 1999, pp. 22- 23

De cette citation nous pouvons déduire qu'El Maleh accorde une grande importance au langage, à la langue et aux mots. C'est peut-être la clef pour déchiffrer l'énigme ou le mystère de son écriture. C'est pourquoi aborder l'œuvre d'El Maleh à partir de la langue, des unités du lexique, des mots eux-mêmes, pourrait se révéler fructueux. Pour ce faire, nous avons choisi de recourir aux méthodes lexicométriques et plus précisément à celles de la statistique linguistique. Cette dernière est au cœur d'un carrefour scientifique vers lequel convergent tout naturellement les principes de la statistique et ceux de la linguistique. A ce noyau dur vient s'ajouter l'informatique qui se veut un outil destiné à parfaire l'analyse, la description et la comparaison en nous offrant des profils statistiques, des graphiques, des tableaux, et ce, le plus rapidement et le plus efficacement possible.

Il est important, par ailleurs, de souligner que cette démarche a le privilège de rapprocher les études littéraires des études linguistiques, un rapprochement qui est assez naturel tel que le souligne Roland Barthes :

« Ce rapprochement paraît aujourd'hui assez naturel. N'est-il pas naturel que la science du langage s'intéresse à ce qui est incontestablement langage, à savoir le texte littéraire ? N'est-il pas naturel que la littérature, technique de certaines formes de langage, se tourne vers la théorie du langage ? N'est-il pas naturel qu'au moment où le langage devient une préoccupation majeure des sciences humaines, de la réflexion philosophique et de l'expérience créative, la linguistique éclaire l'ethnologie, la psychanalyse, la sociologie des cultures ? Comment la littérature pourrait-elle rester à l'écart de ce rayonnement dont la linguistique est le centre ? N'aurait-elle pas dû, même, être la première à s'ouvrir à la linguistique ? »⁴

En effet, la lexicométrie et le traitement informatique des textes permettent une description chiffrée du vocabulaire, ouvrant ainsi la voie à une analyse objective du

⁴ Roland BARTHES, « Linguistique et littérature ». In : **Langage**, Volume 3, N° 12, 1968, « Linguistique et littérature », p. 3.

corpus, et ne laissant pas de place aux impressions et aux effets de subjectivité. Nous ne prétendons pas par ces propos stigmatiser les analyses impressionnistes qui, pour longtemps, ont fait leurs preuves dans le domaine des études littéraires. Aussi n'avons-nous aucunement l'intention de dire que la statistique lexicale se passera du critique littéraire. Bien au contraire, elle lui sera d'une grande aide, dans la mesure où elle lui permettra de déchiffrer le texte, généralement sinueux, sans *a priori*. De plus, le recours à l'outil informatique et aux logiciels de traitement des données textuelles lui feront gagner du temps pour qu'il puisse se focaliser sur l'analyse et l'interprétation des données présentées sous forme de graphiques ou de tableaux.

S'agissant de la problématique qui sera étudiée au cours de ce travail de recherche, elle puise ses racines dans la vie et l'œuvre de notre auteur. En effet, les textes d'El Maleh sont tous hantés par la disparition de la communauté juive marocaine et les retombées de cette tragédie sur le devenir judéo-arabe. Cette absence collective est due à un départ abrupt, auquel fait allusion le titre du roman d'El Maleh *Mille ans, un jour* : en un seul jour une présence millénaire s'est changée en absence, un jour a suffi pour baliser un pan de l'histoire et de la mémoire sur cette terre marocaine ancestrale. Le départ abrupt et massif de cette minorité ethnique et religieuse n'a cessé d'être vécu par l'auteur comme un trauma, une cassure, ou bien « *inchiqaq* »⁵ tel qu'il le mentionne à maintes reprises dans ses textes. En effet, cette disparition ne causait non seulement la dispersion des différentes communautés juives établies au Maroc depuis des siècles, mais elle constitue, pour El Maleh, une véritable hémorragie culturelle. Cependant, cette blessure profonde ou bien ce « choc

⁵ Edmond Amran EL MALEH, *Mille ans, un jour*, Paris, André Dimanche, 2002, p. 16

émotionnel » sera le catalyseur de l'œuvre maléhiennne, tel qu'il le mentionne dans ses Entretiens avec Marie Redonnet :

« ... j'écris à partir d'un choc émotionnel qui déclenche l'écriture (...), le choc est de sentir qu'il y a une sorte d'oubli qui se fait concernant les événements historiques du Maroc, que les personnages disparaissent (...). Je ne travaille pas comme un arpenteur, je ne suis pas un architecte. Evidemment, mes livres s'en ressentent. Les gens peuvent penser que c'est un torrent verbal avec une absence de structure et dans le pire des cas sans ponctuation (...) Ce qu'on peut indiquer, c'est que (...) la quête plonge ses racines dans la nuit intime de l'être. Pour éclairer ces ténèbres, ce n'est pas facile. »⁶

Chagriné par la tournure des événements, il va trouver remède dans l'écriture, parce que, dans l'optique d'El Maleh, c'est elle seule qui peut restituer une certaine vérité de l'histoire et rendre l'existence possible. Ainsi, il commence à s'interroger sur la raison qui a bousculé le sort de toute une communauté vivant en toute sérénité sur sa terre-mère :

« Ils sont partis comme un souffle de vie qui se retire et s'éteint lentement. Mais pourquoi donc ? La question tombe dans un gouffre sans réponse (...) Suivez l'itinéraire des synagogues, ces demeures maintenant désaffectées, portes fermées sur le silence d'une haute ferveur, fenêtres aveugles sur l'éclatante lumière. »⁷

Incapable d'apporter une réponse à son interrogation, il va puiser dans un passé qui s'est définitivement clos et tenter de restituer la mémoire sociale et individuelle d'une nation fondamentalement interculturelle, afin de montrer finalement qu'il n'y avait

⁶ Marie REDONNET, *op.cit.*, 2005, pp. 126-127

⁷ Edmond Amran EL MALEH, « *Essaouira, l'oubliée* », In Revue Autrement, n° 11 (Hors-série), Paris, 1985, pp. 217-218

aucune raison pour que de milliers de juifs marocains quittent leur terre natale, dans le but de retrouver une illusion appelée la terre promise.

Cependant, écrivant en français, il y avait un obstacle qui se dressait devant l'auteur. En effet, après les multiples mutations que le Maroc a subies depuis sa mise sous le protectorat français jusqu'à l'exode des juifs marocains et l'indépendance du pays, la question primordiale pour El Maleh est comment transmettre l'authenticité de ces transformations à travers la langue de l'Autre ? Comment les mots, qui ne sont pas innocents, peuvent-ils faire revivre une mémoire collective vouée à la déliquescence et à l'oubli ? Comment faire en sorte que les valeurs musulmanes, judaïques et arabes parlent au sein d'un texte écrit dans la langue du colonisateur ? comment la langue maternelle, « *lougha el oum*⁸ », parle-t-elle au sein d'une langue étrangère ?

Cela étant dit, en posant la problématique dont l'étude sera mise en œuvre au cours de ce travail, nous chercherons à répondre à ces interrogations. En effet, Edmond Amran El Maleh a toujours célébré dans ses écrits la symbiose culturelle marocaine, ces œuvres sont toutes imprégnées d'une mémoire juive, arabe et berbère, ce qui nous incitera, moyennant les outils statistiques et informatiques, à chercher ***«comment Edmond Amran El Maleh a pu faire de son œuvre, à travers la langue d'écriture et les unités lexicales employées, un outil pour la redécouverte de la mémoire, de l'histoire et de la culture marocaine. »***

Notre étude s'effectuera, rappelons-le, dans le cadre des analyses quantitatives des œuvres littéraires reconnues dans les milieux linguistiques par l'appellation « statistique lexicale » ou « statistique linguistique ».

⁸ Edmond Amrane EL MALEH, 1999, **op. cit.**, p.75.

Signalons que notre intérêt pour la matière a vu le jour au cours d'une journée d'études relative aux « Recherches en statistique lexicale », organisée en 2012 par le Laboratoire de linguistique : Dictionnaire et Corpus multimédia (LABO-LING), de la Faculté des lettres et des Sciences humaines de Dhar El Mehraz (Fès). Cette journée nous a permis de découvrir le fruit des efforts de plusieurs chercheurs qui s'intéressent au domaine de la statistique lexicale.

Dans ce sens, et grâce à l'aide et aux directives de notre professeur, nous avons préparé un mémoire de fin d'étude intitulé « *Statistique lexicale : Le vocabulaire de Flaubert, Étude sémantique du substantif* ». Le contact que nous avons entrepris avec la statistique lexicale nous a permis de constater le rapprochement, voir l'alliance qui unit la littérature à la statistique, une alliance favorable à toute étude qui s'intéresse à l'aspect quantitatif du vocabulaire et cherchant à traduire les unités du langage, à savoir les mots, à des unités numériques susceptibles de subir tout genre d'opérations mathématiques et statistiques.

L'analyse statistique des œuvres d'Edmond Amran El Maleh ne saurait être réalisée sans recourir aux privilèges offerts par l'outil informatique notamment le logiciel de traitement des données textuelles « Hyperbase ». En effet, ce dernier nous facilitera le dépouillement du texte, c'est-à-dire le dénombrement de toutes les unités élémentaires du corpus. De même, il nous permettra de soumettre les œuvres d'El Maleh aux lois et tests de la statistique lexicale telles : le modèle binomial, la loi de Khi-deux, test de Pearson, l'écart réduit, etc., afin d'assurer une meilleure investigation et un examen rigoureux du corpus choisi.

Bien sûr, l'outil informatique ne se substitue point au labeur du chercheur que seul pourra interpréter les résultats obtenus, d'autant plus que cette démarche, bien qu'elle soit quantitative, exige un retour au texte et au contexte, ouvrant ainsi la voie

à une analyse qualitative, à une réflexion sur l'utilisation des mots, à un travail de recherche sur le vocabulaire et la langue.

Avant de présenter les différents axes de notre thèse, nous précisons que cette dernière ne s'intéresse pas uniquement à l'œuvre maléhiennne mais a également une mission méthodologique claire, consistant en l'exploitation et la confrontation de plusieurs principes méthodologique et par là même les différentes fonctionnalités du logiciel Hyperbase.

En termes de clôture, nous signalons que notre thèse se composera de trois parties fondamentales, dont la première sera réservée à la présentation du cadre théorique et méthodologique du travail, la deuxième sera consacrée à l'étude de la structure du vocabulaire, quant à la dernière partie, elle s'intéressera au contenu lexical, à la coloration thématique et aux manifestations du plurilinguisme dans l'œuvre d'El Maleh.

Dans la première partie, qui opérera sur ce qui est théorique et méthodologique, nous jetterons la lumière, dans un premier temps, sur la vie d'Edmond Amran El Maleh. Un projet qui nous semble indispensable dans la mesure où il nous permettra de mieux connaître l'auteur et de mieux comprendre ses écrits. Notons que l'œuvre maléhiennne ne se limite pas uniquement aux essais et aux romans, elle est composée également de plusieurs articles et critiques sur l'art. C'est la raison pour laquelle nous comptons présenter, dans un deuxième temps, son attitude esthétique. S'agissant de ses écrits littéraires, ils interpellent fréquemment des notions ou des concepts qui se rapportent à l'imaginaire judaïque et au vécu d'une minorité juive au sein d'une majorité musulmane. De ce fait, nous jugeons nécessaire de nous attarder sur ces points qui forment la toile de fond de son écriture afin de comprendre l'essence même de son œuvre.

Effectuer une recherche dans le domaine de la statistique lexicale nous mettra certainement face à quelques concepts qui exigent de nous une pause terminologique. Dans ce sens, nous consacrerons le deuxième chapitre à l'explication de la démarche d'analyse entreprise et à la définition des concepts de base de la statistique lexicale, tels : la loi binomiale et la loi normale ainsi que les différents tests comme le calcul des probabilités, l'écart-type, l'écart réduit, etc. A ce stade, nous aborderons également la notion de norme lexicologique, étape obligatoire pour toute étude lexicométrique.

Dans le dernier chapitre, nous présenterons, tout d'abord, le corpus qui sera étudié ainsi que les étapes de la préparation de la base de données (hyperEDM.exe), en commençant par la numérisation des livres, la correction des erreurs, le découpage des textes, etc. Ensuite, nous mettrons en exergue les différentes fonctionnalités du logiciel Hyperbase.

Dans la deuxième partie, nous comptons approcher la structure du vocabulaire, c'est-à-dire tout ce qui est formel et qui touche au traitement statistique. Ce qui nous donnera la possibilité de réaliser une exploration générale de notre corpus. En effet, cette étape prendra en charge l'aspect quantitatif du lexique en l'absence de tout intérêt pour la signification des mots. Nous allons alors traiter les unités lexicales en tant qu'unités numériques susceptibles d'être soumises à tout genre d'opérations statistiques.

Ainsi, dans le premier chapitre, il sera question de *l'étendue du corpus*. À ce stade, nous nous intéresserons à la longueur des mots, afin de connaître le penchant de l'auteur à privilégier les mots courts ou longs. Cette analyse nous renseignera sur le style de chacun des textes ainsi que sur le genre auquel ils sont affectés, d'autant plus qu'El Maleh refuse d'inscrire ses écrits dans un genre bien particulier.

Notons que les différents tests et analyses effectués à ce niveau, servent également à comparer l'ordre chronologique de parution des textes, d'abord, avec l'étendue des occurrences (N), ensuite avec l'étendue du vocabulaire (V). Ce qui nous permettra de dire si l'auteur varie ses thématiques en fonction de sa production ou bien il ne fait que jouer sur la forme.

Le second chapitre s'attardera sur *la richesse et l'accroissement du vocabulaire*. Ce volet est incontournable en lexicométrie et se réalise selon deux visées opposées. La première, la richesse lexicale, considère l'œuvre comme une entité achevée, ce qui permettra de distinguer le vocabulaire le plus riche de celui le moins riche. La deuxième, l'accroissement du vocabulaire, appréhende le corpus selon la progression de ses parties, donnant ainsi une idée sur l'évolution du vocabulaire d'un point de vue chronologique.

Les catégories grammaticales constitueront le troisième et dernier chapitre de cette partie. Ainsi, nous dégagerons, pour l'ensemble du corpus, la distribution des cinq catégories grammaticales majeures, ce qui n'est point évident si nous l'abordons de façon intuitive. Notons que les recherches en statistique lexicale ont démontré que les proportions des catégories grammaticales sont liées au style de l'auteur et au genre des œuvres⁹. D'autres ont prouvé que la distribution pourra être révélatrice de l'époque de production des textes mais aussi du genre¹⁰.

En somme, ladite partie nous invitera à mettre en œuvre les traitements statistiques concourant à explorer l'œuvre d'El Maleh. Le résultat prévu nous

⁹ Etienne BRUNET, *Le vocabulaire de Proust I, Étude quantitative*, Genève, SLATKINE, 1983, p. 836

¹⁰ P. GUIRAUD, *Les Caractères Statistiques du Vocabulaire*, Paris, Presse Universitaire de France, 1954, pp. 34-53

montrera la spécificité de l'écriture maléhiennne et nous assurera certainement une vision claire et objective des caractéristiques formelles des textes soumis à l'analyse.

La troisième partie, quant à elle, est de nature sémantique. Elle prendra en charge l'analyse sémantique des unités lexicales qui demeurent inabordables en l'absence d'une interpellation des textes et d'une étude des mots dans leur contexte. Les unités lexicales feront, au cours de cette étape, l'objet de plusieurs opérations statistiques comme la connexion lexicale, le calcul des spécificités, etc. La présente étape nous permettra essentiellement d'extraire les thèmes majeurs du corpus servant d'indices pour révéler les liens qui se tissent entre les thèmes abordés et les dominants formels adoptés.

A l'instar des deux premières parties, la troisième sera également subdivisée en trois chapitres. L'objectif du premier, intitulé *connexion lexicale*, sera de mesurer la distance entre les différents textes du corpus et de dire s'ils sont proches ou éloignés l'un de l'autre. Plusieurs méthodes seront mises en œuvre, d'abord nous utiliserons celle de Jaccard malgré sa sensibilité aux écarts des étendues ; puis, pour remédier aux insuffisances de la première, nous aurons recours à la loi binomiale. Également dans cette étape les calculs numériques, les représentations graphiques et l'analyse factorielle nous serviront pour identifier les rapprochements et les éloignements ainsi que les ressemblances et les différences entre les partitions de notre corpus.

Nous comptons dans un deuxième chapitre reconnaître l'identité du vocabulaire du corpus étudié, c'est-à-dire faire jaillir la charge sémantique qui détermine l'appartenance thématique de chacun de ses éléments. Une étude de ce genre ne peut que s'inscrire dans le cadre des approches qui touchent au contenu du lexique et qui abordent son aspect sémantique. Dans cette étape, nous chercherons

des éléments de réponse à notre problématique à travers l'extraction des liens qui unissent la structure et le contenu du vocabulaire d'Edmond Amran El Maleh.

Ainsi, nous nous intéresserons, dans un premier temps, aux spécificités du vocabulaire suivant une étude exogène. L'objectif de cette dernière sera de connaître le vocabulaire spécifique de l'œuvre maléhiennne pour identifier les thèmes qu'il privilégie. Notons que ladite étude sera réalisée en nous basant sur la comparaison du vocabulaire spécifique avec un corpus externe, en l'occurrence le *Trésor de la Langue Française informatisé*.

Dans un deuxième temps, nous nous attarderons sur le vocabulaire spécifique suivant une étude endogène. En effet, le logiciel *Hyperbase*, se basant sur des lois et des algorithmes, nous offre automatiquement la liste du vocabulaire spécifique à chaque texte. Toutefois, si dans la première étape la norme de référence est le T.L.F.i, ici, les partitions seront confrontées à l'ensemble du corpus qui servira de norme. Les regroupements en champs sémantiques nous permettront de dégager le profil thématique qui constitue la charpente de base de chaque texte.

Nous avons signalé plus haut que l'un des défis soulevés par El Maleh, dans son projet d'écriture, est comment transmettre la culture et l'univers de sa langue maternelle, « *logha el oum* », à travers un texte écrit en français, d'autant plus que la question de la culture reste inséparable de celle de la langue. D'ailleurs, il a signalé, dans la préface de l'un de ses textes, cette dialectique d'antagonisme qui se tisse entre ces deux langues :

« (...) Je pense que la langue est le lieu même où se manifeste, dans toute sa complexité, cette collision de deux cultures, situées dans une hiérarchie historiquement déterminée, l'une en position de domination par rapport à l'autre. »¹¹

De ce fait, l'auteur souligne clairement la relation conflictuelle entre le français qui souhaiterait imposer sa suprématie à l'image du colonialisme et la langue maternelle des marocains, la langue de l'affect, qui nourrit leur imagination, leurs sensations, leurs passions et leurs rêves.

Ajoutons à ceci un autre constat que tout lecteur d'El Maleh peut soulever, c'est que les textes maléhiens foisonnent d'expressions et de mots arabes, berbères, judéo-arabes, etc., chose qui nous poussera à réaliser une autre étude s'attardant bien évidemment sur cet état de fait.

Cette tâche n'est point aisée, du moment qu'elle nous obligera de constituer un autre « sous corpus » regroupant tous les mots et expressions permettant de faire une étude sur les faits langagiers qui surgissent dans un univers multilingue.

Partant, nous concevrons un troisième chapitre intitulé : ***Le plurilinguisme dans l'œuvre d'EDM***. Nous nous pencherons alors sur l'analyse de phénomènes linguistiques divers et variés. Il s'agit, en effet, de l'emprunt, du calque, de l'alternance codique et de l'interférence phonétique.

Bien sûr nous ne comptons pas cerner tous les faits multilingues présents dans le corpus, parce que l'œuvre d'El Maleh reste une mine inépuisable. De ce fait, nous nous intéresserons à ceux dont l'étude nous semble avantageuse pour l'examen de

¹¹ Edmond Amran EL MALEH, *Parcours immobile*, Paris, André Dimanche, 2000, p. 13

notre problématique de recherche.

Mais signalons d'emblée que, pour éviter l'alourdissement de la partie théorique par une revue de la littérature portant sur les phénomènes langagiers qui seront traités, nous choisirons de dresser un bref aperçu définitoire de chacun des phénomènes soulevés, au fur et à mesure de la progression du chapitre.

La conclusion générale présentera les différents résultats obtenus, que ce soit au niveau de la structure (forme) ou du contenu (fond). Nous mettrons en exergue les particularités de l'écriture maléhiennne sans oublier, bien évidemment, de revenir à notre problématique de recherche.

Après avoir élucidé la perspective dans laquelle s'inscrit notre étude, nous entamons la première partie qui s'intitule *Cadre théorique et méthodologique*

PREMIERE PARTIE
Cadre théorique et méthodologique

La première partie se fixe pour tâche la présentation des principes théorique et méthodologique de notre thèse. Elle est subdivisée en trois chapitres.

Le premier s'attardera sur la présentation de l'auteur et de son œuvre. Nous commencerons par présenter sa biographie ainsi que ses écrits qui recouvrent des critiques sur l'art, des essais et des romans. Aussi, ledit chapitre sera une occasion pour aborder la littérature de l'exil au même titre que l'imaginaire et l'histoire des juifs au Maroc, et ce, dans le but de comprendre l'essence même de l'œuvre étudiée.

Le second chapitre sera consacré aux principes et méthodes de la statistique lexicale. Le but est d'élucider le rapport entre linguistique et statistique sans oublier de présenter les concepts de base et les outils utilisés par l'analyse lexicale, nous faisons allusion, ici, aux techniques de dépouillement et de quantification du corpus. Les principales lois qui seront utilisées lors de la réalisation de notre recherche seront aussi mises en évidence.

Le dernier chapitre s'intéressera au corpus ainsi qu'à l'outil d'analyse, à savoir le logiciel *Hyperbase*. Dans un premier temps, nous commencerons par présenter le corpus de travail en expliquant les étapes de la préparation de notre base de données. Dans un deuxième temps, nous décrirons le logiciel *Hyperbase* qui sera utilisé pour le traitement automatique des œuvres d'El Maleh. Nous y examinerons l'efficacité de ce programme informatique, en nous basant sur un manuel destiné à cette fin, conçu par Étienne BRUNET. Ainsi nous réaliserons une synthèse sélective dudit manuel tout en nous focalisant sur ce que nous estimons utile à cette étude.

CHAPITRE PREMIER

L'homme et l'œuvre

Il est des figures littéraires et intellectuelles qui, de leur vivant, arrivent à marquer indélébilement leur temps. Le sillage laissé par chacune de leurs paroles ou positions acquiert le statut d'une sorte de passage obligé pour les contemporains et les générations futures. C'est le cas d'Edmond Amran El Maleh, écrivain marocain de confession juive, natif du sud du Maroc. El Maleh est un homme de grande sensibilité, de lucidité, d'émotion débordante, d'un engagement contre l'injustice et l'exclusion, attaché à ses valeurs, sincère dans toutes ses idées qu'il développait avec modestie. Cet écrivain éprouvait une admiration particulière pour les arts plastiques en général et pour la peinture en particulier. Il a déjà derrière lui une longue carrière de critique d'art et de nombreux articles abordant les réalisations artistiques nationales.

L'œuvre d'El Maleh est si importante et si exigeante que l'identité narrative de cet auteur prend racine dans les tréfonds anthropologiques et historiques du Maroc. Il opère en observateur prodigieusement attentif aux pratiques et aux émotions, Il dévisage le Maroc comme on dévisage l'être aimé. Il en décrit les mutations sociales à travers des croquis parfois acides, mais le plus souvent empreints d'une tendresse essentielle.

Dans ce qui suit, nous nous intéresserons, tout d'abord, à la biographie d'Edmond Amran El Maleh. Ensuite, nous présenterons sa conception personnelle

de la pratique picturale marocaine, pour terminer ce chapitre avec une présentation générale de la toile de fond de son œuvre littéraire.

A. Notice biographique sur Edmond Amran El Maleh

Edmond Amran El Maleh, qui repose dans le cimetière marin d'Essaouira, voit le jour à Safi, le 30 mars 1917, au sein d'une famille juive intimement mêlée à l'histoire d'Essaouira depuis sa fondation en 1764 et entreprend le grand voyage le 15 novembre 2010.

Militant au sein du parti communiste marocain, il s'engage cœur et âme dans la lutte du peuple marocain pour son indépendance contre le colonialisme français et connaît « *une vie exposée au danger de la répression, de la terreur policière* »¹². En 1948, il est élu membre du comité central et du bureau politique dont il assume les responsabilités. En 1959, étant en désaccord avec les décisions et les méthodes bureaucratiques du parti, il s'en sépare définitivement et renonce à toute activité politique.

Révolté par la répression contre la population, il quitte le pays en 1965 et s'installe en France avec sa femme, Marie-Cécile Dufour. Après avoir enseigné la philosophie au lycée Mohamed V de Casablanca, il continue à enseigner la même discipline à Paris, au collège Sainte-Barbe tout en entamant une thèse sur la violence sous la direction de Paul Ricœur. Malgré son séjour en France, il n'a pas hésité à effectuer des visites régulières au Maroc.

Parallèlement à son métier d'enseignant, il mène une activité de journaliste et de critique d'art. Il publie des articles dans divers journaux. Il collabore au journal

¹² Marie REDONNET, *op.cit.*, 2005, p.67.

Le Monde et réalise des entretiens avec plusieurs philosophes comme : Castoriadis, Claude Lefort, Miguel Adensour, Jacques Rancière, Jürgen Habermas, Kostas Axelos. Il réalise également une étude sur la culture juive marocaine tant pour le *Mémorial du Maroc* que pour la *Grande Encyclopédie du Maroc*.

Sa vocation littéraire ne l'a pas mis à l'abri des charmes des beaux-arts et de la peinture notamment. En effet, El Maleh s'affirme comme un véritable critique d'art. Il publie un livre sur la peinture d'Ahmed Cherkaoui et de nombreux articles sur les arts plastiques. Il noue des amitiés profondes avec les artistes marocains comme Mohamed Melehi, Fouad Bellamine, Mohamed Kacimi, Khalil Gharib, Abderrahim Yamou, etc. En 1999, il s'installe définitivement au Maroc après le décès de Marie-Cécile, qui met fin à son exil.

Malgré l'éloignement de l'écrivain de sa terre natale et de son espace culturel, il apparaît de plus en plus proche du Maroc à travers ses écrits. En effet, dès le début des années 80, vers l'âge de 63 ans, il s'adonne à l'écriture littéraire, il prend la plume avec *Parcours Immobile* et écrit dans différents genres (romans, nouvelles, essais...). Son œuvre suit les méandres d'une pensée sinueuse et profonde, à travers laquelle il célèbre une mémoire aux multiples facettes : arabe, juive et berbère. Il aborde avec justesse le chaos des peuples qui s'étouffent et s'affrontent, s'anéantissent parfois et ressurgissent, pour se donner mutuellement leurs valeurs, les échanger ou les relativiser.

À 92 ans, il est nommé « Doyen des écrivains marocains de langue française ». En 1996, il reçoit le Grand Prix du Maroc pour l'ensemble de son œuvre. En 2004 une fondation a été créée en son nom pour le dialogue entre les cultures plurielles du Maroc. La même année, il reçoit deux récompenses

prestigieuses, l'une, le Wissam Alaouit du roi lui-même, l'autre de la France qui le nomme chevalier de la Légion d'honneur, Marie Redonnet dit à ce propos :

« Ce fut pour lui un moment important, celui de cette double reconnaissance qui le consacre enfin entièrement dans sa double culture tout autant que dans son identité, unique et plurielle en même temps, d'écrivain marocain juif de langue française. Maintenant que cette reconnaissance est enfin venue, des deux côtés, j'ai l'impression qu'il est tranquille et qu'il peut de nouveau prendre ses distances par rapport aux honneurs »¹³.

Après avoir dressé un bref aperçu sur la vie de l'auteur, nous nous intéressons, dans ce qui suit, à la conception artistique d'El Maleh.

B. Edmond Amran El Maleh et l'Art

El Maleh avait un intérêt soutenu et bienveillant pour les arts plastiques en général et pour la peinture en particulier. D'ailleurs, le « premier » livre qu'il publie en 1975, en collaboration avec A. Khatibi, T. Maraini et M. Melehi, était entièrement dédié à l'illustre peintre marocain Ahmed Cherkaoui qui était pour lui « un éveil à la peinture »¹⁴. En effet, leur rencontre dans les milieux parisiens fut un moment crucial dans la découverte de la peinture marocaine contemporaine par El Maleh qui était séduit par les travaux de Cherkaoui. Ce dernier s'inspirait de sa terre natale, à savoir le sud du Maroc, qui se caractérise par sa terre ocre et son climat chaud. La représentation de la région à laquelle appartenait cet artiste se reflétait dans ses œuvres par l'utilisation des dérivés de la couleur ocre et des

¹³ Marie REDONNET, *op.cit.*, 2005, p.10.

¹⁴ Ibid., p. 159.

lettres arabes tout en se référant aux symboles amazighs, ce qui fait qu'El Maleh croisait ce peintre avec chaleur et ferveur.

1. Enfance : esquisse d'un esthéticien

El Maleh a pratiqué les arts plastiques à sa manière. À une certaine époque de sa vie, il s'est adonné d'abord à la photographie, puis il s'est exercé à travailler artistiquement l'argile et la céramique. L'intérêt qu'il accordait à la peinture n'était pas nouveau. D'ailleurs, il a fait à Marie Redonnet la révélation suivante :

« La peinture n'est pas une chose nouvelle pour moi. J'ai commencé à m'intéresser à la peinture aussi loin que je me souviens. Avec mon frère on s'intéressait à la peinture. A la maison on avait des reproductions de Manet. On s'enthousiasmait pour Dali »¹⁵.

De même il décrit agréablement, dans *Parcours immobile*, comment il s'épanouissait pour les images du grand Larousse :

« la première image du sexe de la femme c'était dans ce gros livre rouge à fermoir métallique, le deuxième livre en français après le dictionnaire Larousse, deux portes imposantes à l'entrée d'un monde nouveau (...) Fascination de l'image, privilège sans précédent pour les enfants de ce pays : le sexe en image, l'image du sexe »¹⁶.

De ce fait, l'intérêt d'El Maleh pour l'esthétique est la cristallisation d'une passion enfantine éveillée par l'ambiance douillette du foyer familiale qui était

¹⁵ Ibid., p. 163.

¹⁶ Edmond Amran EL MALEH, *op. cit.*, 1980, p.29

très sensible aux pouvoirs de l'image et de ses virtualités plastiques. Cette passion a marqué le destin d'un garçon qui allait vers une prise de conscience du grand drame humain et de la petite lueur d'espoir que pourraient lui procurer l'art et la littérature.

Sa lucidité et sa bienveillance lui ont permis de comprendre que la peinture marocaine se caractérisait par la pénurie de textes théoriques ou leur légèreté. Les peintres souhaitent continuellement avoir des écrits sur leurs créations artistiques afin de donner du sens à leurs œuvres, mais aussi pour se démarquer et se retrouver dans la scène picturale de leur environnement. Il plaignait le sort des peintres marocains postcoloniaux qui n'ont eu ni l'aide ni le succès qu'ils méritaient malgré leur statut initiateur d'une pratique picturale contemporaine. Pour lui, le peintre postcolonial marocain était saisi « *dans la situation d'un navigateur qui poursuit son périple, tout en s'efforçant dans le même temps d'établir des cartes maritimes, des instruments de navigation qui lui font défaut* »¹⁷.

Donc, son engagement auprès du mouvement artistique marocain contemporain répond à un besoin immédiat d'accompagnement et de promotion de la création plastique nationale dans toute sa fragilité et son originalité conceptuelle, thématique et matérielle.

2. Esthétique moderne

Avant d'aborder la critique d'El Maleh, signalons qu'il adopte la posture de l'esthétique moderne amorcée par le philosophe allemand Alexander Gottlieb Baumgarten qui tentait de délimiter un champ très varié hébergeant une pluralité

¹⁷ Edmond Amran EL MALEH, *La peinture marocaine au rendez-vous de l'histoire*, Wafabank, Casablanca, 1988, P.19.

de thèmes relatifs à la doctrine de la sensibilité, l'imagination productive, les théories de l'art et la question de la réceptivité du goût. Ce philosophe a défini l'esthétique comme « *la science du mode de connaissance et d'exposition sensible* »¹⁸.

Il intègre le problème de la sensibilité dans la sphère des connaissances qui, selon la tradition de la pensée de Leibniz et Wolff se subdivisent en deux branches, à savoir : les connaissances claires et obscures. Les claires sont classées, à leur tour, en confuses et distinctes ; les connaissances confuses entrent dans la sphère de la connaissance du sensible, tandis que les connaissances distinctes entrent dans la sphère de la connaissance rationnelle.

Quoique l'instance de la subjectivité, à savoir l'idée du beau et du sensible, émerge et réclame sa place dans la sphère de la connaissance, Baumgarten l'assimile à une forme de connaissance « inférieure », dans le sens où elle instaure un rapport intuitif avec les choses, sans la médiation de la raison.

Le mérite de Baumgarten est celui de signaler que l'esthétique n'est pas une discipline unitaire mais s'instaure à l'intérieur d'un contexte philosophique bien précis. Ce contexte est formé par une pluralité de thèmes liés autour d'un dénominateur commun représenté par la sensibilité, l'imagination et la sensation. Cela veut dire que l'approche du sensible implique une émancipation créative et poétique de la part de l'esthéticien qui est amené à formuler des discours plastiques, à émettre des jugements de goût et à décrire les procédés subjectifs qui sont à la base de la création artistique.

¹⁸ Zouhir ZIGHIGHI, « *Une poétique de la matière* », In : Anouar Ben Msila. Edmond Amran El Maleh, Art, culture et écriture : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013, p. 272.

3. Attitude esthétique d'Edmond Amran El Maleh

L'attitude esthétique qu'El Maleh développait en faveur de la pratique picturale paraît insaisissable pour le public non averti, accoutumé aux conformités plastiques habituelles. En effet, dans sa stratégie esthétique, il engageait une érudition impressionnante assimilée à l'œuvre de W. Benjamin, K. Malevitch, O. Grabar, J. Genet, Ch. Baudelaire qui nourrissaient sa pensée plastique. En plus de l'appareillage méthodologique qui le distinguait à l'échelle nationale, il jouissait d'une sensibilité esthétique singulière qui s'attachait à une vision contemporaine traversée par une anthropologie de l'art islamique et de sa spiritualité mystique. Il parcourait l'histoire de l'art depuis l'ère antique avec Platon, en passant par l'époque médiévale et l'ère classique jusqu'à l'esthétique moderne développée par Baudelaire, Kant, Hegel et surtout Walter Benjamin dont il partageait l'intérêt croissant avec son épouse Marie-Cécile Dufour. De même, il s'inspirait de plusieurs références, particulièrement le patrimoine tribal et soufi ainsi que les concepts philosophiques orientaux.

Signalons que notre auteur avait pris ses distances par rapport aux thèses occidentales qui réduisent l'art islamique à un art mineur, se basant ainsi sur le bilan des découvertes archéologiques récentes, mais aussi, sur l'énorme production artistique de l'Espagne andalouse, du Maghreb ainsi que d'autres pays musulmans. Par ailleurs, il estimait que l'occident a confisqué à son profit les valeurs artistiques et monopolisé les moyens de la création plastique. Pour lui, l'art occidental a perverti et soumis à la loi un désir pourtant surgi d'un champ relevant du sensible, voire de la sensualité. Ce qui était inacceptable pour El Maleh, c'est que ces lois instituées ont été profondément intériorisées par les intellectuels marocains. C'est la raison pour laquelle il s'est voué corps et âme à l'établissement

d'une approche picturale fondée sur des valeurs esthétiques authentiques, inhérentes à la création marocaine.

De même, il refusait l'idée qu'il n'est d'autre manifestation de l'art que dans l'espace clos de la peinture et des arts plastiques. Il accordait, en effet, une importance énorme à la création artisanale qui nous fait découvrir le paysage esthétique du pays à travers ses formes et ses couleurs. L'artisanat, pour El Maleh, mérite d'être approché pour mieux apprécier l'art marocain, du moment qu'il a exercé une influence indéniable sur plusieurs peintres marocains qui se sont inspirés de ses valeurs esthétiques.

À travers ses écrits, Edmond Amran El Maleh relevait un constat fondamental caractérisant la peinture marocaine : d'une part, il y a la peinture lyrique, plus ou moins influencée par la peinture occidentale ; d'autre part, il y a la peinture du signe, qui fait l'originalité de la création picturale marocaine, du moment qu'elle se trouve ancrée dans les racines de notre vécu culturel et social, parce que le signe utilisé puise dans des formes symboliques et architecturales spécifiques à notre patrimoine, évitant ainsi la standardisation relative à l'art occidental et véhiculant une esthétique de sensibilité différente.

Pour El Maleh, ce qui prime tout d'abord, dans une œuvre picturale, c'est le potentiel plastique et non la technicité du travail, surtout lorsque la technique utilisée est une pure imitation de certaines valeurs occidentales intégrées dans la pratique locale sans connaître les enjeux esthétiques, philosophiques et socioculturels qu'elle peut engendrer. Il cherche la manifestation allégorique de la culture marocaine dans une merveilleuse ellipse picturale synthétisée dans les couleurs locales, les tatouages, la calligraphie, la poterie, les motifs des tapis et dans d'autres éléments transfigurables à l'échelle de la toile.

4. Écrits d'Edmond Amran El Maleh sur la peinture

Lorsqu'il lui arrivait de parler de la peinture, il le faisait de manière très prudente, exprimant ainsi ses craintes et ses hésitations. Son principe suprême pour toutes lectures artistiques était de regarder le travail en silence, de contempler l'œuvre sans dire un mot, de se laisser imprégner par elle jusqu'aux tréfonds de son être. Il estime que la genèse du discours que l'on essaie d'établir est longue à prendre forme :

« Ecrivant cette présentation, et comme chaque fois, je suis assailli par le même doute, le constant sentiment de la quasi impossibilité d'écrire, de parler de la peinture sans tomber dans les généralités, des approximations qui prétendent révéler les significations, le sens d'une œuvre et, dans le bruitage de cette parole parasitaire qui ruine les chances d'y accéder »¹⁹.

Mu par le désir de comprendre, il refusait les généralités ou les lieux communs. Il s'obstinait à saisir l'interprétation la plus proche de la vérité et non celle produite par la juxtaposition des idées de certains critiques d'art. Ses critiques étaient toujours marquées par un sens humain profond et par l'émotion qui suscitait chez lui tout un processus d'analyse dans les profondeurs de la créativité.

Les écrits d'El Maleh sur les peintres marocains sont de valeur inégale, il pousse son investigation au-delà du visible, du palpable, de la couleur et de la composition. Il s'attarde sur la matière comme catalyseur d'autres beautés esthétiques. Il cherche des significations sur la vie, la mort, l'espace et le temps. Il s'intéressait énormément à cette dichotomie espace/temps qu'il considérait

¹⁹ Edmond Amran EL MALEH, « *Flammes et cendres* », In EL MALEH Edmond Amran, *Parcours mobile*, Casablanca, Ed. Galerie 38, Europrint, 2010, p. 124

comme deux facettes d'une seule entité, parce que l'espace est la condensation du temps. Dans ses analyses, il soulevait certains problèmes qui peuvent parfaitement s'inscrire dans les préoccupations actuelles de la peinture nationale. Certes, il avait un penchant naturel pour certains peintres plus que d'autres, pour des expériences plus que d'autres, mais cela n'altérerait en rien sa conviction sur la vitalité et l'originalité de leur travail. Il reconnaissait toute la valeur des peintres marocains et élaborait ces critiques avec la même lucidité et objectivité, mais aussi avec une certaine générosité.

Il disposait d'un goût raffiné et d'une sensibilité esthétique profonde lui permettant d'appréhender les œuvres artistiques qui retenaient son attention. Généralement, les œuvres qui suscitaient son intérêt sont celles qui s'abreuvent à la mémoire tant individuelle que collective. C'est dans cette perspective qu'il a souvent choisi d'aborder des œuvres réalisées à partir de matériaux qui rappellent le patrimoine national tels que le cuir, le bois, la terre glaise, le henné, la chaux, etc, parce que ceux-ci renvoyaient automatiquement aux références de l'artiste et à celles de sa patrie. Toujours à propos de la matière, un concept se confirme dans la sphère de la critique maléhiennne, celui de la poéticité de la matière qui, pour lui « conditionne la création dans l'espace des arts plastiques et détermine les chances de réussite... »²⁰. D'ailleurs, il estimait que la matière n'est pas « *un concept abstrait auquel on serait renvoyé. Elle est là, dans sa concrétude que nous touchons du regard, de la main, que nous respirons et qui, au détour, il arrive qu'elle nous parle et nous retienne* »²¹. Ce concept plastique original met en

²⁰ Edmond Amran EL MALEH, « Bouchta El Hayani », In: EL MALEH Edmond Amran, *Parcours mobile*, op. cit., p. 148

²¹ Aïssa IKKEN, « *Quelques réflexions d'E. A. El Maleh sur l'art et les peintres marocains* », In : BEN MSILA Anouar. *Edmond Amran El Maleh, Art, culture et écriture : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès*, 2013, p. 272.

évidence le rapport intime de l'artiste à l'organique dont il est intrinsèquement lié et qu'il s'attache à manifester dans un supplément de vie artistique. C'est sous cet emblème symboliquement organique que prennent forme les signes, les traits et les couleurs dans la main créative du peintre qui insuffle à cette matière terne et inerte un souffle plastique.

De même, en analysant une toile, El Maleh fait une archéologie des sources, qui engage l'univers autobiographique du peintre, à la recherche d'une correspondance plastique potentielle. C'est ce que Roland Barthes appelle « biographème » :

« j'aime certains traits biographiques qui, dans la vie d'un écrivain, m'enchantent à l'égale de certaines photographies ; j'ai appelé ces traits des biographèmes »²².

En manipulant ces biographèmes, El Maleh tente de percevoir l'originalité de l'œuvre par rapport à un repère autobiographique chargé de sens et de piste d'investigation.

De ce fait, en corrélant ces deux principes de biographème et de poéticité de la matière, El Maleh se réjouissait de chercher, dans l'œuvre de l'artiste, une éventuelle manifestation de sa mémoire visuelle et l'émancipation de son imaginaire créatif par le biais d'une spiritualité originale. Ainsi, lors de son approche de l'œuvre d'Ahmed Cherkaoui, par exemple, il mettait en évidence les traits d'une ellipse plastique capable de représenter un souffle de vie et une

²² Roland BARTHES, *La Chambre claire, Notes sur la photographie*, Cahiers du cinéma, Paris, Gallimard/Seuil, 1980

mémoire collective, celle bien évidemment des cherkaoua. Cette ellipse plastique se manifeste dans une touche chromatique particulièrement allusive. Ainsi, El Maleh louait l'utilisation du signe graphique arabe, le tiffinagh, la couleur rouge des terres arides et la lumière écrasante du sud marocain, du moment que cet ensemble de traits, ainsi que d'autres formes dépouillées, synthétisent l'ambiance maternelle dans la spontanéité d'un geste expéditif et dans l'immédiateté du matériau local.

En abordant la créativité singulière de Chaabia Tallal , El Maleh glorifiait son excentricité plastique réfractaire au traditionalisme patriarcale de la société marocaine. Ainsi, son œuvre est approchée par El Maleh dans sa dimension chromatique marquée par le recours à un rouge explosif susceptible de manifester une situation féminine tragiquement allégorique et le destin d'une femme émancipée dans et par la peinture.

Ailleurs, dans les travaux de Khalil El Ghrib, El Maleh montre que l'utilisation de la chaux, marque emblématique de la cité littorale nord-marocaine, traduit une tradition esthétique et représente la mystérieuse métamorphose de la matière dans un processus naturel durant lequel la subtilité humaine s'y immisce avec finesse, pour déposer sa mémoire sociale tout en apportant une touche artistique.

Cette matière blanche, pour El Maleh, nourrit de ses virtualités plastiques un imaginaire collectif dans l'intimité des demeures jusqu'aux remparts de la ville, de même, elle offre une halte paradisiaque, embrumées de parfum océanique, pour les touristes de passage. El Maleh pense aussi que Khalil El Ghrib détient le secret de la chaux, cette matrice brûlante qu'il manie avec une habileté séculaire, comme le confie la mère de celui-ci à l'auteur dans *Une femme, une mère* :

« Il faut l'entendre parler de la manière d'éteindre la chaux, des soins qu'il faut apporter pour qu'elle ne soit pas trop diluée mais crémeuse, onctueuse comme un beurre, pas trop compacte et qu'elle se bonifie au fur et à mesure, on y ajoute la fameuse nila, ce bleu qui s'incorpore merveilleusement et parfois aussi, chez les juifs surtout, la moghra.....il en venait à penser à khalil, son fils et il voyait alors en quoi il en est génialement l'héritier. Khalil, le maître de l'écriture blanche, le découvreur du code génétique de cette urbanité faite ville»²³.

Le secret de la préparation de la chaux, transmis de père en fils, témoigne de l'importance esthétique de cette matière pour la population littorale, mais aussi du degré de civilisation d'une communauté qui a pu exprimer un désir plastique commun et un plaisir de vivre ensemble.

El Maleh n'écrivait pas uniquement pour une élite mais se faisait un généreux mécène aussi bien pour les créateurs prometteurs que pour les artistes déjà confirmés. Parmi les peintres qui ont eu la chance d'être approchés par El Maleh nous citons : Farid Belkahia, Hassan Bourkia, Bouchta Hayani, Mouna Cherrat, etc. Lors de son approche des performances nationales, il demeurait très attentif et analysait l'objet peint avec beaucoup de subtilité et de finesse, tout en alliant connaissance théorique et plaisir esthétique pour mieux approcher l'œuvre loin de toute motivation lucrative.

5. Amitiés d'Edmond Amran El Maleh

Parmi ses amis proches, il côtoyait des artistes avec qui il chemine et dont il interroge constamment l'œuvre plastique. Il les accompagnait également par l'écriture et tissait avec eux des relations de convivialité, et ce, dans le but d'atteindre la profondeur de leur expérience créatrice. Ainsi, écrivain et peintres cohabitaient et leurs relations d'amitié étaient couronnées par des échanges

²³ Edmond Amran EL MALEH, *Une femme, une mère*, Grenoble, La pensée sauvage, 2004, pp. 9-10.

fructueux : El Maleh adorait les œuvres de ces artistes et, en signe de reconnaissance, ces derniers lui dédiaient certaines de leurs toiles dont chacune possède une histoire et représente l'amitié unissant l'auteur à un artiste. D'ailleurs :

« En franchissant le seuil de l'appartement d'Edmond, le visiteur s'imprégnait instantanément d'une ambiance invisible émanant des multiples œuvres de peinture, accrochées ou déposées contre les murs dans un ordre qui lui était particulier. Il semblait affirmer, par cet ordre indéfinissable, que l'artificialité de l'œuvre peinte, offerte à la contemplation, à la saveur du regard, explose en puissance, anéantit les repères, brise les cadres. Son espace était un lieu privilégié de rencontre où on pouvait apprécier des tableaux, des masques, des statuettes, des spécimens de poterie »²⁴.

De ce fait, ceux qui ont eu la chance de visiter la demeure d'El Maleh la considéraient comme un véritable musée d'art plastique.

6. Peinture et littérature : deux facettes d'une seule entité

Tout le monde s'accorde à dire que peinture et littérature forment deux sources nourricières pour El Maleh. D'ailleurs, à ses yeux les deux cheminent vers le même but, celui de restituer de la mémoire culturelle marocaine dans ses multiples configurations :

« je pense que la meilleure expression de la culture marocaine contemporaine, c'est la peinture. La peinture marocaine est essentielle. Elle est infiniment plus avancée, plus ouverte et plus en mesure de révéler les virtualités créatrices de la culture marocaine que la production littéraire. C'est la pièce

²⁴ Aïssa IKKEN, 2013, *op.cit.*, p. 272.

*maitresse de ma façon de voir les choses. Je me demande parfois si je n'ai pas un œil de peintre. »*²⁵

Dans cette perspective, l'approche esthétique d'El Maleh est le prolongement naturel d'un mode de pensée qui s'investit dans le pictural par-dessus le littéraire. Elle se consacre laborieusement dans la réhabilitation de la mémoire culturelle marocaine dans ses multiples facettes : orale/scripturale, populaire/savante et littéraire/visuelle, pour mettre en valeur un héritage riche en virtualités plastiques.

Ajoutons que la peinture constitue non seulement un art plastique pour El Maleh, mais devient un sujet de prédilection et un mode d'écriture. En effet, formes, couleurs, matières, figures, stéréotypes, portraits... jalonnent l'ensemble de ces récits. D'ailleurs, le professeur Anouar Ben Msila stipule à ce propos :

*« C'est en peinture qu'El Maleh écrit...peinture et écriture se tiennent organiquement liées, et rarement auront entretenu une relation aussi fusionnelle. De cette fusion adviennent le sens, la valeur et l'unité de l'œuvre »*²⁶.

Le lecteur, en parcourant les récits d'El Maleh, a l'impression de contempler un tableau de peinture. De plus, presque tous ses livres portent une illustration de couverture d'ordre picturale, signe indéniable de son attachement à la peinture, mais aussi élément évocateur des tableaux qu'il dresse au fil de la narration. C'est aussi une sorte de métonymie où le contenant, le livre, à travers la représentation

²⁵ Marie REDONNET, *op.cit.*, 2005, p. 162.

²⁶ Anouar BEN MSILA, « *L'écrivain, le peintre et l'allégoricien* », In : Anouar Ben Msila. Edmond Amran El Maleh, Art, culture et écriture : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013, p. 272.

picturale qui l'orne, fait allusion au contenu qui n'est qu'un reflet scriptural de la peinture.

Pour paradoxal que cela puisse paraître, El Maleh, en regardant une toile, écrit. En revanche, en rédigeant ses récits, sa narration semble prendre forme dans des toiles. Ce mode d'écriture reste difficile à apprivoiser parce qu'il ne libère pas un sens plein et définitif, mais invite à lire les effets du sens entre les lignes et offre une signification particulière, plutôt sensorielle et émotionnelle.

Littérature et peinture se déclarent donc consubstantielles et c'est cette consubstantialité qui constitue une originalité esthétique pour El Maleh.

Finalement nous pouvons dire que l'approche esthétique d'El Maleh est à cheval entre l'histoire et la philosophie, ces deux variables se trouvent finalement étoffées par une perception personnelle des multiples virtualités de la matière et de l'image.

Lors de ses lectures picturales, El Maleh se trouvait heurté à un problème sérieux, c'est que les outils d'analyse restent souvent rattachés aux critères de la critique occidentale. Or, pour lui, la lecture d'une peinture requiert des critères qui diffèrent d'une aire culturelle et d'une société à une autre, tout en ayant une dimension universaliste. Il n'était pas contre l'art occidental, d'ailleurs, il revenait souvent à Antoni Tapies qu'il considérait comme une référence majeure dont l'œuvre et les écrits sont d'une valeur incontestable. Mais, il condamnait cette critique qu'il considérait faussée par des matériaux venant d'ailleurs. Il refusait également la critique de complaisance qui ne rendait pas service à l'évolution artistique nationale.

À travers une impressionnante carrière de critique d'art, amorcée par les affinités amicales et une pensée esthétique incontestable, il a essayé de fonder une école esthétique marocaine tout en puisant dans la mémoire visuelle nationale et dans les différentes interrogations plastiques qui jalonnent la peinture contemporaine. Cela veut dire que les arts plastiques ne forment plus seulement un objet d'interprétation pour El Maleh, mais aussi un terrain fertile à une implication effective dans un projet plus large, celui de la sauvegarde des identités.

C. Toile de fond des œuvres d'Edmond Amran El Maleh

Malgré l'exil volontaire d'Edmond Amran El Maleh, il a pu montrer comment l'éloignement géographique attise le lien affectif avec les origines de l'être exilé. Son éloignement lui a offert l'opportunité de surgir des confins du silence et d'être présent sur la scène littéraire par l'acte scriptural. Il a trouvé dans l'écriture une sorte de remède, à travers lequel il reconstitue son identité brisée, le lieu idéal où il va se réfugier afin d'exprimer toute l'horreur inhérente à la question de l'exode des juifs marocains. D'ailleurs, la genèse même de son œuvre littéraire est bâtie sur la nécessité de concevoir une pléiade textuelle autour du traumatisme causé par le départ abrupt des communautés juives marocaines.

Dans les points qui suivent, nous nous intéresserons à ce qui peut constituer la toile de fond de l'œuvre d'Edmond Amran El Maleh. Nous commencerons tout d'abord par la présentation de quelques caractéristiques de l'œuvre maléhienne, en passant par la littérature de l'exil et en terminant finalement par l'imaginaire judaïque et l'histoire des juifs au Maroc

1. Œuvre maléhienne

El Maleh, cet auteur cosmopolite, nous offre une œuvre labyrinthique qui place le lecteur d'emblée sur un terrain enchevêtré. Il nous présente, par son génie d'écrivain, un texte qui résiste à la lecture et qui défie les grilles d'analyse que les intellectuels tentent de lui soumettre.

Son œuvre, d'une grande qualité littéraire, est traversée par la question principale de l'affirmation de son identité marocaine aux racines judéo-arabes et berbères, et par celle de son refus du sionisme qu'il considère comme une idéologie destructrice, ayant eu des conséquences tragiques sur les palestiniens et sur les communautés juives du monde entier et du Maroc notamment. Il écrit dans différents genres, romans, nouvelles, essais, mais c'est l'œuvre romanesque qui fait la qualité et l'originalité de son écriture. Ainsi, nous présentons une liste non exhaustive de ses écrits littéraires

➤ Romans :

- *Parcours immobile*, Paris, Maspero, 1980, réédité par André Dimanche, 2000
- *Aïlen ou la nuit du récit*, Paris, La Découverte, 1983, réédité par André Dimanche, 2000
- *Mille ans, un jour*, Grenoble, La Pensée sauvage, 1986, Le Fennec, 1990, réédité par André Dimanche, 2002
- *Le Retour d'Abou El Haki*, Paris/Grenoble, La Pensée sauvage, 1988

➤ Essais :

- *Jean Genet, Le Captif amoureux et autres essais*, Grenoble/Casablanca, La Pensée sauvage/Toubkal, 1990

- *Le café bleu. Zrerek*, Thionville/Casablanca, La Pensée sauvage/Le Fennec, 1998
- *Une femme, une mère*, Grenoble, La Pensée sauvage, éditions Lixus, 2004
- *La malle de Sidi Maâchou*, Casablanca, Al Manar, 1998

➤ **Recueil de nouvelles :**

- *Abner, Abounour*, Thionville/Casablanca, La Pensée sauvage/Le Fennec, 1995

➤ **Récit épistolaire :**

- *Lettres à moi-même*, Casablanca, Le Fennec, 2010

Une éthique, sans entraves, anime toujours les œuvres maléhiennes, celle d'aujourd'hui, d'hier et de demain : le devoir de dire la vérité, rien que la vérité, l'obligation de ne pas mentir à l'Autre. En effet, El Maleh n'invente pas des histoires mais il les raconte, il exige que leur caractère, leur constance, leurs métamorphoses, leur authenticité soient sauvegardés et discutés. C'est dans ce sens qu'A. Serhane dit qu'El Maleh est « *notre Voix* »²⁷.

C'est un auteur qui veut communiquer des épisodes historiques, idéologiques, sociologiques et culturelles d'une époque donnée, pour permettre au lecteur d'aujourd'hui de faire un voyage dans le temps passé/futur et circumambuler dans les hauts lieux de la philosophie maléhiennne. Une philosophie qui ne cesse d'introduire des bifurcations brusques qui débouchent finalement sur la blessure ouverte par un événement dramatique, celui de « l'exode des juifs marocains ». En fait, les œuvres d'El Maleh disent autrement la blessure ontologique qu'avait pour conséquence le

²⁷ Abdelhak SERHANE, « *Notre voix* », In: Horizons Maghrébins - Le droit à la mémoire, N°27, 1994. Présences d'Edmond Amran El Maleh. pp. 73-76.

déracinement de 90 % des communautés juives marocaines et, parallèlement, l'escamotage de tout un patrimoine culturel ethnologique, sociologique et linguistique.

Chagriné par la tournure des événements, El Maleh ne se lasse pas de revisiter cette thématique qui lui est chère. Depuis *Parcours immobile*, le départ massif des juifs marocains n'a jamais cessé d'être vécu par l'auteur comme un trauma, d'ailleurs, c'est cet événement cicatriciel qui a induit le « choc émotionnel » d'El Maleh et tout au long de ses récits, il porte constamment le regard sur le destin de ses concitoyens juifs dont il aurait suivi l'itinéraire jusqu'au bout.

Outre la dimension judaïque qui fait surgir des questions relatives aux pressions du déracinement, de l'exode, de l'extermination massive et planifiée des juifs européens par les nazis, El Maleh s'intéresse aussi aux thèmes qui jalonnent la littérature marocaine traitant de l'acculturation, de l'aliénation, du traumatisme, la quête de l'identité, le parcours de l'homme colonisé au Maroc, etc. De même, il ne dissimulait pas le fait que le drame palestinien le remplit d'un malaise profond. Il dénonçait avec virulence la violence ethnocidaire exécutée à l'encontre des réfugiés palestiniens dans les camps de Sabra et Chatila. D'ailleurs, à travers ses écrits, il se faisait un défenseur confirmé de la cause palestinienne au point qu'il est présenté dans un dictionnaire établi par les israéliens comme :

« un juif marocain militant arabe nationaliste et antisémite
fanatique »²⁸.

²⁸ Marie REDONNET, *op.cit.*, 2005. p. 78

De même, El Maleh ne cesse d'instaurer une dialectique féconde, celle d'Israël et du Maroc, dont le pari est de démontrer que ce dernier est le noyau irréductible, le centre de gravitation qui attire tout vers lui. À travers ces récits, il met en scène la rivalité de deux discours. D'une part, le discours sioniste qui accorde aux juifs de la diaspora le statut d'« *idolâtres* »²⁹ ou d'exilés et érige Israël en valeur suprême, du moment qu'il constitue l'incarnation de la terre promise. D'autre part, le discours maléhien qui est absolument antisioniste, sacralise le Maroc et le rehausse au rang de valeur absolue. Il met en exergue l'enracinement ancestral des juifs dans la terre marocaine et fait du Maroc un noyau dur, vers lequel tous converge, un symbole de la symbiose harmonieuse des cultures et des religions, un lieu de l'exubérance des sens. Il confère au retour occasionnel des touristes juifs la forme d'un pèlerinage aux origines qui n'est, en fait, que l'expression de l'attachement indéfectible au pays. Ces touristes, retrouvent le Maroc avec beaucoup de nostalgie et de regret, parce qu'ils ont pris conscience, finalement, qu'ils ont été parti à la recherche d'une illusion, à savoir : l'identité juive. Arrivant au territoire sioniste, ces juifs ont compris tragiquement qu'il n'y a pas d'essence juive mais juste des conditions juives (un État hébreu, une langue officiel hébraïque...). Quoiqu'ils aient pu réaliser en quelque sorte leur rêve de vivre dans la terre promise, ils se sont privés de l'harmonie et de la symbiose qui caractérisaient la coexistence entre les communautés musulmanes et juives sur la terre marocaine, la terre perdue. Parce que la nature des rapports qui relient les juifs sépharades aux juifs ashkénazes est d'ordre conflictuel et ce pour plusieurs raisons. En effet, les ashkénazes, vu qu'ils sont issus de l'Europe, croient en leur suprématie par rapport aux autres. De plus, contrairement aux juifs du Maghreb qui ont mené une vie paisible marquée par la tolérance et la fraternité, les ashkénazes, eux, ont vécu dans un climat de manque de confiance et d'intolérance,

²⁹ Edmond Amran EL MALEH, *op.cit.*, 1986 p. 185

connu toutes sortes de discrimination que ce soit politique, économique ou raciale et survécu par miracle au génocide antisémite nazi, qui visait l'extermination des juifs européens lors de la seconde guerre mondiale. Toutes ces contraintes et bien d'autres ont fait que les ashkénazes empruntent une attitude hostile envers l'Autre³⁰.

Par ailleurs, l'auteur considère Israël comme un lieu de la déshumanisation de l'homme qui se transforme en un monstre « ivres de haine »³¹. En effet, cet auteur cosmopolite blâme et stigmatise les vicissitudes et les conjonctures politiques qui ont conduit les communautés juives à chasser le peuple palestinien de sa patrie et bâtir leur existence « l'État d'Israël » sur les ruines de Palestine. D'ailleurs dans les entretiens qu'il a réalisés avec Marie Redonnet, il dit :

« Depuis la fondation de l'Etat d'Israël, il y a ce racisme qui est consubstantiel à l'idéologie sioniste. Il y a le projet d'une terre juive purifiée de toute présence étrangère donc palestinienne, au nom du peuple élu. Quand on analyse le poids des mots, on découvre à la base une position raciste contre laquelle les sionistes ont voulu se défendre »³².

Parallèlement aux juifs marocains qui étaient une proie du prédateur sioniste, les palestiniens sont aussi victimes de cette machine diabolique qui ne fait que reproduire le génocide nazi dans le but d'exterminer les palestiniens et de les chasser de leur patrie, pour finalement s'accaparer de la Palestine et légitimer la création de l'État d'Israël.

³⁰ Notons que cet antisémitisme dont souffraient les ashkénazes était à l'origine même de l'aspect ethnocidaire de l'identité juive appelé communément : sionisme.

³¹ Ibid., p. 32

³² Marie REDONNET, *op.cit.*, 2005, p.78.

Ainsi, le fils de la terre promise et du territoire d'accueil cherche à établir le lien entre l'histoire du passé récent des juifs marocains (au Maroc) et leur positionnement en Israël. Il nous présente une œuvre qui parcourt les territoires les plus divers dont le Maroc constitue une longue et incontournable escale. Ainsi, l'acte scriptural d'El Maleh partage, avec nous lecteurs, une obsession évidente, celle de l'histoire et de l'évolution politique du pays, il arpente les filets et les replis de la mémoire, reconstitue le passé et peint le présent, pour nous offrir ainsi une sorte d'anthropologie historique.

2. Exode et la littérature de l'exil

Selon Yasmina El Haddad et Ieme van der Poel³³, dans les littératures de la diaspora marocaine, le Maroc demeure présent. Mais le regard porté sur ce Maroc suit deux sillages diamétralement opposés : d'une part, il y a des auteurs qui se montrent très critiques à l'égard de leur pays d'origine, tels Rachida Lamrabet et Naima Albdiouni ; d'autre part, il y a ceux qui véhiculent une image nostalgique largement fantaisiste sur le Maroc, citons à titre d'exemple Hafid Bouazza et Abdelkader Benali. Il en découle qu'il existe une contradiction remarquable entre la littérature de la diaspora qui trouve son origine dans l'émigration ouvrière des années 1960 et celle qui se rapporte à l'exode de la communauté juive du Maroc. De ce fait nous pouvons admettre l'hypothèse de Rosemay Marangoly George³⁴, telle qu'il l'a confirmée pour les écritures d'exil anglophones, qu'il existe deux genres littéraires dissemblables, à savoir : la littérature de la migration et la

³³ Yasmina EL HADDAD, Ieme van DER POEL, « *Variations sur la figure du double : la mémoire judéo-marocaine chez E.A. El Maleh et M. Bennaroch* », In : Anouar Ben Msila. Edmond Amran El Maleh, Art, culture et écriture : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013, P. 272.

³⁴ Marangoly George ROSEMARY, *The Politics of Home: postcolonial Relocations and Twentieth-Century Fiction*, Berkeley (cal), University of California Press, 1996

littérature de l'exil. La première se caractérise par le regard péjoratif que l'auteur porte sur le pays d'origine, un regard marqué par un certain désintéressement. Tandis que dans la seconde, le présent rime toujours avec le passé :

« Le récit de l'exilé est rempli de l'injustice brutale et débilitante d'être séparé de son chez-soi »³⁵.

Nous postulons donc que la littérature judéo-marocaine en tant que littérature de l'exil est hantée de la nostalgie d'un monde qui n'existe plus. De ce fait les adeptes de cette littérature ont souvent recours à un passé révolu et cherchent à remédier amèrement à un présent qui les accable en tant que « juifs ».

Selon Marie-Brunette Spire, ce type de littérature est focalisé sur une « cassure », marquant à la fois une rupture sur le plan personnel et collectif³⁶. Du moment que le départ brusque et massif d'une minorité ethnique et religieuse provoque non seulement une hémorragie culturelle mais annonce la fin de la coexistence de musulmans et de juifs sur le continent africain.

Les récits d'EAM, lorsqu'on les considère de ce point de vue, sont de véritables lieux de méditation et de commémoration de la nostalgie dans toutes ces facettes. Ils nous offrent des textes imbus de regret d'un monde perdu, tout en exprimant une volonté de revenir aux origines, pour célébrer l'harmonie qui régnait entre les différentes catégories sociales, ethnologiques et religieuses

³⁵ "The vicious, debilitating injustice of exile that coats the narrative (of exile) is missing in the immigrant novel." Ibid., p. 175

³⁶ Marie-Brunette SPIRE, « *Des écrivains juifs du Maghreb à la recherche de leur identité* », dans Jean-Claude Lasry et Claude Tapia, *Les juifs du Maghreb : diasporas contemporaines*, Paris, L'Harmattan, 1989, pp. 267-268.

composant la société marocaine à une époque donnée. Mais aussi pour tenter de combler le vide provoqué par le départ massif de la communauté juive marocaine :

« Il faut raconter une histoire, qui remplace et supplée le milieu perdu. »³⁷

Cet exode, qui atteint son apogée dans les années 1958-1961, est mal vécu par El Maleh. D'ailleurs, il l'assimile à une tragédie, d'une double solitude, d'une « déchirure »³⁸ ou encore d'« incheqaq »³⁹ accompagné par l'écroulement d'une société plurielle aussi bien sur les plans culturels que religieux.

Toujours dans le même ordre d'idées, nous nous référons au travail de Rachel Nisselson⁴⁰, qui considère l'exode des juifs marocains comme un ethnocide, une véritable destruction de la mémoire de ces derniers. C'est ce qu'Edmond Amran El Maleh essaye de nous transmettre tout au long de ses récits, il pleure la richesse anthropologique et historique qu'avaient emportée avec elles les communautés judéo-berbéro-arabe dans leur départ pour l'orient et présente cet événement comme une réelle lobotomie culturelle infligée à la mémoire du Maroc. Le caractère abrupt de cette perte est amplifié par le fait, comme le

³⁷ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 160

³⁸ Ibid., p. 16

³⁹ Ibid

⁴⁰ Rachel NISSELON, «*Locating the Productive in the Inthinkable*» In: *Expressions maghrébines*, vol. 09, n° 2, hiver 2010, pp. 15-22.

confirme l'historienne Anne Grynberg, que les juifs du Maghreb émigrèrent souvent par communautés entières⁴¹.

Il serait pertinent de revenir sur le concept d'ethnocide. En effet, l'ethnologue Robert Jaulin a formulé ce concept lors de ses études sur les massacres des Indiens de l'Amérique du sud. Contrairement au génocide qui vise l'anéantissement physique d'une minorité raciale, l'ethnocide n'est pas une extermination physique des hommes mais plutôt une destruction de leur mode de vie, de leurs pensées, de leur morale et de leur culture. Il vise à obliger l'Autre, qui est différent, de se transformer et de se conformer au modèle proposé par les praticiens de l'ethnocide. Ainsi, la négation ethnocidaire de l'Autre conduit à une identification à soi. Concernant les romans d'El Maleh, les praticiens de l'ethnocide sont les propagateurs sionistes qui travaillent dans l'ombre en vertu d'une instrumentalisation de la religion :

« ces taupes au travail souterrain, qui pratiquaient la tactique de la terre brûlée »⁴².

Les stratégies persuasives des agents de la Alya sont marquées en permanence par le souffle biblique dans le but de conduire l'autochtone, que fut le juif marocain, par le chemin de la vraie foi, à la lumière de la religion et de la civilisation. Mais, arrivant à la terre promise, le juif marocain se trouve heurté à

⁴¹ Anne GRYNBERG, *Vers la terre d'Israël*, Paris, Gallimard, 2008, p.100

⁴² Edmond Amran EL MALEH, *op. cit.*, 2002, p. 119

une véritable désillusion et conclut finalement qu'il était victime d'un grand mensonge formulé par la propagande sioniste.

Donc, c'est cette disparition chaotique de la communauté juive qui fait le malheur d'Edmond Amran El Maleh et ce dernier se fait un devoir d'interroger ce départ massif afin de l'arracher au silence qui le trouble et le laisse en plein désarroi. Il s'attache à réhabiliter cette mémoire séculaire, par la force de sa conviction et par la force du verbe qui possède une force intrinsèque, dans l'espoir de faire revivre une vie commune.

3. Imaginaire judaïque

L'œuvre d'El Maleh place le lecteur d'emblée sur le terrain de l'imaginaire universel tout en puisant dans l'imaginaire culturel marocain. Par ailleurs, ses récits actualisent souvent des notions qui se rapportent à la dimension judaïque. Nous jugeons donc nécessaire d'y revenir afin d'assimiler certaines idées évoquées par l'auteur. À vrai dire, ne nous n'abordons pas la judéité comme religion ou ethnie, mais en tant qu'imaginaire, autrement dit, en tant qu'ensemble de représentations communes à un peuple ou à un groupe social bien déterminé, d'abord dans son rapport à lui-même, ensuite à autrui. En effet, les représentations juives communes ne se réduisent pas à la religion en commun, ni à la langue en partage, mais constitue un ensemble dans lequel s'imbriquent des éléments religieux, culturels et ethniques et qui s'étendent à l'ensemble des valeurs émanant du vécu historique. Autrement dit, le temporel et le spirituel forment le piédestal de l'imaginaire judaïque.

Le peuple juif se considère privilégié et différent des autres parce que Dieu a fait plusieurs « Alliances » avec lui et promet à Moïse que ce peuple sera élu parmi les nations : « *Vous serez pour moi privilégiés parmi tous les peuples, car toute la terre est à moi. Et vous serez pour moi une dynastie de prêtres et une nation sainte* » (Exode 19, 5-6). Ayant été choisi par Dieu, Israël est un « peuple élu ». Cette idée est omniprésente dans le monde juif et serait pour la conscience juive un dogme non formulé.

Cependant, avant d'apporter un complément d'information à ce sujet, il est bon de souligner que le sort du peuple juif a toujours été celui de l'exil et de la diaspora, d'où la naissance du sentiment d'être opprimé, d'être minoritaire. Néanmoins ceci ne diminue en rien de son attachement à la terre promise. En effet, la conscience d'être juif doit s'accompagner impérativement de l'idée d'émigration en Israël. Cette obsession de la terre, qui est liée à l'histoire de la longue exclusion des juifs, est venue suite à une autre alliance conclue entre Dieu et Abraham ; c'est ce que suggère la lecture suivante de la Genèse :

« A ta descendance j'ai donné cette terre, depuis le fleuve d'Egypte jusqu'au grand fleuve Euphrate. Le pays des Kénites, des Kenizites, et des Kadmonites » (Exode 12,1).

C'est-à-dire le pays de Canaan, la terre d'Israël telle qu'elle sera occupée par les juifs jusqu'à la destruction du second Temple. Donc, comprenons-nous pourquoi le peuple juif appuie sa revendication sur une promesse divine. Notons que pour les juifs orthodoxes, chaque fois qu'un événement dramatique survient et perturbe le cours de la vie (destruction du temple, l'exil de Babylone, Shoah...)

c'est toujours parce que le peuple juif est dans l'état de péché et a, en quelque sorte, rompu l'Alliance. Donc tout malheur est interprété comme punition de Dieu.

L'imaginaire juif se nourrit des traditions religieuses et culturelles étrangères relatives au milieu dans lequel ils vivent. Là où ils sont, ils développent un imaginaire comportant des caractéristiques assez particulières, une vision originale et attachante de se représenter le monde. Leur manière de concevoir la divinité diffère sensiblement de la compréhension de Dieu en milieu chrétien ou musulman, comme leur attachement indéfectible aux valeurs communautaires et à la terre promise est immuable.

4. Histoire des juifs au Maroc

Notre compétence ne saurait parcourir deux mille ans de vie juive sur la terre marocaine, mais elle tentera de toucher quelques aspects de la cohabitation des juifs avec les autres groupes ethniques composant la société marocaine, notamment les relations intercommunautaires et inter-religieuses des juifs et des musulmans.

Dans le Maroc préislamique, Juifs⁴³ et tribus berbères étaient prédominants jusqu' à la conquête des Umayyades (707), puis des Idrissides qui arrivent à répandre définitivement l'Islam dans la région. Les personnes de confession chrétienne commencent à diminuer vers le VII^{ème} siècle, de ce fait, les juifs qui ont refusé de se convertir à l'Islam, constituaient la seule minorité religieuse non-musulmane. Depuis, et grâce au statut canonique de « dhimmi⁴⁴ », les juifs

⁴³ Des légendes font remonter l'établissement des premières colonies juives du Maroc au roi Salomon, d'autres légendes les associent à la destruction du Temple en -586.

⁴⁴ Nom donné par l'Islam aux Gens du Livre : Juifs et chrétiens

menaient une existence communautaire autonome tout en ayant une participation active à la société.

Ceci porte à considérer cette minorité religieuse comme une entité intégrée grâce à un système d'organisation sociale et économique qui permet la sauvegarde et l'entretien de l'identité religieuse spécifique. Les relations judéo-arabes se situeraient de ce fait entre différenciation et complémentarité, entre autonomie et indépendance, ce qui établit un réseau d'échange et d'influence perceptible au niveau de la langue, des us et coutumes, de la musique et de la culture en général.

La communauté juive du Maroc a augmenté par l'arrivée de groupes chassés de la péninsule ibérique lors des persécutions des Wisigoths⁴⁵. Elle s'amplifie encore une fois par l'afflux de plus de 20.000 séfarades en provenance d'Espagne à cause de l'Edith royal du 30 mars 1492.⁴⁶

Les juifs du Maroc⁴⁷ vivaient généralement dans des mellahs⁴⁸, dont le premier fut celui de Fès, fondé en 1438 sous le règne de la dynastie Mérinide. Le schéma du quartier juif était presque identique à celui de la médina : habitations, synagogues, écoles, fours, moulins, boutiques, etc. Soulignons que le terme « mellah » est spécifique pour le Maroc, il tiendrait son appellation de

⁴⁵ G. Camps, « Réflexion sur l'origine des juifs des régions nord-sahariennes », In ABITBOL Michel (éd), « Communautés juives des marges sahariennes », Jérusalem, 1982, p. 63.

⁴⁶ Y. D. Semach, « Une chronique juive de Fès », Hespéris, Institut des Hautes Etudes Marocaines, p. 91-92, in : Mohammed Kenbib, Juifs et Musulmans au Maroc 1859-1948. Contribution à l'histoire des relations inter-communautaires en Terre d'Islam, Rabat, publication de la faculté des Lettres et des Sciences humaines, 1994, pp. 24-25.

⁴⁷ Jane GERBER, *Jewish society in Fez : 1465-1700*, Leiden. E. J. Brill, 1980.

⁴⁸ Avant leur localisation aux mellahs, les juifs vivaient dans les principaux centres urbains, souvent au milieu de leurs voisins musulmans.

l'emplacement du premier quartier juif bâti sur un marais salant (al-millah)⁴⁹. Cette appellation s'étendra à tous les quartiers juifs édifiés plus tard : à Marrakech en 1557, à Meknes en 1682, à Tetouan, Rabat et Mogador de 1805 à 1807. Cette séparation d'habitations et des lieux de cultes des deux communautés n'atteint pas la sphère économique et sociale. En effet, les échanges se poursuivaient et la majorité des juifs travaillaient dans la médina où ils occupaient parfois des rues entières.

La reconnaissance dont bénéficient les juifs sous l'Islam les autorise à une autonomie interne dans les domaines religieux, civil et juridique. De ce fait, ils disposaient d'institutions rabbiniques et communautaires. Aussi, leur vie sociale communautaire était régie par un nombre de pratiques rythmant la vie quotidienne qui se trouve organisée autour de lieux culturels, tels : la synagogue, la yeshivah, etc.

S'agissant des activités économiques, les juifs marocains pratiquaient essentiellement le commerce et l'artisanat. En effet, selon André Chouraqui⁵⁰, la constante mobilité géographique des juifs, due à leur condition politique précaire, leur imposait l'apprentissage de métiers qui peuvent être exercés partout et qui leur permettaient d'amasser de l'argent, seule propriété facilement transportable.

Notons que l'activité commerciale des juifs se développait de façon notable avec les conquêtes de l'Islam qui ouvrait de grandes voies terrestres et maritimes. Ce constat était favorisé par les atouts propres que les juifs possédaient notamment

⁴⁹ André ADAM attribut cette appellation à une légende selon laquelle les juifs avaient la corvée de saler les têtes coupées des rebelles avant leur exposition publique. Voir, ADAM André, casablanca, Aix-en-Provence, t. 1, 1968.

⁵⁰ André CHOURAQUI, *La condition juridique de l'Israélite marocain*, Paris, éd. Presses du Livre français, 1950, p. 38.

une connaissance des langues, du fait de leur dispersion géographique. Ainsi, ils excellaient dans le commerce international surtout celui de longues distances (les juifs de Salé se sont spécialisés dans le textile importé d'Europe, d'autres partaient jusqu'en Afrique noire, en Asie et en Inde d'où ils importaient épices et aromates, etc). La spécialisation des juifs dans l'import et l'export était possible, voire officiel, par le fait que la tradition musulmane était surtout basée sur l'implantation terrienne⁵¹.

Les commerçants juifs importants constituaient la classe aristocratique de leur communauté. Grâce à leur fortune acquise dans le commerce et leurs affinités avec le pouvoir central qui encourageaient leur activité rentable pour le pays, on leur a concédé des fonctions publiques, dont la frappe de monnaie et des missions diplomatiques.

Les principes du judaïsme conduisent les juifs à une opportune complémentarité avec leurs compatriotes arabes, par exemple, le domaine financier dont l'Islam interdit la gérance aux musulmans leur est confié, de ce fait les juifs travaillaient aussi dans la banque et le prêt. Les préjugés et les réticences que le travail des métaux précieux inspirait aux musulmans favorisent également la prépondérance des juifs dans l'orfèvrerie, notamment à Fès⁵².

Ainsi la différence religieuse crée une organisation fonctionnelle complémentaire des deux communautés, contribuant à leur équilibre socio-

⁵¹ Daniel SCHROETER, *Merchants of Essaouira ; urban society and imperialism in south-western Morocco. 1844-1866*. Cambridge, Cambridge University Press, 1988, p. 21.

⁵² Léon l'Africain, « *Description de l'Afrique* », traduit de l'italien par A. Epaulard, Paris. Librairie d'Amérique et d'Orient, 1956, t. 1, p. 234. Cité d'après Mohammed Kenbib, *Juifs et Musulmans au Maroc, 1859-1948*. Rabat. Publication de la Faculté des Lettres et des Sciences humaines, 1994, p. 28.

économique. La participation juive aurait donc été reconnue en tant que composant différent mais nécessaire.

Concernant l'artisanat, les juifs étaient orientés vers toutes les activités et les travaux traditionnels : le cuir, les métaux précieux, la forge, la boucherie, la cordonnerie et les activités artisanales liées au commerce de vêtements. Précisons que ces activités étaient plus spécifiques à la communauté juive de Fès qui concentrait plus d'artisans que de commerçants et de colporteurs⁵³.

Cependant, l'activité économique des juifs ne se limitait pas uniquement au commerce sédentaire, elle s'étalait aussi au commerce itinérant.

Les itinérants juifs étaient nombreux et indispensables puisqu'ils assuraient le relais entre villes et campagnes. Ainsi, les villages étaient approvisionnés par les colporteurs juifs, eux-mêmes approvisionnés par les commerçants juifs et musulmans de la ville.

Ce type de commerce a favorisé le développement des rapports sociaux intercommunautaires. Aussi, il a joué un rôle essentiel dans le Maroc d'avant le protectorat, où une large partie du pays était classée « bled es-siba » (territoire non-soumis). En effet, les commerçants juifs itinérants étaient les seuls intermédiaires entre arabes et berbères, mais aussi entre les différentes tribus berbères. Le maintien de cette fonction d'intermédiaire paraissait indispensable aux tribus éloignées qui nouaient des relations personnelles « symbiotiques » avec

⁵³ Roger LETOURNEAU, *La vie quotidienne à Fès en 1900*, Paris, éd. Hachette, 1965. p. 101

les juifs, ce qui était, selon P. Flamand, à l'origine de la non application des lois anti-juives de Vichy (1940-1941) dans le sud marocain⁵⁴.

La langue, elle aussi, constitue un indice de marquage sociopolitique et culturelle. Au Maroc, il existe chez les juifs deux niveaux d'expressions : l'oral et l'écrit. La langue vernaculaire, le parler, se rapporte à la vie quotidienne. Elle est influencée par l'espace géographique où elle est en usage, mais aussi par les facteurs historiques qui lui donnent sa diversité. Ainsi, nous distinguons le judéo-arabe, le judéo-berbère et le judéo-espagnol relatif aux communautés hispanophones. Ces parlers se trouvent circonscrites à leur cadre géographique, de ce fait, la phonétique et certains traits spécifiques dans le vocabulaire indiquent l'identité propre à laquelle s'identifie le groupe. Notons également que la diversité des parlers est révélatrice de l'ascendance historique, géographique et culturelle des groupes qui les utilisent. À côté de l'oral il y a la langue écrite, l'hébreu, langue du sacré qui est réservée exclusivement pour l'étude de la science religieuse, du Livre et de la tradition, par conséquent n'en use que l'élite traditionnelle de la communauté.

L'influence réciproque des deux communautés s'est progressivement développée au cours des siècles. Le plus étonnant est l'introduction de certaines pratiques dans la tradition religieuse des marocains comme le culte des saints. En effet, juifs et musulmans, mis à part de leur appartenance religieuse, vénéraient des saints à qui ils attribuaient des pouvoirs surnaturels de guérison, de réalisation de vœux et de protection. Mais ce qui est impressionnant, c'est la vénération

⁵⁴ Pierre FLAMAND, *Diaspora en terre d'Islam. Les communautés israélites du sud marocain*, éd. Imprimeries réunies, Casablanca, 1958, p. 98

commune de saints identiques par juifs et musulmans, chose qui a accentué leur « symbiose culturelle »⁵⁵.

Quoique les juifs marocains constituaient une minorité face à une majorité musulmane, ces deux communautés s'influençaient mutuellement tout en évoluant ensemble. Cependant, l'arrivée du protectorat français, par ses desseins coloniaux, ses visées politiques, sa gestion à base de privilèges accordés, l'incitation et l'entretien de divisions populaires, l'exaspération des hiérarchies sociales troublaient l'équilibre national judéo-musulman et berbère.

Ce constat était accentué par l'implantation des écoles de l'alliance israélite universelle⁵⁶ dans plusieurs villes marocaines, notons que la langue et la civilisation française étaient au cœur de leur programme scolaire. De ce fait, dans l'espace d'une génération, les disparités culturelles entre juifs et musulmans deviennent très apparentes :

*« La francisation avait évidemment joué de façon différente suivant les classes chez les juifs et les musulmans. Cependant, les juifs y perdirent la maîtrise de l'arabe, qui avait été leur langue ».*⁵⁷

D'ailleurs, au recensement des populations, en 1960, 57 % de juifs étaient alphabétisés surtout en français, contre 13,5 % de musulmans. 82 % de jeunes juifs

⁵⁵ Ben-Ami ISSACHAR, **Culte des saints et pèlerinages judéo-musulmans au Maroc**, Paris, Maisonneuve et Larose, 1990, p. 207.

⁵⁶ Ces écoles avaient pour objectif de promouvoir la mission « hautement civilisatrice » de la France « bienfaisante » qui envisage la libération des sociétés du tiers-monde de leur condition « arriérée », et ce, en leur offrant une langue et une culture « moderne et supérieure ».

⁵⁷ Keneth BROWN, **People of Sale. Tradition and change in a Moroccan, City 1830-1930**, Manchester, M.U.P., 1976, XVI-265, p. 194.

âgés de 10 à 14 ans savaient lire et écrire le français, contre 30 % de musulmans dans la même catégorie d'âge. Notons que la francisation de l'élite juive lui a permis l'ascension vers les fonctions et carrières administratives, industrielles et du commerce international. De ce fait, à travers l'acculturation et la promotion de minorités locales, la France assurait non seulement son implantation économique et culturelle, mais troublait l'équilibre ancestral qui régnait entre juifs et musulmans.

El Maleh le souligne également dans ces récits. En effet, la projection qu'il fait dans l'histoire moderne du Maroc le ramène nécessairement au colonialisme par lequel le mécanisme du rapport communautaire juif et intercommunautaire judéo-arabe se déstabilise, notamment par le surgissement de l'europhisme, l'introduction de nouveaux critères de référence sociale, l'instauration du dahir berbère, l'éclatement des structures économiques traditionnelles..., mais le facteur le plus conséquent pour le devenir de la communauté juive fut la propagation des aspirations antisionistes.

Au terme de ce chapitre, consacré à l'exposition du cadre de travail, nous nous sommes intéressée, dans un premier volet, à la biographie de l'auteur où nous avons découvert qu'il était un militant nationaliste et un défenseur confirmé de la cause palestinienne. Nous avons signalé également qu'il a commencé sa carrière en tant que professeur de philosophie avant de mener une activité de journaliste et de critique d'art.

Dans le second volet, nous avons montré qu'El Maleh avait une vision singulière pour les arts plastiques en général et pour la peinture en particulier. En effet, l'attitude esthétique qu'il développait en faveur de la pratique picturale engageait une érudition impressionnante et une sensibilité esthétique singulière qui s'attachait à une vision contemporaine jumelée à la spiritualité mystique de l'art islamique et des concepts philosophiques orientaux.

Dans le dernier volet, nous avons commencé par présenter certaines caractéristiques de ses écrits littéraires. Ainsi, nous avons montré que son œuvre est hantée par la question principale de l'affirmation de son identité marocaine aux racines judéo-arabes et berbères. En plus, nous avons précisé que l'exode massif des juifs marocains était le catalyseur de ses écrits littéraires. De ce point de vue nous pouvons considérer que l'œuvre maléchienne constitue un éveil culturel à leur égard, éveil orienté plus particulièrement vers le thème des relations judéo-musulmanes. En effet, à travers les thèmes de la mémoire, de l'identité, de la nostalgie, des rapports aux musulmans et aux berbères, il tente de pénétrer au cœur de problèmes et de situations vécues, il retrace de l'intérieur de sa communauté, les effets d'une intégration sociétale profonde des juifs au sein d'une communauté majoritairement musulmane.

Ensuite nous avons présenté la littérature dans laquelle l'auteur s'inscrit, à savoir la littérature de l'exil. Nous nous sommes intéressée également à l'imaginaire judaïque et à l'histoire des juifs au Maroc. Nous avons alors signalé que juifs et musulmans entretenaient des relations « civiles » malgré ce qui les séparaient. D'ailleurs, les échanges commerciaux entraînaient un apport culturel manifeste à plusieurs niveaux (us et coutumes, traditions, valeurs, langue, music, etc), ceci découlait d'un partage de vie et d'une participation commune à des coutumes que les différences religieuses n'empêchaient pas complètement.

Finalement nous pouvons dire que ce chapitre nous a permis de comprendre la toile de fond qui se tisse derrière l'écriture d'El Maleh et d'assimiler, en quelque sorte, la vocation de l'œuvre maléhiennne, chose qui nous servira forcément dans la recherche d'une réponse à la problématique que nous avons formulé au début de ce travail.

CHAPITRE 2

De la statistique à la statistique lexicale

L'analyse que nous comptons réaliser s'appuie sur un courant aux dénominations changeantes : statistique lexicale, statistique linguistique, linguistique quantitative, etc. Il s'agit d'une science qui est née de la rencontre de plusieurs disciplines : l'étude des textes, la linguistique, l'analyse du discours, la statistique, l'informatique, le traitement des enquêtes, pour ne citer que les principales.

Cela étant dit, nous ferons un tour d'horizon sur la statistique lexicale pour approcher la notion. Un tel objectif nécessite, tout d'abord, d'établir la relation entre la linguistique et la statistique, où nous allons retracer les circonstances qui ont abouti à un tel métissage. Nous verrons, également, que l'alliage statistique/linguistique a été favorisé largement par le développement de l'outil informatique du moment qu'il a permis la conversion de données textuelles énormes en données chiffrées, représentées sous forme de schémas, de graphiques et/ou de tableaux.

Nous nous intéresserons par la suite à la statistique lexicale proprement dite, sans oublier d'aborder le dépouillement et la quantification du corpus ainsi que les principales lois qui seront utilisées lors de la réalisation de notre travail de recherche.

A. Analyse statistique en linguistique

La statistique linguistique est à la croisée de deux disciplines absolument distantes que ce soit dans leurs fondements que dans leurs vocations, nous sommes donc au cœur d'un carrefour scientifique vers lequel convergent tout naturellement les principes de la statistique et ceux de la linguistique. À ce noyau dur vient s'ajouter l'informatique qui se veut un outil destiné à parfaire l'analyse, la description, la comparaison et, par conséquent, l'interprétation des données étudiées.

Nous aborderons tout d'abord la notion de statistique et sa relation avec la linguistique, puis nous enchaînerons sur l'informatique et son apport indéniable pour le développement de la statistique linguistique.

1. Statistique et Linguistique

Pour établir la relation entre la linguistique et la statistique nous nous proposons, tout d'abord, de donner un aperçu sur la statistique. Nous allons, par la suite, reconstituer les circonstances qui ont abouti à l'introduction de lois mathématiques et statistiques dans la linguistique.

1.1. Statistique

Il s'agit d'une activité indispensable à la gestion des sociétés modernes. Elle joue un rôle de premier plan dans d'innombrables domaines, en sciences pures comme en sciences appliquées. Elle est définie comme une science qui vise le groupement méthodique des données qui se prêtent à une évaluation mathématique. Elle englobe un ensemble de lois et de techniques devant établir des synthèses générales. Ces dernières sont le fruit d'une démarche qui commence

par la recherche et la collecte des données, leur traitement, leur analyse et prend fin par leur interprétation.

Bien que le terme « statistique » n'ait été utilisé pour la première fois qu'en 1589 -par l'historien italien Girolamo Ghilini- l'activité de recueil de données demeure très ancienne et répond aux besoins d'organisation et de gouvernement des grands empires. D'ailleurs, la signification première du mot « state-istique» est l'ensemble des informations indispensables à l'État (dans la langue latine *Statisticum* signifie : ce qui a trait à l'État) ⁵⁸.

Au cours de l'histoire, la collecte d'observations et la méthodologie de leur traitement se sont développées de façon largement indépendantes. Au XIX^e siècle, grâce à l'alliance de la statistique et des sciences expérimentales, cette activité connaît une véritable révolution. En effet, des règles précises sur la collecte et le traitement des données sont promulguées et l'on voit apparaître, dans l'interprétation des données, des méthodes extrêmement puissantes.

Au XX^{ème} siècle la statistique acquiert le statut de discipline, vu la richesse et la diversité des méthodes qu'elle renferme. Ce statut a été renforcé par l'avènement de l'informatique, dans la mesure où elle a permis le traitement d'un nombre important de données.

Notons qu'il existe une certaine nuance entre les statistiques et la statistique. En effet, la première désigne un ensemble de données numériques concernant un ou plusieurs phénomènes ou caractères. Par contre, la statistique est un ensemble de méthodes scientifiques permettant de collecter, d'organiser, de

⁵⁸ Maurice LETHIELLEUX, *Statistique descriptive*, Paris, Dunod, 8^{ème} édition, 2016, p.52.

synthétiser et d'analyser les données ou les informations relatives à un ou plusieurs phénomènes.

Ainsi, l'objectif de la démarche statistique⁵⁹ est de tirer des conclusions et de prendre des décisions à partir de l'information recueillie. Cette démarche comporte deux niveaux, la statistique descriptive et la statistique mathématique. Nous allons désormais présenter ces différents niveaux.

a. **Statistique descriptive**

Fondée sur l'observation de données relatives à toutes sortes de phénomènes (économiques, financiers, géographiques, biologiques, littéraires etc.), la statistique descriptive⁶⁰ englobe un ensemble de méthodes qui ont pour but d'organiser et de synthétiser les informations collectées sous forme de tableaux, de graphiques et d'indicateurs numériques. Ces derniers résument les propriétés les plus importantes des données observées, telles que leur moyenne ou leur degré de variation, de sorte que l'on puisse, par exemple, identifier les caractéristiques d'un auteur ou d'un genre particulier. Les indicateurs numériques les plus importants sont les paramètres de position et les paramètres de dispersion:

Les paramètres de position, aussi appelés indicateurs de valeur centrale, ils servent à caractériser l'ordre de grandeur des données. Nous pouvons en citer, la médiane, le mode et la moyenne.

⁵⁹ Ibid., p.1

⁶⁰ Ibid.

- La médiane

La Médiane, notée M , d'une série statistique, est la valeur de la série qui partage la population en deux parties de même effectif. C'est à dire le nombre d'observations inférieures à M est égal au nombre d'observations supérieures à M .

Si le nombre d'observations n est impair, la médiane sera unique et elle sera égale à la valeur de la série ordonnée de rang $\left(\frac{n+1}{2}\right)$. Si le nombre d'observations n est pair, alors nous aurons un intervalle médian : $\left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$

Remarque : parfois nous avons $x_{\frac{n}{2}} = x_{\frac{n}{2}+1}$ dans ce cas la médiane est unique⁶¹.

Exemple : on considère le nombre de mot des phrases d'un corpus donné :

Série brute : 16 , 20, 12, 13, 17, 20, 11, 13, 14, 15, 15, 16, 20.

Série ordonnée : 11, 12, 13, 13, 14, 15, 15, 16, 16, 17, 20, 20, 20.

Dans notre cas le nombre d'observations n est égal à 13, donc la médiane sera la longueur de la phrase de la série ordonnée du rang $\left(\frac{n+1}{2}\right) = \left(\frac{13+1}{2}\right) = 7$

La médiane est égale alors à $x_7 = 15$, Cela veut dire que 50 % des phrases du corpus ont un nombre de mot inférieur à 15 et 50 % restant ont un nombre de mot supérieur à 15.

⁶¹ Michael OAKES, *Statistics for corpus linguistics edinburgh textbooks*, In: empirical linguistics, Cambridge, Edinburgh university press, 1998, p. 3

- Le mode

Le mode d'une série statistique est la valeur dont l'effectif (ou la fréquence) est plus grand que les effectifs (ou les fréquences) des valeurs voisines⁶², il est noté M_0 .

Par exemple, un corpus peut être constitué de phrases contenant les nombres de mots suivants : 16, 20, 15, 14, 12, 16, 13. Le mode est donc 16, car il constitue la longueur de phrase la plus répandue.

- La moyenne arithmétique

La moyenne arithmétique d'une distribution statistique, que l'on note $\bar{x}$, est une valeur qui résume toute la série statistique, elle est donnée par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k f_i x_i$$

En général, elle est calculée par le rapport du total de toutes les observations par rapport au nombre des observations composant la série statistique⁶³ :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Dans l'exemple de longueur de phrase, la somme de tous les mots dans toutes les phrases est 106, tandis que le nombre de phrases est 7. Ainsi, la longueur moyenne

⁶² Hatch EVELYN et Lazaraton. ANNE, *The research manual: Design and Statistics for Applied Linguistics*, Los Angeles, Heinle Publishers, 1991, p. 160

⁶³ Butler CHRISTOPHER, *Statistics*, In: Linguistics, Oxford, Blackwell Publishers, 1985, p. 27

de la phrase est de $106/7 = 15,1$ mots. Cela veut dire qu'en moyenne les phrases du corpus présentent une longueur de 15 mots.

Le choix d'une mesure de tendance centrale par rapport à une autre est particulièrement important lorsque l'on veut comparer les caractéristiques d'écriture de plusieurs auteurs ou la fluctuation changeante d'un même auteur au fil du temps.

Dans des situations, nous sommes face à des distributions statistiques ayant le même mode et la même médiane, nous nous trouvons donc dans l'incapacité de faire des conclusions. D'où la nécessité d'introduire des paramètres qui caractérisent la dispersion des valeurs d'une série statistique.

Les paramètres de dispersion : Ces paramètres⁶⁴, comme leur nom l'indique, mesurent la dispersion des données autour d'une valeur centrale. Nous pouvons en citer la variance, l'écart type et l'étendue.

- La variance

La variance⁶⁵ d'une distribution statistique, $X = (x_i; n_i) \ 1 \leq i \leq k$, que l'on note : $V(X)$ ou σ^2 , est donnée par :

$$V(X) = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2$$

⁶⁴ Alalouf SERGE, *Méthodes statistiques*, Québec, Edition Loze-Dion, 2014, p.13

⁶⁵ Elisabeth OLIVIER, *L'essentiel de la statistique descriptive*, Issy-les-Moulineaux, Gualino, 2008, p.62

- L'écart-type

L'écart-type⁶⁶ d'une distribution statistique $X = (x_i; n_i) \ 1 \leq i \leq k$, que l'on note σ_x est donnée par :

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2} = \sqrt{\sum_{i=1}^k f_i (x_i - \bar{x})^2}$$

L'écart type mesure le degré d'éparpillement des observations par rapport à la moyenne. Reprenons l'exemple précédent pour lequel nous calculerons la variance et l'écart type :

Nous considérons un corpus constitué de phrases contenant les nombres de mots suivants : 16, 20, 15, 14, 12, 16, 13.

$$V(X) = \frac{1}{7} [2 (16 - 15,1)^2 + (20 - 15,1)^2 + (15 - 15,1)^2 + (14 - 15,1)^2 + (12 - 15,1)^2 + (13 - 15,1)^2]$$

$$V(X) = 6,80952381$$

Et par conséquent $\sigma_x =$

$$\sqrt{\frac{1}{7} [2 (16 - 15,1)^2 + (20 - 15,1)^2 + (15 - 15,1)^2 + (14 - 15,1)^2 + (12 - 15,1)^2 + (13 - 15,1)^2]}$$

$$\sigma_x = \sqrt{6,80952381}$$

$$\sigma_x = 2,60950643$$

⁶⁶ Ibid., p. 64

L'écart type n'est pas faible, ce qui montre l'ampleur de la dispersion des longueurs des phrases par rapport à la longueur moyenne des phrases.

- L'étendue

L'étendue⁶⁷ d'une série statistique, que l'on note e , est donné par :

$$e = x_{\max} - x_{\min}$$

Où $x_{\max}$ et $x_{\min}$ sont respectivement la plus grande et la plus petite valeur de la série.

Nous reprenons le même exemple précédent pour lequel nous allons calculer l'étendue :

$$e = x_{\max} - x_{\min} = 20 - 12 = 8$$

Cela signifie que l'écart entre la longueur maximale des phrases et celle minimale est de 8 mots.

b. Statistique théorique ou mathématique

On oppose généralement la statistique descriptive, dont l'objectif est de décrire une réalité statistique, à la statistique mathématique dont l'objectif est de modéliser cette réalité et d'apporter des outils d'aide à la décision.

⁶⁷ Mathé ARMELLE, *L'essentiel de la statistique descriptive*, Issy-les-Moulineaux, Gualino, 2016, p.43

Contrairement à la théorie des probabilités qui étudie des variables aléatoires de loi connue, la statistique mathématique étudie des variables aléatoires de loi inconnue à partir d'observations du phénomène aléatoire⁶⁸.

La statistique mathématique est fondée sur deux piliers principaux, à savoir : la théorie de l'estimation et la théorie des tests ou inférence⁶⁹.

L'objet de la théorie de l'estimation est de savoir comment passer d'une information portant sur un échantillon à une information portant sur la population ? Pour ce faire, le statisticien⁷⁰ cherche à révéler la valeur d'un ou de plusieurs paramètres, associés à la distribution de la caractéristique d'intérêt dans la population. Il construit pour cela un estimateur qui se définit comme une fonction des variables de l'échantillon.

La démarche du statisticien est alors la suivante : il commence par étudier les propriétés de l'estimateur. Cela revient à analyser certaines caractéristiques de sa distribution : son espérance, sa variance, etc. L'idée générale est de vérifier théoriquement si les réalisations de cette variable aléatoire ont de grandes chances d'être « proches » de la vraie valeur du paramètre que l'on souhaite estimer. Nous pouvons aussi comparer différents estimateurs afin de choisir le plus performant : pour cela, le statisticien introduit les notions d'estimateur optimal et d'estimateur efficace. Une fois que l'on dispose d'un « bon » estimateur, on l'utilise pour obtenir une estimation.

⁶⁸ Valérie GIRARDIN & Nikolaos LIMNIOS, *Probabilités en vue des applications : introduction aux processus et à la statistique*, Paris, Vuibert, 2008, p. 123

⁶⁹ Christophe HURLIN et al, *Statistique et probabilités en économie-gestion*, Paris, Dunod, 2015, p. 252

⁷⁰ Ibid., p. 256

Il est aussi possible de fournir un intervalle de confiance, c'est-à-dire un encadrement sur la valeur du paramètre que l'on souhaite estimer. Cet encadrement permet de rendre compte de l'incertitude autour de la prévision ponctuelle.

Cette démarche de l'estimation se situe au cœur de très nombreux domaines : sondages politiques, enquêtes d'opinion, enquêtes économiques, méthodes de scoring, analyses marketing quantitatives, modèles de prévision, etc.

Soulignons que la méthode d'estimation la plus utilisée est la méthode du maximum de vraisemblance⁷¹ qui peut être appliquée à l'estimation de paramètres de modèles linéaires ou non linéaires. Une des raisons de ce succès est que, sous des hypothèses relativement générales dites hypothèses de régularité, l'estimateur du maximum de vraisemblance présente de très bonnes propriétés. Notons que le principe de cet estimateur consiste à déterminer la valeur des paramètres qui rend l'échantillon observé le plus vraisemblable.

La théorie des tests (ou inférence) étudie la construction et les propriétés d'un test statistique. Ce dernier fait appel à une règle de décision permettant de rejeter ou de ne pas rejeter une hypothèse, appelée hypothèse nulle, en fonction des observations d'un échantillon. La théorie des tests⁷² est la théorie fondamentale de ce que l'on appelle aujourd'hui la statistique décisionnelle ou business intelligence. Elle constitue le fondement de tous les outils statistiques modernes d'aide à la décision. Au-delà de la règle de décision, le principal

⁷¹ Ibid., p. 292

⁷² Ibid., p. 328

avantage d'un test statistique est qu'il permet de minimiser ou de contrôler les risques associés à la décision. C'est pourquoi ces tests sont toujours associés à trois éléments :

- Une hypothèse nulle et une hypothèse alternative : Une hypothèse est une assertion concernant la population. Un test, en tant que règle de décision, se réfère toujours à une hypothèse de référence dite hypothèse nulle, permettant la connaissance de la loi de la variable aléatoire, contre une hypothèse alternative. Ces hypothèses sont respectivement notées H_0 et H_1 .
- Une région critique : c'est un ensemble délimité par une ou plusieurs valeurs critiques, suivant les cas. Elle correspond à la règle de décision du test statistique. Cette règle est extrêmement simple : si la réalisation de la statistique de test, obtenue à partir des observations appartient à la région critique, on rejette l'hypothèse nulle H_0 .
- Des risques de première espèce et de seconde espèce : le risque I ou risque de première espèce, correspond au risque de rejeter l'hypothèse nulle H_0 alors qu'elle est effectivement vraie dans la population. Tandis que le risque II ou risque de seconde espèce correspond au risque de ne pas rejeter l'hypothèse nulle H_0 alors que l'hypothèse alternative H_1 est valide. Pour une règle de décision donnée, on cherche à quantifier les probabilités associées à ces deux types de risque. Par convention, la probabilité associée au risque de première espèce est notée α . Dans la pratique, le niveau α est fixé par le statisticien à un seuil relativement faible, typiquement $\alpha = 5\%$ ou $\alpha = 10\%$.

On distingue, généralement, deux types principaux de tests statistiques :

Les tests paramétriques : ce sont des tests statistiques⁷³ qui imposent des conditions d'application relatives aux paramètres de la population. C'est aussi un processus qui utilise des échantillons statistiques pour tester une affirmation sur un paramètre de la population. Des chercheurs, dans des domaines allant de la médecine en passant par l'économie jusqu'à la politique, comptent sur les tests d'hypothèse pour prendre des décisions éclairées sur de nouveaux médicaments, des marchés et des issues d'élections.

Les tests non paramétriques : il s'agit de tests d'hypothèse pour lesquels il n'est pas nécessaire de spécifier la forme de la distribution de la population étudiée. Les méthodes non paramétriques requièrent peu d'hypothèses, elles ignorent notamment l'hypothèse classique de la normalité de la population. Ces tests peuvent être appliqués à de petits échantillons, à des caractères qualitatifs ou à des paramètres de la population (moyenne, variance, écart type, proportion...). nous pouvons les utiliser également dans le cas des données incomplètes ou imprécises.

Parmi les tests non paramétriques⁷⁴ les plus célèbres nous pouvons citer le test de Khi-deux (il étudie l'existence d'une liaison entre deux variables quantitatives) et le test de coefficient de corrélation des rangs de Spearman (si l'une des conditions du test de coefficient de corrélation de Pearson n'est pas vérifiée, on utilise alors un test non paramétrique : le coefficient de corrélation des rangs de Spearman, qui est basé sur le rang des variables et non sur leurs valeurs).

⁷³ Hassene SIBY, *Introduction à la statistique et aux probabilités*, Québec, Edition Loze-Dion, 2017, p.220

⁷⁴ Ibid., p. 258

Un des objectifs des tests statistiques est d'étudier le lien ou la dépendance entre deux variables statistiques, par exemple : vérifier la concordance entre une distribution expérimentale et une distribution théorique, tester l'indépendance entre deux variables qualitatives, etc.

Pour éviter la redondance des exemples, nous soulignons que nous avons mis en application la majorité de ce que nous avons évoqué jusqu'à l'instant au niveau de la deuxième partie de cette thèse.

1.2. Statistique linguistique

Récemment, la statistique fut adoptée dans les sciences humaines et plus précisément en linguistique. Les commencements de cette rencontre datent de 1935 avec la publication du grand ouvrage *The Psycho-Biology of Language, an Introduction to Dynamic Philology* de George Kingsley Zipf, qui a joué le rôle de pionnier dans le domaine de la statistique linguistique. Le titre de l'ouvrage indique que l'auteur ne voulait pas se conformer à la manière classique pour aborder les problèmes linguistiques. En effet, il les a étudiés «de la même façon qu'un physiologiste étudie le rythme cardiaque»⁷⁵. Cette idée devait l'amener à introduire dans l'étude du langage les principes de la statistique. Alors en 1949 dans *Human Behavior and the Principle of Least Effort*, son œuvre maîtresse, moyennant des faits statistiques, Zipf a pu démontrer que « la longueur d'un mot est très étroitement liée à la fréquence de son emploi. Plus grande est cette dernière, plus bref est le mot. Par ailleurs, plus un élément du langage est complexe du point de vue phonétique, il apparaît moins fréquemment dans le

⁷⁵ Philippe BULLY Zipf, *Créateur de la linguistique statistique*. In : Communication et langages. N°2, 1969, p. 23.

discours »⁷⁶. Alors zipf, En attirant l'attention sur le caractère stochastique des processus qui caractérisent les sciences humaines, finira par élaborer une loi permettant le traitement d'un texte littéraire, en multipliant les statistiques, les courbes et les graphiques. C'est ainsi que le linguiste américain a édifié un nouvel instrument dont les recherches actuelles montrent encore l'indiscutable fécondité.

En 1956, G. Herdan a fourni aux linguistes de nombreux éclaircissements sur les fondements et les méthodes de cette nouvelle étude quantitative des textes qu'est la statistique linguistique.

Quoique, les deux disciplines en question sont diamétralement distantes, cependant ils ont un point commun, à savoir l'observation des tendances ; et d'ailleurs il existe quelques linguistes qui considèrent le langage comme étant un phénomène statistique régit par des lois numériques et sujet à des interprétations d'ordre quantitatif. Ce rapprochement a donné lieu à une harmonisation entre ces deux disciplines. C'est ce à quoi Ali Hefied fait allusion lorsqu'il dit : « *la statistique formule des hypothèses dont la linguistique établit le contenu et dégage les causes thématique ou stylistique...elle propose des modèles mathématiques des faits linguistiques, compatibles avec une théorie linguistique et fondés sur l'expérience des textes; ensuite elle cherche à établir les écarts entre ces modèles et la réalité des textes, et détermine si ces écarts ont une explication stylistique cohérente* »⁷⁷.

⁷⁶ Ibid., p. 24.

⁷⁷ Ali HEFIED, *Statistique linguistique : Aspects stylo-statistique du vocabulaire dans quinze « Voyages Extraordinaires » de Jules Verne*, Thèse de doctorat d'Etat, Université Sidi Mohamed Ben Abdallah, Faculté des lettres et des sciences humaines, Dhar El Meraz , Fès, 1999, Vol 1, p. 61.

2. Linguistique et informatique

L'histoire des ordinateurs numériques a commencé il y a plusieurs années et continue à s'améliorer de jours en jours. La percée technologique dans le matériel et les logiciels s'est traduite par un système informatique de plus en plus performant, ce qui a contribué largement au développement de plusieurs domaines, notamment ceux qui se rapportent aux sciences exactes. En effet, le traitement des données est devenu plus sophistiqué, plus rapide et moins volumineux.

Les spécialistes de langue et de littérature se trouvaient séduits, pareillement à leurs collègues dans les sciences exactes, par les privilèges des systèmes informatiques qui, idéalement, s'occupent de la besogne lourde pendant que le chercheur se réserve un travail plus satisfaisant. Notons également que certains linguistes ont déjà utilisé l'ordinateur dans des travaux de linguistique, de traduction automatique et de production automatique de bibliographies ou d'éditions critiques.

Le maniement de nombreuses petites unités est vite fait par machine; Baudelaire, par exemple, qui détestait intensément la société mécanisée naissante, cite avec approbation une observation de Ste-Beuve: « *Pour deviner l'âme d'un poète, ou du moins sa principale préoccupation, cherchons dans ses œuvres quel est le mot ou quels sont les mots qui s'y représentent avec le plus de fréquence. Le mot traduira l'obsession* »⁷⁸.

Suivant le conseil de Baudelaire et cherchant la clé des obsessions d'un auteur en observant la concentration des mots, le chercheur peut recourir à l'index

⁷⁸ Charles BAUDELAIRE, *l'Art Poétique*, Paris, Conard, 1925, p.352.

de mots qui se veut un outil conséquent pour réaliser ce type de recherche. Quoique le format varie quelque peu, un index de mots⁷⁹ est une liste par ordre alphabétique de tous les mots dans une œuvre; la liste est accompagnée d'une série de chiffres indiquant chaque lieu où le mot apparaît dans le texte. Alors pour former un index quelconque, le chercheur peut choisir entre deux procédés : il peut s'astreindre à un travail fastidieux, dressant lui-même les listes de mots et de fréquences, ou bien il peut confier cette tâche purement mécanique à une machine.

Dans le premier cas, celui de l'index artisanal, le résultat sera moins fiable, et ce, pour des raisons très faciles à deviner : écrire à la main littéralement des milliers de fiches, les trier par ordre alphabétique et finalement dresser une liste des résultats reste un travail très long et peu intéressant ; celui qui le fait est donc extrêmement sujet à la distraction, d'où viennent les erreurs. Dans le deuxième cas, celui de l'index mécanique ou informatisé, le processus du travail sera plus rapide, moins ennuyeux et le résultat sera plus fiable. Donc, le deuxième choix semble préférable.

Louis Delatte, le directeur du Laboratoire d'Analyse Statistique des Langues Anciennes à Liège, définit précisément le rôle de l'ordinateur dans l'étude de la langue :

« Toutes les recherches, tous les calculs, dont je viens de parler, pourraient être faits manuellement, artisanalement, sans recourir aux machines. L'ordinateur n'ajoute rien qualitativement à l'homme : il lui donne seulement une nouvelle dimension d'ordre quantitatif et si l'ordinateur fournit les éléments d'une réponse, il reste tout de même à l'homme l'interprétation et la décision

⁷⁹ Fortier PAUL, *Etat présent de l'utilisation des ordinateurs pour l'étude de la littérature française*, Computers and the Humanities, Vol.5, N° 3, January 1971

finales. »⁸⁰

Donc l'ordinateur apporte une aide cruciale au chercheur mais ne supprime point l'apport de l'homme dans son travail de recherche littéraire ou linguistique.

L'avènement de l'informatique a offert de nouvelles perspectives à une linguistique qui s'appuyait sur des méthodes et des objectifs plus classiques. Elle a permis au chercheur d'utiliser des méthodes et des outils qui permettent l'exploitation et la comparaison de grandes quantités de données textuelles sur support électronique.

Moyennant le logiciel Hyperbase, l'informatique nous sera d'une grande utilité. En effet, ledit logiciel nous permettra, d'une part, de préparer la base de données. D'autre part, en se basant sur des lois statistiques, il nous donnera la possibilité d'interroger les textes, d'obtenir des graphiques et de calculer des pourcentages qui nous aideront à mieux étudier notre corpus et à bien interpréter les résultats.

B. Statistique lexicale

Les origines de la statistique lexicale remontent aux années 50. Le Centre d'Étude du Vocabulaire Français de Besançon avait entrepris le dépouillement des œuvres de Corneille sur support informatique. Cette véritable mine mécanographique encouragea, Charles Muller à en faire une analyse

⁸⁰ « L'Ordinateur et la création littéraire : Communication faite au colloque International d'Esthétique et des Sciences connexes, Liège, octobre 1966, » Revue de l'organisation Internationale pour l'étude des langues anciennes par ordinateur, 1967, pp. 19-20

lexicométrique. Il s'agit, en fait, d'une étude quantitative qui ne peut être faite qu'à travers le dénombrement de toutes les unités élémentaires d'un corpus quelconque.

Depuis Ch. Muller, la statistique lexicale⁸¹ a connu, une importante expansion, surtout avec les travaux, les développements, les réaménagements, les critiques et les refontes successives d'Etienne Brunet, André Salem, Maurice Tournier, Pierre Lafon, Charles Bernet, Dominique Labbé, Benoît Habert et de beaucoup d'autres ; et elles ne cessent de se développer.

Nous signalons que la stylistique des œuvres littéraires constitue le champ d'investigation le plus réussi quant à ce nouveau genre d'analyse qu'est la statistique lexicale, notamment pour des recherches de paternité d'œuvres. Charles Muller dit à ce propos : « La statistique lexicale, dans la majorité de ses applications, est orientée vers l'analyse stylistique des œuvres littéraires ; elle recherche les différences plutôt que les faits constants, la variété plutôt que l'homogénéité. »⁸²

En effet, elle nous acquiert une précision idéale permettant d'apprécier les variations stylistiques chez un même écrivain. De même l'examen global du vocabulaire nous permet d'aborder l'individualisation lexicale des personnages que l'auteur fait parler. En plus, l'analyse statistique des vocables nous renseigne sur la distance qui sépare deux œuvres d'un même auteur, ou même celle de deux auteurs différents.

⁸¹ Zoubeïr MOUELHI, *Essai de lexicométrie d'une œuvre arabe : le vocabulaire de Tawhîdî dans al- 'Imtâ' Y wa-l-Mu'ânasa*, Thèse de Doctorat, Université Lumière-Lyon 2, 2008, Vol 1, p. 36.

⁸² Charles MULLER, « *Statistique lexicale et théorie du lexique* », In : *Langue Française et linguistique quantitative*, Genève, Slatkine, 1979, p. 449.

Nous citons, à titre d'exemple, l'étude du professeur Frautschi à l'Université de La Caroline du Nord qui s'intéressait à l'attribution des articles anonymes de l'Encyclopédie⁸³. Les historiens de la littérature croyaient d'abord que Diderot avait écrit un grand nombre de ces articles. Pour vérifier ce constat Frautschi et ses assistants ont fait recenser par ordinateur des articles signés par Diderot aussi bien que des contributions de ses collaborateurs : D'Alembert, Jaucourt, Rousseau. Ils ont trouvé que des groupes de mots ayant une même fonction grammaticale apparaissent dans les articles avec une fréquence assez bien délimitée. Par exemple, les prépositions "à, dans, en, pour, sans" constituent une proportion élevée du langage de Diderot, comparé au langage de ses collaborateurs ; les expressions temporelles "jamais, lors, toujours" se trouvent moins souvent chez Diderot que chez ses collaborateurs.

À partir des données ainsi développées, Frautschi a pu inventer dix-huit épreuves susceptibles d'évaluer la similarité entre le langage d'un article anonyme et le langage de Diderot. Au cours de l'été 1969, il a mis à l'épreuve vingt-six articles anonymes communément attribués à Diderot. Un seul s'est montré conforme aux structures de la langue utilisée par cet auteur. Ces résultats suggèrent⁸⁴ que, dans le cas des articles étudiés pour le moins, Diderot avait moins écrit pour l'Encyclopédie que l'on n'a cru.

Frautschi et ses assistants ont braqué la lumière sur un vrai problème et ils ont produit des résultats étonnants. Leur travail peut servir comme modèle de l'utilisation heureuse de l'ordinateur et de la statistique lexicale pour l'étude de la

⁸³ R. L. Frautschi, « A Project for Author Discrimination, In : the Encyclopédia », SAMLA Bulletin 32, Novembre 1967, 14- 17, communication par Frantschi, « Diderot in Computerland », 23 rd Kentucky Conference.

⁸⁴ Frautschi lui-même dit "suggèrent" non pas "prouvent"

littérature française.

Une autre application de l'approche lexicométrique⁸⁵ est de repérer des corrélations/dispersions entre phénomènes au niveau de la structure lexicale pour un auteur, un genre ou une époque donnée. Ce qui permettra par exemple, de résoudre le problème de datation ou d'ordre chronologique de tel ou tel texte d'origine mal établie ou de période d'écriture mal définie.

Outre la datation des textes, la lexicométrie peut déceler les premières apparitions des unités lexicales, leurs actualisations dans le discours ou le passage des unités de ce dernier dans le lexique de la langue ; ce qui fournirait des informations inestimables pour la confection, par exemple, du dictionnaire historique d'une langue quelconque.

Notons que les recherches en statistiques lexicales ont porté aussi sur la richesse, la spécificité, l'accroissement du vocabulaire, etc. La démarche statistique adoptée dans ces cas consiste à faire une comparaison entre les données obtenues et les données calculées à partir d'un modèle théorique, une référence externe comme celle que représente le Trésor de la Langue Française. Tacitement, le texte analysé est perçu comme un échantillon représentatif de la langue par l'analyse duquel nous pourrions inférer des informations générales.

Ajoutons que la statistique lexicale renferme toute une série de « *méthodes qui permettent d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire d'un corpus de textes* »⁸⁶.

⁸⁵ Zoubeïr MOUELHI, *op.cit.*, 2008, p. 6.

⁸⁶ André SALEM, *Pratique des segments répétés. Essai de statistique textuelle*, Paris, INafl, coll. "SAINT-CLOUD", 1987, p. 315.

Autrement dit, elle regroupe des procédures et des lois permettant d'approcher un texte donné et d'étudier l'organisation de son vocabulaire d'un point de vue quantitatif. Nous parlerons ainsi plutôt de forme lexicale (ou graphique) que de mot qui fait davantage référence au sens. Ayant le texte au centre de son objet d'étude, la lexicométrie partage de ce fait, des zones plus ou moins vastes avec d'autres disciplines, à savoir : la linguistique, l'analyse de discours, la recherche documentaire, l'analyse de contenu, l'informatique, l'intelligence artificielle, la statistique, etc.

Avant d'aborder toute analyse lexicométrique, le chercheur doit impérativement commencer par le dépouillement du corpus. Donc, à quoi consiste cette opération et quelles en sont les limites ?

1. Dépouillement et quantification du corpus

Du moment que l'étendue du texte, objet de l'analyse, est souvent importante le linguiste recourt à l'outil informatique pour effectuer le dépouillement du texte ; mais pour que cette démarche soit méthodique et fructueuse, il est nécessaire que les opérations de dépouillement⁸⁷ préalables soient réalisées selon des règles claires et stables assorties d'une réflexion minutieuse autour des notions de segmentation, de lemmatisation, de désambiguïsation, etc. Et ce n'est qu'après ces étapes de dépouillement et de quantification que les décomptes obtenus seront soumis aux lois statistiques et à l'interprétation du chercheur pour pouvoir juger, finalement, des variations des différentes unités linguistiques du corpus dont on veut étudier la structure.

⁸⁷ Zoubeïr MOUELHI, *op.cit.*, 2008, p. 5.

Donc, l'aide de l'outil informatique dans les opérations de dépouillement et de quantification ne supprime point le labeur du linguiste qui doit établir la norme lexicologique des unités élémentaires, matière première de tout raisonnement statistique. Loin des différentes connotations que peut avoir, surtout en linguistique, le terme norme, la notion de « norme lexicologique » ou « norme de dépouillement⁸⁸ » doit être comprise comme un besoin de standardisation ou de normalisation provisoire des textes contenus dans un corpus. Cette standardisation est destinée avant tout à les rendre comparables, à les stabiliser le temps d'une étude donnée.

Mais l'établissement d'une telle norme lexicologique ne va pas sans susciter des difficultés de nature différente. L'une de ces difficultés est de satisfaire aux exigences parfois contradictoires de la linguistique et de la statistique pour lesquelles :

« La norme devrait être acceptable à la fois pour le linguiste, pour ses auxiliaires, et pour le statisticien. Mais leurs exigences sont souvent contradictoires. L'analyse linguistique aboutit à des classements nuancés, qui comportent toujours des zones d'indétermination ; la matière sur laquelle elle opère est éminemment continue, et il est rare qu'on puisse y tracer des limites nettes [...]. La statistique, dans toutes ses applications, ne va pas sans une certaine simplification des catégories ; elle ne pourra entrer en action que quand le continu du langage a été rendu discontinu »⁸⁹.

Autrement dit, toute opération statistique a pour objet d'étudier un certain

⁸⁸ B. HABERT, A. NAZARENKO et A. SALEM, *Les linguistiques de corpus*, Paris, Armand Colin, 1997, p. 187

⁸⁹ Muller CHARLES, *Initiation aux méthodes de la statistique linguistique*, Paris, Honoré Champion, 1992, p. 113

nombre de caractères bien définis, relatifs à une population donnée, à travers un échantillon plus ou moins représentatif. Or, l'analyse linguistique quant à elle a pour objet d'étudier des mots dont il est difficile de tracer des limites nettes. La population cible peut être constituée soit par une liste de mot soit par une suite continue de mots qui forment un texte. Ajoutons à cela, la notion de vocable qui constitue une autre unité élémentaire nécessaire pour étudier n'importe quel corpus d'un point de vue quantitatif.

Donc, pour réaliser une étude en statistique lexicale il faut que l'objet d'analyse soit assujetti aux exigences des deux disciplines. Il faut que les limites des unités lexicales étudiées soient bien définies. Dans cette perspective, une question importante s'impose : Quels sont, dans chaque étude lexicométrique, les choix arrêtés et les décisions prises au moment du dépouillement lexical du corpus à étudier ?

Pour répondre à cette interrogation, il faudra définir préalablement une norme lexicologique afin de lever toute ambiguïté quant à la délimitation du mot et du vocable.

1.1. Norme lexicologique

Dans le domaine de la statistique lexicale, les langues sont présentées généralement quant à leur système d'écriture. En effet, nous distinguons deux familles différentes : les langues « avec séparateurs » et les langues « sans séparateurs ». Les premières sont celles qui disposent de systèmes d'écritures segmentées, où les mots sont nettement délimités par des séparateurs (espace, signes de ponctuation, caractères spéciaux, ...). Le français et l'anglais sont les représentants parfaits de cette famille de langue. Les secondes sont celles qui

présentent des systèmes d'écritures « non segmentées » où les limites des mots ne sont pas claires. Le chinois et le japonais sont les exemples types de cette famille.

La statistique lexicale, comme nous l'avons signalé, traite de la partie la plus vulnérable de la langue, à savoir le lexique et dont le segment élémentaire est le « mot ».

Quoique le français appartienne aux langues « avec séparateurs », le dépouillement du texte demeure difficile. Du moment que le « mot », qui semble être « *assez clair pour le sens commun, est scientifiquement suspect* »⁹⁰. En effet, la polémique qui s'est instaurée entre les linguistes voulant définir cette entité n'avait pas trouvé une définition assurée du mot⁹¹ comme unité du texte et ne pouvait pas établir une norme conventionnelle prévoyant tous les cas douteux, Charles Muller dit à ce propos :

*« Que l'on songe seulement aux « expressions », aux lexies plus ou moins figées par l'usage, et qui se présentent au locuteur comme une seule unité : clin d'œil est lexicalisé au point que clin n'a plus aucune existence propre ; on sera tenté d'y voir une seule unité; fera-t-on de même pour coup d'œil, coup d'État, coup d'essai, coup d'épée? Où mettre la limite entre celles des combinaisons qui seront admises pour une unité et celles qui en vaudront trois ? Et on se posera des questions semblables pour quantité de locutions grammaticales : au lieu de, au-devant de, afin que, etc. Sur le plan syntaxique, faudra-t-il traiter il a été vu comme une unité, ou deux (il /a été vu), ou trois (il /a été/vu), ou quatre ? Et fera-t-on une différence entre au (combat) et à la (guerre), ou bien analysera-t-on au et les autres formes contractées ? »*⁹².

⁹⁰ Muller CHARLES, *Langue française. Débats et bilans. Recueil d'articles*, Paris, Champion, 1993, p. 3.

⁹¹ Muller admet, en première approximation, que le mot est une unité graphique, séparée des unités voisines par un blanc ou une ponctuation.

⁹² Muller CHARLES, *La statistique lexicale*, In : *Langue française*. N°2, 1969, P. 31.

Ainsi, la délimitation du mot pose relativement quelques problèmes aux linguistes par son orthographe et sa typologie. En effet, Il est souvent question des cas où :

- L'unité graphique correspond à plusieurs mots, le cas des formes contractées : au, aux, du, des, etc.
- Les signes de ponctuation dont la présence à l'intérieur d'un groupe graphique peut créer un doute, il s'agit de l'apostrophe et du trait d'union. Dans ce sens Ch. MULLER⁹³ se demande s'il faut compter différemment loup et l'agneau ? en effet, le considérer comme deux mots c'est briser des unités comme aujourd'hui et presque. Le trait d'union⁹⁴, lui aussi, présente des ambiguïtés pas moins importantes que celles de l'apostrophe. Il s'agit notamment de la difficulté de distinguer entre le trait d'union de chauve-souris ou de porte-clé d'une part et celui de vient-il ou de moi-même d'autre part.
- Il peut y avoir, également, des unités graphiques qui forment un seul mot, soit pour des raisons morphosyntaxiques, soit pour des raisons lexicales. Au niveau syntaxique, « a été servi » peut-être considéré comme une seule occurrence du verbe servir. Au niveau lexical, le lexique nous propose des groupes qui devront être identifiés en unité de texte, malgré la séparation graphique. Citons à titre d'exemple, le cas des locutions verbales comme

⁹³ Muller CHARLES, *op.cit.*, 1979, p.129

⁹⁴ Muller CHARLES, *Initiation à la statistique linguistique*, Paris, Larousse, coll. "langue et langage", 1968, pp.147-151.

avoir peur ou avoir faim, le cas des composés nominaux comme pomme de terre ou chemin de fer.

Ayant survolé quelques problèmes liés à la délimitation du mot, passons à celle du vocable. Contrairement au mot, la définition du vocable ne fait pas de polémique. En effet, L'ensemble des lexèmes qui ont des occurrences dans un texte est dit vocabulaire de ce texte ; les éléments de cet ensemble seront dits vocables. Il y a en principe un certain degré de synonymie ou plutôt une isomorphie entre le lexème et le vocable, mais l'emploi de ces deux termes permet d'isoler un type particulier de vocable : ceux qui n'appartiennent pas au lexique, nous faisons allusion ici aux noms propres, aux mots étrangers introduits dans un texte, etc.

Autrement dit, tout mot est une forme d'un vocable et les vocables, dans leur majorité, sont identiques à des lexèmes du lexique ; ceux qui ne réalisent pas cette relation forment un sous-ensemble particulier du vocabulaire et fournissent une sorte de supplément au lexique.

Lors des opérations du dépouillement, il doit y avoir une cohérence parfaite entre le découpage syntagmatique du texte et le groupement paradigmatique des mots en vocables.

La délimitation des vocables consiste au regroupement des formes fléchies et la séparation des formes homographes, également cette étape pose d'énormes problèmes qu'il faut résoudre.

Pour ce qui est de la première opération, il s'agit de réunir toutes les formes flexionnelles du verbe sous l'infinitif, les différentes formes du substantif sont regroupées sous la rubrique du substantif masculin singulier. Les flexions d'un

adjectif sont réunies sous la forme de l'adjectif masculin singulier. Notons que cette opération ne peut être faite sans difficultés. Il existe, par exemple, un très grand nombre de formes verbales qui sont homographes à des formes nominales, etc.

Pour ce qui est de la seconde opération, il s'agit de séparer les formes homographes et les rattacher à un seul vocable. L'homonymie de deux lemmes peut se manifester de deux façons assez différentes : soit par une polyvalence sémantique (polysémie), soit par une polyvalence syntaxique (la dérivation impropre).

La polyvalence sémantique⁹⁵ concerne les lexèmes qui, sans changer de catégorie grammaticale, peuvent apparaître dans le discours avec des significations différentes ; suivant le cas, nous parlerons d'homonymie ou de polysémie. Les spécialistes de langue font la distinction entre l'homonymie (quand deux mots d'origines différentes s'écrivent et se prononcent de la même façon, par exemple : grève (bord de mer fait de sable) et grève (arrêt temporaire et collectif du travail)) ; et la polysémie au sens étroit (quand un même mot présente des sens différents mais apparentés, par exemple : se rejoindre au *café* (le lieu) et la plantation de *café* (la plante). Généralement, la distinction entre polysémie et homonymie est impossible dans une étude synchronique. Seule une étude de la langue en diachronie permet de résoudre le problème.

La polyvalence syntaxique est l'existence de deux ou plusieurs lemmes ayant la même forme graphique, mais distingués par leurs catégories grammaticales.

⁹⁵ Muller CHARLES, *Principes et méthodes de statistique lexicale*, Paris, Hachette, 1977, p. 21

« C'est le cas, en français moderne, pour la substantivation de l'infinitif ou l'adjectivation du participe présent, bien marquées par des critères morphologiques. Les difficultés commencent avec l'emploi adverbial des adjectifs (boire frais, parler fort). Elles sont gênantes pour les échanges entre substantifs et adjectifs (un ingrat, un fils ingrat, traiter d'ingrat- une robe rouge, le rouge du sang, se mettre du rouge-un domestique, le personnel domestique, etc.). La limite à partir de laquelle il faut séparer un couple masculin-féminin est difficile à tracer (vendeur, vendeuse-servant, servante-époux, épouse-maître, maîtresse-prince, princesse-mari, femme..., sans parler des noms d'animaux.) »⁹⁶

Muller⁹⁷ a signalé que le fonctionnement d'une unité dans deux ou plusieurs catégories grammaticales n'entraîne pas nécessairement la reconnaissance de lexèmes indépendants. Il s'agit souvent de faits syntaxiques plutôt que lexicaux et on devra accepter l'existence de lexèmes communs à deux catégories quand la répartition de leurs occurrences entre ces catégories laisse trop de place à l'arbitraire.

Généralement, l'analyse des lexèmes homographes peut conduire à réunir dans un même effectif des unités dont le rôle est très variable. Cependant, il faut bien admettre que le même vocable, suivant son contexte, traduit des réalités multiples. Muller⁹⁸ stipule à ce propos qu'il faut en être conscient et se dire qu'on opère sur des signifiants non sur des signifiés.

Pour faire face à ces problèmes, les linguistes ont établi un certain nombre de règles qui constituent ce que nous appelons la norme lexicale, il s'agit d' :

⁹⁶ Ibid., p.149

⁹⁷ Ibid., p. 26

⁹⁸ Ibid., p. 23

« un ensemble de règles qui, dans la lemmatisation d'un texte, décident de la délimitation des mots et des vocables, afin de soustraire le plus possible de cas douteux à l'appréciation momentanée et subjective de l'opérateur. »⁹⁹

Il faudra donc, pour des besoins de la statistique lexicale, adopter des conventions bien précises qui fixent les limites des unités étudiées.

Notons que la statistique lexicale n'est pas la première à découper un texte en mots ou bien à regrouper les mots en unités : index, concordances, glossaires, dictionnaire ; tous ces ouvrages sont fondés sur ce principe. Le lexicographe découpe sa nomenclature en articles et, par-là, crée des unités à la fois syntaxiques et sémantiques. C'est à lui de juger quand il faut grouper en une unité des emplois syntaxiques différents, ou de rassembler des variantes sémantiques allant jusqu'à l'homonymie. Muller¹⁰⁰ se demande, à ce propos, s'il est raisonnable d'établir à grand frais une norme lexicologique pour des travaux statistiques. Pour lui, le plus sage serait d'adopter celle des dictionnaires, ou celle d'un dictionnaire bien particulier. Les nombreuses décisions qui composent une norme ne se prennent pas en examinant dans leur ordre alphabétique tous les lexèmes de A à Z, mais en traitant les cas douteux à mesure qu'ils se présentent. Donc il serait facile de vérifier si ces cas litigieux ont été résolus par le dictionnaire et si sa décision peut être retenue. Il suffira par la suite de dresser une liste des désaccords et c'est cette liste qui tiendra lieu de norme lexicologique.

Ainsi pour élaborer une norme, il serait commode de le faire à partir d'un dictionnaire, mais en soulevant des désaccords plutôt que d'adopter aveuglément son découpage. Notons que le choix du dictionnaire de référence reste une tâche

⁹⁹ Ibid., pp. 13-14.

¹⁰⁰ Ibid., p. 29

délicate. Il est souhaitable qu'il soit contemporain, ou mieux postérieur aux textes qu'on se propose de quantifier et qu'il englobe l'époque de ces textes.

Ajoutons que Muller¹⁰¹ a pu vérifier expérimentalement qu'en appliquant à deux textes deux normes très différentes, on obtenait certes des données quantitatives très différentes, mais qu'en valeur relatives ces données restaient sensiblement égales. Ce qui importe aussi dans l'élaboration d'une norme lexicologique c'est sa simplicité, c'est-à-dire qu'elle sera beaucoup plus simple si elle est analytique pour le décompte des mots (N) et synthétique pour celui des vocables (V).

Finalement, il ne faut pas nier qu'une norme lexicologique ne satisfait personne à commencer par celui qui l'a élaborée ; mais l'important n'est pas tant dans la qualité scientifique des décisions prises, car presque toutes pourraient trouver leur justification, mais dans leur constance.

Pour ce qui est de notre étude, il n'est pas question de garder ou de rejeter la solution du dictionnaire, parce que le problème de choix d'une norme lexicologique est partiellement résolu grâce à Etienne brunet qui a mis, gratuitement, à notre disposition le logiciel *Hyperbase*. En effet, les dépouillements réalisés par ledit logiciel obéissaient déjà à une norme préétablie pour la délimitation du mot et du vocable. Il suffit donc d'exploiter les ressources mises à notre disposition en s'éloignant le plus possible de quelques erreurs de la machine.

¹⁰¹ Ibid., p. 27

Après avoir passé en revue les problèmes liés à la délimitation d'une norme lexicologique, nous présentons, dans ce qui va suivre, les mécanismes du découpage d'un texte : la lemmatisation.

1.2. Lemmatisation du corpus

L'indexation du texte se réalise suite à deux opérations distinctes : d'une part un découpage de la suite naturelle du texte en unités élémentaires, dont le comptage fournira la valeur N (nombre d'occurrence). D'autre part, le classement et le regroupement des unités considérées comme identiques sous un même lemme. Le nombre de ces unités différentes donnera la valeur V (effectif des vocables). Muller dit à ce propos :

*« Quantifier le vocabulaire d'un texte, c'est procéder à deux opérations distinctes, qu'elles soient successives ou simultanées : Le dénombrement des **mots** qui composent le texte et dont le nombre, représenté par N donnera une mesure de l'**étendue du texte**. Le dénombrement des **vocables** employés dans le texte, et dont le nombre, représenté par V, mesure l'**étendue du vocabulaire**. Il faut donc des règles précises pour la délimitation du mot unité de texte, et d'autres pour celle du vocable unité de lexique. »¹⁰²*

En effet, la première de ces opérations nécessite une définition du mot pris comme unité élémentaire de texte. La seconde exige une définition du vocable comme unité de lexique. Notons que la deuxième opération consiste à réunir sous une même rubrique les mots qui représentent un même vocable. C'est ce qu'on appelle communément la lemmatisation.

¹⁰² Muller CHARLES, *op.cit.*, 1968, p.142.

À ce sujet, Ali HEFIED stipule que :

« Le phénomène de la lemmatisation consiste à considérer les mots du texte [formes, occurrences] comme ensemble de départ, et à chercher leurs correspondants parmi les lemmes [vedettes, vocables] de l'ensemble d'arrivée ou lexique. [...] le lexème est donc noté par une seule forme graphique qui en compose le lemme, l'adresse ou le constituant vedette: l'infinitif pour le verbe, le singulier pour le substantif, le masculin singulier pour l'adjectif, etc. »¹⁰³

De ce fait, la lemmatisation du vocabulaire d'un texte donné consiste à :

1) le segmenter en un certain nombre d'unités dites « mots » ; dont la somme constitue l'étendue du texte, elle se représente généralement par N;

2) rassembler chacun de ces mots sous un vocable (lemme). Cette opération se traduit par :

- La séparation des unités identiques désignant des vocables distincts (par ex. ferme adj. et ferme, forme de verbe);
- Le regroupement des formes hétérographes d'un même vocable, en ramenant les formes verbales à l'infinitif, les substantifs au singulier, les adjectifs au masculin singulier et les formes élidées à la forme sans élision. Notons que les adverbes et les mots grammaticaux ne subissent aucun changement.

¹⁰³ Ali HEFIED, 1999, *op. cit.*, p. 68.

Au terme de ces deux opérations nous obtenons, au lieu d'un ensemble continu de caractère, un index plus ou moins détaillé. Si le dénombrement a été fait tout en ayant recours à l'outil informatique, ce dernier fournit un dictionnaire lemmatisé. Dans ce cas, le linguiste peut se référer au texte à l'aide des références qui accompagne chaque forme et examiner tous les contextes d'un même vocable pour pouvoir prendre une décision concernant les cas douteux. Ainsi, il s'agit d'une étape de prise de position où il faut trancher sur un certain nombre d'ambiguïtés. Toutefois, pour réaliser les deux opérations susmentionnées :

« Il faut des règles précises pour la délimitation du mot unité de texte, et d'autres pour celle du vocable unité de lexique. »¹⁰⁴

Donc la lemmatisation, ainsi présentée, se heurte encore à l'écueil d'une norme lexicologique apte à lever le maximum d'ambiguïtés concernant la délimitation du mot et du vocable.

Muller¹⁰⁵ a signalé que pour effectuer une opération statistique, rien n'interdit de prendre d'autres bases que le mot et le vocable, si l'on en trouve de moins suspectes, mais toute option de ce genre comporte ces difficultés spécifiques sur le plan pratique.

Pour ce qui est de la lemmatisation de notre corpus, nous avons eu recours au logiciel *Hyperbase* qui nous a présenté un dictionnaire lemmatisé distingué par les codes suivants :

1. Les verbes (infinitif, formes conjuguées ainsi que les participes) ont pour

¹⁰⁴ Muller CHARLES, *op.cit.*, 1968, p.142.

¹⁰⁵ Muller CHARLES, *op.cit.*, 1977, p.11

constituant vedette l'infinitif :

Chantes, chantait, chanterons ----- *chanter*

2. Les substantifs sont regroupés sous le singulier :

Chaises, chaise ----- *chaise*

3. Les adjectifs qualificatifs, qui englobent aussi les participes adjectivés, sont représentés par le masculin singulier :

Calmes ----- *calme*

4. Les adjectifs numéraux, cardinaux et ordinaux :

Quatrième, mille... ----- *quatrième, mille...*

5. Les adverbes :

Après, bien... ----- *après, bien...*

6. Les adverbes en « -ment » :

Évidemment, lentement... ----- *évidemment, lentement...*

7. Les articles et pronoms :

Le, un... ----- *le, un...*

8. Les conjonctions et les prépositions :

Afin, tandis, puisque... ----- *afin, tandis, puisque...*

9. Les adverbes :

Dedans, contre... ----- *dedans, contre...*

10. les noms propres :

Zina...----- Zina...

Après avoir abordé les notions de dépouillement, de lemmatisation et de norme lexicologique, nous passons maintenant à la présentation de la méthode d'analyse.

2. Présentation de la méthode d'analyse

L'approche sur laquelle se base notre étude est celle de la statistique lexicale, ladite méthode repose en grande partie sur des lois statistiques de probabilités. En effet, la méthode lexicométrique fait appel, le plus souvent, à la loi binomiale pour l'étude de la richesse lexicale et l'accroissement du vocabulaire. La loi normale y est aussi utilisée dans le calcul des spécificités, etc. Nous tâcherons, dans ce qui va suivre, de définir les lois qui vont être les plus utilisées lors de la réalisation de notre travail de recherche.

2.1. Probabilités

La théorie des probabilités a pour objet l'étude des phénomènes aléatoires à travers un traitement mathématique de la notion intuitive de hasard. Ce phénomène aléatoire peut être une expérience de la vie quotidienne, une expérimentation scientifique ou autre. Les exemples de telles situations sont innombrables ; nous pouvons penser aux jeux de hasard, au sexe d'un enfant à naître, au point d'impact d'un projectile, à la répartition des catégories grammaticales dans un texte ou un ensemble de textes, etc.

Le résultat de l'expérience aléatoire se distingue par deux caractéristiques¹⁰⁶ :

¹⁰⁶ Daniel REVUZ, *Probabilités*, Paris, Hermann, Collection Méthodes, 1997, p. 1

Premièrement, c'est un élément d'un ensemble bien déterminé de résultats possibles connus à l'avance. Deuxièmement, on ne sait pas, à priori, avec certitude quel est le résultat que l'on va obtenir.

Nous présenterons, par la suite, un bref aperçu historique sur l'évolution des probabilités, puis nous enchaînerons sur quelques méthodes de calcul.

a. Notes historiques

La notion d'aléatoire et le concept intuitif de probabilité¹⁰⁷ remontent à l'antiquité, mais ce n'est qu'au XVI^{ème} et XVII^{ème} siècle que des règles élémentaires de calcul des probabilités se sont développées. En effet, la fameuse correspondance entre les mathématiciens français Pascal et Fermat en 1654, relative à un problème de jeu de hasard proposé par un noble de l'époque, le Chevalier de Méré, est considérée comme le point de départ de « calcul » des probabilités.

Parmi les grands savants qui ont abordé, par la suite, des problèmes de probabilités nous pouvons mentionner Jacob Bernoulli avec son œuvre *Ars Conjectandi* (1713) ainsi que d'autres membres de cette famille unique de mathématiciens suisses¹⁰⁸. Nous citons également de Moivre avec son œuvre *The doctrine of chances* (1718), Laplace, Euler, Gauss, Lagrange et Legendre, etc.

¹⁰⁷ Eva CANTONI, Philippe HUBER, Elvezio RONCHETTI, *Maitriser l'aléatoire*, Paris, Springer, Collection statistique et probabilités appliquées, 2006, p. 2

¹⁰⁸ Nous faisons allusions à son frère Jean Bernoulli et ses deux neveux Daniel Bernoulli et Nicolas Bernoulli qui sont tous des mathématiciens et physiciens.

Jusqu'au début du XX^{ème} siècle le domaine des probabilités¹⁰⁹ resta un champ relevant des mathématiques et constitué d'un ensemble de résultats intéressants et utiles mais sans aucune base axiomatique¹¹⁰. En 1900 Hilbert énonça son fameux programme qui contenait comme sixième problème le développement d'une structure axiomatique pour les probabilités. En 1933 le mathématicien russe A.N. Kolmogorov releva le défi en publiant un article présentant les fameux axiomes de calcul des probabilités qui devenaient alors un domaine des mathématiques à part entière comme la géométrie, l'algèbre ou encore l'analyse.

Notons que les probabilités ont pu s'infiltrer dans plusieurs sciences. Elles ont prouvé leur utilité non pas uniquement en statistique lexicale mais aussi dans de nombreux domaines relevant des sciences du langage¹¹¹. Nous citons à titre d'exemple les grammaires probabilistes (aussi appelées grammaires d'arbres adjoint stochastiques) qui sont utilisées pour décrire la nature probabiliste d'un grand nombre de phénomènes linguistiques, tels que l'acceptabilité phonologique, les alternances morphologiques, l'interprétation sémantique, la désambiguïsation de phrases et variation sociolinguistique.

Ce progrès vient du fait que la théorie des probabilités permet aux chercheurs de changer l'angle de vue lors de l'exploration de problèmes théoriques et pratiques en linguistique. De plus, les probabilités ne sont pas uniquement un

¹⁰⁹ Ibid.

¹¹⁰ Un axiome ne se démontre pas, il est posé a priori, et tous les autres éléments, propriétés, théorèmes en découlent.

¹¹¹ Voir notamment, Bod RENS & al, *Probabilistic Linguistics*, The MIT Press, London, 2003. pp. 464 qui étudie l'application des probabilités pour un ensemble de sous-domaines de la linguistique (phonologie, morphologie, syntaxe, sémantique, psycholinguistique, linguistique historique et sociolinguistique), chacun couvert par un spécialiste.

outil d'analyse, ils opèrent même dans l'acquisition, la perception et la production du langage. D'ailleurs, la théorie des probabilités¹¹² a été inventée à l'origine comme un modèle cognitif du raisonnement humain dans l'incertitude, du moment qu'elle joue un rôle très important dans la modélisation du fonctionnement du cerveau et conditionne sa capacité à modéliser avec précision les phénomènes langagiers.

Pratiquement, la langue affiche toutes les caractéristiques d'un système probabiliste. De ce fait, l'exploration de la langue¹¹³ doit être comprise non comme un ensemble de règles catégorielles, mais comme un ensemble de règles graduelles (éventuellement redondantes) qui peuvent être caractérisées par une distribution statistique. Donc, l'approche probabiliste peut faire avancer les limites de la théorie linguistique en enrichissant substantiellement l'état actuel des connaissances.

b. Calcul des probabilités

Actuellement nous faisons recours au calcul des probabilités dans plusieurs domaines, notamment en sciences humaines et sociales, vu que nous sommes souvent confrontés à des phénomènes aléatoires, c'est à dire des phénomènes dont nous ne pouvons pas connaître ou prévoir avec certitude les résultats.

Avant toute formalisation, le résultat d'une expérience aléatoire¹¹⁴ s'appelle événement. La quantification des « chances » qu'un tel événement a de se réaliser correspond à la notion intuitive de probabilité. Pour réaliser cette quantification, il

¹¹² Bod RENS & al, *Probabilistic Linguistics*, London, The MIT Press, 2003, p. 40

¹¹³ Ibid., p. 10

¹¹⁴ Jean-Pierre LECOUTRE, *Statistique et probabilités*, Paris, Dunod, 6^{ème} édition, 2016, p. 5

est nécessaire de décrire au préalable, très précisément, l'ensemble des résultats possibles, appelés événements élémentaires. Cet ensemble expérimental s'appelle ensemble fondamental ou univers, il est noté traditionnellement Ω . Par exemple, nous tirons une boule dans une urne contenant une boule noire, deux blanches et cinq rouges, l'ensemble fondamental retenu sera $\Omega = \{\text{noire, blanche, rouge}\}$.

Lorsque les résultats sont numériques, nous parlons de variable aléatoire c'est-à-dire que Ω est identique à tout ou une partie de l'ensemble des nombres réels $\mathbb{R}$. Nous distinguons habituellement¹¹⁵ :

- les variables aléatoires discrètes pour lesquelles l'ensemble Ω des résultats possibles est un ensemble discret de valeurs numériques fini ou infini (typiquement : l'ensemble des entiers naturels) ;
- les variables aléatoires continues pour lesquelles l'ensemble Ω est tout $\mathbb{R}$ (ou un intervalle de $\mathbb{R}$ ou, plus rarement, une union d'intervalles).

Généralement, les événements aléatoires¹¹⁶ sont représentées par des lettres majuscules comme A, B, C, E ... ; et les différentes épreuves ou réalisations d'une expérience par la lettre grecque minuscule ω . Ainsi, quand un événement se réalise par les épreuves $\omega_1, \omega_2, \omega_3, \dots \dots \dots \omega_k$ on le représentera souvent par :

$$A = \{\omega_i, 1 \leq i \leq k\}$$

$$= \{\omega_1, \omega_2, \omega_3, \dots \dots \dots \omega_k\}$$

¹¹⁵ Michel LEJEUNE, *Statistique : La théorie et ses applications*, Paris, Springer, 2^{ème} édition, Collection statistique et probabilités appliquées, 2010, p. 1

¹¹⁶ Corina REISCHER, *Théories des probabilités : problèmes et solutions*, Sainte-foy, Presses de l'Université du Québec, 2002, P. 3

Notons que les événements qui sont réalisables par une seule épreuve s'appellent événements élémentaires et correspondent aux singletons ; les autres événements s'appellent « événements composés », ou tout simplement « événements ».

Dans beaucoup de situations, l'ensemble¹¹⁷ fondamental Ω peut être très grand et ce n'est pas tellement le résultat précis ω de l'expérience qui est intéressant (il peut d'ailleurs être concrètement impossible à atteindre) que de savoir si certains événements sont réalisés ou pas.

La probabilité d'un événement est le rapport du nombre de cas favorables à sa réalisation sur le nombre de tous les cas possibles pour l'expérience aléatoire étudiée¹¹⁸. Cela suppose que tous les cas possibles ont la même chance de se produire (équiprobabilité) et que leur nombre est fini.

$$P(A) = \frac{\text{Nombre de cas favorable à sa réalisation}}{\text{Nombre de tous les cas possibles}}$$

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$$

Notons que le calcul des probabilités repose sur quatre propriétés¹¹⁹ fondamentales à savoir :

Propriété 1 : pour tout événement A de l'espace fondamentale Ω , on a toujours $0 \leq P(A) \leq 1$.

¹¹⁷ Daniel REVUZ, *op.cit.*, 1997, p. 1

¹¹⁸ Siby HASSENE, *Introduction à la statistique et aux probabilités*, Québec, Edition Loze-Dion, 2017, p. 99

¹¹⁹ Ibid., p. 102

Propriété 2 : la probabilité de l'événement impossible est toujours nulle, elle est notée $P(\emptyset) = 0$

Propriété 3 : la probabilité de l'événement certain est toujours unitaire et notée par $P(\Omega) = 1$

Propriété 4 : on a $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ et si les deux événements A et B sont disjoint, c'est-à-dire qu'ils ne se réalisent pas simultanément, alors, $P(A \cup B) = P(A) + P(B)$

Dans le cas de notre travail de recherche, le calcul des probabilités nous sera d'une grande utilité. En effet, il nous permettra de vérifier les chances de tomber sur une forme quelconque dans le corpus d'El Maleh. Par exemple, nous pouvons chercher la répartition des catégories grammaticales (substantifs, adverbe, adjectifs...) dans un texte ou un ensemble de textes.

Prenons un petit corpus composé de 50 mots non ambigus dont 25 sont des noms, 20 sont des verbes et 5 sont des adjectifs. Considérons l'expérience de la sélection aléatoire d'un mot A à partir de ce corpus. Quelle est la probabilité de sélectionner un verbe, un nom ou un adjectif ?

L'espace fondamental Ω de cette expérience aléatoire est l'ensemble des mots du corpus.

Soit A l'événement associé à l'expérience aléatoire avoir un verbe, alors :

$$A = \{\omega_i, 1 \leq i \leq k\}$$

$$A = \{verbe_1, verbe_2, \dots \dots \dots verbe_{20}\}$$

$$P(A) = \frac{Card(A)}{Card(\Omega)} = \frac{20}{50} = 40\%$$

Soit B l'évènement associé à l'expérience aléatoire avoir un Nom, alors :

$$B = \{nom_1, nom_2, \dots \dots \dots nom_{25}\}$$

$$P(A) = \frac{Card(B)}{Card(\Omega)} = \frac{25}{50} = 50\%$$

Soit C l'évènement associé à l'expérience aléatoire avoir un adjectif, alors :

$$C = \{adjectif_1, adjectif_2, \dots \dots \dots adjectif_5\}$$

$$P(C) = \frac{Card(5)}{Card(\Omega)} = \frac{5}{50} = 10\%$$

Dans les points suivants, nous présenterons quelques lois de probabilités regroupées selon le type de la variable aléatoire.

2.2. Lois de probabilités discrètes usuelles

Assez souvent, nous sommes confrontés dans la pratique à des phénomènes dont les variables aléatoires associées peuvent prendre des valeurs isolées d'un intervalle de $\mathbb{R}$, il s'agit souvent de nombre entier, par exemple : le nombre de mots dans les phrases d'un corpus, le nombre d'occurrence dans un texte, le nombre d'enfant par ménage, etc). Ce type de variable est appelée variable aléatoire discrète, elle prend ses valeurs sur un ensemble fini $x_1, x_2, \dots, x_n$ ou sur un ensemble infini dénombrable. Nous présenterons, dans ce qui va suivre, quelques lois de probabilité discrètes.

a. Loi de Bernoulli

C'est la loi¹²⁰ la plus simple que l'on puisse envisager. Elle est utilisée lorsqu'une expérience aléatoire n'a que deux résultats possibles : le succès avec une probabilité p et l'échec avec une probabilité $q = 1 - p$. Les applications de cette loi sont nombreuses. Nous pouvons citer, à titre d'exemple, l'étude de la composition d'une population (masculin-féminin) ou le contrôle de la qualité de certaines marchandises (bonne ou défectueuse). Notons que dans la littérature, lorsque nous parlons de variables dichotomiques, il est d'usage de dénoter $1 - p$ tous simplement par la lettre q .

En pratique, la variable¹²¹ aléatoire X sera utilisée comme fonction indicatrice d'un événement donné au cours d'une expérience aléatoire. Par convention, X prend la valeur 1 si l'événement se produit et 0 s'il ne se produit

¹²⁰ Dodge YADOLAH, Melfi GIUSEPPE, *Premiers pas en simulation*, Paris, Springer, Collection Statistique et probabilités appliquées, 2008, p.17

¹²¹ Michel LEJEUNE, *op.cit.*, 2010, p. 47

pas à l'issue de l'expérience. Dans ce contexte, p représente la probabilité de l'événement considéré.

Soit un événement A de probabilité $p = P(A)$ connue, nous définirons la variable aléatoire X associée à la réalisation ou non de A , par :

$X = 1$ si A est réalisé (succès) et

$X = 0$ si A est non réalisé (échec)

On a donc pour $D_X = \{0,1\}$:

$$P(A) = P(X = 1) = p$$

et

$$P(X = 0) = 1 - p = q$$

De ce fait, nous disons que la variable aléatoire X suit une loi de Bernoulli de paramètre p et nous la notons ainsi : $X \sim B(p)$.

avec :

l'espérance mathématique d'une variable bernoullienne : $E(X) = p$

la variance : $V(X) = \sigma^2 = p(1 - p) = pq$

Une suite d'épreuves de Bernoulli¹²², répétées au cours du temps, définit alors un processus de Bernoulli. Selon les conditions de réalisation les épreuves (indépendamment les unes des autres ou non) et la variable aléatoire à laquelle

¹²² Mathé ARMELLE, *L'essentiel des probabilités*, Issy-les-Moulineaux, Gualino, 2016, p.58

nous nous intéressons, nous pouvons modéliser la distribution de probabilité de ces variables à l'aide de différentes lois de probabilités : lois binomiale, géométrique, hypergéométrique, Pascal, binomiale négative. Seule la loi binomiale sera exposée dans ce qui va suivre vu son utilité en statistique lexicale.

b. Loi binomiale

L'expérience¹²³ consiste en une suite de répétitions, successives et indépendantes, de l'épreuve aléatoire de Bernoulli.

Soit :

A un événement de probabilité $p = P(A)$ connue et n un entier non nul,

X une variable aléatoire discrète, associée au nombre de réalisations de A, au cours des n expériences.

Nous aurons, pour tout $k \in D_x$ et $D_x = \{0, 1, \dots, n\}$:

$$P(X = k) = C_n^k p^k q^{n-k}$$

Avec :

L'espérance mathématique d'une variable binomiale : $E(X) = n.p$

La variance : $V(X) = \sigma^2 = np(1 - p) = npq$

¹²³ Michel LEJEUNE, *op.cit.*, 2010, p. 47

La variable aléatoire X correspond au nombre de succès au cours de n répétitions du processus de Bernoulli. C'est une variable aléatoire discrète à valeurs entières et suit une loi binomiale de paramètres n et p .

Notons que la suite des épreuves doit impérativement obéir aux conditions¹²⁴ suivantes :

1. Le nombre n des épreuves est déterminé à l'avance ;
2. L'épreuve est une expérience de Bernoulli avec une seule alternative : succès ou échec
3. Les épreuves sont identiques, indépendantes et effectuées dans les mêmes conditions : la probabilité p de réalisation de l'événement A est donc constante.

La loi binomiale est omniprésente en statistique. Elle est surtout utilisée dans le domaine des sondages. Ayant sélectionné au hasard n individus dans une grande population, nous pouvons « estimer » la proportion p d'individus ayant un caractère donné (succès).

En statistique lexicale, Charles Muller a proposé d'utiliser cette loi pour mesurer la richesse lexicale et étudier l'accroissement du vocabulaire. Elle permet également de déterminer la probabilité d'obtenir l'une des catégories grammaticales (verbe, substantif, adjectif, etc.) en procédant à un tirage au hasard de n mots dans un corpus donné.

¹²⁴ Hassène SIBY, *op. cit.*, 2017, p. 151

2.3. Lois continues

Nous rencontrons dans la pratique des phénomènes aléatoires dont les variables aléatoires associées peuvent prendre toute valeur de $\mathbb{R}$ ou d'un intervalle de $\mathbb{R}$ (revenus, tailles, poids, ...). Donc les ensembles des valeurs possibles de telles variables aléatoires sont infinis et non dénombrables, elles sont dites variables aléatoires continues.

Contrairement aux variables aléatoires discrètes, les lois des variables aléatoires continues ne sont pas caractérisées par les probabilités discrètes des réalisations $P(X = x_i)$ mais elles sont définies par la donnée d'une fonction continue f , appelée densité de probabilité de la variable aléatoire continue.

a. Loi uniforme

La loi uniforme¹²⁵ est la loi continue la plus simple. Une variable aléatoire X suit une loi uniforme sur l'intervalle fini $[a, b]$, $X \sim U(a, b)$, si sa fonction de densité est constante sur $[a, b]$ et nulle à l'extérieur de cet intervalle, soit :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

La loi uniforme de référence est la loi $U[0,1]$ correspondant aux générateurs de nombres au hasard des logiciels¹²⁶. À partir d'un tel générateur nous pouvons produire des nombres au hasard sur $[a,b]$ par la transformation :

$$y = (b - a) x + a$$

¹²⁵ Eva CANTONI, Philippe HUBER, Elvezio RONCHETTI, *op.cit.*, 2006, p. 46

¹²⁶ Tels que les fonctions «RANDOM» ou «ALEA»

Donc, la loi $U [0,1]$ joue un rôle particulier dans les algorithmes qui permettent la génération efficace par ordinateur de réalisations (« nombres aléatoires ») d'une variable aléatoire qui suit cette loi. Cela permet ensuite de générer des réalisations de variables aléatoires qui suivent d'autres lois discrètes ou continues. Ceci est à la base des techniques de simulation de processus aléatoires par ordinateur, qui demeure un outil indispensable pour la modélisation statistique moderne.

b. Loi normale (ou loi de LAPLACE GAUSS)

La loi normale¹²⁷ est la loi de probabilité la plus courante pour les variables aléatoires continues. Son importance vient du fait qu'une grande partie de l'inférence statistique repose sur elle. Appelée également loi de Gauss ou loi de Laplace Gauss, la loi normale a été formulée pour la première fois en 1733 par De Moivre dans ses recherches sur la forme limite de la loi binomiale. En 1816, Gauss a pu confirmer que la loi normale s'applique à des variables aléatoires qui peuvent prendre toutes les valeurs réelles possibles entre moins l'infini ($-\infty$) et plus l'infini ($+\infty$).

C'est une loi¹²⁸ d'une portée générale. Elle peut être exploitée dans plusieurs domaines : l'agriculture, l'industrie, la médecine, la démographie, etc. Elle permet de modéliser¹²⁹ le plus grand nombre de phénomènes aléatoires

¹²⁷ Hassène SIBY, *op. cit.*, 2017, p. 168

¹²⁸ P. SALAT, VERBOROM RATIO, *Exemples d'Études Statistiques portant sur le Vocabulaire Latin*, Clermont-Ferrand, Faculté des lettres et sciences humaines de l'université de Blaise-Pascale, 1991, pp. 33-53

¹²⁹ Mathé ARMELLE, *op.cit.*, 2016, p.82

(phénomènes possédant de nombreuses causes indépendantes dont les effets s'additionnent, sans que l'un d'eux domine).

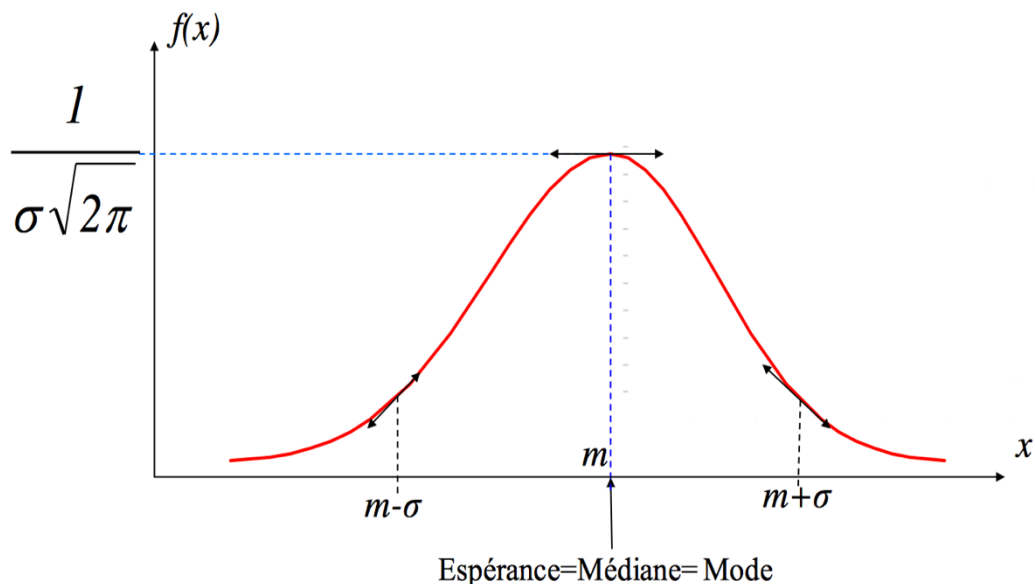
Soient μ et σ deux nombres réels avec ($\sigma > 0$). nous disons qu'une variable aléatoire continue X suit une loi normale générale de paramètres μ et σ si sa densité de probabilité est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall x \in \mathbb{R}$$

Les paramètres¹³⁰ sont notés μ et σ^2 du fait qu'ils correspondent respectivement à la moyenne et à la variance de la loi, σ étant donc son écart-type. Le graphe de la densité est la fameuse courbe en cloche symétrique autour de la valeur μ .

¹³⁰ Michel LEJEUNE, *op.cit.*, 2010, p.57

Figure I. 1 : Fonction de répartition de la loi normale générale



Si X suit une loi¹³¹ normale $N(\mu, \sigma^2)$, il est facile de montrer que $E(X) = \mu$ et $V(X) = \sigma^2$. Le paramètre μ ne change pas la forme du graphe de la f . Sa variation déplace uniquement la « cloche » le long de l'axe des x . En revanche, la variation de σ change la forme de la courbe. En effet, ce paramètre représente la dispersion de la variable aléatoire autour de sa moyenne μ . Plus σ est petit, plus la courbe sera « serrée » autour de μ .

Les paramètres $\mu = 0$ et $\sigma = 1$ correspondent à la loi normale centrée réduite¹³², il est d'usage de la dénoter avec la lettre Z . Le passage de la variable aléatoire $X \sim N(\mu, \sigma^2)$ à la variable aléatoire $Z \sim N(0, 1)$ s'effectue par la transformation de la variable aléatoire X à Z qui est égale $\frac{X-\mu}{\sigma}$.

¹³¹ Dodge YADOLAH, Melfi GIUSEPPE, *op. cit.*, 2008, p.25

¹³² Ibid., p.26

Ainsi la probabilité $(X \leq \alpha) =$ La probabilité $\left(\frac{X-\mu}{\sigma} \leq \frac{\alpha-\mu}{\sigma}\right)$

$$= \text{La probabilité } \left(Z \leq \frac{\alpha-\mu}{\sigma}\right)$$

De ce fait, nous disons qu'une variable aléatoire continue Z suit une loi normale centrée réduite si la densité de probabilité est donnée par :

$$f_x(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \forall x \in \mathbb{R}$$

Avec :

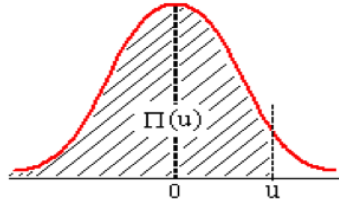
L'espérance $(Z) = 0$ et la variance $(Z) = 1$

On note que $Z \sim N(0; 1)$

Les résultats de ces calculs sont présentés sous forme de tables, appelées tables de la loi normale centrée réduite.

Tableau I. 1 : Extrait de la table de la loi normale centrée réduite

*Table de la loi normale centrée réduite qui donne
 $\pi(u) = P(U < u)$ pour $u \geq 0$ [avec $\pi(-u) = 1 - \pi(u)$]*



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214

La loi normale reste sans aucun doute la loi la plus répandue et la plus utilisée en statistique classique. Elle permet très souvent de décrire la loi des erreurs de mesure. De plus, de nombreuses autres lois statistiques peuvent être approchées par la loi normale, tout spécialement dans le cas des grands échantillons.

Dans ce chapitre, nous avons essayé, tout d'abord, de dégager la relation entre la linguistique et la statistique. Nous avons montré qu'il s'agit de deux disciplines éloignées dans leurs fondements mais rapprochées par l'avènement de l'outil informatique. Cela a confirmé l'idée que l'unité lexicale a aussi une dimension statistique et qu'elle peut être abordée d'un point de vue chiffré. Aussi, nous nous sommes intéressée à la fonction informatique et son apport indéniable pour le développement de la statistique linguistique.

Ensuite, nous nous sommes focalisée sur la statistique lexicale, où nous avons approché, dans un premier temps, le dépouillement et la quantification du corpus. Dans un second temps, il était question de présenter la méthode d'analyse en rappelant quelques présupposés théoriques qui nous serviront de base pour l'analyse du corpus. Nous précisons, par ailleurs, que notre intention n'était pas de nous faire passer pour un spécialiste en la matière. Il s'agit uniquement de nous arrêter sur quelques notions et règles classiques que nous jugeons nécessaires à l'étude, comme : les lois de probabilités discrètes (loi de Bernoulli, loi binomiale), les lois continues (loi uniforme, loi de LAPLACE GAUSS), etc.

Néanmoins, avant de se consacrer à l'application de ces lois, nous avons jugé utile de présenter, dans le chapitre suivant, le corpus qui fera l'objet de notre recherche ainsi que le logiciel de travail.

CHAPITRE 3

Corpus et méthodes de découpage et d'analyse

Au niveau du dernier chapitre de cette première partie nous nous intéresserons, tout d'abord, à la présentation du corpus ainsi qu'à la préparation de la base de données. Il s'agit, en effet, d'une étape indispensable pour l'exploitation des œuvres d'Edmond Amrane El Maleh via un logiciel hypertextuel.

Ensuite nous réaliserons une présentation détaillée du logiciel de travail Hyperbase. Notons que le concepteur de ce logiciel, le professeur Étienne BRUNET, a mis à la disposition des utilisateurs un manuel de référence, contenant un mode d'emploi, pour une meilleure application des fonctions destinées à l'analyse lexicométrique des œuvres littéraires. C'est ce manuel qui nous servira de guide pour faire une démonstration des principales fonctionnalités de ce logiciel. Nous soulignons, cependant, que notre présentation ne sera pas exhaustive, elle s'intéressera uniquement aux fonctionnalités que nous jugeons importantes dans la présente étude.

A. Délimitation du corpus et préparation de la base de données

Comme nous l'avons signalé précédemment, notre corpus est constitué de l'œuvre de l'écrivain marocain juif Edmond Amrane El Maleh. Il s'agit d'une dizaine de textes recouvrant deux genres littéraires : le roman et l'essai. Nous nous proposons de l'étudier en respectant l'ordre de parution des textes, ce qui nous

permettra d’analyser l’évolution de quelques aspects du vocabulaire de l’auteur durant quarante ans.

1. Présentation du corpus

Notre corpus englobe 10 œuvres qui, après lemmatisation, totalisent 522421 occurrences et 28638 vocables. Le tableau suivant présente l’étendue de chaque texte ainsi que celui du corpus pris dans son intégralité :

Tableau I. 2 : Étendue des textes et du corpus

TEXTES	Nombre de vocables	Nombre d'occurrences
PARCOURS	9467	64566
AILEN	10322	75357
MILLE	10109	75802
ELHAKI	13054	109549
GENET	8139	53949
ABOUNOUR	5520	26350
ZRIREK	10622	88200
FEMME	1755	6083
LETTRES	4681	22565
Total	28638	522421

La préparation documentaire et statistique des données s’appuie sur le tableau présenté ci-haut. Ce dernier montre que l’étendue des textes composant notre corpus varie d’un ouvrage à l’autre. Cependant, nous constatons que le classement des textes reste immuable quel que soit le critère retenu (nombre d’occurrences ou nombre de vocables). En effet, selon les deux critères précédemment cités, le classement est le suivant :

**ELHAKI > ZRIREK > MILLE > AILEN > PARCOURS > GENET >
ABOUNOUR > LETTRES > FEMME**

Il reste à signaler que nous avons jugé optimale de fusionner le texte *La Mâlle de Sidi Maachou*, qui présente une étendue excessivement faible par rapport aux autres segments du corpus, avec un autre texte à savoir *Lettres à moi-même* pour ne pas fausser l'interprétation des résultats. Une fois notre corpus choisi, il faut le préparer pour qu'il puisse être exploitable à l'aide du logiciel Hyperbase.

**2. Préparation des données textuelles : Base de données «
HYPEREDM.EXE »**

Pour élaborer la base de données, le chercheur doit suivre une démarche bien déterminée. D'abord, il est nécessaire de préparer la version numérisée du corpus. Pour ce faire, nous avons eu recours à l'outil du scanner pour convertir la version papier en fichier électronique JPEG, puis en fichier électronique modifiable via OPCR. Cependant, la dernière version obtenue nécessitait une correction, vue que la machine ne reconnaît pas tous les caractères de l'alphabet. Pour cela, le retour au texte original était indispensable pour remédier aux anomalies.

Par la suite, les textes sont assemblés selon leurs dates de publication et enregistrés dans un fichier format ASCII (Américain Standard Code for Information Interchange) ou format texte, il s'agit d'une norme de codage informatique internationalement reconnue des caractères. Chaque partie du corpus doit être précédée d'une ligne où l'on indiquera la nomenclature des textes un à un en employant le symbole composite &&& avant et après le titre qui est accompagné d'une abréviation et d'un code.

Exemple : &&&Abner Abounour, ABOUNOUR, AA&&&

Une fois la préparation de la base hypertextuelle « HYPEREDM.EXE » est terminée, il reste à choisir un logiciel hypertexte pour le traitement documentaire et statistique des grands corpus. Nous avons alors opté pour le logiciel Hyperbase mis en œuvre par le professeur Etienne Brunet. À cette étape notre professeur M. Ali HEFIED nous a aidée à installer le logiciel sur le système d'exploitation Windows version 8.1, installé en parallèle avec macOS High Sierra, ainsi que pour la création de la base.

Une fois la base de travail est préparée, elle est prête, alors, pour l'exploration et l'exploitation lexicométrique. Reste à définir le logiciel de travail.

B. Hyperbase

Hyperbase est un logiciel conçu pour le traitement documentaire et statistique des corpus textuels. Il a été créé par le professeur Etienne BRUNET en 1989. Ce logiciel est rédigé en langue française mais s'applique à toutes les langues qui utilisent l'alphabet latin (français, anglais, portugais...). Il autorise un ensemble d'opération sur des corpus préalablement préparés et traités selon les normes établies par Etienne Brunet. Il est utilisé en linguistique de corpus, en statistique textuelle et en philologie.

Notons que le créateur d'Hyperbase a mis à la disposition des utilisateurs un manuel de référence, contenant un mode d'emploi, pour une meilleure application des fonctions destinées à l'analyse lexicométrique des œuvres littéraires. Alors, nous nous proposons, dans les pages qui suivent, de présenter le processus de création de la base et une démonstration des principales fonctionnalités de ce logiciel, à travers une synthèse du manuel sus-cité.

1. Création de la base

La création de la base commence, tout d'abord, par la préparation des données. Comme nous l'avons signalé plus haut, les données doivent être présenter sous format texte brut. S'agissant des titres, nous soulignons que nous pouvons proposer plusieurs titres pour un même texte, à condition de les séparer par une virgule.

Exemple : **&&&Abner Abounour, ABOUNOUR, AA&&&**

Le premier titre servira de référence explicite ; le second, réduit à un mot, sera utilisé dans la légende des graphiques ; le troisième, qui ne dispose que de deux lettres, indiquera le code qui symbolise chaque texte dans les concordances.

Le logiciel entame par la suite le découpage du texte en parties de longueur voisine. Notons que le fichier de type ASCII ne préjuge pas de la segmentation du corpus : en mots, en lignes, en paragraphes, en pages, en textes distincts. D'où quelques explications qui précisent les options retenues :

Les pages : Si les pages n'ont pas été distinguées à l'avance, le programme procède de lui-même au découpage des pages à raison de 200 mots au moins par page (en s'abstenant de couper les phrases).

Les paragraphes : Ils sont délimités par le retour de chariot. Si les paragraphes ainsi définis sont trop longs, le programme permet de les découper en unités plus petites.

Les phrases : La segmentation en phrases n'est appliquée que lorsqu'un paragraphe est jugé trop long. Pour distinguer les phrases, le programme prend appui sur une ponctuation forte, principalement le point.

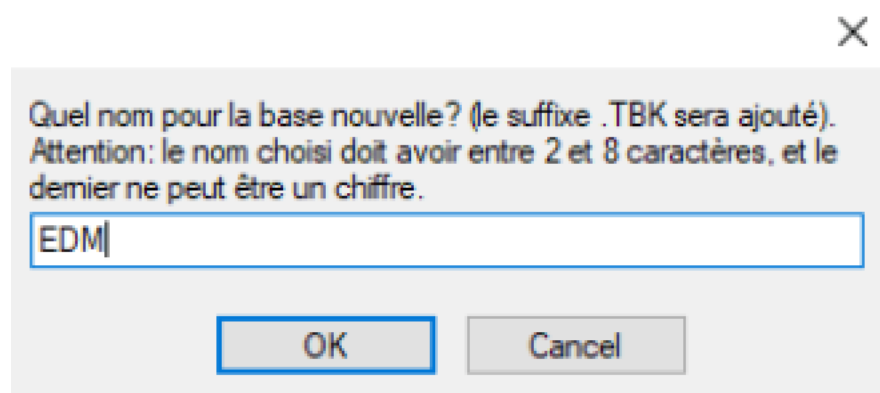
Les lignes : La segmentation en lignes est rarement pertinente, sauf en poésie versifiée. Si les données sont de ce type, nous avons intérêt à supprimer les retours de chariot intempestifs, en veillant en outre à recoller les mots coupés en fin de ligne. Cependant, si cette précaution n'a pas été prise, nous pouvons rectifier le tir en fin de traitement, en sollicitant le bouton "OPTION VERS" du menu principal.

Les mots : Le découpage des mots et leur classement obéissent aux exigences du français. Les mots accentués prennent place au rang qu'ils ont dans le dictionnaire. Notons que la distinction entre minuscule et majuscule est gardée dans le texte initial ainsi que le dictionnaire pour distinguer les noms propres, mais elle est abolie dans les classements.

Finalement, notons que le programme accepte 75 textes. Il est préférable que leur étendue, d'un texte à l'autre, ne soit pas trop déséquilibrée. Si les "textes" correspondent à des segments trop courts ou si leur nombre dépasse la limite de 75, nous procéderons à des regroupements, en satisfaisant non seulement aux contraintes du programme mais aussi aux impératifs méthodologiques de la statistique.

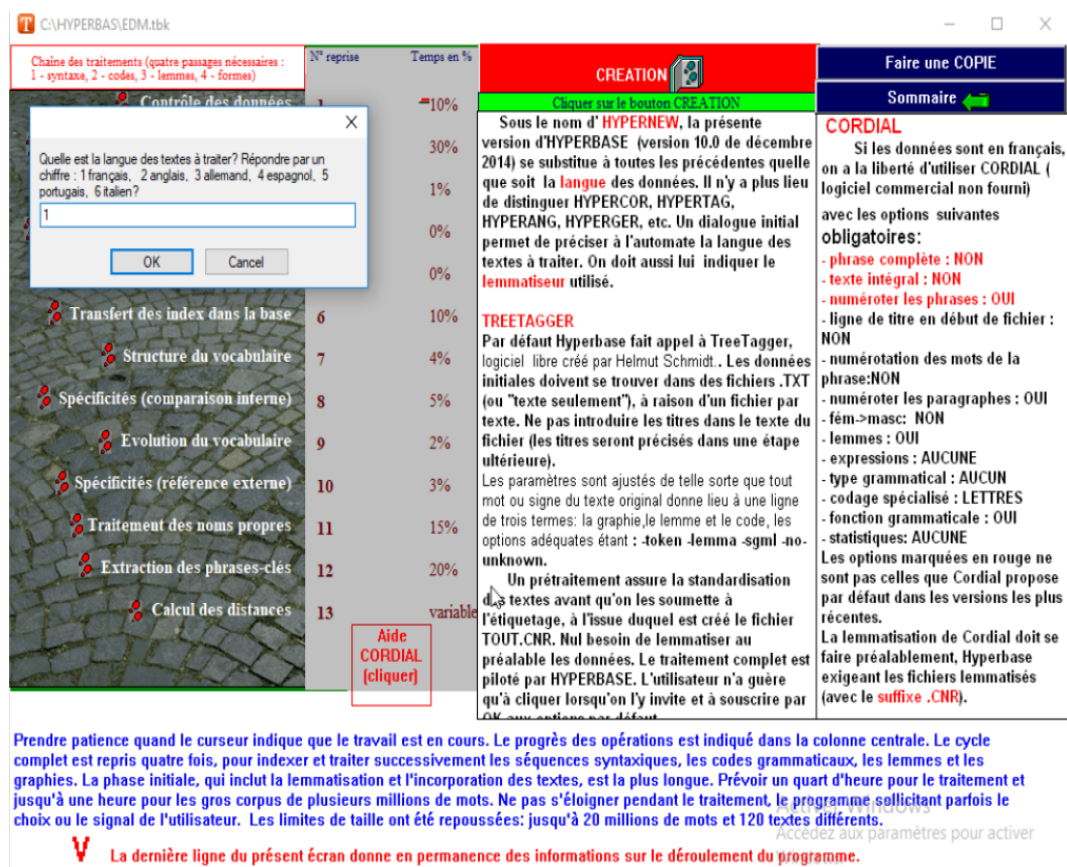
Cela étant dit, nous présenterons, dans ce qui va suivre, le fonctionnement du logiciel pour la création de la base de données. La première étape consiste à exécuter Hyperbase, une boîte de dialogue s'affiche qui nous demande d'insérer le nom de la base qu'on veut créer comme la montre l'image ci-après :

Figure I. 2 : Création de base de données



Après avoir cliqué sur le bouton CREATION, nous choisissons la langue du texte à traiter, qui est dans notre cas la langue française, associée au chiffre 1.

Figure I. 3 : Sélection de la langue des textes



Par la suite nous passons au choix du lemmatiseur comme le montre la photo suivante :

Figure I. 4 : Sélection du lemmatiseur



Pour arriver à la principale étape qui consiste à importer les textes formant notre corpus dans le logiciel Hyperbase et ce en respectant l'ordre de leurs parutions.

Figure I. 5 : Introduction des titres longs des textes

Chaine des traitements (quatre passages nécessaires : 1 - syntaxe, 2 - codes, 3 - lemmes, 4 - formes)

	N° reprise	Temps en %
Contrôle des données	1	10%
Importation et formatage des textes	2	30%
Tri et Indexation des textes	3	1%
Interclassement des index de textes	4	0%
Interclassement (niveau 2)	5	0%
Transfert des index dans la base	6	10%
Structure du vocabulaire	7	4%
Spécificités (comparaison interne)	8	
Evolution du vocabulaire	9	
Spécificités (référence externe)	10	
Traitement des noms propres	11	
Extraction des phrases-clés	12	20%
Calcul des distances	13	variable

CREATION

Cliquez sur le bouton CREATION

Sous le nom d'HYPERNEW, la présente version d'HYPERBASE (version 10.0 de décembre 2014) se substitue à toutes les précédentes quelle que soit la langue des données. Il n'y a plus lieu de distinguer HYPERCOR, HYPERTAG, HYPERANG, HYPERGER, etc. Un dialogue initial permet de préciser à l'automate la langue des textes à traiter. On doit aussi lui indiquer le lemmatiseur utilisé.

TREETAGGER

Par défaut Hyperbase fait appel à TreeTagger, logiciel libre créé par Helmut Schmidt. Les données initiales doivent se trouver dans des fichiers .TXT

Un fichier par ligne dans le texte du corpus

Commencer par introduire les titres (long, court et abrégé) pour chaque texte du corpus.

OK

Faire une COPIE

Sommaire

CORDIAL

Si les données sont en français, on a la liberté d'utiliser CORDIAL (logiciel commercial non fourni) avec les options suivantes obligatoires:

- phrase complète : NON
- texte intégral : NON
- numéroté les phrases : OUI
- ligne de titre en début de fichier : NON
- numérotation des mots de la phrase:NON
- numéroté les paragraphes : OUI
- fém->masc: NON
- lemmes : OUI
- expressions : AUCUNE
- type grammatical : AUCUN
- codage spécialisé : LETTRES
- fonction grammaticale : OUI
- statistiques: AUCUNE

Les options marquées en rouge ne sont pas celles que Cordial propose par défaut dans les versions les plus récentes. La lemmatisation de Cordial doit se faire préalablement, Hyperbase exigeant les fichiers lemmatisés (avec le suffixe .CNR).

Un prétraitement assure la standardisation des textes avant qu'on les soumette à l'étiquetage, à l'issue duquel est créé le fichier TOUT.CNR. Nul besoin de lemmatiser au préalable les données. Le traitement complet est piloté par HYPERBASE. L'utilisateur n'a qu'à cliquer lorsqu'on l'y invite et à souscrire par OK au traitement.

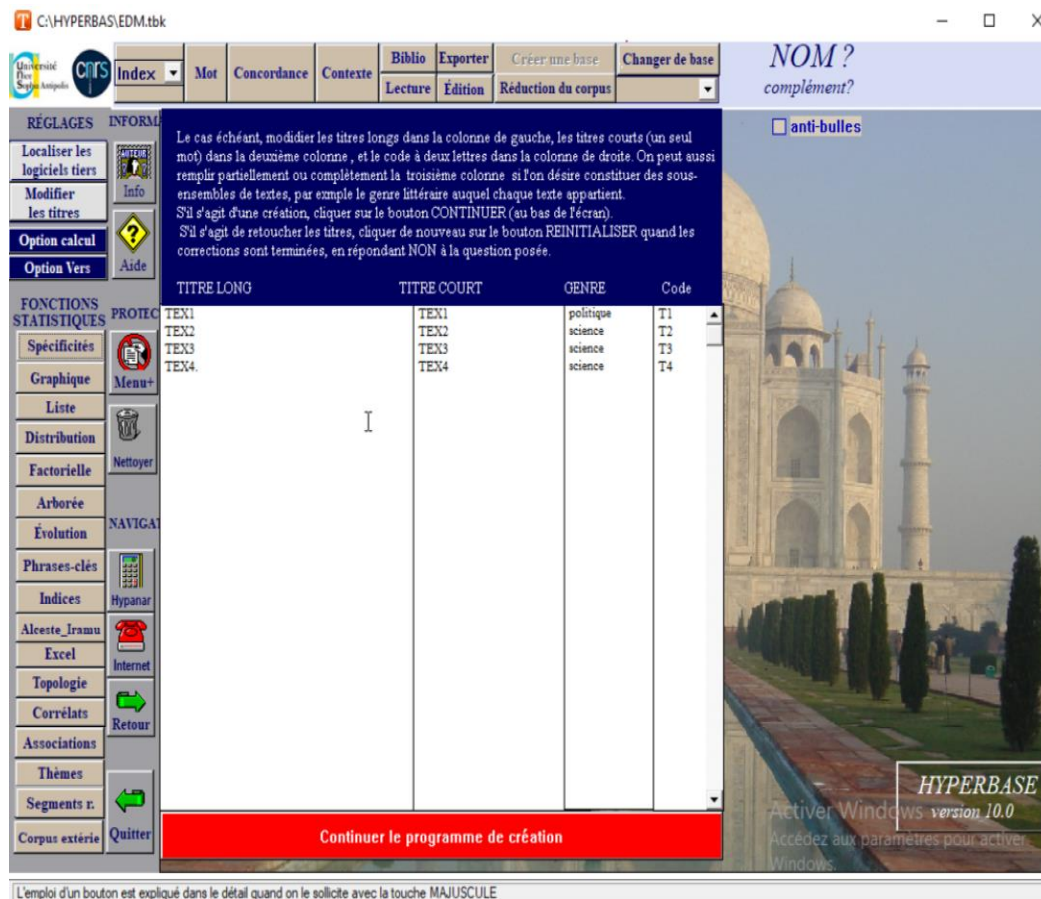
Prendre patience quand le curseur indique que le travail est en cours. Le progrès des opérations est indiqué dans la colonne centrale. Le cycle complet est repris quatre fois, pour indexer et traiter successivement les séquences syntaxiques, les codes grammaticaux, les lemmes et les graphies. La phase initiale, qui inclut la lemmatisation et l'incorporation des textes, est la plus longue. Prévoir un quart d'heure pour le traitement et jusqu'à une heure pour les gros corpus de plusieurs millions de mots. Ne pas s'éloigner pendant le traitement, le programme sollicite parfois le choix ou le signal de l'utilisateur. Les limites de taille ont été repoussées: jusqu'à 20 millions de mots et 120 textes différents.

Accédez aux paramètres pour activer

V La dernière ligne du présent écran donne en permanence des informations sur le déroulement du programme.

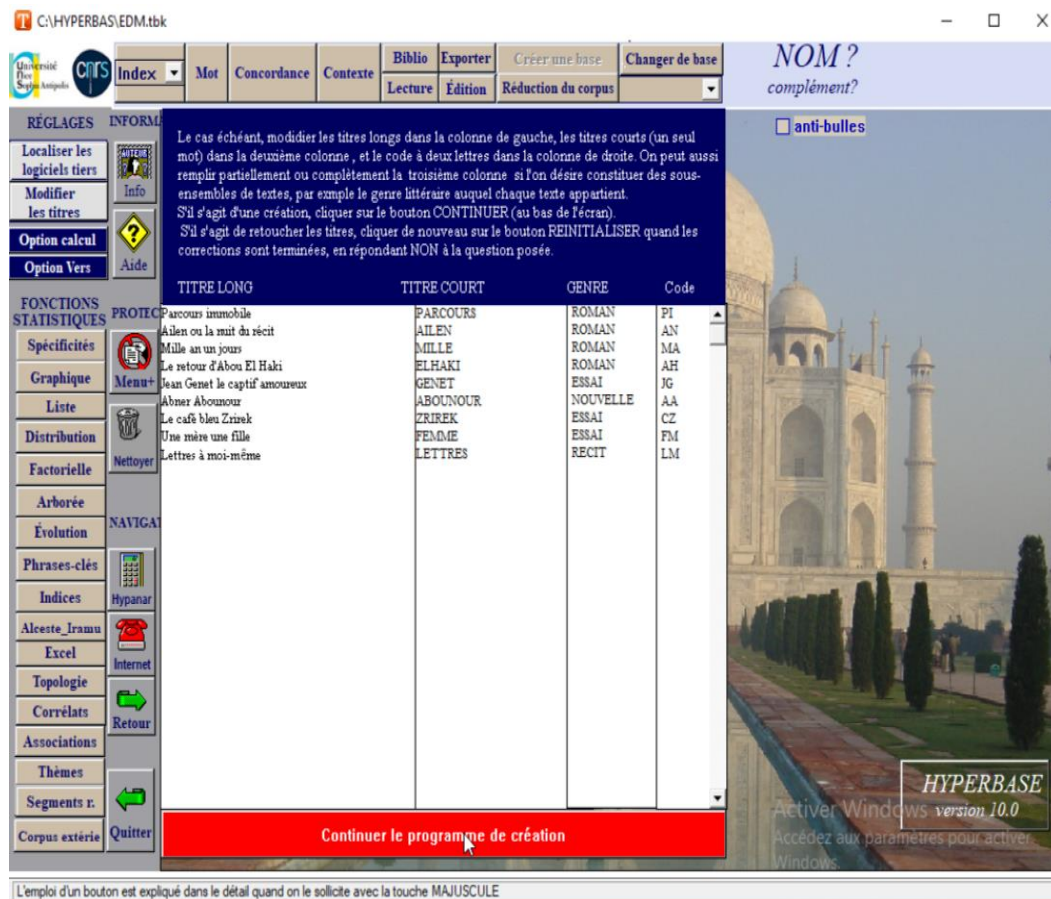
Cette phase s'intéressera à l'introduction des titres (longs, courts et abrégé) pour chaque texte du corpus. Ainsi la page de dialogue suivante s'affiche :

Figure I. 6 : Saisie des titres longs, courts et abrégé



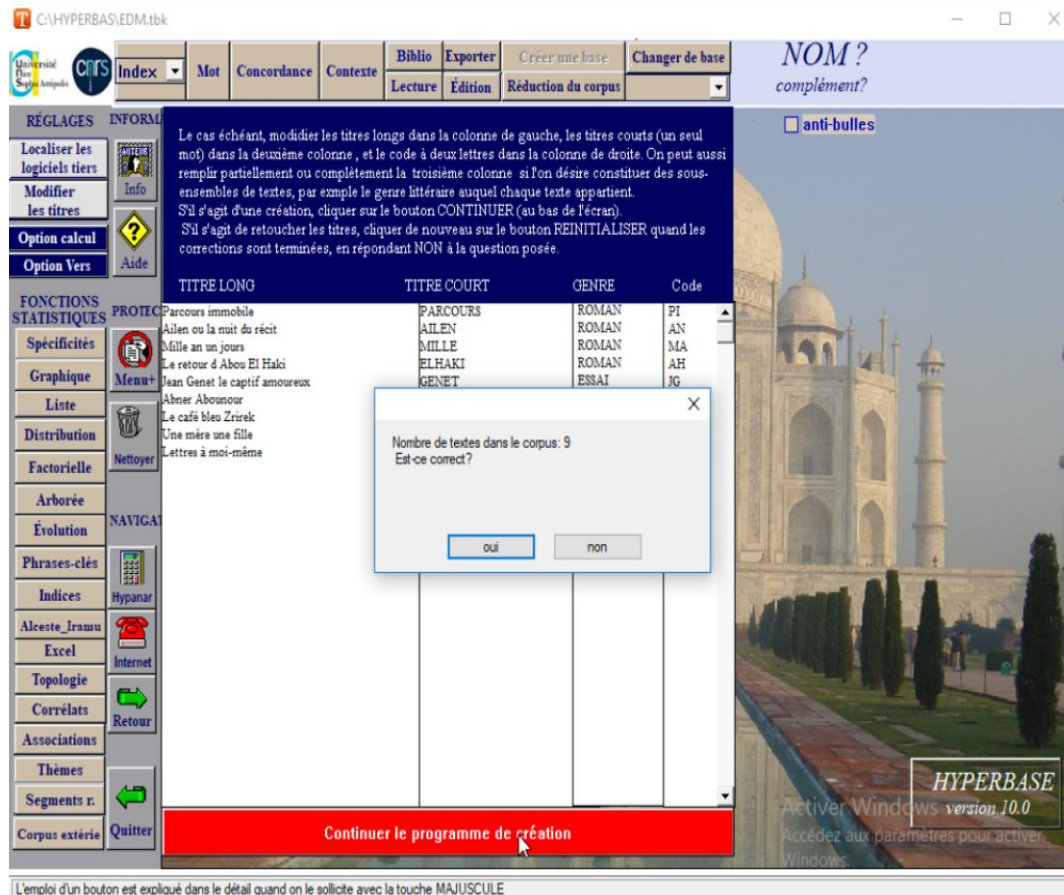
Nous procéderons alors à l'introduction des titres des différents textes ainsi que leurs genres, puis nous déboucherons sur l'image suivante :

Figure I. 7 : Aperçu sur les textes



En cliquant sur « Continuer le programme de création », une boîte de dialogue s'affiche qui nous rappelle le nombre des textes constitutifs de notre corpus.

Figure I. 8 : Boite de dialogue rappelant le nombre des textes constitutifs du corpus



En cliquant sur OUI, une boîte de dialogue nous propose de choisir « la lemmatisation à la volée », comme c’est affiché dans l’image ci-après :

Figure I. 9 : Lemmatisation à la volée

The screenshot shows the HYPERBASE software interface. The main window is titled 'CA\HYPERBASE\EDM.tbk'. It features a sidebar on the left with a list of processing steps: Contrôle des données, Importation et formatage des textes, Tri et Indexation des textes, Interclassement des index de textes, Interclassement (niveau 2), Transfert des index dans la base, Structure du corpus, Spécificités (comparaison), Evolution du corpus, Spécificités (références), Traitement des données, Extraction des données, and Calcul des statistiques. The main area is divided into three columns. The first column shows the 'N° reprise' and 'Temps en %' for each step. The second column, titled 'CREATION', contains instructions for using the software, including a note about the 'HYPERNEW' version and a warning about the 'CORDIAL' dialog box. The third column, titled 'Faire une COPIE', contains a 'Sommaire' button. A 'CORDIAL' dialog box is open, displaying a list of options for text processing, such as 'phrase complète', 'texte intégral', 'numéroter les phrases', and 'ligne de titre en début de fichier'. The dialog box also includes a 'Suspendre' button and a 'Lemmatiser à la volée' button. A red box highlights the 'Aide CORDIAL (cliquer)' button. Below the dialog box, there is a note about the 'CORDIAL' dialog box and a warning about the 'CORDIAL' dialog box.

Chaine des traitements (quatre passages nécessaires : 1 - syntaxe, 2 - codes, 3 - lemmes, 4 - formes)

	N° reprise	Temps en %
Contrôle des données	1	10%
Importation et formatage des textes	2	30%
Tri et Indexation des textes	3	1%
Interclassement des index de textes	4	0%
Interclassement (niveau 2)	5	0%
Transfert des index dans la base	6	10%

CREATION

Cliquez sur le bouton CREATION

Sous le nom d'HYPERNEW, la présente version d'HYPERBASE (version 10.0 de décembre 2014) se substitue à toutes les précédentes quelle que soit la langue des données. Il n'y a plus lieu de distinguer HYPERCOR, HYPERTAG, HYPERANG, HYPERGER, etc. Un dialogue initial permet de préciser à l'automate la langue des textes à traiter. On doit aussi lui indiquer le lemmatiseur utilisé.

TREETAGGER

Par défaut Hyperbase fait appel à TreeTagger

Faire une COPIE

Sommaire

CORDIAL

Si les données sont en français, on a la liberté d'utiliser CORDIAL (logiciel commercial non fourni) avec les options suivantes obligatoires:

- phrase complète : NON
- texte intégral : NON
- numéroter les phrases : OUI
- ligne de titre en début de fichier : NON

otation des mots de la NON

oter les paragraphes : OUI

nas: NON

es : OUI

ssions : AUCUNE

rammatical : AUCUN

e spécialisé : LETTRES

on grammaticale : OUI

ques : AUCUNE

ions marquées en rouge ne s celles que Cordial propose aut dans les versions les plus

Suspendre

Lemmatiser à la volée

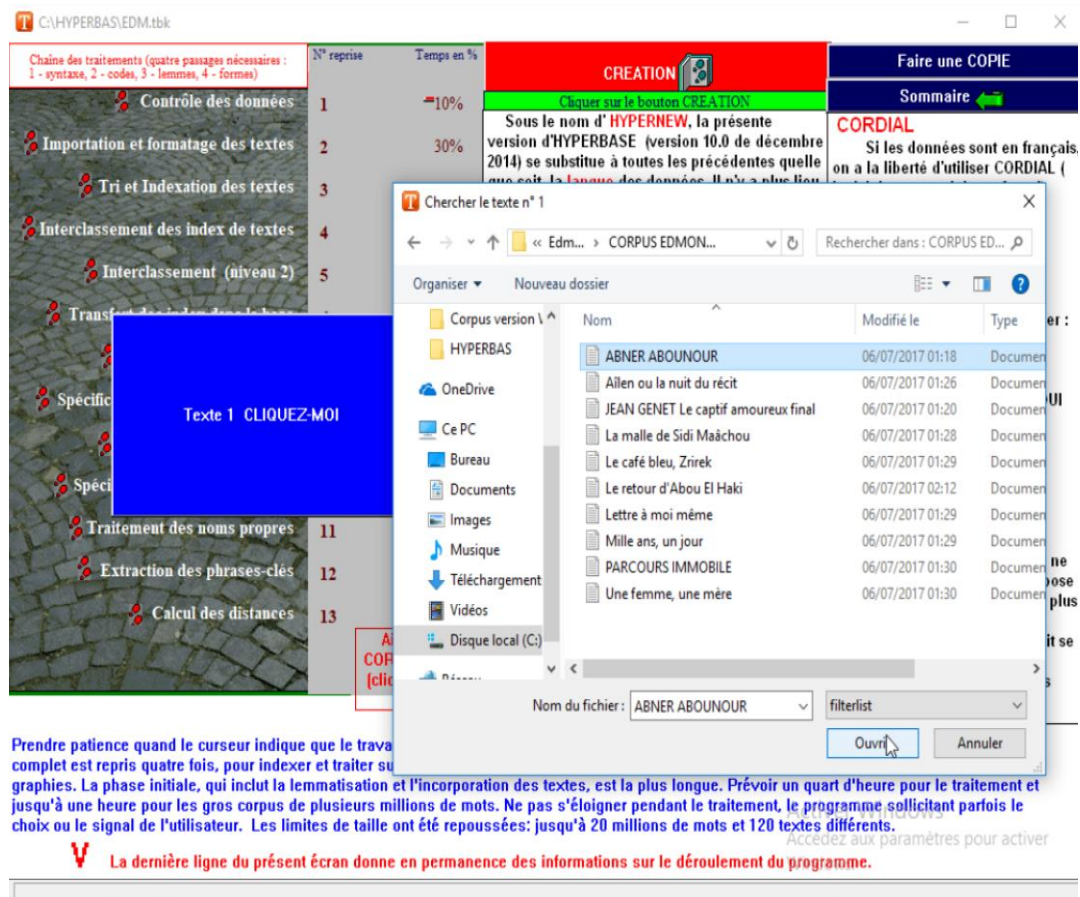
Aide CORDIAL (cliquer)

Le traitement complet est repris quatre fois, pour indexer et traiter successivement les séquences syntaxiques, les codes grammaticaux, les lemmes et les graphies. La phase initiale, qui inclut la lemmatisation et l'incorporation des textes, est la plus longue. Prévoir un quart d'heure pour le traitement et jusqu'à une heure pour les gros corpus de plusieurs millions de mots. Ne pas s'éloigner pendant le traitement, le programme sollicitant parfois le choix ou le signal de l'utilisateur. Les limites de taille ont été repoussées: jusqu'à 20 millions de mots et 120 textes différents.

La dernière ligne du présent écran donne en permanence des informations sur le déroulement du programme.

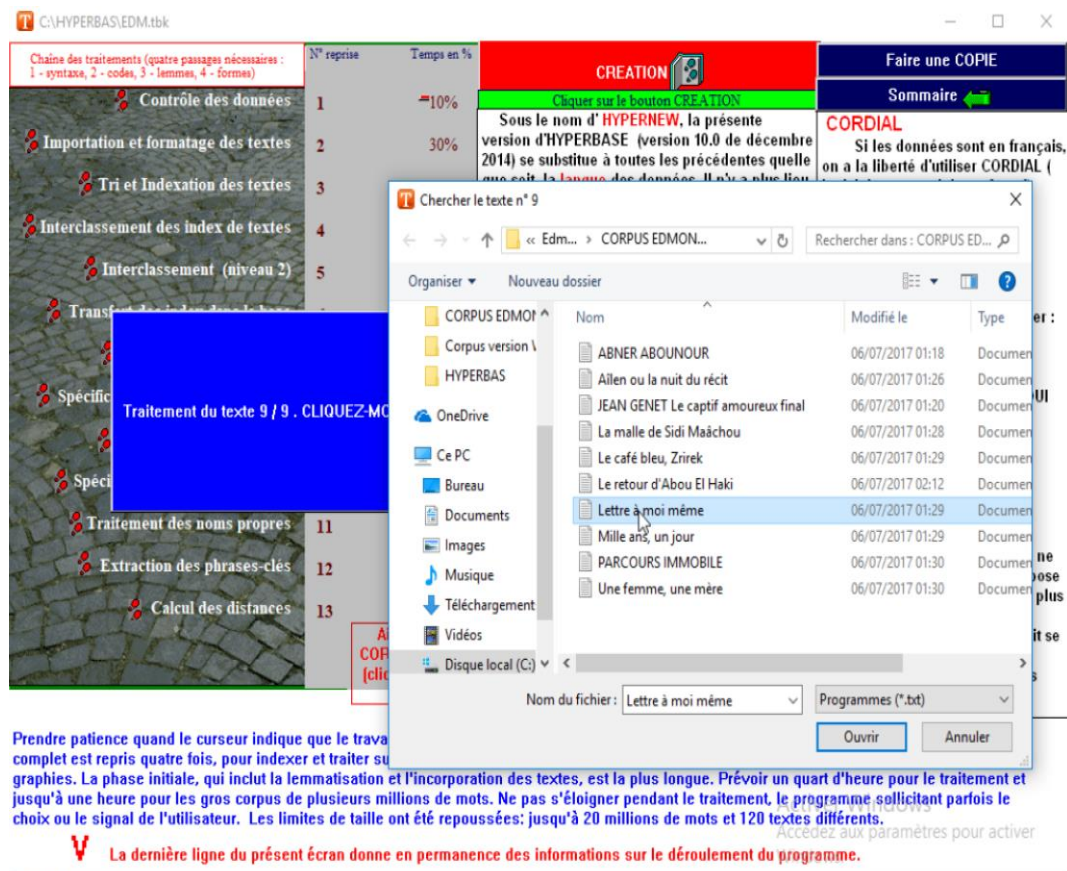
Après avoir effectué notre choix, nous commençons à importer les textes en respectant leur ordre chronologique d'apparition. Ci-après l'image mettant en évidence cette étape :

Figure I. 10 : Importation des textes



Arrivant au neuvième et ultime texte, à savoir Lettre à moi-même, l'interface du logiciel est la suivante :

Figure I. 11 : Importation du dernier texte



À ce stade, la lemmatisation est terminée et l'indexation du texte commencera, en prenant en compte successivement les structures syntaxiques, les codes grammaticaux, les lemmes et enfin les formes.

Figure I. 12 : Fin de la lemmatisation et démarrage de l'indexation

The screenshot shows the HYPERBASE software interface. On the left, a table lists 13 steps of the process, with the progress percentage for each. The 'CREATION' button is highlighted in red. The 'CORDIAL' section on the right lists various options for the indexing process. A dialog box is open in the center, asking if the user wants to continue with the indexing process.

Chaine des traitements (quatre passages nécessaires : 1 - syntaxe, 2 - codes, 3 - lemmes, 4 - formes)	N° reprise	Temps en %
Contrôle des données	1	10%
Importation et formatage des textes	2	30%
Tri et Indexation des textes	3	1%
Interclassement des index de textes	4	0%
Interclassement (niveau 2)	5	0%
Transfert des index dans la base	6	10%
Structure du vocabulaire	7	4%
Spécificités (comparaison interne)	8	
Evolution du vocabulaire	9	
Spécificités (référence externe)	10	
Traitement des noms propres	11	
Extraction des phrases-clés	12	
Calcul des distances	13	

CREATION

Cliquez sur le bouton CREATION

Sous le nom d' **HYPERNEW**, la présente version d'HYPERBASE (version 10.0 de décembre 2014) se substitue à toutes les précédentes quelle que soit la **langue** des données. Il n'y a plus lieu de distinguer HYPERCOR, HYPERTAG, HYPERANG, HYPERGER, etc. Un dialogue initial permet de préciser à l'automate la langue des textes à traiter. On doit aussi lui indiquer le **lemmatiseur** utilisé.

TREETAGGER

Par défaut Hyperbase fait appel à TreeTagger, logiciel libre créé par Helmut Schmidt. Les données initiales doivent se trouver dans des fichiers .TXT

CORDIAL

Si les données sont en français, on a la liberté d'utiliser CORDIAL (logiciel commercial non fourni) avec les options suivantes obligatoires:

- phrase complète : **NON**
- texte intégral : **NON**
- numéroté les phrases : **OUI**
- ligne de titre en début de fichier : **NON**
- numérotation des mots de la phrase: **NON**
- numéroté les paragraphes : **OUI**
- fém->mas: **NON**
- lemmes : **OUI**
- expressions : **AUCUNE**
- type grammatical : **AUCUNE**
- codage spécialisé : **LETTRES**
- fonction grammaticale : **OUI**
- statistiques : **AUCUNE**

Les options marquées en rouge ne sont pas celles que Cordial propose par défaut dans les versions les plus récentes. La lemmatisation de Cordial doit se faire préalablement, Hyperbase exigeant les fichiers lemmatisés (avec le **suffixe .CNR**).

Lemmatisation terminée. L'indexation du texte va commencer en prenant en compte successivement les structures syntaxiques, les codes grammaticaux, les lemmes et enfin les formes. Continuer?

Un prétraitement assure la standardisation des textes avant qu'on les soumette à l'étiquetage, à l'issue duquel est créé le fichier TOUT.CNR. Nul besoin de lemmatiser au préalable les données. Le traitement complet est piloté par HYPERBASE. L'utilisateur n'a guère qu'à cliquer lorsqu'on l'y invite et à souscrire par OK aux options par défaut.

Aide CORDIAL (cliquer)

Prendre patience quand le curseur indique que le travail est en cours. Le progrès des opérations est indiqué dans la colonne centrale. Le cycle complet est repris quatre fois, pour indexer et traiter successivement les séquences syntaxiques, les codes grammaticaux, les lemmes et les graphies. La phase initiale, qui inclut la lemmatisation et l'incorporation des textes, est la plus longue. Prévoir un quart d'heure pour le traitement et jusqu'à une heure pour les gros corpus de plusieurs millions de mots. Ne pas s'éloigner pendant le traitement, le programme sollicite parfois le choix ou le signal de l'utilisateur. Les limites de taille ont été repoussées: jusqu'à 20 millions de mots et 120 textes différents.

V La dernière ligne du présent écran donne en permanence des informations sur le déroulement du programme.

Ci-après la boîte de dialogue affichée lors de la phase d'extraction des passages spécifiques.

Figure I. 13 : Phase d'extraction des passages spécifiques

Chaine des traitements (quatre passages nécessaires : 1 - syntaxe, 2 - codes, 3 - lemmes, 4 - formes)

N° reprise	Temps en %	DESCRIPTION
1	terminé 10%	Contrôle des données
2	terminé 30%	Importation et formatage des textes
3	terminé 1%	Tri et Indexation des textes
4	terminé 0%	Interclassement des index de textes
5	terminé 0%	Interclassement (niveau 2)
6	terminé 10%	Transfert des index dans la base
7	terminé 4%	Structure du vocabulaire
13	variable	Extraction des phrases-clés

CREATION
Cliquez sur le bouton CREATION

Sous le nom d' **HYPERNEW**, la présente version d'HYPERBASE (version 10.0 de décembre 2014) se substitue à toutes les précédentes quelle que soit la **langue** des données. Il n'y a plus lieu de distinguer HYPERCOR, HYPERTAG, HYPERANG, HYPERGER, etc. Un dialogue initial permet de préciser à l'automate la langue des textes à traiter. On doit aussi lui indiquer le **lemmatiseur** utilisé.

TREETAGGER
Par défaut Hyperbase fait appel à TreeTagger, logiciel libre créé par Helmut Schmidt. Les données

Phase d'EXTRACTION des passages spécifiques...
Les paramètres proposés pour la sélection des extraits portent sur le nombre minimum de mots spécifiques contenus dans un extrait et sur le score minimum exigé. Ces paramètres pourront être modifiés ultérieurement

CORDIAL
Si les données sont en français, on a la liberté d'utiliser CORDIAL (logiciel commercial non fourni) avec les options suivantes obligatoires:
- phrase complète : **NON**
- texte intégral : **NON**
- numéroté les phrases : **OUI**
- ligne de titre en début de fichier : **NON**
- numérotation des mots de la phrase: **NON**
- numéroté les paragraphes : **OUI**
- fém->masc: **NON**
- lemmes : **OUI**
- expressions : **AUCUNE**
- type grammatical : **AUCUN**
- codage spécialisé : **LETTRES**
- fonction grammaticale : **OUI**
- statistiques : **AUCUNE**
Les options marquées en rouge ne sont pas celles que Cordial propose par défaut dans les versions les plus récentes. La lemmatisation de Cordial doit se faire préalablement, Hyperbase exigeant les fichiers lemmatisés (avec le **suffixe .CNR**).

Aide CORDIAL (cliquez)

Prendre patience quand le curseur indique que le travail est en cours. Le progrès des opérations est indiqué dans la colonne centrale. Le cycle complet est repris quatre fois, pour indexer et traiter successivement les séquences syntaxiques, les codes grammaticaux, les lemmes et les graphies. La phase initiale, qui inclut la lemmatisation et l'incorporation des textes, est la plus longue. Prévoir un quart d'heure pour le traitement et jusqu'à une heure pour les gros corpus de plusieurs millions de mots. Ne pas s'éloigner pendant le traitement, le programme sollicitant parfois le choix ou le signal de l'utilisateur. Les limites de taille ont été repoussées: jusqu'à 20 millions de mots et 120 textes différents.

V La dernière ligne du présent écran donne en permanence des informations sur le déroulement du programme.

Par la suite le logiciel Hyperbase propose à l'utilisateur de choisir entre deux corpus de référence : Frantext ou Google.

Figure I. 14 : Sélection du corpus de référence

The screenshot shows the HYPERBASE software interface. On the left is a sidebar with a list of processing steps. The main window is titled 'CREATION' and contains instructions for using the software. A dialog box is open in the center, asking the user to select a reference corpus between FRANTEXT (117 millions of words) and GOOGLE (44 billions of words). The 'FRANTEXT' button is highlighted.

Chaine des traitements (quatre passages nécessaires : 1 - syntaxe, 2 - codes, 3 - lemmes, 4 - formes)	N° reprise	Temps en %
Contrôle des données	1	terminé 10%
Importation et formatage des textes	2	terminé 30%
Tri et Indexation des textes	3	terminé 1%
Interclassement des index de textes	4	terminé 0%
Interclassement (niveau 2)	5	terminé 0%
Transfert des index dans la base	6	terminé 10%
Structure du vocabulaire	7	terminé 4%
Spécificités (comparaison interne)	8	
Evolution du vocabulaire	9	
Spécificités (référence externe)	10	
Traitement des noms propres	11	
Extraction des phrases-clés	12	20%
Calcul des distances	13	variable

CREATION

Cliquer sur le bouton CREATION

Sous le nom d'HYPERNEW, la présente version d'HYPERBASE (version 10.0 de décembre 2014) se substitue à toutes les précédentes quelle que soit la langue des données. Il n'y a plus lieu de distinguer HYPERCOR, HYPERTAG, HYPERANG, HYPERGER, etc. Un dialogue initial permet de préciser à l'automate la langue des textes à traiter. On doit aussi lui indiquer le lemmatiseur utilisé.

TREETAGGER

Par défaut Hyperbase fait appel à TreeTagger, logiciel libre créé par Helmut Schmidt. Les données initiales doivent se trouver dans des fichiers .TXT

Deux corpus de référence sont proposés : celui de FRANTEXT (117 millions de mots) et celui de GOOGLE (44 milliards de mots)

FRANTEXT GOOGLE annuler

CORDIAL

Si les données sont en français, on a la liberté d'utiliser CORDIAL (logiciel commercial non fourni) avec les options suivantes obligatoires :

- phrase complète : NON
- texte intégral : NON
- numéroté les phrases : OUI
- ligne de titre en début de fichier : NON
- numérotation des mots de la phrase : NON
- numéroté les paragraphes : OUI
- fém->mas : NON
- lemmes : OUI
- expressions : AUCUNE
- type grammatical : AUCUNE
- codage spécialisé : LETTRES
- fonction grammaticale : OUI
- statistiques : AUCUNE

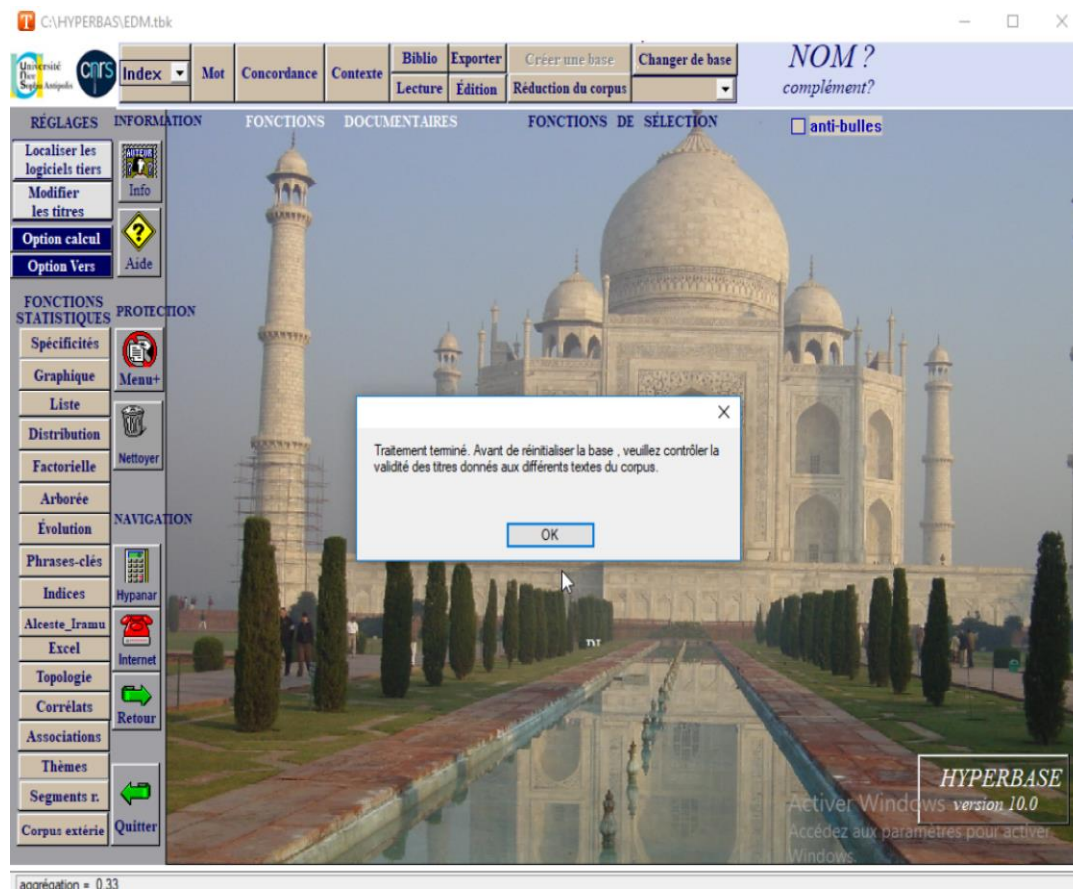
Les options marquées en rouge ne sont pas celles que Cordial propose par défaut dans les versions les plus récentes. La lemmatisation de Cordial doit se faire préalablement, Hyperbase exigeant les fichiers lemmatisés (avec le suffixe .CNR).

Prendre patience quand le curseur indique que le travail est en cours. Le progrès des opérations est indiqué dans la colonne centrale. Le cycle complet est repris quatre fois, pour indexer et traiter successivement les séquences syntaxiques, les codes grammaticaux, les lemmes et les graphies. La phase initiale, qui inclut la lemmatisation et l'incorporation des textes, est la plus longue. Prévoir un quart d'heure pour le traitement et jusqu'à une heure pour les gros corpus de plusieurs millions de mots. Ne pas s'éloigner pendant le traitement, le programme sollicite parfois le choix ou le signal de l'utilisateur. Les limites de taille ont été repoussées : jusqu'à 20 millions de mots et 120 textes différents.

La dernière ligne du présent écran donne en permanence des informations sur le déroulement du programme.

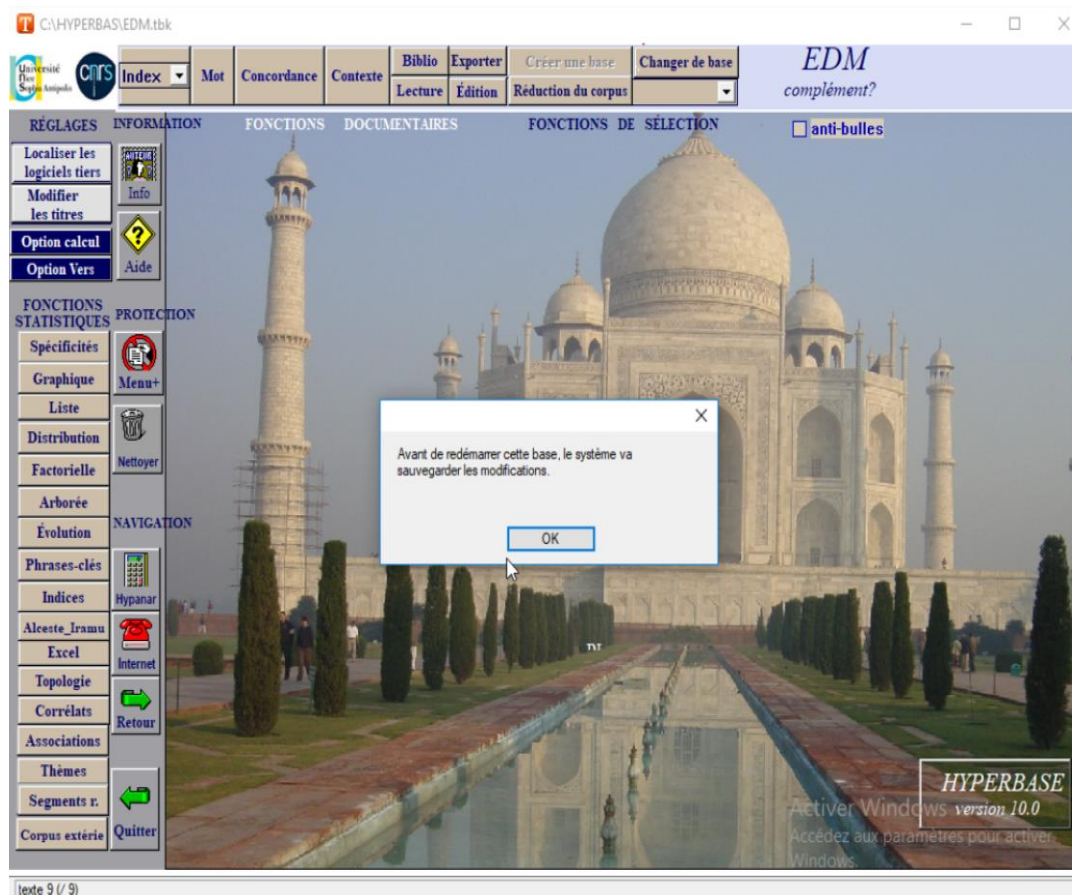
Avant de réinitialiser la base, le logiciel propose de valider les titres donnés aux différents textes du corpus.

Figure I. 15 : Validation des titres des textes



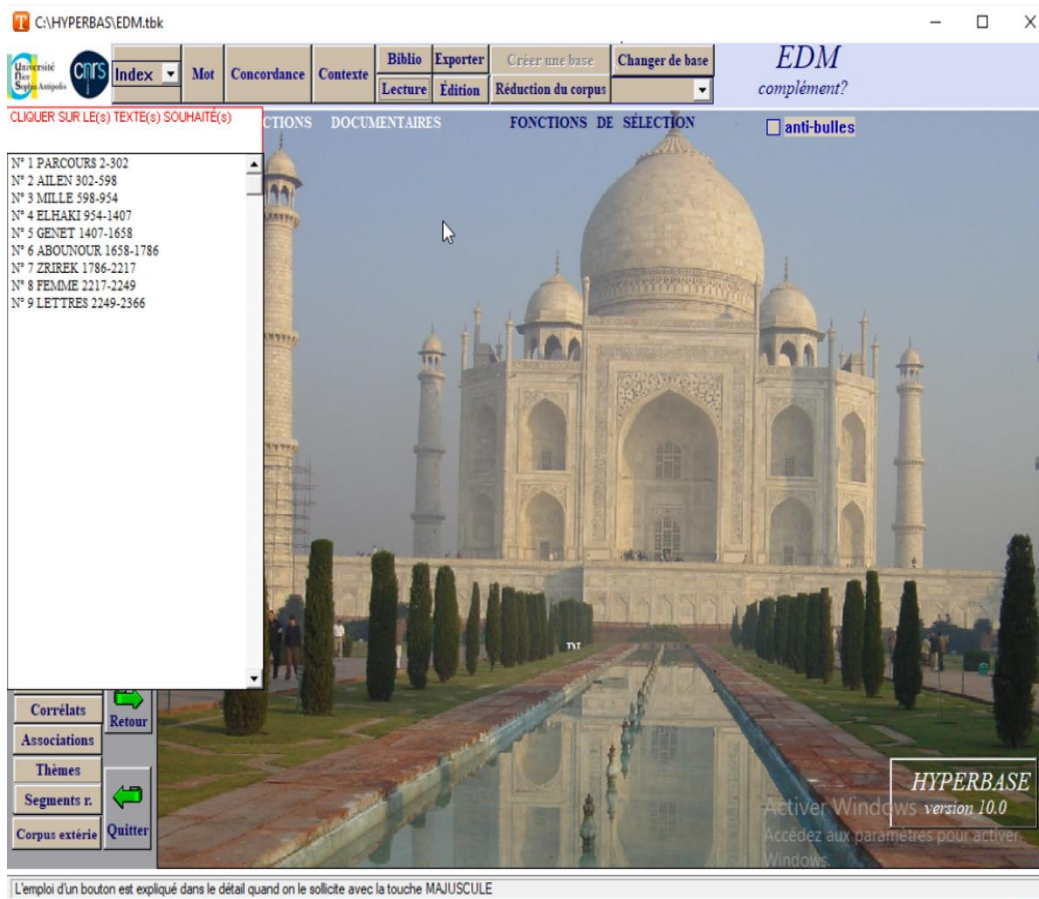
Par la suite, vient l'étape de sauvegarde des modifications.

Figure I. 16 : Étape de sauvegarde des modifications



Et finalement la réinitialisation de la base.

Figure I. 17 : Réinitialisation de la base



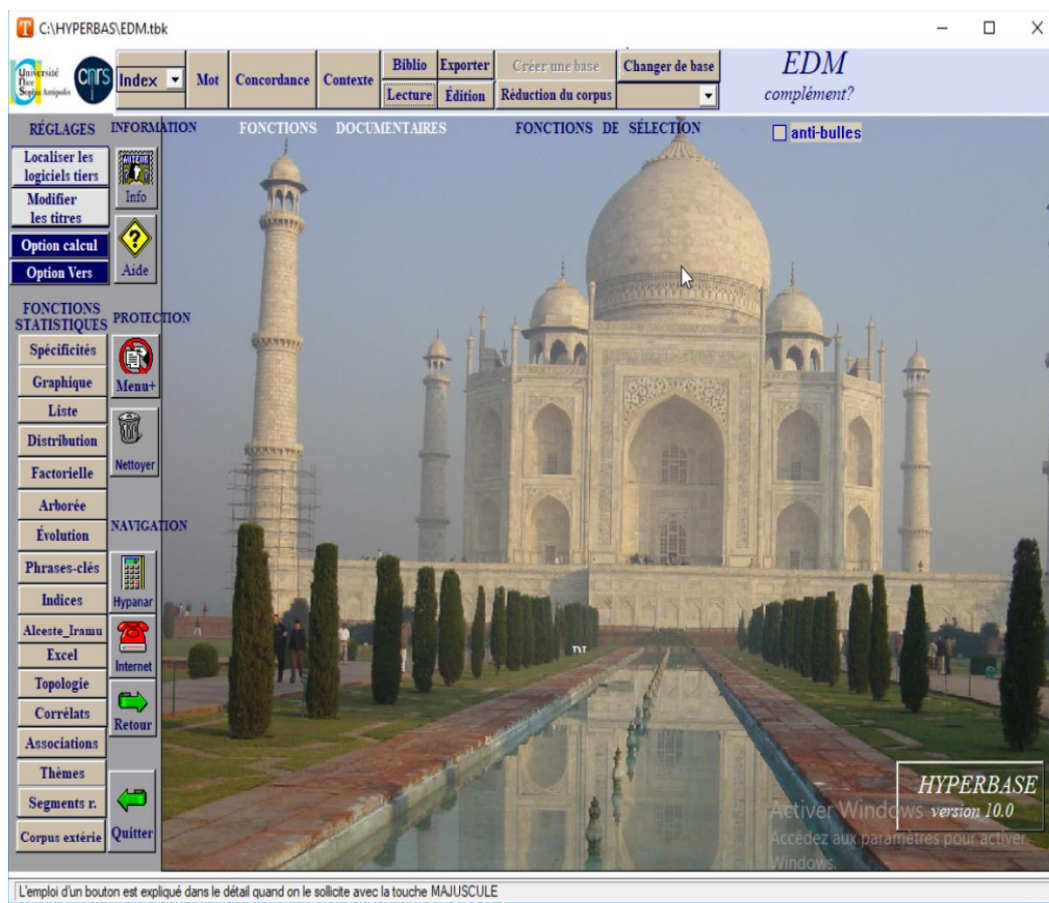
Au terme de ce stade, la base est créée et son exploitation sera explicitée dans ce qui suit.

2. Menu principal

Le menu principal est le carrefour où l'on revient généralement après une opération. Il propose deux sortes de fonctions :

- les fonctions documentaires proposées horizontalement, au haut de l'écran.
- les fonctions statistiques placées verticalement, à gauche de l'écran.

Figure I. 18 : Menu principal



À cela s'ajoutent quelques commandes de contrôle à savoir : les fonctions de navigation, les fonctions de réglage, les fonctions d'information et les fonctions de sélection.

➤ **Les fonctions de navigation :**

Ces fonctions se rapportent aux opérations élémentaires, routinières comme le bouton QUITTER pour sortir, le bouton RETOUR pour retourner à la page que nous venons de quitter. Il y a aussi une dérivation qui conduit du côté d'INTERNET, une autre ouvre une base vide (HYPANAR.EXE) dans lesquels les calculs statistiques peuvent être réalisés à partir de tableaux de nombres, il s'agit

des mêmes fonctions de calcul que nous trouvons dans HYPERBASE sauf qu'elles s'appliquent à des données extérieures, issues ou non d'un corpus.

➤ **Les fonctions de réglage :**

Nous pouvons solliciter le bouton INITIALISER dans la version 9 et celles qui la précèdent ou le bouton MODIFIER LES TITRES dans la dernière version d'Hyperbase (version 10 paru en 2014), pour modifier les titres que nous avons donnés aux textes du corpus au moment de la création de la base.

D'autres réglages, portant sur le choix du mode de calcul, sont permis en appuyant sur le bouton OPTION CALCUL, qui nous offre la possibilité de choisir entre la loi normale ou la loi hypergéométrique.

➤ **Les fonctions d'information :**

Copyright (bouton en forme de timbre- poste) et bouton AIDE.

➤ **Les fonctions de sélection :**

Elles permettent de choisir la base à exploiter, mais offrent aussi la possibilité de créer une nouvelle base avec d'autres données.

Notons qu'il y a une nouvelle fonction qui donne accès non seulement à la base active, mais aussi à toutes celles qui ont été précédemment créées, dès lors que leur nom et leur adresse sont connus. Pour établir ces liens nous devons activer le bouton du haut de l'écran CHOIX DES BASES. Ce bouton permet aussi les ajouts, les suppressions et la remise à zéro.

3. Exploration

On distingue entre deux types d'exploration, libre et documentaire :

➤ **Exploration libre :**

Celle-ci a recours aux méthodes de l'hypertexte, de deux façons symétriques :

La navigation dans le dictionnaire : En sollicitant le bouton MOT, nous sommes renvoyés au dictionnaire des fréquences, à l'endroit où se trouve le mot proposé. Nous pouvons aussi utiliser le menu déroulant INDEX pour choisir la lettre de l'alphabet où le dictionnaire doit être ouvert. Une fois que le dictionnaire est sous les yeux, nous pouvons parcourir la liste alphabétique en nous aidant de l'ascenseur et des flèches gauche et droite qui font défiler les pages. En cliquant sur une forme jugée intéressante, nous repérons tous les passages où cette forme est employée.

Chaque ligne du dictionnaire est consacrée à une forme (ou à un signe) et précise dans l'ordre : la fréquence du mot dans le corpus, la forme elle-même et enfin la liste des sous-fréquences de la forme dans les différentes parties du corpus, séparée du reste par une virgule.

Par exemple pour chercher le mot « aller », nous avons activé la commande MOT et saisi le mot que l'on souhaite trouver. Comme c'est affiché dans la figure ci-après, le logiciel subdivise l'interface en trois parties où il y a la présentation hiérarchique des formes, des lemmes ainsi que les codes et les structures. Nous constatons alors que la fréquence totale du verbe « aller » est de 131, c'est-à-dire le nombre de fois où ce mot est employé dans le corpus. Nous pouvons voir également sa répartition dans les textes formant le corpus selon l'ordre de classement, par exemple :

Figure I. 19 : Mot « aller » dans le dictionnaire

Formes	Lemmes	Codes	Structures
131 aller	N° 1 PARCOURS 18	N° 2 ALEN 18	1 cavsnpn
18 allez	N° 3 MILLE 21	N° 4 ELHAKI 24	1 env
3 alli	N° 5 GENET 13	N° 6 ABOUNOUR 8	2 cavsnpnpn
39 alliance	N° 7 ZIREK 23	N° 8 FEMME 2	1 cavsnpvn
5 alliances	N° 9 LETTRES 4		40 cp
5 alliant	TOUS LES TEXTES		2 cpa
1 allianza			4 cpan
3 allie			1 cpanacrsvpa
4 allions			1 cpanapvp
17 allong			1 cpanndnansn
1 allonge			1 cpannnasn
2 allonger			1 cpav
17 allons			1 cpavrsn
1 allslam			1 cpavrsn
9 allum			1 cpavrsn
1 alluma			1 cpavrsn
4 allumaient			1 cpavrsn
3 allumait			1 cpavrsn
2 allumant			1 cpavrsn
1 allume			1 cpavrsn
2 allument			1 cpavrsn
6 allumer			1 cpavrsn
3 allumette			1 cpavrsn
1 allumettes			1 cpavrsn
23 allure			1 cpavrsn
3 allures			1 cpavrsn
1 allusif			1 cpavrsn
1 allusifs			1 cpavrsn
10 allusion			1 cpavrsn
3 allusions			1 cpavrsn
1 allusives			1 cpavrsn
1 alluvions			1 cpavrsn
1 aln			1 cpavrsn
2 alna			1 cpavrsn
9 almanach			1 cpavrsn
1 almeria			1 cpavrsn
1 almighty			1 cpavrsn
3 almohade			1 cpavrsn
6 alogie			1 cpavrsn
1 alogique			1 cpavrsn
302 alors			1 cpavrsn
2 alourdi			1 cpavrsn
1 alourdie			1 cpavrsn

Il existe une présentation hiérarchique du dictionnaire, par fréquences décroissantes, à laquelle nous avons accès par le bouton INDEX HIÉRARCHIQUE.

Figure I. 20 : Dictionnaire hiérarchique

Les hautes fréquences				Codes			
Formes		Lemmes					
rang	frq mot	rang	frq mot	rang	frq mot		
1	47779 ,	1	47779 ,	1	109583 nom		
2	21563 de	2	39881 le_7	2	59663 prp		
3	14567 la	3	29479 de_9	3	52563 detart		
4	11430 le	4	12659 un_7	4	47773 ,		
5	10405 l'	5	10121 .	5	36284 adj		
6	10219 .	6	9285 du_9	6	31836 proper		
7	8211 à	7	9083 ce_5	7	25313 adv		
8	7963 d'	8	8910 être_1	8	18776 verpres		
9	7275 il	9	8297 à_9	9	16649 con		
10	7274 un	10	7902 il_5	10	15629 verpper		
11	7039 et	11	7006 et_8	11	13407 verimpf		
12	6877 les	12	6241 se_5	12	12556 verinfi		
13	5897 des	13	5610 son_5	13	12231 prpdet		
14	5885 une	14	4965 dans_9	14	10541 prodém		
15	5614 en	15	4808 avoir_1	15	9540 .		
16	4984 dans	16	4614 -	16	8000 prorel		
17	4614 -	17	4467 en_9	17	7719 detpos		
18	3746 est	18	4462 ne_6	18	4612 -		
19	3678 qui	19	3667 qui_5	19	3201 npr		
20	3443 se	20	3535 que_8	20	3156 proind		
21	3441 ce	21	2941 au_5	21	3059 verppre		
22	3403 du	22	2759 le_5	22	2288 :		
23	2872 s'	23	2548 pas_6	23	2269 versubp		
24	2844 que	24	2412 sur_9	24	1888 versimp		
25	2709 pas	25	2323 on_5	25	1885 abr		
26	2704 qu'	26	2290 :	26	1885 !		
27	2636 ne	27	2115 pour_9	27	1783 num		
28	2427 sur	28	2072 par_9	28	1628 «		
29	2384 pour	29	1969 lui_5	29	1327 »		
30	2339 on	30	1885 !	30	1113 vercond		
31	2337 au	31	1869 que_5	31	1099 verfutu		
32	2327 cette	32	1769 je_5	32	755 ?		
33	2290 :	33	1746 elle_5	33	107)		
34	2255 son	34	1735 pouvoir_1	34	107 (		
35	2086 par	35	1658 «	35	80 '		
36	2052 était	36	1619 tout_5	36			
37	1930 lui	37	1507 dire_1	37			
38	1885 !	38	1495 plus_6	38			
39	1849 sa	39	1463 mais_8	39			
40	1831 n'	40	1429 faire_1	40			
41	1682 a	41	1392 vous_5	41			
42	1653 e	42	1378 a	42			

Tableau de l'accroissement du vocabulaire (ordre chrono)

La navigation d'un mot à l'autre, d'une page à l'autre et du dictionnaire au texte, permet d'autres actions comme : l'impression de la page portée à l'écran et la représentation graphique du mot sélectionné.

La navigation dans le texte : La seconde démarche propre à l'hypertexte consiste à feuilleter les pages du texte. Dans cette exploration, nous pouvons suivre la séquence des pages (en suivant les flèches DROITE ou GAUCHE) ou même rejoindre un autre texte du corpus (en sollicitant le bouton TEXTES). Deux boutons dont le symbolisme est clair sont prévus pour l'impression ou l'enregistrement de la page montrée à l'écran.

Quand nous sollicitons le bouton ÉCARTS, au haut de la page, nous obtenons une mise en relief (en rouge) des mots qui, à l'intérieur de la page, reflètent les caractéristiques du texte considéré (par rapport au corpus d'ensemble).

Il existe d'autres fonctions proprement hypertextuelles, qui sont celles de la recherche. Le bouton CHERCHE correspond à une recherche locale sur la page courante et montre en rouge les occurrences du mot proposé, s'il est présent dans la page. Une fonction plus puissante est attachée à chacun des mots de l'écran. Un clic sur l'un d'entre eux permet de montrer sa fréquence et sa répartition dans le corpus de travail. Un second clic à l'intérieur du tableau de répartition amorce la recherche et la visualisation des autres contextes où le même mot se rencontre dans le texte choisi ou dans l'ensemble du corpus, avant et après la page lue.

➤ **L'exploration documentaire**

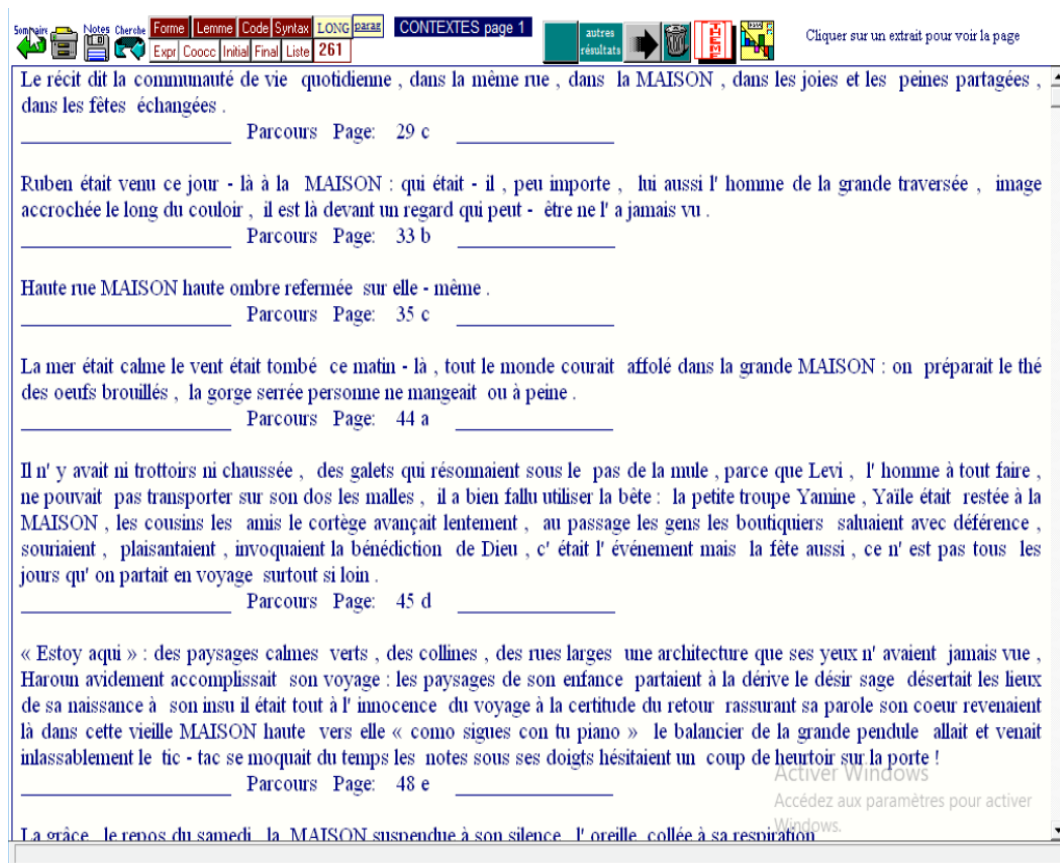
En plus de l'exploration libre du texte et du dictionnaire, Hyperbase offre la possibilité d'une exploitation méthodique de la documentation. À travers deux programmes, CONTEXTE et CONCORDANCE.

La fonction contexte : En sollicitant le bouton CONTEXTE, chaque occurrence de ce que nous cherchons est située et montrée dans le contexte naturel du paragraphe, par exemple, si nous cherchons le mot « maison », une fois nous appuyons sur le bouton CONTEXTE, la page suivante s'ouvrira. Dans cette dernière, nous avons le titre du texte, le numéro de la page et l'indice de la position a,b,c... c'est-à-dire le début, le milieu ou la fin.

Exemple : _____ Parcours, la page 29 c _____

Nous allons trouver le mot « maison » dans le texte *Parcours immobile*, à la fin de la page 29.

Figure I. 21 : Contexte du mot « maison » dans les textes du corpus



Le récit dit la communauté de vie quotidienne , dans la même rue , dans la MAISON , dans les joies et les peines partagées , dans les fêtes échangées .
 _____ Parcours Page: 29 c _____

Ruben était venu ce jour - là à la MAISON : qui était - il , peu importe , lui aussi l' homme de la grande traversée , image accrochée le long du couloir , il est là devant un regard qui peut - être ne l' a jamais vu .
 _____ Parcours Page: 33 b _____

Haute rue MAISON haute ombre refermée sur elle - même .
 _____ Parcours Page: 35 c _____

La mer était calme le vent était tombé ce matin - là , tout le monde courait affolé dans la grande MAISON : on préparait le thé des oeufs brouillés , la gorge serrée personne ne mangeait ou à peine .
 _____ Parcours Page: 44 a _____

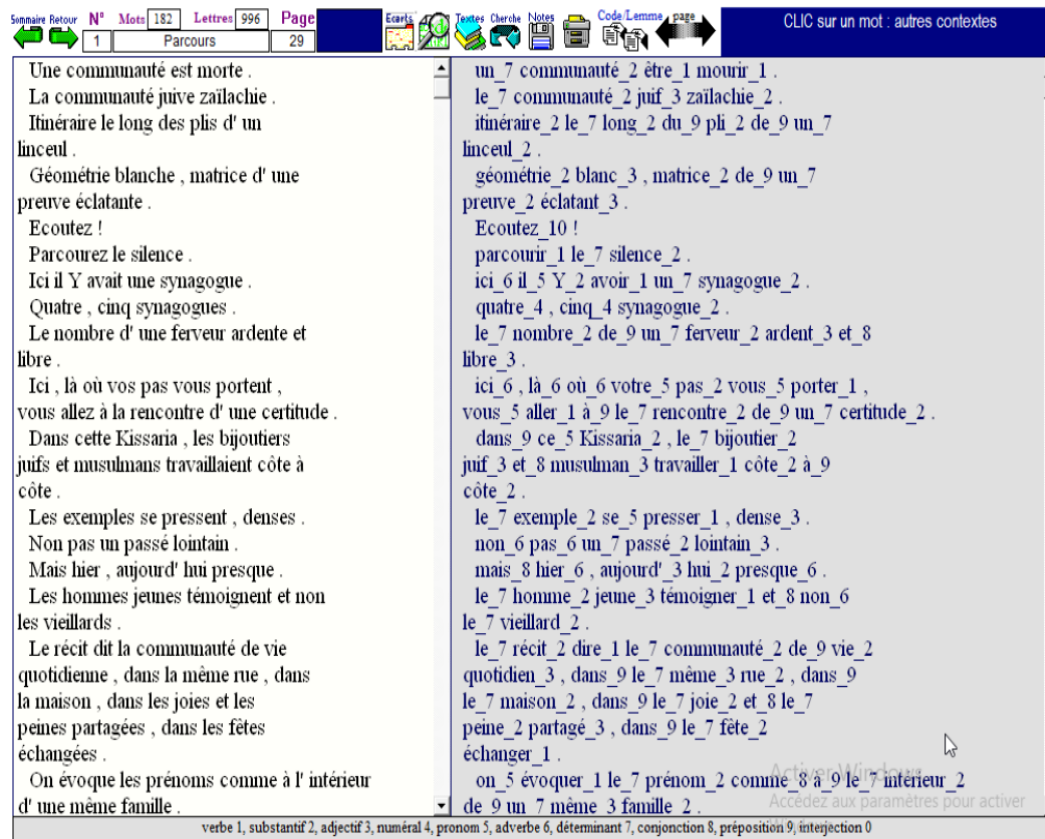
Il n' y avait ni trottoirs ni chaussée , des galets qui résonnaient sous le pas de la mule , parce que Levi , l' homme à tout faire , ne pouvait pas transporter sur son dos les malles , il a bien fallu utiliser la bête : la petite troupe Yamine , Yaïle était restée à la MAISON , les cousins les amis le cortège avançait lentement , au passage les gens les boutiquiers salueaient avec déférence , souriaient , plaisantaient , invoquaient la bénédiction de Dieu , c' était l' événement mais la fête aussi , ce n' est pas tous les jours qu' on partait en voyage surtout si loin .
 _____ Parcours Page: 45 d _____

« Estoy aqui » : des paysages calmes verts , des collines , des rues larges une architecture que ses yeux n' avaient jamais vue , Haroun avidement accomplissait son voyage : les paysages de son enfance partaient à la dérive le désir sage désertait les lieux de sa naissance à son insu il était tout à l' innocence du voyage à la certitude du retour rassurant sa parole son coeur revenaient là dans cette vieille MAISON haute vers elle « como sigues con tu piano » le balancier de la grande pendule allait et venait inlassablement le tic - tac se moquait du temps les notes sous ses doigts hésitaient un coup de heurt sur la porte !
 _____ Parcours Page: 48 e _____

La grâce le renos du samedi la MAISON suspendue à son silence l' oreille collée à sa respiration

Avec un simple clic sur le numéro de la page indiquée, nous sommes renvoyée à l'extrait du texte original à la page demandée ; par exemple, nous avons dans la figure ci-dessous la page 29 du premier texte du corpus « Parcours » où figure le mot « maison ».

Figure I. 22 : Texte d'origine de présence du mot « maison » dans l'ouvrage « Parcours »



La fonction concordance : Si nous faisons appel à la fonction CONCORDANCE du menu principal, nous obtiendrons un contexte étroit qui tient en une ligne et qui montre la forme cherchée, en position centrale, avec une demi-douzaine de mots à gauche et à droite. Juste à côté, il y a une fonction de tri (bouton TRIER), qui est destinée à fournir une autre présentation des contextes. Au lieu de suivre l'ordre normal qui respecte la suite des textes, les contextes sont groupés selon leur environnement immédiat, à gauche ou à droite du mot-pôle. Cela souligne la résurgence de syntagmes répétitifs qui ressortissent souvent aux contraintes syntaxiques mais révèlent parfois aussi les tendances phraséologiques de l'auteur.

Figure I. 23 : Concordance du mot « maison »

CONCORDANCE									
Forme	Lemme	Code	Syntaxe	Expr.	Initial	Final	Chain	Liste	Tout
Nb	261								
PI 29c	nne ,			dans la même rue ,			dans la maison ,		dans les joies et les peines
PI 33b	ben était			venu ce jour - là à la maison :			qui était - il ,		peu importe
PI 35c	evant sous			le blanc .			Haute rue maison		haute ombre refermée sur elle
PI 44a	de courait			affolé dans la grande maison :			on préparait le thé des oeuf		
PI 45d	Yamine ,			Yaile était restée à la maison ,			les cousins les amis le cort		
PI 48e	revenaient			là dans cette vieille maison haute			vers elle « como sigues		
PI 49b	grâce ,			le repos du samedi ,			la maison suspendue à son silence ,		l' or
PI 50b	e est restée			elle - même ,			cette maison n' a pas bougé ,		les fenêtres
PI 69d	l' entrée de			Dar El Mzouka ,			la maison décorée et hantée ,		celle où l
PI 70c	dire s' il			priaient chez lui à la maison ou non ,			en tout cas s' il le		
PI 71d	it rendre			visite ,			emplissait la maison de sa vitalité ,		faisait vibrer
PI 73f	l' asthme de			Josua le soir où la maison			avait failli brûler le rideau		
PI 76e	venaient d' habiter			la nouvelle maison			Dar El Ghali une superbe maiso		
PI 76e	maison Dar El Ghali			une superbe maison			à deux étages à Trab Sini ,		en
PI 80b	t la prière de			l' islam dans une maison			juive cela lui semblait si nat		
PI 87a	était à Safi			dans leur première maison			celle de sa naissance un jour		
PI 90c	entré à Essaouira			dans la grande maison ,			avait admiré les grands boug		
PI 104c	ilence pas			exactement puisque la maison			était très gaie et que les enf		
PI 116b	e sur la mer			tout en haut d' une maison			on y accédait par un escalier		
PI 175a	, dans la			chambre d' une modeste maison			ouvrant sur une cour intérieur		
PI 187c	ille la rue			des Menuisiers cette maison			à balcon où il était né bercé		
PI 188a	deux pas			de la vieille ville une maison			neuve qui sentait encore la ch		
PI 188c	ières sans			frontières dans cette maison			neuve ce visage de femme de je		
PI 211b	' avait			embrassée « Attention la maison			n' est pas loin il peut nous v		
PI 213c	endait			souvent chez Houmrani une maison			en médina maison à la fois ouv		
PI 213c	ez Houmrani			une maison en médina maison			à la fois ouverte par ses brui		
PI 213d	appels			le souffle du rabouz - et maison			fermée sur le secret de la déc		
PI 216a	é seul			dans une chambre de cette maison			où ils avaient fait halte les		
PI 235d	attendais à			la fenêtre de cette maison			près du jardin public dans la		
PI 246d	rrasse du			consulat d' Espagne la maison			d' à côté une nuit on entendit		
PI 249a	age cette			fois on avait remis la maison			à neuf chaulé les murs à l' ex		
PI 249d	espagnole			était mitoyenne de la maison			de grand - mère dans cette rue		
PI 252b	racieux .			Il habitait la grande maison			et avait suffisamment de biens		
PI 252d	vait duré			plus d' une semaine la maison			n' avait pas désempli les pare		
PI 253b	t apposée			sur la porte de chaque maison			où une telle naissance venait		
PI 253b	un signe			funeste ne désignait la maison			où il venait de naître lui le		
PI 260d	e sont			anonymes , sans nom , une maison ,			un cube à un étage dans l' a		
PI 262a	Le			propriétaire de la maison			était un Algérien musulman fra		
PI 262b	on la			rumeur , des biens , cette maison			dont il s' était réservé le pr		

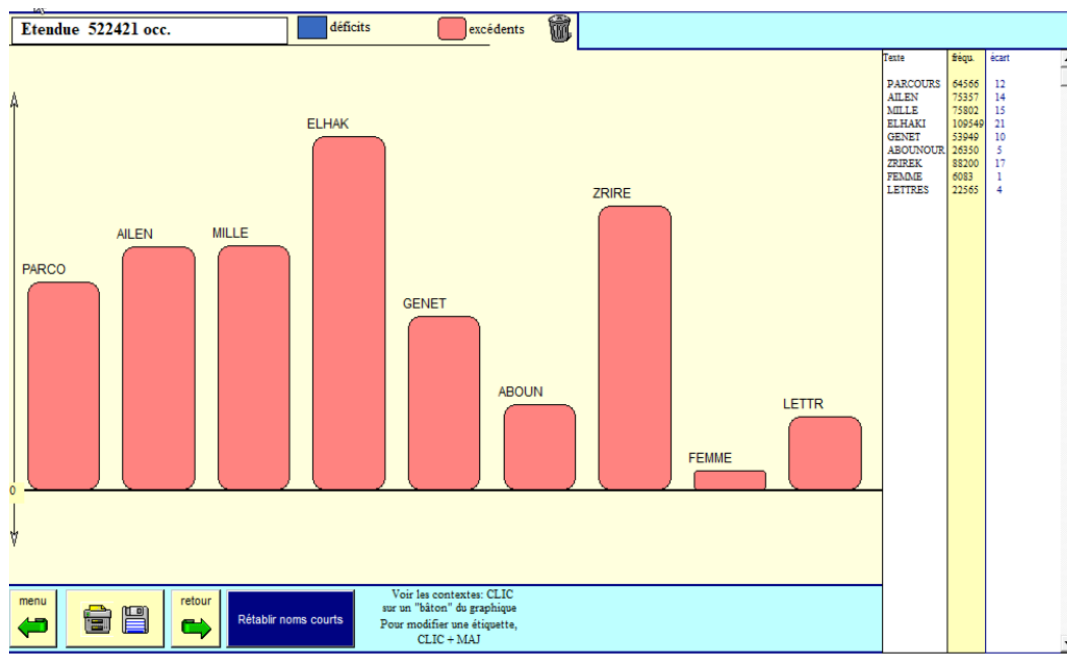
4. Exploitation statistique

Les calculs obéissent aux mêmes principes et s'appuient sur les lois discrètes et continues de la statistique probabiliste, notamment la loi binomiale et la loi normale. Les probabilités p (probabilité de succès) et q (probabilité d'échec) que nous lisons dans le tableau statistique ci-dessous servent à tous les calculs ultérieurs. Quand nous sollicitons le bouton DISTRIBUTION nous pouvons vérifier les calculs des différents paramètres et contrôler les résultats empiriques des tests statistiques, ainsi que l'étendue et les caractéristiques de chaque texte. Sans oublier les boutons ÉTENDUE et PROB, qui nous présentent les résultats sous forme de liste ou sous forme de graphique.

Figure I. 24 : Étendue relative des textes du corpus

Occurrences, vocables, étendue							
Etendue et prob.	N°	TITRE	OCCURRENCES	VOCABLES	Prob P	Prob Q	ABREGE CODE
Richesse et hapax	1	PARCOURS	64566	9467	.1236	.8764	PARCOURS PI
	2	AILEN	75357	10322	.1442	.8558	AILEN AN
	3	MILLE	75802	10109	.1451	.8549	MILLE MA
Acroiss. chrono.	4	ELHAKI	109549	13054	.2097	.7903	ELHAKI AH
	5	GENET	53949	8139	.1033	.8967	GENET JG
Acroiss. inverse	6	ABOUNOUR	26350	5520	.0504	.9496	ABOUNOUR AA
	7	ZRIREK	88200	10622	.1688	.8312	ZRIREK CZ
	8	FEMME	6083	1755	.0116	.9884	FEMME FM
	9	LETTRES	22565	4681	.0432	.9568	LETTRES LM
	TOTAL		522421	28638			
Hautes fréq.							
Distrib. fréq.							
Distance							
EVOL. alphab.							
EVOL. hiérarch.							

Figure I. 25 : Répartition des textes du corpus selon l'étendue relative

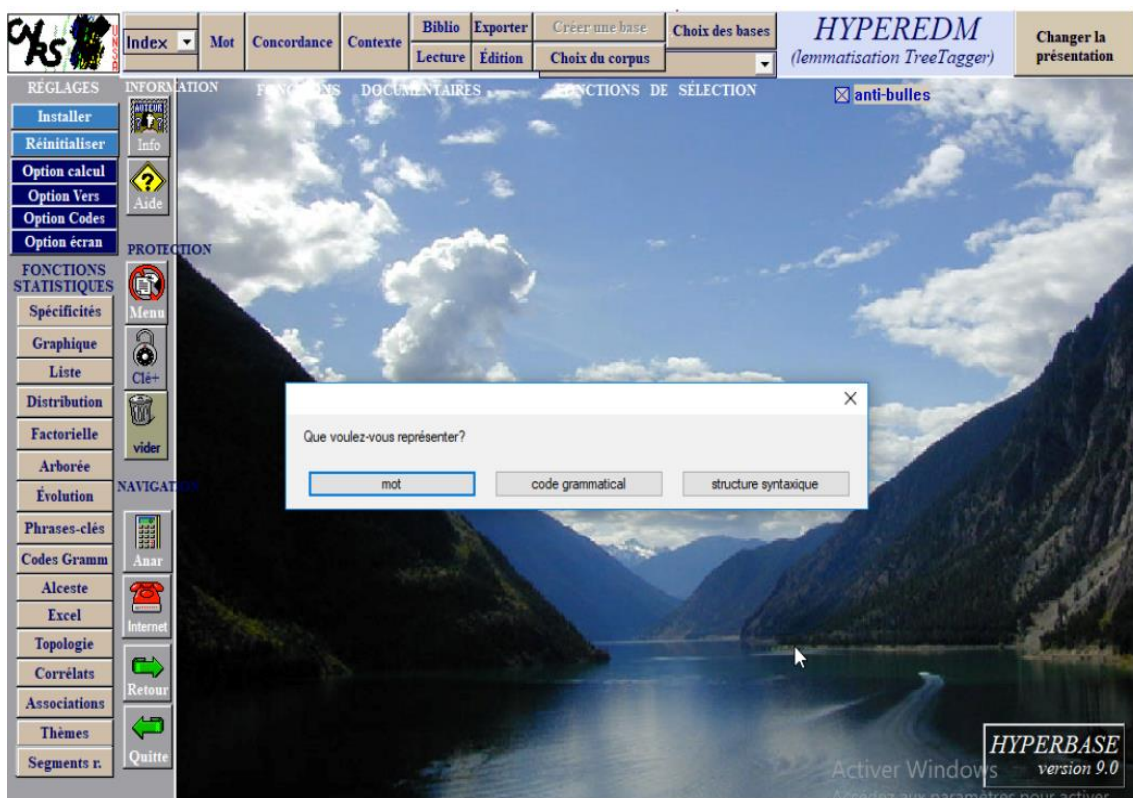


5. Calcul hypergéométrique et écart réduit

Assez souvent, nous sommes confrontée au phénomène d'irrégularité d'apparition d'un mot dans un corpus, ce qui est à l'origine des écarts entre la fréquence observée d'un mot dans un texte et la fréquence théorique que nous attendions, en prenant en considération la proportion du texte dans l'ensemble.

Quand nous cliquons sur le bouton GRAPHIQUE à gauche de l'interface de représentation, une fenêtre apparaît pour demander à l'utilisateur de préciser l'objet de la représentation : un mot (nuit, maison,), un code grammatical (_2, _3...) ou une structure syntaxique (verbe, substantif...)

Figure I. 26 : Sélection de l'objet de la représentation



En choisissant l'objet que l'on cherche à représenter, un graphe apparaît pour traduire l'écart observé entre la fréquence d'un mot dans un texte et sa fréquence théorique attendue. Cet écart est calculé selon la formule de l'écart réduit suivante :

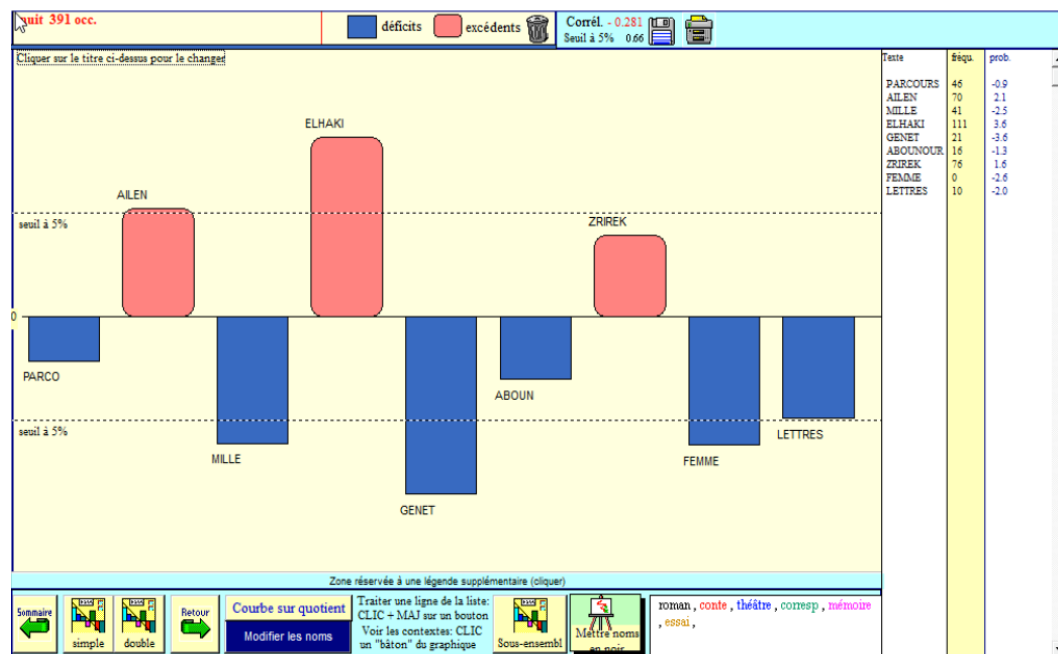
$$z = \frac{\text{fréquence observée} - \text{fréquence théorique}}{\sqrt{\text{fréquence théorique} \times q}}$$

Hyperbase emploie la loi hypergéométrique pour éviter les effets de trompe-l'œil auxquels donne lieu l'écart réduit dans les faibles effectifs. Elle fait automatiquement le choix, du modèle hypergéométrique lorsque la sécurité le recommande et que le corpus est de dimension restreinte. Quand les effectifs sont

suffisamment amples, les calculs convergents et même les graphiques sont superposables.

Nous présentons ci-dessous une illustration de la distribution du mot « nuit » :

Figure I. 27 : Distribution du mot « nuit »



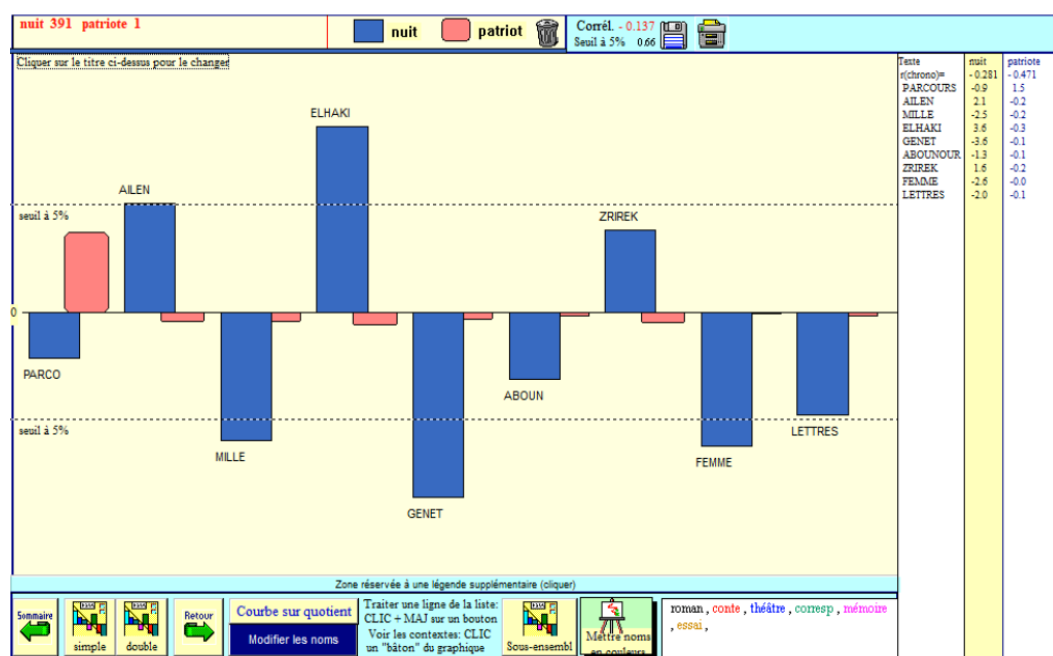
Le diagramme à rectangle présente une ligne médiane qui représente la valeur 0 de l'écart réduit. Les « rectangles » se répartissent de part et d'autre de cette dernière. Les titres des textes sont présentés en haut de chaque « rectangle » correspondant.

A droite, en haut s'affiche l'indice de corrélation qui évalue l'impact de la chronologie sur l'évolution du terme étudié. Cet indice de corrélation, qui varie de -1 à +1, nous renseigne sur le seuil de signification qui est dans notre cas de 95%.

En bas de l'histogramme, nous avons le bouton « double » qui sert à représenter deux séries sur le même graphique.

Exemple : nous avons cherché à avoir l'histogramme du terme « nuit » conjointement avec celui de « patriote ».

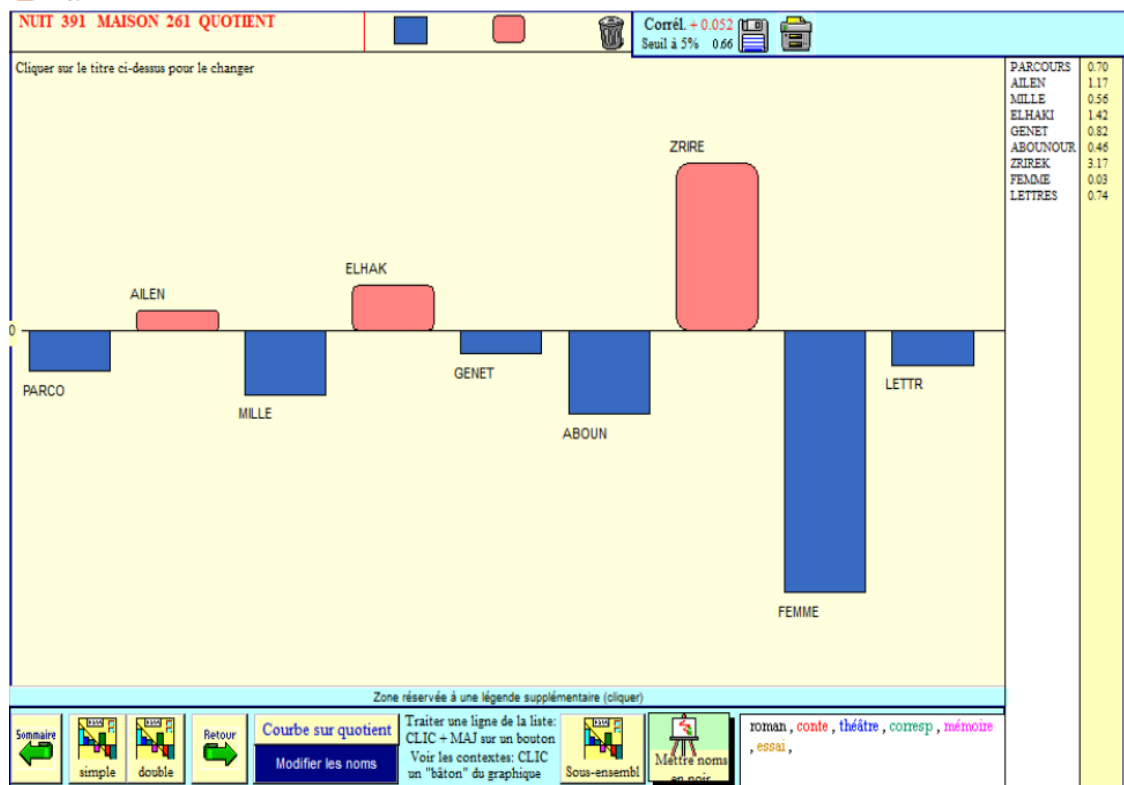
Figure I. 28 : Distribution conjointe des deux substantifs « nuit » et « patriote ».



Les « rectangles » en bleu présentent la première série, « nuit », et les seconds, en rouge, celle de « patriote ». L'indice de corrélation évalué à ce moment la distribution des deux termes en parallèle. Dans cet exemple, le coefficient de corrélation est de - 0,137 au-dessous du seuil requis 0,66, ce qui signifie que ces deux mots sont loin d'être toujours ensemble dans les passages étudiés.

Nous pouvons vouloir vérifier l'écart entre deux formes et non pas le parallélisme, à ce moment-là, nous sollicitons le bouton CONTIENT. Cela est bien clair pour le terme « nuit » et « maison ».

Figure I. 29 : Représentation de l'écart entre les deux occurrences « nuit » et « maison »



6. Corrélation

Le coefficient de corrélation peut être calculé quand les textes du corpus s'échelonnent dans le temps, dans l'espace ou dans quelque succession logique. Ledit coefficient procède par comparaison des valeurs de l'écart réduit au rang de chaque élément et cela pour chaque mot.

Le bouton EVOLUTION du menu principal nous permet d'obtenir tous les mots qui atteignent un seuil approprié au nombre de textes considérés. Ils sont catalogués dans deux pages où les résultats peuvent être lus dans l'ordre alphabétique ou dans l'ordre hiérarchique.

Les figures I.30 et I.40 ci-après montrent l'évolution du lexique dans l'œuvre d'El Maleh par ordre alphabétique et par ordre hiérarchique.

Figure I. 30 : Évolution du lexique par ordre alphabétique

-> hiérar.		L'évolution du lexique (alphabétique)			seuil			Sommaire			Cliquer sur un mot pour voir les contextes (Touche ALT + Clic pour arrêter)					
		Progression	Fréquence	Forme			Régression	Fréquence	Forme							
FORMES	Etendus et prob.	+ 0.892	26	accès			- 0.751	46	allaient							
		+ 0.784	17	advient			- 0.751	25	allié							
	Richesse et hapax	+ 0.845	15	affinités			- 0.841	36	ample							
		+ 0.840	267	ainsi			- 0.798	19	ancestrale							
		+ 0.791	303	alors			- 0.771	611	aux							
	Acroiss. chrono.	+ 0.770	20	apparences			- 0.902	53	aveugle							
		+ 0.774	96	art			- 0.787	16	bizarre							
	Acroiss. inverse	+ 0.911	89	autant			- 0.841	19	casser							
		+ 0.769	239	avant			- 0.822	35	champs							
		+ 0.879	29	ayant			- 0.774	191	chez							
L E M E S		+ 0.780	729	bien			- 0.774	30	clair							
	Hautes fréq.	+ 0.906	82	car			- 0.794	21	confiance							
		+ 0.790	35	cherche			- 0.753	54	coups							
	Distrib. fréq.	+ 0.769	48	culture			- 0.800	16	courant							
		+ 0.756	7963	d'			- 0.815	4984	dans							
		+ 0.787	65	demeure			- 0.795	5897	des							
	Distance	+ 0.751	598	dire			- 0.774	30	devenait							
		+ 0.823	15	disposition			- 0.789	142	eau							
		+ 0.835	288	donc			- 0.761	80	eaux							
		+ 0.899	5614	en			- 0.786	161	enfants							
E V O L U T I O N		+ 0.762	35	essentiel			- 0.775	24	étranges							
		+ 0.762	94	_10			- 0.763	19	affoler_1							
		+ 0.892	26	accès_2			- 0.774	55	aile_2							
		+ 0.843	15	adieu_2			- 0.766	972	aller_1							
		+ 0.863	70	advenir_1			- 0.870	46	ample_3							
		+ 0.887	20	affinité_2			- 0.762	57	angoisse_2							
		+ 0.824	219	ainsi_6			- 0.905	95	appel_2							
		+ 0.863	43	ainsi_8			- 0.787	17	arrêté_2							
		+ 0.859	33	ajouter_1			- 0.765	41	bateau_2							
		+ 0.808	295	alors_6			- 0.759	29	bête_2							
		+ 0.911	89	autant_6			- 0.773	17	bizarre_3							
		+ 0.843	198	avant_9			- 0.756	64	boire_1							
		+ 0.761	702	bien_6			- 0.814	48	casser_1							
		+ 0.917	69	car_8			- 0.774	191	chez_9							
		+ 0.797	58	cesse_2			- 0.789	38	clair_3							
		+ 0.774	37	circonstance_2			- 0.754	344	coeur_2							
		+ 0.858	35	consacrer_1			- 0.794	21	confiance_2							
		+ 0.787	41	constituer_1			- 0.789	21	conquérant_3							
		+ 0.817	20	convenir_1			- 0.815	4965	dans_9							
		+ 0.772	54	culture_2			- 0.779	9285	du_9							
		+ 0.754	40	décrire_1			- 0.860	222	eau_2							
		+ 0.823	15	disposition_2			- 0.752	151	écouter_1							
		+ 0.770	30	dissimuler_1			- 0.799	37	écraser_1							

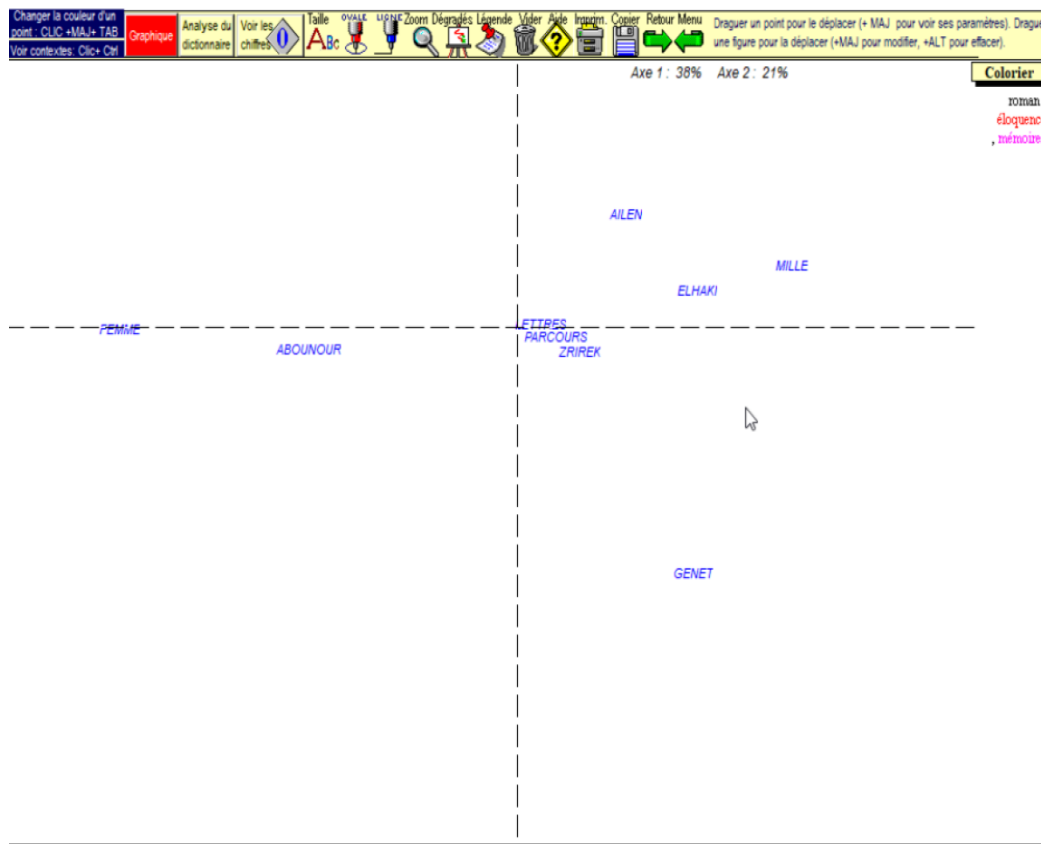
module ici mis en œuvre est celui de l'analyse de correspondance, qui suit l'algorithme proposé par Jean-Paul Benzécri.

Il peut y avoir des inégalités entre les lignes, alors pour les atténuer, le programme calcule les écarts réduits (ou ce qui en tient lieu, quand le modèle hypergéométrique est appliqué), puis les translate dans la zone positive, le plus grand nombre négatif s'alignant sur zéro et les autres éléments gardant leurs distances respectives.

La représentation graphique signale l'emplacement des points doubles où se produit un recouvrement ainsi que la désignation de chaque point occupant quatre lettres au maximum dans le fichier des résultats. Pour faciliter la localisation et l'interprétation, le programme ANCORR fournit les coordonnées de chaque individu (de chaque ligne) et de chaque texte (de chaque colonne), ainsi que la contribution de la variable considérée à chacun des facteurs isolés.

La figure I.32 ci-dessous, nous présente l'analyse factorielle des textes qui font l'objet de notre étude :

Figure I. 32 : Graphe de l'analyse factorielle



L'interprétation du graphe se fait en prenant en considération, d'une part, les distances entre les points, d'autre part, leur distribution par rapport aux deux axes. En effet, le premier facteur d'analyse oppose la droite à la gauche. Le second facteur symbolise l'opposition entre le haut et le bas du graphe. L'interprétation de ces deux facteurs permet de fournir une bonne part de l'information.

8. Richesse lexicale, hapax, accroissement

Le programme de préparation constitue le tableau de distribution des classes de fréquences, le relevé des hapax (ou mots employés une seule fois) et bien d'autres résultats qui intéressent la structure du vocabulaire. En sollicitant le bouton RICHESSE le dénombrement des différentes formes relevées dans chaque texte sera

fait. Et en s'appuyant sur le tableau de distribution des fréquences et sur l'étendue relative des textes, un calcul est exécuté par le programme, qui suit la loi binomiale.

Un calcul plus classique est appliqué aux hapax et conséquemment dans un seul texte. La méthode est ici plus simple et se rattache à la loi normale. Nous aboutissons pareillement à des écarts réduits qui servent d'ordonnées au programme de courbe.

Figure I. 33 : Vocabulaire et Hapax

C:\hyperbas\HYPEREDM.tbk

←

→

Richesse du vocabulaire et hapax

Sommaire

Etendue et prob.

Richesse et hapax

Acroiss. chrono.

Acroiss. inverse

Hautes fréq.

Distrib. fréq.

Distance

EVOL. alphab.

EVOL. hiérarch

n°	réel	théo	écart réduit	Hapax réduit	Titre
1	9467	10226	-759 -7.51	1551 4.84	PARCOURS
2	10322	11192	-870 -8.22	1914 8.08	AILEN
3	10109	11230	-1121 -10.58	1714 2.44	MILLE
4	13054	13798	-744 -6.33	2690 7.99	ELHAKI
5	8139	9182	-1043 -10.88	496 -20.48	GENET
6	5520	5812	-292 -3.83	517 -2.04	ABOUNOUR
7	10622	12242	-1620 -14.64	1548 -8.60	ZRIREK
8	1755	2082	-327 -7.17	140 0.86	FEMME
9	4681	5236	-555 -7.67	617 6.22	LETTRES
Tot	28638			11187	

n°	réel	théo	écart réduit	Hapax réduit	Titre
1	7081	7624	-543 -6.22	998 5.07	PARCOURS
2	7643	8253	-610 -6.71	1241 8.13	AILEN
3	7270	8277	-1007 -11.07	1059 1.71	MILLE
4	8855	9914	-1059 -10.64	1454 -0.12	ELHAKI
5	6288	6940	-652 -7.83	416 -11.91	GENET
6	4331	4631	-300 -4.41	292 -3.22	ABOUNOUR
7	7840	8928	-1088 -11.51	1053 -3.88	ZRIREK
8	1539	1824	-285 -6.67	85 0.45	FEMME
9	3697	4219	-522 -8.04	355 3.21	LETTRES
Tot	19069			6953	

Figure I. 34 : Accroissement du lexique dans l'ordre normal

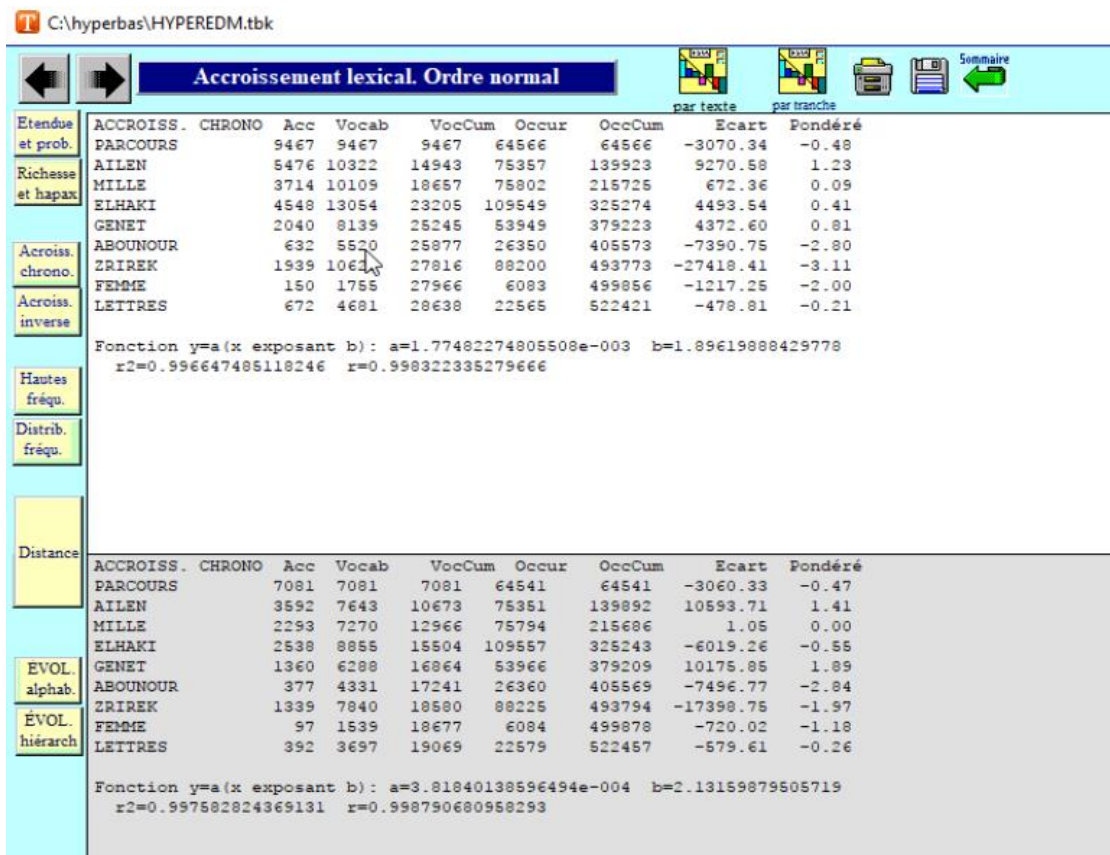


Figure I. 35 : Distributions des fréquences

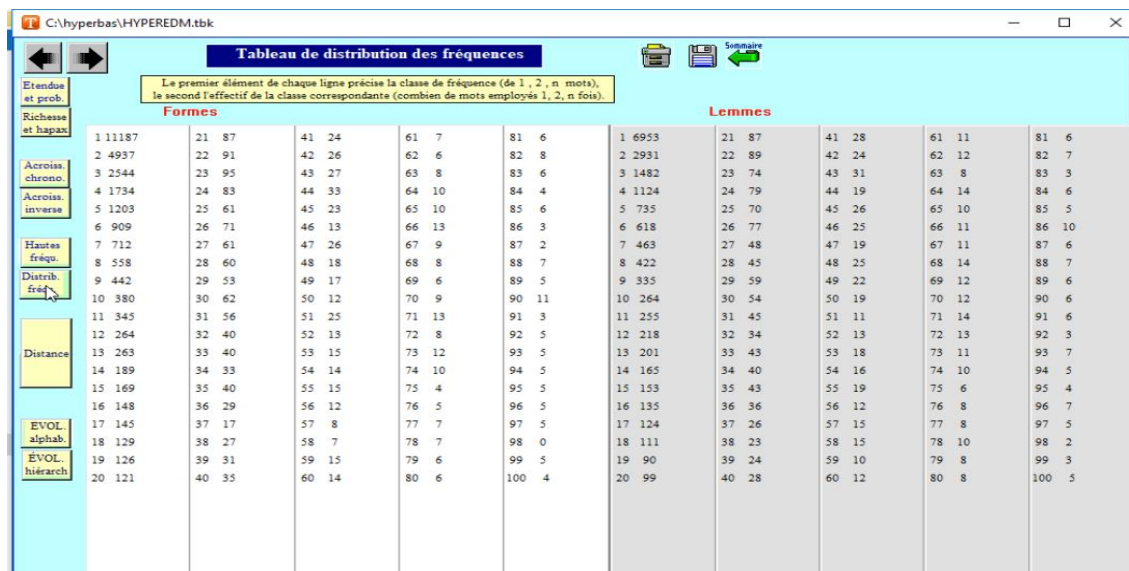


Figure I. 36 : Accroissement inverse du lexique

C:\hyperbas\HYPEREDM.tbk

Accroissement lexical. Ordre inverse								
	ACCROISS. INVERSE	Acc	Vocab	VocCum	Occur	OccCum	Ecart	Pondéré
Etendue et prob.	LETTRES	4681	4681	4681	22565	22565	-1173.97	-0.52
	FEMME	876	1755	5557	6083	28648	1400.36	2.30
Richesse et hapax	ZRIREK	7313	10622	12870	88200	116848	8335.04	0.95
	ABOUNOUR	2052	5520	14922	26350	143198	10676.27	4.05
	GENET	981	8139	15903	53949	197147	-34817.85	-6.45
Acroiss. chrono.	ELHAKI	5450	13054	21353	109549	306696	12857.03	1.17
	MILLE	2776	10109	24129	75802	382498	-3368.88	-0.44
Acroiss. inverse	AILEN	2619	10322	26748	75357	457855	-1035.23	-0.14
	PARCOURS	1890	9467	28638	64566	522421	-7408.46	-1.15
Fonction $y=a(x \text{ exposant } b)$: $a=8.16417849904446e-003$ $b=1.74869943608406$ $x2=0.9969949827566$ $x=0.998496360913048$								
Hautes fréq.								
Distrib. fréq.								
Distance								
	ACCROISS. INVERSE	Acc	Vocab	VocCum	Occur	OccCum	Ecart	Pondéré
	LETTRES	3697	3697	3697	22579	22579	-1328.67	-0.59
	FEMME	634	1539	4331	6084	28663	1571.34	2.58
	ZRIREK	4939	7840	9270	88225	116888	9753.36	1.11
	ABOUNOUR	1265	4331	10535	26360	143248	9459.44	3.59
	GENET	759	6288	11294	53966	197214	-30406.45	-5.63
EVOL. alphab.	ELHAKI	3025	8855	14319	109557	306771	-380.43	-0.03
	MILLE	1739	7270	16058	75794	382565	-2059.83	-0.27
EVOL. hiérarch.	AILEN	1757	7643	17815	75351	457916	7213.87	0.96
	PARCOURS	1254	7081	19069	64541	522457	-680.19	-0.11
Fonction $y=a(x \text{ exposant } b)$: $a=2.46617813790046e-003$ $b=1.94384359381591$ $x2=0.996895622930257$ $x=0.998446604947033$								

9. Connexion lexicale (ou distance intertextuelle)

Pour deux textes dont nous cherchons à apprécier la connexion, un mot contribue à rapprocher ces deux textes s'il en est commun et à augmenter la distance s'il est privatif et ne se rencontre que dans un seul. La collection des données est réalisée dans la phase d'indexation et le résultat auquel nous aboutissons est délivré par le bouton DISTANCE de la page DISTRIBUTION. Le tableau montre cette distance dans sa partie supérieure, les éléments du calcul (parties communes et privatives) étant détaillés dans la suite (du moins lorsque la place est suffisante). Quand les distances lexicales sont visibles à l'écran, le

programme GRAPHIQUE permet la représentation de la distance variable qu'un texte établit avec tous les autres (il y a donc autant de profils que de textes), tandis que le programme ANALYSE envisage l'ensemble de ces distances et propose une typologie des textes selon ce critère.

Le programme pour chaque mot apprécie la distribution réelle des fréquences dans les deux textes en la comparant non plus à la répartition théorique mais à l'écart maximal possible. Le nombre de lignes correspond donc au nombre de mots et les colonnes aux textes distingués dans le corpus. Dans un grand corpus, ces deux dimensions peuvent s'étendre au-delà de la mémoire disponible. Pour y remédier nous irons dans la page « DISTANCE » : et nous appuierons soit sur le bouton Factorielle sur N, soit sur le bouton Analyse arborée (N/V), soit sur le bouton Histogramme (ligne/colonne). Ce calcul ne se fait qu'une fois et ses résultats sont enregistrés dès qu'ils sont acquis (si la base est sur un support inscriptible). Quant à la distance précédemment établie (V), elle est disponible dès que la base est constituée et aucun calcul complémentaire n'est nécessaire, lorsqu'on lance sur de telles données un programme de représentation graphique. Les résultats peuvent être représentés graphiquement, texte par texte, sous forme d'histogramme. Chaque bâton de la figure a une longueur proportionnelle à la distance qui sépare le texte correspondant du texte pris pour pôle d'observation.

Figure I. 37 : Analyse factorielle de la distance lexicale établie sur N

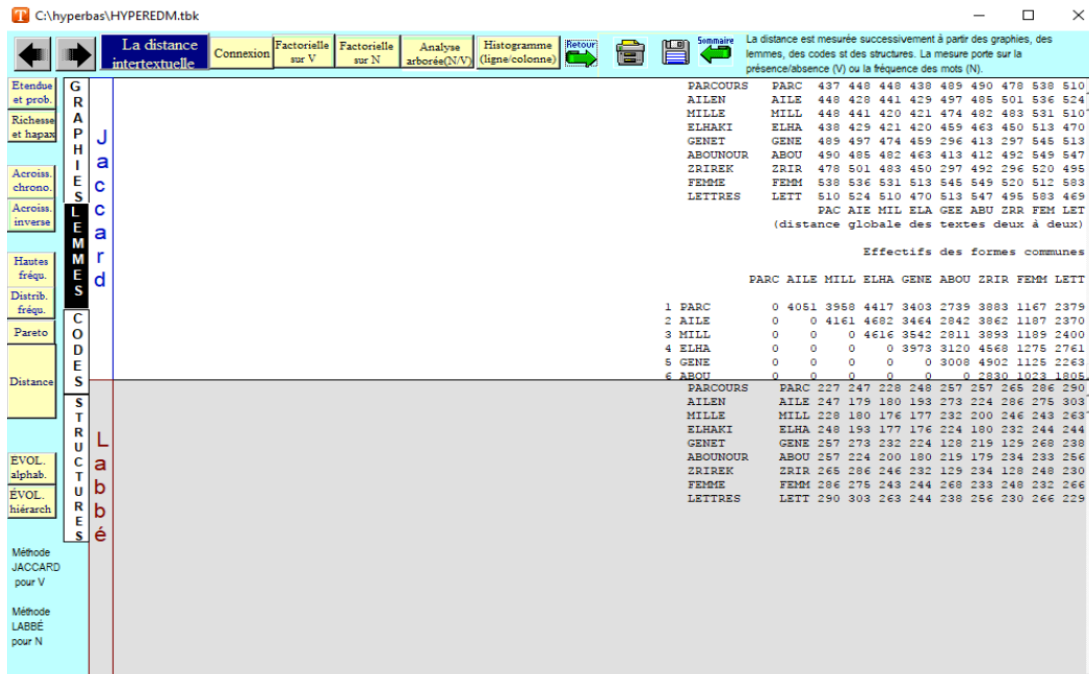
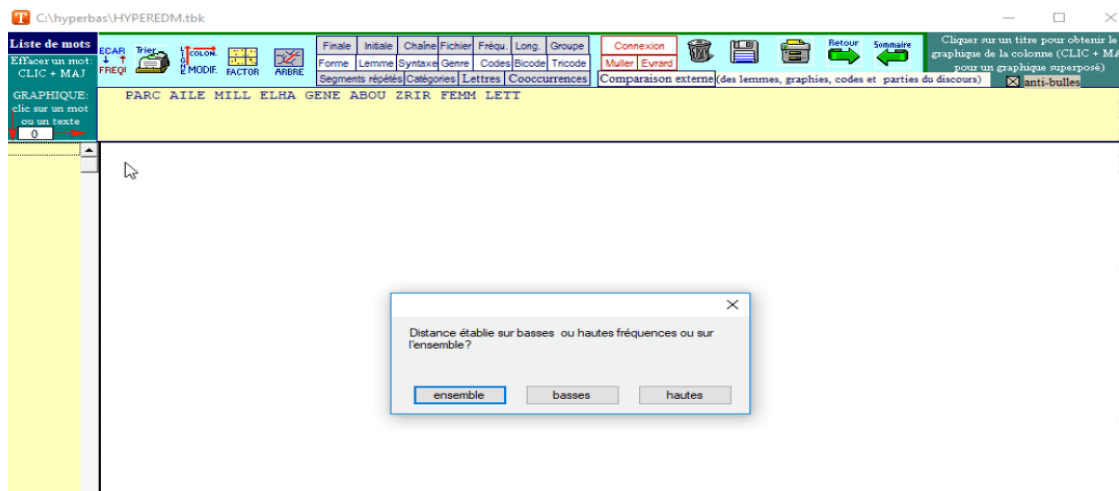


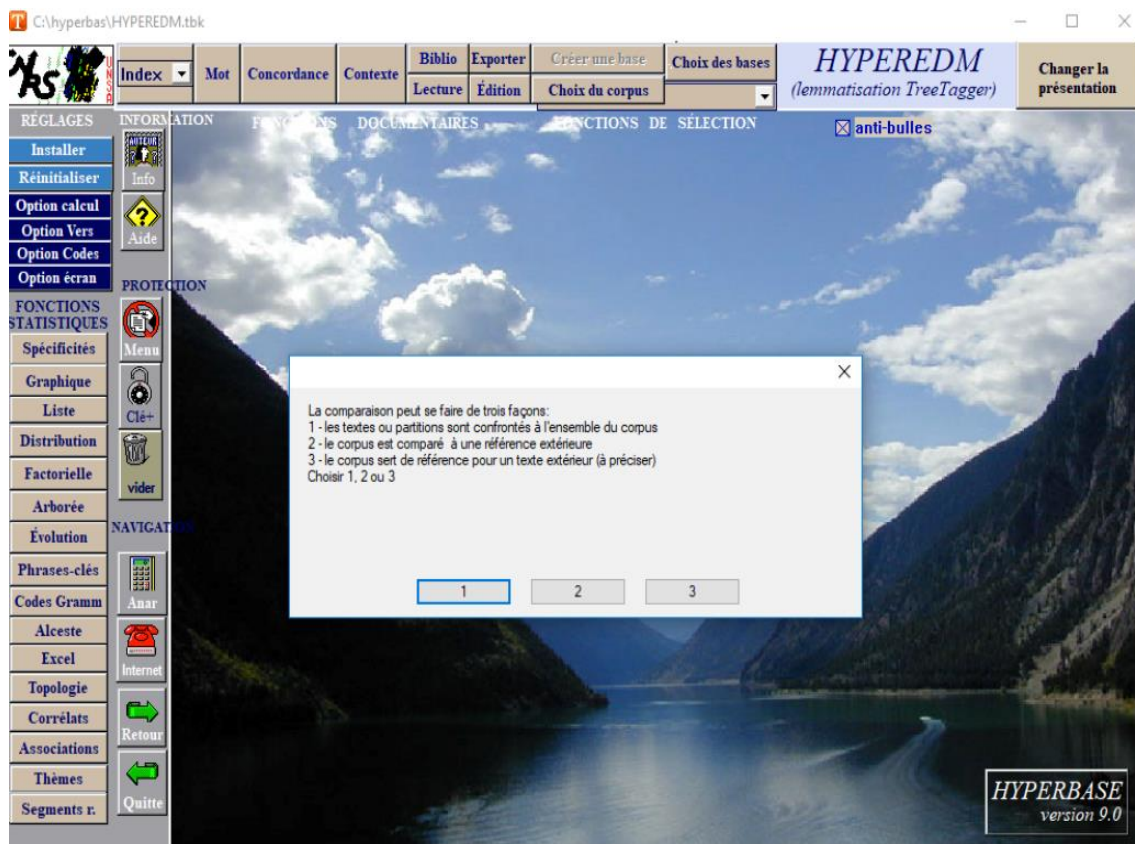
Figure I. 38 : Connexion lexicale sur l'ensemble



10. Spécificités

Le bouton SPÉCIFICITÉS du menu principal invite à choisir entre plusieurs options, selon qu'on envisage une référence interne, externe ou mixte.

Figure I. 39 : Comparaison des textes



La comparaison interne : Les textes sont traités par rapport à l'ensemble du corpus, c'est-à-dire quand le corpus comporte une segmentation en textes ou parties le programme d'indexation calcule le vocabulaire spécifique de chaque texte du corpus en se fondant sur la loi normale et en prenant pour norme l'ensemble du corpus.

Figure I. 40 : Comparaison interne

C:\hyperbas\HYPEREDM.tbk

Refaire résumé PARCOURS (forme) Mots Phrases Codes Syntaxe Cherche Texte Outils Sauvegarde

Choix du texte > Choix du calcul

PARCOURS (lemme) CLIC sur un mot: Recherche du mot dans les spécifications: CLIC+MAJ: Recherche du mot dans les textes

N°	écart	corpus	texte	mot	N°	écart	corpus	texte	mot
1	26.5	283	205	aïssa	1	20.6	136	113	part_2
1	19.8	2052	556	était	1	17.3	135	97	aïssa_1
1	19.8	158	118	parti	1	12.9	84	59	aïssa_10
1	18.0	92	81	jossua	1	12.7	1328	324	>
1	12.2	1351	322	avait	1	12.3	64	48	aïssa_2
1	11.4	59	44	communiste	1	12.1	68	50	camarade_2
1	10.5	44	35	camarades	1	11.1	49	39	communiste_2
1	10.5	26	26	houmran	1	11.0	35	32	jossua_1
1	10.3	1637	345	>	1	10.3	25	25	houmran_3
1	10.1	31	28	haroun	1	10.1	31	28	jossua_10
1	10.0	1659	345	<	1	10.0	1658	345	<
1	9.8	23	23	jacquet	1	9.6	64	40	militant_3
1	9.0	50	33	casablanca	1	9.6	22	22	jacquet_3
1	8.7	23	21	staline	1	9.0	50	33	casablanca_2
1	7.8	5897	925	des	1	8.7	23	21	staline_2
1	7.7	30	22	groupe	1	8.4	9285	1410	du_9
1	7.7	23	19	moscou	1	8.2	39881	5425	le_7
1	7.7	23	19	communistes	1	8.1	21	19	haroun_3
1	7.2	258	75	était	1	8.0	33	24	communiste_3
1	6.9	141	49	famille	1	7.9	45	28	policier_3
1	6.8	71	32	ligne	1	7.7	25	20	jossua_3
1	6.8	357	91	histoire	1	7.7	23	19	moscou_2
1	6.8	28	19	camarade	1	7.6	19	17	responsabilité_2
1	6.7	1639	296	comme	1	7.4	214	67	français_3
1	6.4	22	16	militants	1	7.2	383	99	histoire_2
1	6.4	13	12	responsabilités	1	7.1	178	58	famille_2
1	6.3	23	16	militante	1	6.8	12	12	hram_10
1	6.2	307	78	allait	1	6.6	124	44	politique_3
1	6.2	16	13	biographie	1	6.5	39	22	groupe_2
1	6.2	134	44	julf	1	6.5	17	14	part_3
1	6.1	136	44	politique	1	6.4	37	21	fic_2
1	6.0	87	33	m	1	6.4	11	11	ligne_3
1	5.9	99	35	juste	1	6.3	8910	1296	être_1
1	5.9	20	14	heureusement	1	6.2	29	18	responsable_3
1	5.9	17	13	hram	1	6.2	269	71	julf_3
1	5.9	17	13	clandestinité	1	6.2	18	14	biographie_2
1	5.8	66	27	immobile	1	6.1	770	155	grand_3
1	5.8	86	31	simple	1	6.1	69	29	immobile_3
1	5.6	277	68	savait	1	6.1	127	42	peur_2
1	5.6	19	13	secrétaire	1	5.9	99	35	juste_3

La comparaison externe : Elle consiste à faire une comparaison entre le corpus et le TLF. La spécificité du corpus est détaillée par rapport au dictionnaire de référence choisi. S'il s'agit du français, le choix de la norme peut être précisé : telle ou telle période du TLF, du 16^{ème} siècle à l'époque contemporaine. Nous avons le choix entre une présentation alphabétique ou hiérarchique.

La comparaison avec l'usage observé dans le Trésor de la langue française doit être interprétée prudemment. D'une part le TLF reflète l'usage littéraire de la langue dans un registre relevé, si jamais le corpus qu'on traite se trouve éloigné de ce niveau de langue, la valeur de la comparaison en est amoindrie. D'autre part

toutes les formes n'ont pas été soumises à la comparaison, parce que le calcul de l'écart réduit perd de sa légitimité quand la fréquence théorique est trop faible.

La comparaison mixte : Elle consiste à comparer un texte extérieur au corpus déjà constitué. Pour ce faire nous devons indexer le texte extérieur en le réduisant à une liste de fréquences. En comparant cette liste à celle du corpus, nous obtenons le résultat souhaité.

Cependant, pour analyser des textes du corpus et des textes extérieurs au corpus, nous devons d'abord exporter les textes à partir de la base et solliciter le bouton EXPORTER du menu principal, en choisissant l'option un par un qui reconstitue autant de fichiers qu'il y a de textes, sous les noms tex1.txt, tex2.txt, tex3.txt, etc. Par défaut le nom des fichiers traités apparaît dans le champ du titre.

Dans ce troisième chapitre, nous nous sommes penchée sur la présentation du corpus ainsi qu'à la préparation de la base de données. Il s'agit bien évidemment d'une étape primordiale pour l'exploitation des œuvres d'Edmond Amran El Maleh via un logiciel hypertextuel. Ainsi, la lemmatisation des textes a montré que le corpus d'analyse totalise 522421 occurrences et 28638 vocables.

Ensuite nous nous sommes focalisée sur la présentation du logiciel de travail Hyperbase. En nous basant sur un manuel de référence élaboré par le concepteur dudit logiciel, nous avons mis en évidence les fonctionnalités que nous jugeons importantes dans la présente étude, telles : l'exploration, l'exploitation statistique, le calcul hypergéométrique, la corrélation, la richesse lexicale, la connexion lexicale, etc. Nous avons montré, toutefois, que ce produit se distingue des autres logiciels existant sur le marché par une orientation statistique. En effet, il permet une comparaison interne s'intéressant aux différentes partitions du corpus et, une autre externe pouvant être réalisée avec le Trésor de la langue française, ce qui génère des listes de spécificités et des représentations graphiques diverses : des courbes, des analyses arborées, des analyses factorielles, etc.

La présente partie a pour but l'encadrement théorique de notre travail de recherche qui vise l'analyse quantitative et sémantique du vocabulaire d'Edmond Amran El Maleh. L'étude que nous envisageons effectuer est le produit croisé de trois composantes : un matériau, une méthode et un outil. Ce qui a donné naissance à trois chapitres.

Le premier qui s'intéresse au matériau, à savoir les œuvres d'Edmond Amran El Maleh, a commencé par une biographie de l'auteur. Celle-ci nous a permis de postuler que son identité aux racines judéo-arabes et berbères occupe une place centrale dans sa vie. Nous avons aussi découvert que l'auteur avait un intérêt tout particulier pour la peinture. D'ailleurs, il a déjà derrière lui une longue carrière de critique d'art et de nombreux articles abordant les réalisations artistiques nationales, ce qui nous a poussé à présenter sa conception personnelle de la pratique picturale marocaine.

En abordant son œuvre littéraire qui est marquée par la « désintégration textuelle » (absence de linéarité chronologie, une ponctuation essoufflée ou encore absence de ponctuation dans le cas de *Parcours Immobile*, une narration fragmentée traversée par des images et des souvenirs venant des tréfonds de la mémoire) et la « co-spacialisation » (projection du lecteur dans les confins d'une géographie multiple et diversifiée, enchevêtrement des temps et des instants), nous avons montré qu'il essaye d'apporter des éclaircissements sur la question juive, afin qu'il n'y ait pas d'amalgame entre sionisme et judaïsme tout en condamnant l'asservissement politique des consciences et des passions. Il condamne aussi le déracinement corollaire à la création de l'Etat d'Israël. Ce déracinement concerne toutes consciences qui, à un titre ou à un autre l'ont subi, nous faisons allusion ici aux palestiniens qui ont été chassés de leur terre natale, mais aussi aux juifs du

monde qui ont été convoqués et manipulés par les mystificateurs du sionisme pour délaissé leur pays d'origine.

S'agissant du deuxième chapitre, il se focalise sur la méthode de travail, à savoir la statistique lexicale. Nous avons dévoilé que cette dénomination est le fruit d'une hybridation entre les statistiques et la lexicologie, mais qui a été favorisée largement par le développement de l'outil informatique. Ainsi, nous avons fait un tour d'horizon sur ses origines et ses fondements théoriques tout en abordant quelques notions de base, comme : les paramètres de position (la médiane, le mode, etc), les paramètres de dispersion (la variance, l'écart-type, l'étendue), les tests paramétriques, etc. Notons, que nous avons abordé également, au niveau de ce chapitre, les principes de dépouillement et de quantification du corpus.

Une fois le matériau et la méthode clarifiés, nous nous sommes penchée, au niveau du troisième chapitre, sur l'outil qui est constitué par un logiciel destiné au traitement hypertextuel et statistiques des textes littéraires (Hyperbase). Ainsi, ledit chapitre a commencé par la préparation de la base de données « HYPEREDM ». Il s'agit bien évidemment d'une étape indispensable pour l'exploitation statistiques des œuvres d'Edmond Amran El Maleh via un logiciel hypertextuel. Ainsi, la lemmatisation des textes a montré que le corpus d'analyse totalise 2039789 occurrences et 522421 vocables. Soulignons que la préparation du corpus numérisé était l'une des phases les plus délicates de notre travail, du moment que le scannage des livres était marqué par d'énormes erreurs typographiques. Le seul moyen d'y remédier était de relire sur écran tous les livres tout en les comparant mot à mot avec ceux que nous avons sur format papier. Ce chapitre a pris fin par une démonstration des performances du logiciel quant au

traitement automatique des données : exploration, exploitation statistique, calcul hypergéométrique, richesse lexicale, connexion lexicale, etc.

Tous ce que nous avons exposé au niveau de cette première partie, nous servira de base pour vérifier la convergence des résultats que nous obtiendrions au niveau de la structure du vocabulaire ainsi qu'au niveau du contenu enfoui dans le symbolisme des signifiants. Ce sera donc l'objet des deux parties suivantes.

DEUXIÈME PARTIE

Structure du lexique

Dans la première partie, nous avons souligné que la volonté de l'auteur à transmettre ses idées et ses préoccupations se heurtait à la langue de l'autre. En effet, le défi primordial qu'il cherchait à affronter et qu'il finit par gagner est comment transmettre l'authenticité des événements vécus par toute une nation à travers la langue de l'Autre ? Comment les mots, qui ne sont pas innocents, peuvent-ils faire revivre une mémoire collective vouée à la déliquescence et à l'oubli ? Comment faire en sorte que les valeurs musulmanes, judaïques et arabes parlent au sein d'un texte écrit dans la langue du colonisateur ? Ceux-ci sont autant de questions que nous nous sommes posées au niveau de l'introduction générale de ce travail de recherche.

Dans la présente partie, nous essayerons de trouver une réponse à ce « comment » du point de vue de la **structure**. D'ailleurs, cette investigation nous permettra de dégager objectivement ce qui échappe à l'analyse intuitive. En effet, si l'étude du contenu lexical permet de pénétrer au cœur des textes envisagés d'un point de vue sémantique, stylistique et thématique, la structure du vocabulaire, fondée sur l'examen de plusieurs paramètres structurels, permettra de se faire une idée sur la structure formelle de l'œuvre en dégagant des propriétés linguistiques intéressantes, indépendamment du contenu lexical.

De ce fait, cette analyse s'attardera sur la longueur des mots, la richesse lexicale, l'accroissement du vocabulaire et les parties du discours. Elle nous offrira la possibilité de comprendre la structure formelle des différentes partitions du corpus dans le but de comparer, au niveau endogène, les textes entre eux et, au niveau exogène, les diverses segmentations du corpus avec une ou plusieurs références extérieures. En bref, dans la présente partie, nous essayerons de savoir comment Edmond Amran El Maleh construit son œuvre.

CHAPITRE PREMIER

Étendue du corpus

Le présent chapitre jettera la lumière sur deux aspects fondamentaux de notre analyse, à savoir la longueur du mot et l'étendue du corpus.

Nous nous proposons, dans un premier temps, de mesurer la longueur des mots. Cette longueur, estimée par le nombre de lettres, donnera naissance à trois classes de longueurs : mots longs, moyens et courts. Le logiciel Hyperbase mettra à notre disposition toutes les données nécessaires pour réaliser les calculs requis. Il s'agira ensuite d'étudier les classes de longueurs tout en faisant le lien avec leurs distributions pour, enfin, voir si cette répartition est le résultat d'un simple jeu du hasard ou le produit d'un effet du style, du genre, du temps, de la thématique abordée, etc.

Dans un deuxième temps, nous examinerons l'étendue du corpus, où nous établirons la relation entre le nombre d'occurrences N et le nombre de vocables V . A ce stade également, *Hyperbase* nous sera d'une grande utilité en nous proposant plusieurs options de calculs et de traitements statistiques.

A. Longueur du mot

Ch. MULLER¹³³ indique dans l'un de ses travaux que le calcul de la longueur des mots est considéré comme un outil permettant, d'une part, d'étudier

¹³³ Muller CHARLES, *Essai de Statistique Lexicale, l'illusion Comique de Pierre Corneille*, Paris, Librairie

l'homogénéité entre un ensemble de textes d'un même auteur et, d'autre part, de mener une comparaison entre les productions d'auteurs différents.

Aussi, Etienne Brunet¹³⁴ a-t-il signalé dans son étude entreprise sur le *vocabulaire français de 1789 à nos jours* que : « le mot est plus court en poésie (4,2788 en prose poétique, 4,2502 en vers), plus long dans la prose littéraire (4,3535), et plus long encore dans le corpus technique (4,6486) ». Il stipule ensuite qu' :

« à l'intérieur de chacun des genres, la distribution subit l'influence de la chronologie : dans le roman le mot s'allonge progressivement pour décroître dans la dernière période, sans doute soumise à la contamination du théâtre. Mais dans le théâtre la chronologie voit ses effets contrariés ou tout au moins brouillés par le mélange des pièces. Celles dont l'inspiration est tragique et/ou ontique se distinguent de celle dont l'argument est moderne et/ou comique »¹³⁵

De ce fait, la longueur des mots pourrait être influencée par plusieurs facteurs, entre autres nous trouvons le genre, la période de production, le thème, etc. Alors qu'en est-il de l'œuvre d'El Maleh ? Afin de répondre à cette interrogation, nous nous intéresserons, dans les points suivants, à mesurer la longueur des mots afin d'étudier la corrélation entre cette variable et les textes formant notre corpus. Pour ce faire, nous aurons recours au logiciel *Hyperbase* qui nous fournira le nombre de caractères dans le corpus dans son intégralité mais aussi dans chaque texte pris isolément.

Klincksieck, 1964, p. 15

¹³⁴ Etienne BRUNET, *Le vocabulaire français de 1789 à nos jours, d'après les données du Trésor de la langue française*, Genève-Paris, Slatkine-Champion, 1981, Volume I, p.239.

¹³⁵ Ibid., pp. 163-165

L'étude de la longueur des mots sera complétée par l'examen de la distribution des groupes de longueurs dans l'œuvre en question et prendra fin par l'analyse factorielle des correspondances.

1. Mesure de la longueur moyenne

La longueur moyenne des mots s'obtient en divisant le nombre de caractères dans chacun des textes par le nombre de mots du même texte. Nous commencerons, tout d'abord, par le rassemblement des données nécessaires à cette opération. Ensuite, nous passerons aux calculs proprement dits.

1.1. Calcul du nombre de lettres dans le corpus

Le calcul du nombre de lettres par Hyperbase nous est présenté en cliquant sur le bouton « lettres » situé dans la fonction « Liste ». Ainsi, nous déboucherons sur les résultats étalés dans le tableau suivant :

Tableau II. 1 : Répartition des lettres dans le corpus

Œuvre/	PARC	AIL	MILL	ELHA	GEN	ABOU	ZRIR	FEM	LETT	TOTAL
a	23948	24325	24773	34026	15748	8117	25924	1834	6647	165342
b	2688	3250	3082	4414	2020	1151	3214	207	768	20794
c	9229	9310	9085	13107	6997	3359	11871	819	2997	66774
d	10244	11077	11332	16159	7865	3970	13336	905	3284	78172
e	38483	41308	42921	61187	31380	15273	52483	3466	13239	299740
f	2905	3151	3177	4817	2137	1045	3519	224	930	21905
g	2886	3209	3244	4598	2217	1170	3727	251	879	22181
h	2697	3164	3152	4298	1812	1025	2708	241	703	19800
i	20993	21734	22340	30867	14320	7405	25139	1853	6600	151251
j	1063	1096	1139	1794	978	399	1402	82	428	8381
k	186	294	247	238	126	78	232	13	22	1436
l	16362	18111	17907	24511	12450	5841	20426	1435	4669	121712
m	8386	8285	8685	11801	5955	2907	9733	740	2553	59045
n	18936	19290	20482	30530	14783	7244	25347	1777	6628	145017
o	13931	14350	14668	21523	10790	5139	17960	1233	4892	104486
p	7193	7169	7283	10885	5541	2691	9065	589	2428	52844
q	2513	2234	2502	3562	2137	925	3818	251	990	18932
r	17918	19009	19315	27903	14260	7111	23688	1542	5965	136711
s	21662	23975	24337	34989	15673	8282	25484	1671	6914	162987
t	18999	18851	20620	28902	14509	7071	24970	1723	6363	142008
u	15816	16394	17119	25193	12487	6091	20506	1459	5453	120518
v	4169	4684	4909	6810	2971	1601	4971	346	1701	32162
w	21	82	31	18	30	8	61	0	9	260
x	971	1104	1019	1454	810	317	1301	74	289	7339

y	691	641	976	978	555	242	923	46	212	5264
z	277	341	367	492	221	122	360	21	222	2423
à	1269	1238	1245	1942	1003	542	1783	129	478	9629
â	108	162	125	223	69	46	142	13	40	928
é	5931	6307	6029	8923	4403	2186	7997	528	1936	44240
è	873	956	937	1391	727	339	1134	94	310	6761
ê	639	655	736	941	454	227	745	40	239	4676
ë	3	11	25	63	13	1	17	1	1	135
î	110	135	128	236	112	60	177	5	55	1018
ï	313	130	64	173	63	22	108	12	20	905
ô	99	101	122	171	69	48	102	6	40	758
ù	194	124	198	238	144	34	236	11	57	1236
û	90	114	108	143	60	47	100	10	23	695
ç	261	212	245	259	102	53	138	13	41	1324
TOT	273057	286583	294674	419759	205991	102189	344847	23664	89025	2039789

La dernière colonne de ce tableau expose la somme de chaque lettre dans la totalité du corpus. Tandis que la dernière ligne présente la somme des lettres dans chaque partition. Néanmoins, pour calculer la longueur moyenne des mots, ces données restent insuffisantes, du moment que la valeur recherchée s'obtient en divisant le nombre de caractères dans chacun des textes par le nombre de mots du même texte. Il faut donc connaître le nombre de mots et leur répartition dans chaque subdivision du corpus. Cette donnée nous est fournie en cliquant sur le bouton « DISTRIBUTION » puis, sur la fonction « Étendue et Probabilité » du logiciel Hyperbase, ainsi nous obtenons les résultats suivants :

Tableau II. 2 : Répartition du corpus selon le nombre d'occurrences

TEXTES	Nombre d'occurrences
PARCOURS	64.566
AILEN	75.357
MILLE	75.802
ELHAKI	109.549
GENET	53.949
ABOUNOUR	26.350
ZRIREK	88.200
FEMME	6083
LETTRES	22.565
Total	522.421

Le corpus étudié compte neuf textes qui totalisent 522.421 occurrences. Le tableau ci-dessus, met en évidence la distribution statistique de ces occurrences. Dans la première colonne nous avons le titre abrégé des partitions, tandis que dans la deuxième nous présentons le nombre d'occurrences. À première vue, nous constatons qu'il y a une différence significative concernant le nombre de mots dans chaque texte. Les valeurs varient entre 6.083 occurrences pour le texte le plus court *Une femme, une mère* (FEMME) et 109.549 occurrences pour le texte le plus long *Le retour d'Abou El Haki* (ELHAKI). Nous relevons aussi qu'il y a des textes dont la longueur est proche comme *Aïlen ou la nuit du récit* (AILEN), *Mille ans, un jour* (MILLE) ainsi que *Abner Abounour* (ABOUNOUR) et *Lettre à moi-même* (LETTRES).

Ces données ainsi regroupées, nous pouvons alors calculer la longueur moyenne des mots de chaque texte, puis nous procéderons à leur classement.

1.2. Calcul de la longueur moyenne des mots

La formule que nous adopterons est la suivante :

$$\text{Longueur moyenne} = \frac{\text{Nombre de caractères}}{\text{Nombre de mots}}$$

Le tableau ci-dessous expose les résultats obtenus avec une indication sur le classement décroissant des textes selon la longueur moyenne des mots de chaque segment.

Tableau II. 3 : Longueur moyenne des mots dans les textes du corpus

TEXTES	Nombre de caractères	Nombre de mots	Longueur moyenne	Rang
PARCOURS	273057	64566	4,23	1
AILEN	286583	75357	3,80	9
MILLE	294674	75802	3,88	5
ELHAKI	419759	109549	3,83	7
GENET	205991	53949	3,82	8
ABOUNOUR	102189	26350	3,88	6
ZRIREK	344847	88200	3,91	3
FEMME	23664	6083	3,89	4
LETTRES	89025	22565	3,95	2
Corpus	2039789	522421	3,9	-

La longueur moyenne du mot dans le corpus pris intégralement est de 3,90. Cette valeur n'atteint pas la moyenne calculée¹³⁶ pour le T.L.F. (4,3853) ni celle de la prose littéraire (4,3535). Nous pouvons noter également que seuls les textes

¹³⁶ Ibid., p.239.

Parcours immobile (PARCOURS) ; *Lettre à moi-même* (LETTRES) ; *Le café bleu*, *Zrirek* (ZRIREK) ont une valeur supérieure à la longueur moyenne, avec un maximum de 4,23 et les six textes restants *Une femme, une mère* (FEMME) ; *Mille ans, un jour* (MILLE) ; *Abner Abounour* (ABOUNOUR) ; *Le retour d'Abou El Haki* (ELHAKI) ; *JEAN GENET Le captif amoureux* (GENET) ; *Aïlen ou la nuit du récit* (AILEN) ont une valeur inférieure qui atteint un minimum de 3, 8.

De même, si nous nous référons aux conclusions émises par Brunet¹³⁷, nous ne pouvons trancher, ni pour les uns ni pour les autres, quant à leur appartenance à la prose littéraire ou à la prose poétique.

De ce fait, la longueur en tant que paramètre de tendance centrale ou de position ne peut aboutir à des conclusions fiables dans les comparaisons entre les textes formant notre corpus, elle n'a qu'une apparence trompeuse, d'où la nécessité de faire intervenir un paramètre de dispersion à savoir l'écart-type qui mesure le degré d'éparpillement des observations (le cas échéant, la longueur des mots) par rapport à la moyenne.

2. Comparaison des moyennes

Dans ce qui suit, le premier point à aborder sera le calcul de l'écart-type. Certes, ce paramètre permet de voir le degré de dispersion de la longueur des mots dans un texte donné par rapport à la longueur moyenne du même texte, mais, il ne permet pas de faire une comparaison intertextes. Donc, au moins pour notre propos, l'écart type n'a pas de valeur en soi, sa valeur réside dans le fait qu'il est pris en compte dans les opérations de calcul de l'écart réduit qui permet, bien

¹³⁷ Ibid., p. 239

évidemment, de mener des comparaisons entre les différents segments du corpus. Tel sera l'objectif du deuxième point.

2.1. Calcul de l'écart-type

Pour calculer l'écart-type, il faut absolument calculer la variance de la distribution statistique (x_i, n_i) quelque soit i allant de 1 jusqu'à 11 ($1 \leq i \leq 11$), où les x_i représentent la longueur des mots et n_i l'effectif correspondant.

➤ **Étape 1 :** Calcul de la moyenne arithmétique de la longueur

$$\bar{x} = \frac{\sum_{i=1}^k n_i x_i}{n}, \quad 1 \leq i \leq k \quad \text{avec} \quad n = \sum_{i=1}^k n_i$$

k étant le nombre de modalité ; dans notre cas, il est de 11 modalités (valeur prise par le caractère étudié à savoir la longueur du mot).

➤ **Étape 2 :** Déduction de la moyenne auxiliaire

Généralement, le calcul de la moyenne nous donne des valeurs décimales. Étant donné que la nature du caractère étudié, à savoir « la longueur des mots » est une variable statistique discrète (ses modalités prennent des valeurs isolées c'est-à-dire des nombres entiers naturels), il s'avère donc indispensable d'arrondir la moyenne calculée à l'entier le plus proche, nous déboucherons donc sur la moyenne dite auxiliaire $\bar{x}'$.

➤ **Étape 3 :** Calcul de la variance auxiliaire

La variance auxiliaire est un paramètre de dispersion qui mesure le degré de dispersion de la longueur des mots par rapport à la moyenne auxiliaire. Elle est obtenue à travers la formule suivante :

$$\sigma'^2 = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x}')^2$$

➤ **Étape 4** : Calcul de la variance

La variance s'obtient par la différence entre la variance auxiliaire et l'écart au carrée de la moyenne auxiliaire à la moyenne, sa formule est la suivante :

$$\sigma^2 = \sigma'^2 - (\bar{x}' - \bar{x})^2$$

➤ **Étape 5** : Déduction de l'écart-type.

L'écart type n'est rien d'autre que la racine carrée de la variance, sa formule est donnée par :

$$\sigma = \sqrt{\sigma^2}$$

Appliquons maintenant ces formules à *Parcours Immobile*

Tableau II. 4 : Calcul de la variance dans PARCOURS

x_i	n_i	$n_i x_i$	$(x_i - \bar{x}')$	$(x_i - \bar{x}')^2$	$n_i(x_i - \bar{x}')^2$
1	1402	1402	-3	9	12618
2	16223	32446	-2	4	64892
3	7524	22572	-1	1	7524
4	6750	27000	0	0	0
5	7097	35485	1	1	7097
6	5410	32460	2	4	21640
7	4616	32312	3	9	41544
8	3619	28952	4	16	57904
9	4297	38673	5	25	107425
10	1684	16840	6	36	60624
11	278	3058	7	49	13622
Total	58900	271200	22	154	394890
$\bar{x}$	$\frac{\sum_{i=1}^k n_i x_i}{n}$ avec $n = \sum_{i=1}^k n_i = 58900$			4,604414261	
$\bar{x}'$	4 correspond à la moyenne auxiliaire				
σ'^2	$\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x}')^2$			6,704414261	
σ'	$\sqrt{\sigma'^2}$			2,482964761	
σ^2	$\sigma'^2 - (\bar{x}' - \bar{x})^2$			2,58928837	
σ	$\sqrt{\sigma^2}$			2,517756474	

Tableau II. 5 : Calcul de la variance dans AÎLEN

x_i	n_i	$n_i x_i$	$(x_i - \bar{x'})$	$(x_i - \bar{x'})^2$	$n_i(x_i - \bar{x'})^2$
1	1433	1433	-3	9	12897
2	17468	34936	-2	4	69872
3	8758	26274	-1	1	8758
4	7669	30676	0	0	0
5	7166	35830	1	1	7166
6	6291	37746	2	4	25164
7	5254	36778	3	9	47286
8	3615	28920	4	16	57840
9	4039	36351	5	25	100975
10	1442	14420	6	36	51912
11	197	2167	7	49	9653
Total	63332	285531	22	154	391523
$\bar{x}$	$\frac{\sum_{i=1}^k n_i x_i}{n}$ avec $n = \sum_{i=1}^k n_i$			4,508479126	
$\bar{x'}$	4 correspond à la moyenne auxiliaire				
σ'^2	$\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x'})^2$			6,182072254	
σ'	$\sqrt{\sigma'^2}$			2,482964761	
σ^2	$\sigma'^2 - (\bar{x'} - \bar{x})^2$			2,486377335	
σ	$\sqrt{\sigma^2}$			2,433828513	

Après avoir présenté la démonstration du calcul de l'écart-type, nous passons maintenant au calcul de l'écart réduit qui sert de base au test paramétrique de comparaison des moyennes.

2.2. Calcul de l'écart réduit

Avant de percer dans la comparaison des moyennes, il est crucial de mettre en évidence les différentes notations adoptées lors de cette comparaison.

- μ est la moyenne de la longueur du texte,
- σ_1 et σ_2 sont respectivement les écarts types du texte 1 et texte 2
- n_1 et n_2 sont respectivement les effectifs du texte 1 et texte 2
- H_0 est l'hypothèse nulle selon laquelle il n'y a pas de différence significative entre les moyennes des deux textes
- H_1 est l'hypothèse alternative selon laquelle la différence entre les moyennes des deux textes est significative.
- Z correspond à l'écart réduit

Nous pouvons alors comparer les moyennes des deux textes en faisant appel au test unilatéral de comparaison de deux moyennes :

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 > \mu_2 \end{cases} \quad \begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{cases}$$

nous proposons $\overline{X_1} - \overline{X_2}$ comme estimateur de $\mu_1 - \mu_2$

puisque σ_1 et σ_2 sont connus et n_1 et n_2 sont supérieurs ou égale à 30.

La règle de décision sera la suivante :

$$\begin{cases} \text{si } |\bar{x}_1 - \bar{x}_2| > Z_\alpha \sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}} & \text{on rejette } H_0 \\ \text{si } |\bar{x}_1 - \bar{x}_2| \leq Z_\alpha \sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}} & \text{on ne rejette } H_0 \end{cases}$$

$$\begin{cases} \text{si } \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}}} > Z_\alpha & \text{on rejette } H_0 \\ \text{si } \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}}} \leq Z_\alpha & \text{on ne rejette } H_0 \end{cases}$$

Application numérique :

$$Z = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{\sigma_a^2}{n_a-1} + \frac{\sigma_b^2}{n_b-1}}} = \frac{|4,23 - 3,80|}{\sqrt{\frac{6,3391}{58900-1} + \frac{5,665}{63332-1}}} = 30,6302639197$$

Nous devons comparer Z avec le fractile d'ordre α (que l'on note Z_α) lu sur la table de la loi normale centrée et réduite, sachant que α est le risque de rejeter l'hypothèse nulle alors qu'elle est vraie.

D'après la table de la loi normale centrée et réduite, nous avons :

Tableau II. 6 : Table de la loi normale centrée et réduite

Table de la loi normale centrée réduite (unilatérale à droite : $1-F_Z : \alpha=P(Z>z_\alpha)$)

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064

Nous constatons, donc, que :

$$P(Z > z_{1,65}) = 0,0495 \text{ et } P(Z > z_{1,64}) = 0,0505$$

Nous devons chercher le fractile $Z_{0,05}$ c'est à dire :

$$P(Z > z_{0,05}) = 0,05$$

Par interpolation linéaire nous aurons :

$$0,0495 < 0,05 = \alpha < 0,0505$$

$$1,64 < Z < 1,65$$

Donc :

$$\frac{Z - 1,64}{0,05 - 0,0495} = \frac{1,65 - 1,64}{0,0505 - 0,0495}$$

Par conséquent $Z = 1,64 + (0,05 - 0,0495) \frac{1,65 - 1,64}{0,0505 - 0,0495} = 1,645$

Ceci veut dire que l'écart réduit qui correspond au risque de 5% de rejeter l'hypothèse nulle alors qu'elle est vraie, est :

$$Z_{0,05} = 1,645$$

Donc : $z = 30,6302639197 > Z_{0,05} = 1,645$

Et par conséquent, nous rejetons H_0 en faveur de H_1 , c'est à dire qu'il y a une différence significative entre les moyennes des textes PARCOURS et AÎLEN. Autrement dit, la longueur moyenne dans PARCOURS est supérieure à celle de AÎLEN.

De la même manière, nous avons renouvelé les calculs pour tous les binômes qui composent notre corpus et qui correspondent au nombre de combinaison de 2 parmi 9 :

$$C_9^2 = \frac{A_9^2}{2!} = \frac{9 \times 8}{2 \times 1} = 36$$

Donc, nous aurons 36 combinaisons, les résultats obtenus sont présentés dans le tableau suivant :

Tableau II. 7 : Écart réduit des textes pris deux à deux

TEXTES	$\bar{x}_a$	$\bar{x}_b$	$\bar{x}_a - \bar{x}_b$	$ \bar{x}_a - \bar{x}_b $	$\frac{\sigma_a^2}{n_a - 1} + \frac{\sigma_b^2}{n_b - 1}$	$\sqrt{\frac{\sigma_a^2}{n_a - 1} + \frac{\sigma_b^2}{n_b - 1}}$	z
(PARCOURS, AILEN)	4,23	3,8	0,43	0,43	0,00019708	0,0140384	30,63
(PARCOURS, MILLE)	4,23	3,88	0,35	0,35	0,0001987	0,01409625	24,83
(PARCOURS, ELHAKI)	4,23	3,83	0,40	0,40	0,00017416	0,01319714	30,31
(PARCOURS, GENET)	4,23	3,82	0,41	0,41	0,00024423	0,01562798	26,23
(PARCOURS, ABOUNOUR)	4,23	3,88	0,35	0,35	0,00038417	0,01960037	17,86
(PARCOURS, ZRIREK)	4,23	3,91	0,32	0,32	0,00019434	0,01394061	22,95
(PARCOURS, FEMME)	4,23	3,89	0,34	0,34	0,07321634	0,27058519	1,26
(PARCOURS, LETTRES)	4,23	3,95	0,28	0,28	0,00	0,02084883	13,43
(AILEN, MILLE)	3,8	3,88	-0,08	0,08	0,00018053	0,01343606	5,95
(AILEN, ELHAKI)	3,8	3,83	-0,03	0,03	0,00015599	0,01248952	2,40
(AILEN, GENET)	3,8	3,82	-0,02	0,02	0,00022606	0,0150352	1,33
(AILEN, ABOUNOUR)	3,8	3,88	-0,08	0,08	0,000366	0,01913108	4,18
(AILEN, ZRIREK)	3,8	3,91	-0,11	0,11	0,00017616	0,01327269	8,29
(AILEN, FEMME)	3,8	3,89	-0,09	0,09	0,00127059	0,03564538	2,52
(AILEN, LETTRES)	3,8	3,95	-0,15	0,15	0,0004165	0,02040826	7,35

(MILLE, ELHAKI)	3,88	3,83	0,05	0,05	0,00015762	0,0125545	3,98
(MILLE, GENET)	3,88	3,82	0,06	0,06	0,00022768	0,01508923	3,98
(MILLE, ABOUNOUR)	3,88	3,88	0,00	0,00	0,00036763	0,01917357	0,00
(MILLE, ZRIREK)	3,88	3,91	-0,03	0,03	0,00017779	0,01333386	2,25
(MILLE, FEMME)	3,88	3,89	-0,01	0,01	0,00127222	0,0356682	0,28
(MILLE, LETTRES)	3,88	3,95	-0,07	0,07	0,00041812	0,0204481	3,42
(ELHAKI, GENET)	3,83	3,82	0,01	0,01	0,00020315	0,0142529	0,70
(ELHAKI, ABOUNOUR)	3,83	3,88	-0,05	0,05	0,00034309	0,01852258	2,70
(ELHAKI, ZRIREK)	3,83	3,91	-0,08	0,08	0,00015325	0,0123795	6,46
(ELHAKI, FEMME)	3,83	3,89	-0,06	0,06	0,00124768	0,03532253	1,70
(ELHAKI, LETTRES)	3,83	3,95	-0,12	0,12	0,00039358	0,01983897	6,05
(GENET, ABOUNOUR)	3,82	3,88	-0,06	0,06	0,00041316	0,02032622	2,95
(GENET, ZRIREK)	3,82	3,91	-0,09	0,09	0,00022332	0,01494394	6,02
(GENET, FEMME)	3,82	3,89	-0,07	0,07	0,00131775	0,03630083	1,93
(GENET, LETTRES)	3,82	3,95	-0,13	0,13	0,00046365	0,02153263	6,04
(ABOUNOUR, ZRIREK)	3,88	3,91	-0,03	0,03	0,00036326	0,01905944	1,57
(ABOUNOUR, FEMME)	3,88	3,89	-0,01	0,01	0,00145769	0,03817972	0,26
(ABOUNOUR, LETTRES)	3,88	3,95	-0,07	0,07	0,0006036	0,02456817	2,85

(ZRIREK, FEMME)	3,91	3,89	0,02	0,02	0,00126786	0,03560698	0,56
(ZRIREK, LETTRES)	3,91	3,95	0,04	0,04	0,00041376	0,02034112	1,97

Les résultats obtenus présentent, pour 28 couples sur 36, un écart réduit supérieur à $Z_{0,05} = 1,645$, ce qui nous conduit à rejeter l'hypothèse nulle au profit de l'hypothèse alternative selon laquelle la supériorité de la longueur moyenne des mots entre les 28 couples des textes est significative. Autrement dit, lesdits textes « *ne peuvent être considérés comme des échantillons aléatoires d'une même population* »¹³⁸. Tandis que les 8 autres binômes, à savoir :

$$\left\{ \begin{array}{l} (\text{PARCOURS, FEMME}); (\text{AILEN, GENET}); (\text{MILLE, ABOUNOUR}); (\text{MILLE, FEMME}); \\ (\text{ELHAKI, GENET}); \\ (\text{ABOUNOUR, ZRIREK}); (\text{ABOUNOUR, FEMME}); (\text{ZRIREK, FEMME}) \end{array} \right\}$$

Présentent un écart réduit inférieur à $Z_{0,05} = 1,645$. Dans ce cas, nous ne saurons trancher si les similitudes entre les paires sont dues à leur ressemblance stylistique, thématique ou s'il s'agit de l'intervention d'un autre facteur.

Il s'avère, donc, difficile d'expliquer la nature des variations pour les binômes qui ont un faible écart réduit. Pour combler cette limite, nous chercherons des réponses à travers d'autres tests. En adoptant le même raisonnement que Brunet, nous devons compléter cette analyse en faisant appel aux tests basés sur les effectifs théoriques, car :

« Les comparaisons de ces moyennes ne donnent qu'une vision incomplète du phénomène. [...] Il est préférable d'abandonner les moyennes et de s'en tenir à la méthode la plus sûre qui consiste à calculer

¹³⁸ Etienne BRUNET, *Le Vocabulaire de Jean Giraudoux, Structure et Evolution*, Genève, Slatkine, 1978, p.161.

*les effectifs théoriques des mots de 1, 2, n lettres, puis à les confronter aux effectifs observés. »*¹³⁹

Abordons maintenant la question des effectifs théoriques.

3. Effectifs théoriques

Les effectifs théoriques seront approchés via le test de khi-deux. Rappelons que ce test a été proposé par le statisticien Karl Pearson en 1900. Il a pour objectif de vérifier dans quelle mesure les effectifs relatifs à un ou plusieurs caractères sont conformes aux effectifs attendus sous l'hypothèse nulle. Autrement dit, il consiste à mesurer l'écart entre la distribution des effectifs théoriques et celle des effectifs observés (réels).

L'application du test de khi-deux, exige tout d'abord de mettre en évidence le principe qui régit le calcul des effectifs théoriques :

Sachant que le corpus dans son intégralité comprend 2039789 lettres et que le texte *Parcours immobile* représente une proportion de $\frac{64566}{522421} = 0,12359 = 12,359\%$ de ce corpus, (taux calculé à partir des occurrences), selon E. Brunet, si le hasard était seul en jeu, le nombre de lettres attendues dans PARCOURS serait : $2039789 \times 0,12359 = 252097,4780$. Comme nous en avons 273057, l'excédent est de 20959,5220. Le χ^2 correspondant est de l'ordre de 1742,5868. Le χ^2 se calcule de la manière suivante :

¹³⁹ Ibid., p.163.

Si nous notons la valeur observée par O et celle théorique par C, nous aurons :

$$\chi^2 = \sum \frac{(O-C)^2}{C}$$

Le même procédé appliqué à toutes les parties du corpus donne :

Tableau II. 8 : Application du test de Khi-deux

TEXTES	Nombre de mots	Probabilité	Lettres réelles	Lettres théoriques	Écart	Khi-deux
PARCOURS	64566	0,12359	273057	252097,4780	20959,5220	1742,5861
AÎLEN	75357	0,14425	286583	294230,8592	-7647,8592	198,7886
MILLE	75802	0,14510	294674	295968,3584	-1294,3584	5,6606
ELHAKI	109549	0,20969	419759	427733,2748	-7974,2748	148,6652
GENET	53949	0,10327	205991	210643,4787	-4652,4787	102,7592
ABOUNOUR	26350	0,05044	102189	102883,3836	-694,3836	4,6866
ZRIREK	88200	0,16883	344847	344376,2594	470,7406	0,6435
FEMME	6083	0,01164	23664	23751,0293	-87,0293	0,3189
LETTRES	22565	0,04319	89025	88104,8786	920,1214	9,6093
Corpus	522421	1	2039789	2039789,000		2213,7180

Le nombre de lettres auquel nous nous attendons après le tirage aléatoire dans PARCOURS, ZRIREK et LETTRES est inférieur à la réalité. Selon ce test, l'écart entre le nombre de lettres réel et le nombre théorique est positif dans les textes précédemment cités. Par contre dans AILEN, MILLE, ELHAKI, ABOUNOUR et FEMME, il est négatif.

De la table de la loi de khi-deux, à un degré de liberté, nous dégageons le fractile d'ordre α qui est de 3,841. Nous remarquons d'après les résultats obtenus,

dans le tableau ci-dessus, que la majorité des textes présentent un χ^2 supérieur au khi-deux théorique¹⁴⁰ (3,841). Donc la distribution des lettres dans ces textes exclut toute explication d'ordre aléatoire. Exception faite de deux textes, à savoir : ZRIREK et FEMME qui enregistrent un khi-deux inférieur à celui théorique et pour lesquels l'hypothèse nulle ne peut être rejetée. Cela montre que la distribution des lettres dans presque tous les textes n'est pas le résultat d'un simple hasard, mais au contraire, elle peut s'expliquer par des effets du style, du thème et/ou autres. D'ailleurs dans lesdits textes l'auteur ne s'éloigne pas trop des thèmes qu'il y aborde. En l'occurrence les thèmes de la mémoire, de l'histoire et de la littérature.

En nous basant sur l'écart réduit, nous aurons :

¹⁴⁰ Il s'agit du Khi-deux théorique lu sur la table de la loi de Khi-deux à un degré de liberté et au risque de 5%. Il se distingue du Khi-deux observé que l'on nomme aussi Khi-deux calculé ou réel.

Tableau II. 9 : Calcul des écarts réduits

TEXTES	Nombre de caractères	Nombre de mots	Probabilité	Lettres théoriques	Écart	khi-deux	Écart réduit	Rang
PARCOURS	273057	64566	0,12359	252097,4780	20959,522	1742,5861	41,7443	1
AILEN	286583	75357	0,14425	294230,8592	-7647,8592	198,7886	-14,0992	9
MILLE	294674	75802	0,14510	295968,3584	-1294,3584	5,6606	-2,3792	6
ELHAKI	419759	109549	0,20969	427733,2748	-7974,2748	148,6652	-12,1928	8
GENET	205991	53949	0,10327	210643,4787	-4652,4787	102,7592	-10,1370	7
ABOUNOUR	102189	26350	0,05044	102883,3836	-694,3836	4,6866	-2,1648	5
ZRIREK	344847	88200	0,16883	344376,2594	470,7406	0,6435	0,8022	3
FEMME	23664	6083	0,01164	23751,0293	-87,0293	0,3189	-0,5647	4
LETTRES	89025	22565	0,04319	88104,8786	920,1214	9,6093	3,0999	2
Corpus	2039789	522421	1	2039789		2213,7180		

D'après ce test, nous remarquons que le classement selon l'écart réduit confirme celui obtenu à partir de la longueur moyenne dans 55,55% des cas et ce pour les textes suivants : PARCOURS, AÎLEN, LETTRES, FEMME, ZRIREK. Donc la superposition n'est pas parfaite dans les deux tableaux.

Cela étant dit, nous passerons, par la suite, à la distribution des groupes de longueur dans notre corpus, afin d'apprécier la tendance de l'auteur dans le choix des mots (courts, moyens, longs).

4. Groupes de longueurs

La distribution des classes de mots de 1 à n lettres s'inscrit dans la logique des démarches précédentes. Le logiciel *Hyperbase* répartit le nombre d'occurrences

en onze groupes de longueur. Le résultat de cette répartition est présenté dans le tableau ci-dessous :

Tableau II. 10 : Distribution des classes de n lettres (vocables)

	PARCOURS	AILEN	MILLE	ELHAKI	GENET	ABOUNOUR	ZRIREK	FEMME	LETTRE	TOTAL	Rang
lg1	1402	1433	1386	2314	1264	615	2229	155	553	11351	10
lg2	16223	17468	18471	27540	13746	6565	22933	1532	5838	130316	1
lg3	7524	8758	8812	12366	5916	2993	9777	663	2341	59150	2
lg4	6750	7669	7659	10779	5425	2494	8455	638	2639	52508	3
lg5	7097	7166	7479	9799	4896	2446	7698	610	1954	49145	4
lg6	5410	6291	6550	8747	4225	2130	6697	470	1657	42177	5
lg7	4616	5254	5363	7298	3263	1767	5373	379	1494	34807	6
lg8	3619	3615	3615	5460	2703	1413	4709	295	1115	26544	8
lg9	4297	4039	4234	6351	3144	1548	5460	374	1399	30846	7
lg10	1684	1442	1525	2389	1299	572	2482	138	602	12133	9
lg11	278	197	180	320	147	74	335	18	88	1637	11
Total	58900	63332	65274	93363	46028	22617	76148	5272	19680	450614	

Une lecture succincte du tableau montre que les premières places sont occupées par les mots courts (lg2, lg3 et lg4), exception faite de la classe lg1 qui prend le dernier rang ; cela s'explique par le fait que la langue française contient en réalité peu de mots à une seule lettre. Les positions médianes sont occupées par les classes moyennes lg5, lg6 et lg7 qui se positionnent dans le 4^{ème}, 5^{ème} et 6^{ème} rang. De plus, nous constatons une substitution entre les deux positions lg8 et lg9 qui prennent respectivement les 8^{ème} et 7^{ème} places. Quant aux dernières places, nous y trouvons les mots de longueurs extrêmes à savoir lg1 et lg11 qui occupent respectivement le 10^{ème} et 11^{ème} position.

Pour une meilleure lecture de ce tableau, nous nous proposons de classer ces groupes de longueurs dans un ordre décroissant :

$$lg2 > lg3 > lg4 > lg5 > lg6 > lg7 > lg9 > lg8 > lg1 > lg11$$

Nous constatons, alors, que les mots courts sont les plus abondants. Cette remarque reste valable pour la totalité des parties de notre corpus, chose qui n'est pas étonnante du moment que les mots les plus courts sont généralement les plus fréquents dans la langue.

Dans ce sens, Ali Hefied stipule que :

« Si les mots courts sont abondants, cela n'a rien d'étonnant du moment que généralement les mots les plus fréquents dans la langue se trouvent être ceux qui sont en même temps les plus économiques [...] il s'agit le plus souvent de mots grammaticaux comme de, et, le, la, ce...qui occupent toujours les premiers rangs quant à leur fréquence. »¹⁴¹

¹⁴¹ Ali HEFIED, *op. cit.*, 1999, p.116.

L'étude que nous venons d'effectuer pourrait être affinée par une autre approche afin de corroborer les conclusions tirées. Il s'agit, bien évidemment, de l'analyse factorielle des correspondances, qui nous donnera une vue d'ensemble sur le regroupement des classes de longueur à travers le graphique proposé.

5. Analyse factorielle

Mise au point par J.-P. Benzécri¹⁴², l'analyse factorielle des correspondances permet de représenter graphiquement, de manière synthétique, de grandes tables de contingences. Elle permet d'analyser un grand nombre de variables contenues dans les lignes et les colonnes d'un tableau de données et de faire ressortir les correspondances qui peuvent exister entre ces deux ensembles. Son développement a été favorisé par l'essor de l'outil informatique. En effet, l'association ADDAD a incorporé cette méthode dans un logiciel statistique d'analyse de données, avant d'être exploitée par André Salem qu'il l'a adaptée à Windows et E. Brunet qui l'a intégrée dans le logiciel *Hyperbase*.

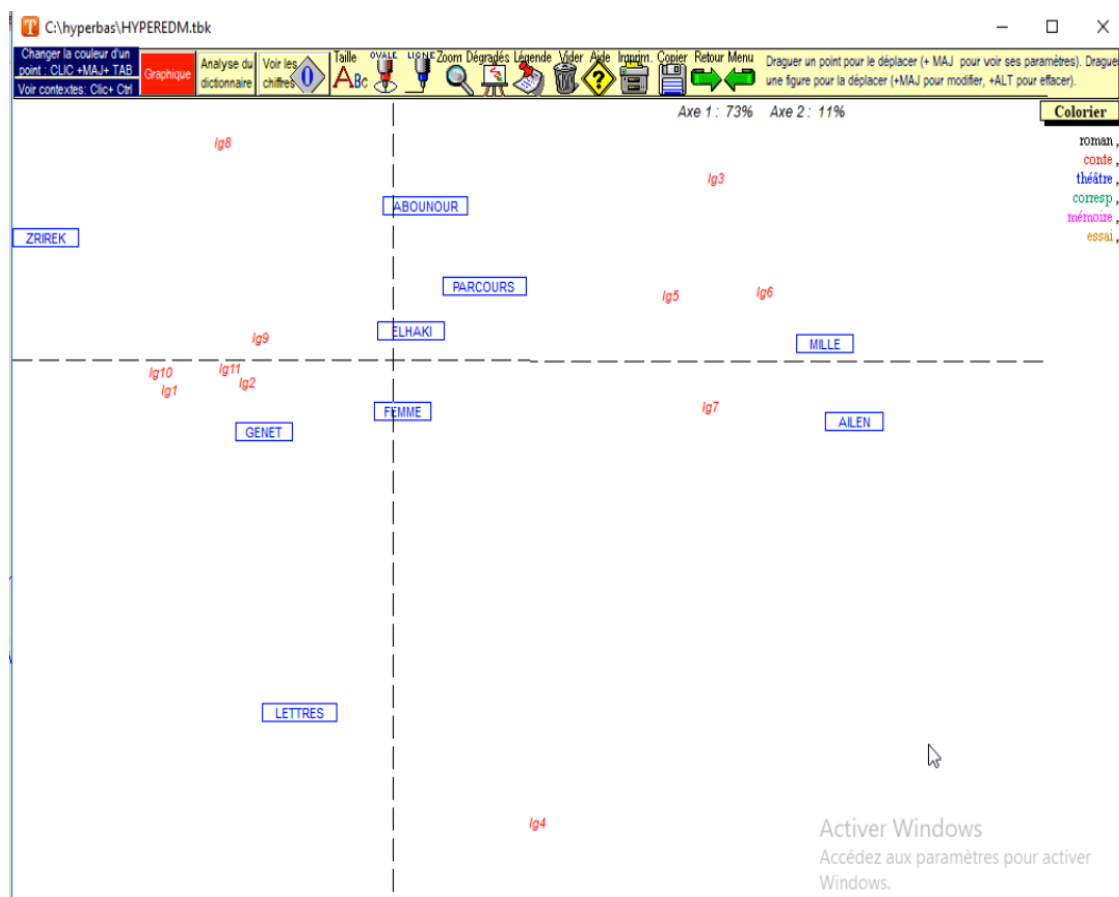
Avant d'enchaîner sur ce qui suit, nous signalons que, lors de l'élaboration de l'analyse factorielle, le logiciel *Hyperbase* suggère à l'utilisateur de choisir entre deux programmes à savoir : « Tabet » qu'il définit comme étant une méthode simple et « Coran » qui serait une méthode évoluée, du moment qu'elle prend en considération des données pondérées. Mais, dans ce dernier cas les résultats sont moins sensibles aux inégalités dans les lignes ou les colonnes et le graphique, plus arrondi, résiste mieux aux effets centripètes ou centrifuges dus à la taille. Finalement, le logiciel recommande d'utiliser la méthode « Tabet » car elle transforme les fréquences en écarts réduits et fait intervenir des probabilités issues

¹⁴² J.P. Benzécri, *L'analyse des données*, Tome I et II, Paris, Dunod, 1973.

du corpus entier. Tandis que la deuxième méthode traite le tableau de contingence comme un tout détaché du reste du corpus.

En appliquant l'analyse factorielle des correspondances à notre corpus, suivant le programme « Tabet » nous obtenons le graphique suivant :

Figure II. 1 : Analyse factorielle des classes de longueurs



A première vue, nous remarquons que le premier facteur récupère 73 % de l'information globale, tandis que le deuxième en récupère 11 %. Donc, les représentations graphiques obtenues sont satisfaisantes, car le premier plan factoriel formé des deux premiers axes principaux restitue 84 % de l'inertie totale,

c'est-à-dire, de l'information globale contenue dans les variables (les textes et les longueurs de mots). La lecture de la distribution des groupes de longueur selon l'AFC se fait selon la position par rapport aux axes vertical et horizontal.

D'après la représentation, nous constatons que les mots courts et moyens (de lg3 à lg7), excepté ceux de longueur 1 et 2, se placent à droite de l'axe vertical, tandis que ceux qui sont longs (lg8 à lg11) se trouvent à gauche. L'ensemble des points représentant chaque classe de longueur évolue suivant une forme géométrique ovale opposant ainsi les (lg3, lg4, lg5, lg6, lg7) aux (lg8, lg9, lg10, lg11) et (lg1, lg2).

Les textes ZRIRZK et AÎLEN se trouvent sur deux positions extrêmement opposées par rapport à cet axe ; par contre ABOUNOUR, ELHAKI et FEMME sont placés sur la même ligne verticale. De même, l'ensemble des points représentant chaque texte s'arrange suivant une forme ovale. En effet, du côté droit nous avons AÎLEN, MILLE, PARCOURS, ELHAKI auxquels les critiques attribuent l'étiquette de roman ; du côté gauche nous trouvons ZRIREK, GENET, FEMME que les critiques rangent du côté des essais.

Si nous voulons prendre en considération à la fois la répartition des textes et la distribution des longueurs de mots, nous pouvons nous référer à E. Brunet qui stipule que :

« (...) parmi ces caractères [caractère lié à la longueur du mot] s'impose la fréquence ou rareté des mots, mais aussi la catégorie grammaticale et le recours plus au moins large à la dérivation. Nous verrons plus loin que le discours intellectuel se repaît de suffixes et de préfixes et s'accommode des catégories encombrantes comme les adverbes en -ment. Le discours intellectuel gonfle les mots de sèmes et de spécifications multiples. La prose littéraire et la

poésie au contraire ont un bagage léger et dans leur boîte à outils on ne trouve guère que des instruments à tout faire et des mots à tout dire. »¹⁴³

En nous basant sur la citation de Brunet, nous pouvons interpréter les résultats comme suit :

- Les textes GENET et ZRIREK, attirent les mots courts (lg1, lg2), cependant, ils favorisent un emploi fort des mots longs (lg9, lg10 et lg11), ce qui pourrait s'expliquer par le fait que ces textes forment un discours intellectuel avec un style recherché¹⁴⁴. Ajoutons que ce constat renforce l'idée avancée par les critiques, selon laquelle les textes GENET et ZRIREK sont bien des essais.
- MILLE, AÏLEN et PARCOURS sont des textes attirés par les mots courts et moyens (lg3, lg5, lg6 et lg7). Pourrions-nous alors épouser l'affirmation de José Gregorio PARADA-RAMIREZ qui dit : « *Plus de dialogue, moins de description, un langage familier et beaucoup moins technique sont des paramètres qui peuvent expliquer de fortes concentrations de mots courts* »¹⁴⁵? De toute façon, nous pouvons dire que l'utilisation des mots courts et moyens dans ces textes est due au genre dans lequel ils s'inscrivent.

¹⁴³ Etienne BRUNET, *op.cit.*, 1789, p.246.

¹⁴⁴ Quoique nous avons signalé l'attraction des deux textes pour la lg1 et lg2, ça n'a aucune influence sur l'interprétation que nous avons suggéré, parce que les mots courts sont les plus fréquents dans la langue, et c'est généralement les mots grammaticaux qui font que la fréquence de cette classe de longueur soit élevée.

¹⁴⁵ José Gregorio PARADA RAMIREZ, *Lecture documentée et analyse textométrique de l'œuvre de Jules Verne. Les influences de la Franc-maçonnerie dans son œuvre*, Thèse rédigée pour l'obtention du doctorat en langue et littérature françaises, Nice, 2013, p. 53

Les textes restant ABOUNOUR (Nouvelle), ELHAKI (Roman), FEMME (Essai) se placent au centre, marquant ainsi une symétrie entre deux groupes de longueur : les mots longs et courts (lg9, lg10, lg11, lg1, lg2) et les mots courts et moyens (lg3, lg5, lg6 et lg7). Le texte LETTRE (récit épistolaire) n'appartient à aucune constellation.

Notons que le texte ELHAKI qui est à cheval entre les mots longs et les mots courts et moyens est aussi un roman. Cependant son attraction pour les mots longs pourrait se justifier par l'utilisation de l'auteur d'un discours soutenu abordant les écritures mystiques de certaines figures intellectuelles prestigieuses qui acheminent le lecteur dans des espaces et des périodes pluriels.

Pour ce qui est de l'axe horizontal les deux textes ABOUNOUR (Nouvelle) et LETTRES (récit épistolaire) se repoussent en s'écartant de l'origine de l'axe factorielle. Nous pouvons également noter que les classes de longueur et les textes évoluent selon une forme géométrique ovale entre (lg3, lg5, lg6) et (lg8, lg9) en haut pour les cinq textes, ZRIREK, ABOUNOUR, ELHAKI, PARCOURS et MILLE, avec (lg2, lg1), (lg10, lg11) et (lg7, lg4) en bas, pour les textes GENET et FEMME.

L'analyse que nous avons faite est en faveur d'une variation régulière des classes de longueur. Notons que la période de production n'a pas d'effet, vu que les textes regroupés en constellations sont produits pendant des périodes distancées. La question qui se pose alors est la suivante : quels sont les autres facteurs qui ont contribué à ce classement et qui ont ordonné ces textes selon ces classes de longueur ? La seule explication qui peut être avancée est en relation avec les effets du genre et du style. En effet, nous avons constaté des régularités entre les romans et l'utilisation des mots courts et moyens qui sont en faveur d'un

registre familier, et entre les essais et la tendance de l'auteur à utiliser des mots longs qui renvoient souvent au style soutenu.

Après avoir étudié le phénomène de la longueur du mot qui a révélé des régularités dans la distribution des mots selon leur longueur, nous nous pencherons par la suite sur l'analyse de l'étendue du texte qui est fonction de l'étendue du vocabulaire.

B. Étendue du corpus

Le dépouillement automatique du corpus a pour but de nous fournir des données numériques. Rappelons que nous avons déterminé deux ensembles qui existent pour un texte donné : le premier est formé par le nombre total de mots ou le nombre d'occurrences, que nous représentons par N ; le second est constitué par le nombre de vocables qui ont au moins une occurrence dans le texte. Ce dernier ensemble est symbolisé par V . Le rapport V/N nous donne une première impression sur l'étendue du lexique. Notons qu'il existe une relation étroite entre le nombre de vocables et l'étendue N . En effet, V_1 est l'ensemble de vocables utilisés 1 fois, V_2 ceux utilisés 2 fois et V_n les vocables utilisés n fois, ce qui donne naissance à la formule¹⁴⁶ suivante :

$$N = 1V_1 + 2V_2 + 3V_3 + \dots + nV_n$$

$$N = \sum_i^n iV_i$$

¹⁴⁶ Etienne BRUNET, « *les Rougon- macquart : Aspects quantitatifs* », Extrait de la revue : *Informatique et Statistique dans les sciences Humaines XXI*, de 1 à 4, université de liège, C.I.P.L, 1985, pp.83-84

Dans ce qui va suivre, nous entamerons l'exploration de l'étendue N, puis nous soumettrons à l'étude l'étendue V.

1. Étendue N (occurrences)

Parmi les premières données quantitatives que nous livre le dépouillement automatique du corpus il y a le nombre des mots qui le composent. Cette opération consiste à déterminer le nombre d'occurrences N pour tout le corpus, mais aussi le nombre d'occurrences qui existent dans chaque texte pris isolément. Rappelons que notre corpus se compose de neuf œuvres, publiées entre 1980 et 2010. Nous commencerons, tout d'abord, par classer les textes selon l'étendue N, puis nous étudierons la corrélation entre le rang de l'étendue N et l'ordre chronologique de parution des textes.

1.1. Classement des textes selon l'étendue N

En soumettant notre corpus au logiciel Hyperbase, ce dernier nous fournit le tableau suivant :

Tableau II. 11 : Classement des textes selon le nombre d'occurrence

TEXTES	Nombre d'occurrences N	Rang croissant N	Rang décroissant N
PARCOURS	64566	5	5
AILEN	75357	6	4
MILLE	75802	7	3
ELHAKI	109549	9	1
GENET	53949	4	6
ABOUNOUR	26350	3	7
ZRIREK	88200	8	2
FEMME	6083	1	9
LETTRES	22565	2	8
Total	522421		

Le tableau ci-dessus expose la distribution des occurrences dans les différents textes composant le corpus. La première colonne présente les œuvres selon l'ordre chronologique de parution. La deuxième colonne fournit le nombre d'occurrences de chaque texte. Les deux dernières colonnes affichent respectivement le rang croissant et le rang décroissant de l'étendue N relative aux textes. Notons que nous avons procédé aux classements des textes afin de les soumettre au principe de corrélation des rangs.

Une lecture succincte du tableau nous permet de constater que le nombre d'occurrences évolue dans un intervalle situé entre 6083 comme limite basse ; et 109549 comme limite haute.

Si nous procédons au classement décroissant des textes sur la base de l'étendue N, le résultat pourrait se présenter de la manière suivante :

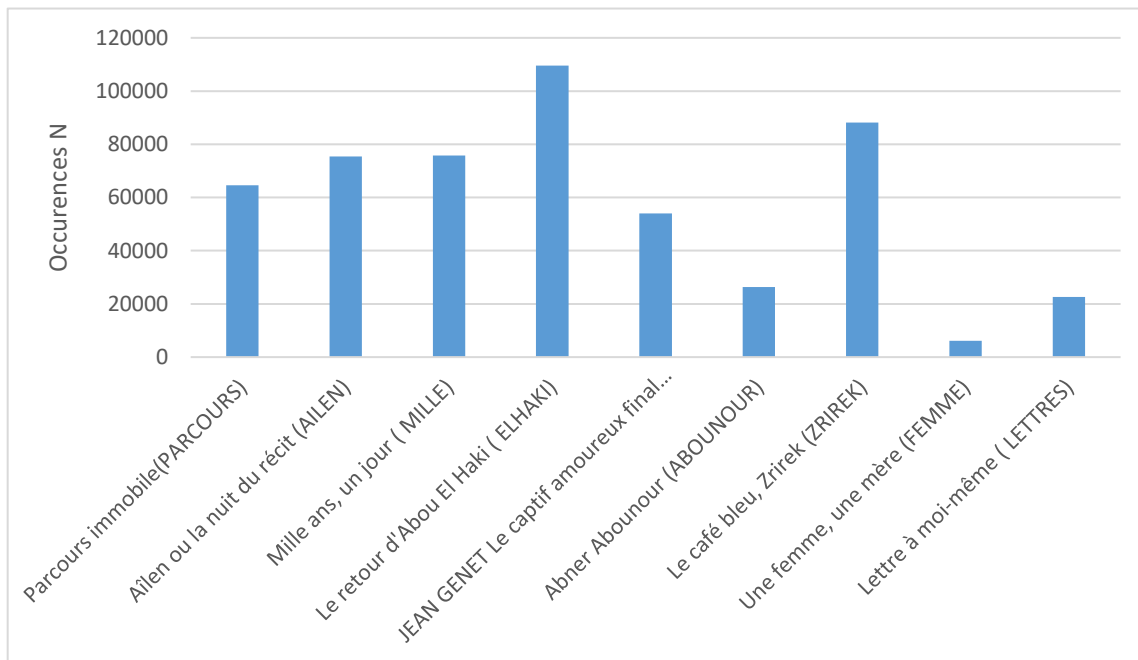
ELHAKI> ZRIREK> MILLE > AÎLEN> PARCOURS > GENET >
ABOUNOUR> LETTRE> FEMME

Si nous nous basons sur l'ordre croissant, les textes pourraient se présenter ainsi :

FEMME > LETTRE> ABOUNOUR > GENET> PARCOURS >
AÎLEN> MILLE > ZRIREK ELHAKI

Nous avons traduit ces résultats en un histogramme qui offre une vue synthétique et permet de mieux apprécier les constatations que nous avons faites.

Figure II. 2 : Répartition des textes selon le nombre d'occurrences



Ce graphique présente deux modes. En effet, ELHAKI enregistre l'étendue la plus élevée par rapport à ses voisines MILLE et GENET. Le texte ZRIREK également est marquée par l'étendue la plus élevée, sauf que ELHAKI correspond au mode absolu et ZRIREK n'est qu'un mode relatif. Donc la série est bimodale ce qui montre l'hétérogénéité de la distribution statistique des textes.

Nous pouvons remarquer, également, qu'il y a trois phases. Dans la première, l'étendue des quatre premiers textes évoluent selon une tendance haussière, pour atteindre le pic dans ELHAKI, puis nous assistons à une phase de tendance baissière qui prend fin par ABOUNOUR, avant d'effectuer un second pic dans ZRIREK. Finalement nous assistons à une décroissance relative aux textes FEMME et LETTRES, ceux dont l'étendue est la plus faible de toutes les partitions composant notre corpus.

Soulignons que le caractère élevé de l'étendue des quatre premiers textes d'El Maleh, peut-être expliqué par son acharnement pour l'écriture. En effet, notre auteur était très sensible à la variable « temps », il se voyait comme une matière vouée à la déliquescence et voulait étaler dans ses textes, avant que la fin ne le surprenne, tous ce qu'il a envie d'exprimer. C'est la raison pour laquelle les premiers écrits d'El Maleh se caractérisent par une étendue élevée comparativement aux autres textes qui sont publiés après. Abstraction faite de ZRIREK qui se démarque par une étendue aussi élevée que les quatre premières productions. Ceci est peut-être dû à des raisons d'ordres thématiques. C'est vrai que les productions d'El Maleh abordent plus ou moins les mêmes thèmes, cependant dans *Le Café bleu*, *Zrerek* nous assistons à une concentration d'idées sur la littérature et l'écriture allégorique. À notre humble avis, c'est ce qui peut justifier la hausse de l'étendue au niveau de la deuxième phase.

Après avoir présenté le classement des textes selon le nombre d'occurrences, nous procéderons maintenant au test d'indépendance entre l'ordre chronologique de parution des textes et le nombre d'occurrences de chaque segment.

1.2. Test sur le coefficient de corrélation des rangs (Analyse bivariée)

Généralement pour tester l'indépendance entre deux variables nous pouvons faire appel à deux tests de nature différentes et de conditions d'application différentes :

Il y a le test paramétrique appelé Test de Pearson qui ne peut être appliqué que si les variables étudiées répondent aux conditions suivantes :

- La liaison entre les deux variables est supposée linéaire.
- Les lois des deux variables sont supposées normales.
- La loi conditionnelle des deux variables est normale.

Si l'une des conditions du calcul du coefficient de corrélation de Pearson n'est pas vérifiée, nous utilisons alors un test non paramétrique, à savoir : le coefficient de corrélation des rangs de Spearman. Ce dernier test s'intéresse au rang des variables et non à leurs valeurs, c'est-à-dire, il permet de comparer la concordance du classement de deux variables et de mesurer leur degré de dépendance.

Pour obtenir ce coefficient, il faut adopter la formule suivante :

$$\rho_s = 1 - \frac{6 \sum (xi' - yi')^2}{n(n^2 - 1)}$$

Avec :

- X et Y sont les deux variables étudiées ;
- x'_i le rang des valeurs observées de X
- y'_i le rang des valeurs observées de Y.
- ρ_s le coefficient de corrélation des rangs de Spearman.

Dans notre corpus, au moins l'une des conditions du test du coefficient de corrélation de Pearson n'est pas vérifiée, à savoir la normalité, nous aurons recours alors au test de coefficient de corrélation des rangs de Spearman. Notons que ce test étudie l'existence d'une liaison entre deux variables statistiques quantitatives. Ces dernières sont définies comme étant des variables qui ont des modalités

mesurables. Dans notre cas, les valeurs que peuvent prendre la chronologie ou bien le classement des étendues N sont des entiers naturels (1, 2, 3n), il s'agit donc de variables statistiques quantitatives discrètes¹⁴⁷.

Avant de procéder au calcul, nous commençons par émettre les hypothèses suivantes :

$$\begin{cases} H_0: \text{l'ordre chronologique et l'étendue sont indépendants} \\ H_1: \text{l'ordre chronologique et l'étendue sont liés} \end{cases}$$

Si $|\rho_s| \text{ observé} \geq \rho_{s,\alpha} \text{ théorique (lu sur la table de Spearman)}$, nous pouvons rejeter H_0 en faveur de H_1 qui stipule qu'il y a une liaison significative entre les deux variables.

Nous calculons ρ_s le coefficient de corrélation des rangs de Spearman en nous référant au tableau des rangs suivant :

¹⁴⁷ Une variable statistique quantitative peut être discrète ou continue. Une variable discrète ne peut prendre que des valeurs entières (1, 2, 3, ..., n), elle a une valeur finie. Une variable continue peut prendre des valeurs décimales, formant un ensemble continu.

Tableau II. 12 : Calcul de la corrélation des rangs des textes

TEXTES	Ordre chronologique	Nombre d'occurrence N	Rang croissant	Distance	Distance au carrée
PARCOURS	1	64566	5	-4	16
AILEN	2	75357	6	-4	16
MILLE	3	75802	7	-4	16
ELHAKI	4	109549	9	-5	25
GENET	5	53949	4	1	1
ABOUNOUR	6	26350	3	3	9
ZRIREK	7	88200	8	-1	1
FEMME	8	6083	1	7	49
LETTRES	9	22565	2	7	49
			$\sum d_i^2$		182

Table A18 Table of Critical Values for Spearman's Rho

<i>n</i>	One-tailed level of significance			
	.05	.025	.01	.005
	Two-tailed level of significance			
	.10	.05	.02	.01
4	1.000	—	—	—
5	.900	1.000	1.000	—
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.700	.783	.833
10	.564	.648	.745	.794
11	.536	.618	.709	.755
12	.503	.587	.671	.727
13	.484	.560	.648	.703
14	.464	.538	.622	.675
15	.443	.521	.604	.654

Donc, nous obtenons :

$$\rho_s = 1 - \frac{6 \times 182}{9(9^2 - 1)} = -0,517$$

Et par conséquent :

$$|\rho_s|_{\text{observé}} = 0,517$$

Une telle valeur observée devra être comparée avec la valeur théorique lue sur la table de Spearman bilatérale au risque de 5%.

Sur la table de Spearman à 9 degrés de liberté et à risque de 5% bilatérale, nous aurons :

$$\rho_{s;0,05} = 0,7$$

Donc :

$$|\rho_s| \text{ observé} = 0,517 < \rho_{s;0,05} \text{ théorique} = 0,7$$

De ce fait, nous constatons que l'ordre chronologique de parution et l'étendue des textes sont dépendants. Après avoir examiné l'étendue N, nous nous intéresserons à l'étendue du vocabulaire.

2. Étendue de V (vocabulaire)

L'étendue du vocabulaire¹⁴⁸ désigne le nombre de vocables. Il est symbolisé par V comme le nombre de mots est symbolisé par N. Selon Ch. MULLER¹⁴⁹, l'étendue du vocabulaire change en fonction de l'étendue des textes, c'est-à-dire quand ce dernier s'allonge le nombre de vocable croît aussi. Elle est également fonction du style, car elle est déterminée par le lexique actualisé dans la situation stylistique. Aussi, le vocabulaire est :

« le reflet du lexique mental d'où il est tiré, de l'ensemble des mots dont l'auteur dispose pour la composition du texte, mots dont chacun est affecté d'une probabilité qui est déterminée à la fois par le lexique mental et par les limites sémantiques et stylistiques de son sujet. »¹⁵⁰

D'une autre manière, l'auteur puise dans son lexique mental pour sélectionner les préférences concernant le vocabulaire et bien évidemment, ce choix se fait en fonction du thème traité et du style dans lequel il s'inscrit.

¹⁴⁸ P. GUIRAUD, *Problèmes et Méthodes de la Statistique Linguistique*, Dordrecht, D.Reidel Publishing Company, 1959, pp. 83-88

¹⁴⁹ Charles MULLER, *Étude de Statistique Lexicale, le Vocabulaire du Théâtre de Pierre de Corneille*, Paris, Larrousse, 1967, p.53.

¹⁵⁰ P. GUIRAUD, *op. cit.*, 1959, p. 84

En fonction de l'étendue du texte, nous procéderons à l'exploration de l'étendue V suivant des calculs relativement compliqués mais qui aboutissent à des résultats crédibles. Tout d'abord, nous commencerons par calculer le vocabulaire théorique de chaque texte, en nous appuyant sur la loi binomiale, pour rendre compte de l'écart existant entre la distribution théorique des vocables (distribution à laquelle nous nous attendons) et la distribution réelle ; et par ricochet, identifier la nature déficitaire ou excédentaire du vocabulaire des différents textes. Ensuite, nous procéderons au calcul de l'écart réduit de la distribution des vocables pour tester si cette distribution est régie par une loi ou non.

2.1. Modèle binomial d'une distribution théorique

En nous basant sur les principes de la loi binomiale, nous allons calculer le vocabulaire théorique de V'. Nous commencerons tout d'abord par l'explication de la démarche de calcul.

Connaissant la distribution des classes de fréquence dans un corpus donné, nous pouvons alors calculer le vocabulaire théorique de l'un des segments qui le compose.

Nous symbolisons par p la probabilité pour qu'une occurrence quelconque du corpus (N) figure dans l'un de ses segments (N') et q la probabilité complémentaire pour ne pas y figurer, nous aurons donc :

$$p = \frac{N'}{N} \quad , \quad q = 1 - p = \frac{N - N'}{N}$$

L'effectif théorique de chaque texte nous donne la possibilité de calculer le nombre des mots qui manquent pour chaque segment du corpus. Ainsi, le

vocabulaire théorique absent du texte N' , représenté par V'_0 , se calcule de la manière suivante :

$$V'_0 = q^1 V_1 + q^2 V_2 + \dots + q^n V_n = \sum_i^n q^i V_i$$

Autrement dit, c'est la somme des vocables de fréquence 1 qui ont une probabilité q^1 d'être absent, plus les vocables de fréquence 2 qui ont une probabilité q^2 d'être absent et ainsi de suite jusqu'à ceux qui ont une fréquence n fois et une probabilité q^n . Le vocabulaire restant serait alors :

$$V'_{théorique} = V - \sum_i^n q^i V_i$$

Appliquons maintenant cette démarche de calcul au premier texte PARCOURS, sachant que :

$$\text{Pour } N = 522421 \text{ et } N' = 64566$$

$$\text{On a : } p = \frac{N'}{N} = \frac{64566}{522421} = 0,12359$$

$$q = 1 - p = 1 - 0,12359 = 0,87641$$

Tableau II. 13 : Application de la loi binomiale à PARCOURS

f_i	V_i	q^i	$q^i V_i$	f_i	V_i	q^i	$q^i V_i$
1	11187	0,876	9804,399	51	25	0,001	0,030
2	4937	0,768	3792,083	52	13	0,001	0,014
3	2544	0,673	1712,534	53	15	0,001	0,014
4	1734	0,590	1023,007	54	14	0,001	0,011
5	1203	0,517	622,017	55	15	0,001	0,011
6	909	0,453	411,915	56	12		
7	712	0,397	282,769	57	8		
8	558	0,348	194,220	58	7		
9	442	0,305	134,831	59	15		
10	380	0,267	101,591	60	14		
11	345	0,234	80,835	61	7		
12	264	0,205	54,212	62	6		
13	263	0,180	47,332	63	8		
14	189	0,158	29,810	64	10		
15	169	0,138	23,361	65	10		
16	148	0,121	17,930	66	13		
17	145	0,106	15,395	67	9		
18	129	0,093	12,004	68	8		
19	126	0,082	10,276	69	6		
20	121	0,071	8,648	70	9		

21	87	0,063	5,450	71	13		
22	91	0,055	4,996	72	8		
23	95	0,048	4,571	73	12		
24	83	0,042	3,500	74	10		
25	61	0,037	2,254	75	4		
26	71	0,032	2,300	76	5		
27	61	0,028	1,732	77	7		
28	60	0,025	1,493	78	7		
29	53	0,022	1,156	79	6		
30	62	0,019	1,185	80	6		
31	56	0,017	0,938	81	6		
32	40	0,015	0,587	82	8		
33	40	0,013	0,515	83	6		
34	33	0,011	0,372	84	4		
35	40	0,010	0,395	85	6		
36	29	0,009	0,251	86	3		
37	17	0,008	0,129	87	2		
38	27	0,007	0,180	88	7		
39	31	0,006	0,181	89	5		
40	35	0,005	0,179	90	11		
41	24	0,004	0,107	91	3		
42	26	0,004	0,102	92	5		
43	27	0,003	0,093	93	5		

44	33	0,003	0,099	94	5		
45	23	0,003	0,061	95	5		
46	13	0,002	0,030	96	5		
47	26	0,002	0,053	97	5		
48	18	0,002	0,032	98	0		
49	17	0,002	0,026	99	5		
50	12	0,001	0,016	100	4		

$$V = \sum_i^{100} V_i = 28198 \quad V'_0 = \sum_i^{100} q^i V_i = 18412,227$$

A ce niveau, nous procédons au calcul du nombre des mots manquants PARCOURS, nous avons, alors :

N'	$V'_{théorique} = V - \sum_i^n q^i V_i$	$V'_{réel}$	Ecart
64566	28638 - 18412 = 10226	9467	-759

Nous remarquons que pour un fragment du corpus de 64566 occurrences, 18412 vocables devraient échapper au tirage, tandis que le reste qui est de 10226 vocables y serait compris. Nous pouvons déduire, alors, que pour un vocable pris au hasard dans le vocabulaire du corpus d'Edmond Amran El Maleh, la probabilité d'échapper au tirage est :

$$P = 18412 / 28638 = 0,6429219918 \text{ soit } 64,2921\%$$

Tandis que la probabilité complémentaire est :

$$q = 1 - p = 1 - 0,65295 = 0,3570780082 \text{ soit } 35,7079\%$$

Nous constatons également que l'écart entre la valeur réelle et la valeur théorique est négatif. Donc le vocabulaire dans *Parcours* est déficitaire.

Après avoir explicité la démarche de calcul du vocabulaire théorique, nous appliquerons le même calcul à tous les fragments de notre corpus. Le résultat obtenu est présenté dans le tableau suivant :

Tableau II. 14 : Écart entre la valeur réelle et la valeur théorique

N°	TEXTES	Valeur réelle	Valeur théorique	Écart
1	PARCOURS	9467	10226	-759
2	AILEN	10322	11192	-870
3	MILLE	10109	11230	-1121
4	ELHAKI	13054	13798	-744
5	GENET	8139	9182	-1043
6	ABOUNOUR	5520	5812	-292
7	ZRIREK	10622	12242	-1620
8	FEMME	1755	2082	-327
9	LETTRES	4681	5236	-555

D'après le tableau ci-dessus, nous découvrons que l'ensemble des textes présentent un écart négatif, c'est-à-dire que la valeur réelle est toujours inférieure

à la valeur théorique, cela montre que nous sommes en présence d'un vocabulaire déficitaire dans l'ensemble des textes.

Dans le point suivant nous procéderons au calcul de l'écart réduit de la distribution des vocables pour voir si cette distribution est aléatoire ou pas.

2.2. Test de l'écart réduit ou z score

L'écart réduit¹⁵¹ constitue un indice qui nous donne la possibilité d'évaluer la spécificité d'une unité linguistique dans un corpus ou dans l'une des parties qui le composent : nous comparons ainsi la fréquence d'une unité dans une partie, à la fréquence à laquelle nous pourrions s'attendre, étant donné sa distribution que nous supposons normale et la composition du corpus global.

Nous commencerons par calculer l'écart réduit de *Parcours Immobile* pour exemplifier la démarche de calcul, avant de présenter l'ensemble des résultats obtenus dans un tableau récapitulatif.

Pour obtenir l'écart réduit, nous devons, tout d'abord, calculer l'écart type :

Sachant que :

$$P = 0,6429219918$$

$$q = 0,3570780082$$

$$V = 28638$$

¹⁵¹ Céline POUDAT, Frédéric LANDRAGIN, *Explorer un corpus textuel : Méthodes – pratiques – outils*, Louvain-la-Neuve, De Boeck Supérieur, 2017, p.156

Nous pouvons déduire aisément la variance théorique :

$$\sigma^2 = Vpq = 28638 \times 0,642921 \times 0,357079 = 6574,528406$$

Et l'écart type sera :

$$\sigma = \sqrt{\sigma^2} = \sqrt{6574,528406} = 81,08346568$$

Nous passons de l'écart déjà calculé qui est de -759 à un écart réduit en le divisant par l'écart type. De ce fait l'écart réduit sera :

$$z = -759/81,083 = -9,36$$

Si nous comparons l'écart réduit que nous avons obtenu pour *Parcours Immobile*, en appliquant cette démarche de calcul, aux résultats fournis par *Hyperbase*, nous allons nous rendre compte que les résultats sont légèrement différents. Cela s'explique du fait que le logiciel *Hyperbase* utilise une autre méthode de calcul beaucoup plus simple. Nous dévoilons par la suite le principe de calcul de l'écart réduit selon la méthode adoptée par *Hyperbase* :

L'écart réduit¹⁵², comme son nom l'indique, se base sur une réduction de la distribution normale. Nous signalons que dans une distribution normale, plus des deux tiers des valeurs se situaient à $\pm$ un écart type de la moyenne et 95% des valeurs se situent à $\pm$ deux écarts type de la moyenne. De ce fait, nous pouvons

¹⁵² Ibid, p.158

considérer que les valeurs qui se situent dans les 5% restants sont significativement sur ou sous représentées.

Nous procédons à la réduction de la distribution afin de rendre les valeurs comparables. Les unités ou les phénomènes linguistiques connaissent en effet des variations sensibles de fréquences. Par exemple, certaines unités grammaticales comme les déterminants définis *le* ou *la* ou encore la préposition *de*, sont généralement parmi les unités les plus fréquentes que nous observons quel que soit le corpus considéré. Elles sont sensiblement plus représentées mais nous voudrions malgré tout les comparer à des unités moins fréquentes.

Nous transformons donc toutes les distributions en des distributions normales réduites dont la moyenne est 0 et l'écart type est unitaire, ce qui permet d'évaluer les valeurs en terme d'écart (type) à la moyenne.

Pour effectuer cette transformation, Nous nous référons à la formule suivante représentant toutes les valeurs de la distribution observée :

$$\text{écart réduit} = \frac{X - \text{moyenne}}{\text{écart type}}$$

En ce qui concerne les données textuelles, la formule a dû être adaptée puisqu'à moins de travailler sur des textes de longueur comparable, la fréquence d'une unité dépend de la longueur du texte, ce qui rend le calcul de la moyenne peu pertinent.

C'est donc l'écart entre la fréquence observée (X) et la fréquence théorique (ou attendue étant donnée la taille de la partie et la taille du corpus) qui est mesuré et l'écart réduit se calcule au moyen de la formule suivante¹⁵³ :

$$Ecart\ réduit = \frac{fréquence\ observée - fréquence\ théorique}{\sqrt{fréquence\ théorique}}$$

Calculons donc l'écart réduit de *Parcours immobile* en suivant cette nouvelle démarche :

$$\begin{aligned} Ecart\ réduit &= \frac{fréquence\ observée - fréquence\ théorique}{\sqrt{fréquence\ théorique}} = \\ &= \frac{9467 - 10226}{\sqrt{10226}} = -7,5 \end{aligned}$$

Nous avons reproduit le même calcul pour tous les autres textes, les résultats obtenus sont présentés dans le tableau suivant :

¹⁵³ Il s'agit d'une formule simplifiée, puisqu'on admet certaines approximations, comme le fait de considérer que la variance est égale à la fréquence théorique. Voir notamment : Ch. Muller (1977/1992), **op.cit.**, p.p. 51-52

Tableau II. 15 : Nombres de mots qui manqueront dans le corpus

N°	TEXTES	Valeur	Valeur théorique	Écart	Écart
1	PARCOURS	9467	10226	-759	-7,51
2	AILEN	10322	11192	-870	-8,22
3	MILLE	10109	11230	-1121	-10,58
4	ELHAKI	13054	13798	-744	-6,33
5	GENET	8139	9182	-1043	-10,88
6	ABOUNOUR	5520	5812	-292	-3,83
7	ZRIREK	10622	12242	-1620	-14,64
8	FEMME	1755	2082	-327	-7,17
9	LETTRES	4681	5236	-555	-7,67

Avant d'interpréter le résultat, nous commencerons par le test paramétrique où nous formulerons les hypothèses suivantes :

$$\begin{cases} H_0: \text{les vocables se répartissent au hasard dans le texte} \\ \# \\ H_1: \text{la répartition des vocables n'est pas due au hasard} \end{cases}$$

D'où la règle de décision suivante :

$$\begin{cases} \text{Si} & Z > Z_\alpha & \text{on rejette } H_0 \\ \text{Si} & Z \leq Z_\alpha & \text{on ne peut rejeter } H_0 \end{cases}$$

Nous constatons que les écarts réduits sont tous négatifs ; nous devons donc comparer la valeur absolue de Z avec le fractile d'ordre α (que l'on note Z_α) lu sur la table de la loi normale centrée et réduite, sachant que α est le risque de rejeter l'hypothèse nulle alors qu'elle est vraie.

D'après la table de la loi normale centrée et réduite, nous avons :

$$P(Z > z_{1,65}) = 0,0495 \text{ et } P(Z > z_{1,64}) = 0,0505$$

Nous cherchons le fractile $Z_{0,05}$ c'est à dire :

$$P(Z > z_{0,05}) = 0,05$$

Par interpolation linéaire, nous aurons :

$$0,0495 < 0,05 = \alpha < 0,0505$$

$$\text{Et} \quad 1,64 < Z < 1,65$$

Donc :

$$\frac{Z - 1,64}{0,05 - 0,0495} = \frac{1,65 - 1,64}{0,0505 - 0,0495}$$

$$\text{Par conséquent} \quad Z = 1,64 + (0,05 - 0,0495) \frac{1,65 - 1,64}{0,0505 - 0,0495} = 1,645$$

Ceci veut dire que l'écart réduit qui correspond au risque de 5% de rejeter l'hypothèse nulle alors qu'elle est vraie, est :

$$Z_{0,05} = 1,645 \text{ (Voir Annexe « Table de la loi normale centrée et réduite »)}$$

Et puisque la valeur absolue de Z de tous les segments du corpus est supérieure à 1,645, nous pouvons rejeter H_0 en faveur de H_1 , c'est à dire que la répartition des vocables dans notre corpus n'est pas le simple jeu du hasard, mais elle pourrait être le produit des effets du style, du thème et/ou autres.

Une autre remarque s'impose, c'est que l'ensemble des textes présentent un écart négatif, c'est-à-dire la valeur réelle est toujours inférieure à la valeur

théorique, cela montre que nous sommes en présence d'un vocabulaire déficitaire dans l'ensemble des textes. Nous pourrions, *a fortiori*, souligner les moins déficitaires d'entre eux, à savoir : ELHAKI, ANOUNOUR, FEMME, PARCOURS. Cela dit que l'auteur aurait usé de presque tout son lexique dans ces textes.

En nous basant sur la valeur des écarts, nous pouvons classer les textes de la manière suivante :

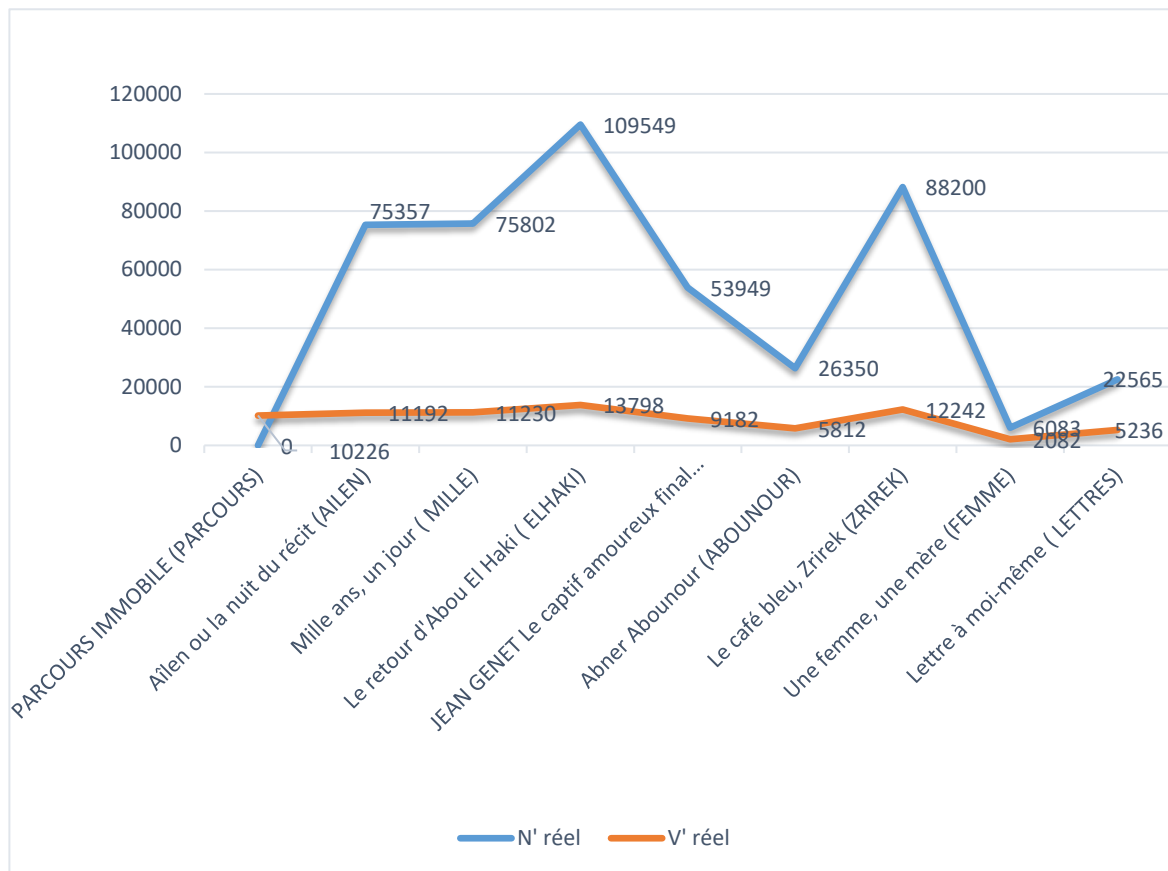
ABOUNOUR > FEMME > LETTRES > *ELHAKI* > PARCOURS > *AILEN* >
GENET > *MILLE* > *ZRIREK*

Si nous prenons en considération seulement l'effectif réel, nous aurons le classement suivant qui est tout à fait différent du précédent :

ELHAKI > *ZRIREK*) > *AILEN* > *MILLE* > PARCOURS > *GENET* >
 ABOUNOUR) > LETTRES > FEMME

Le graphique présenté ci-dessous met en évidence l'évolution chronologique à la fois de N'et de V' réel des textes du corpus.

Figure II. 3 : Progression chronologique de N'et V' réel



Nous constatons que les deux courbes croissent et décroissent de la même manière même si les valeurs de N'et de V' demeurent très éloignées. Appliquons, par ailleurs, le test de Spearman afin de dégager la corrélation entre l'ordre chronologique des textes et l'étendue du vocabulaire.

Avant de procéder au calcul, nous commençons par émettre les hypothèses suivantes :

- $\{ H_0: \text{l'ordre chronologique et l'étendue } V' \text{ sont indépendants}$
- $\{ H_1: \text{l'ordre chronologie et l'étendue } V' \text{ sont liés}$

Si $|\rho_s| \text{ observé} \geq \rho_{s;\alpha} \text{ théorique}$ (lu sur la table de Spearman), nous pouvons rejeter H_0 en faveur de H_1 qui stipule qu'il y a une liaison significative entre les deux variables.

Nous calculons ρ_s le coefficient de corrélation des rangs de Spearman en nous référant au tableau des rangs suivant :

Tableau II. 16 : Test de Spearman

TEXTES	Ordre chronologique	V' réel	Rang de l'étendue	Distance	Distance au carrée
PARCOURS	1	9467	5	-4	16
AILEN	2	10322	3	-1	1
MILLE	3	10109	4	-1	1
ELHAKI	4	13054	1	3	9
GENET	5	8139	6	-1	1
ABOUNOUR	6	5520	7	-1	1
ZRIREK	7	10622	2	5	25
FEMME	8	1755	9	-1	1
LETTRES	9	4681	8	1	1
			$\sum di^2$		56

Donc, nous obtenons :

$$\rho_s = 1 - \frac{6 \times 56}{9(9^2 - 1)} = 0,533$$

Et par conséquent :

$$|\rho_s| \text{ observé} = 0,533$$

Une telle valeur observée devra être comparée avec la valeur théorique lue sur la table de Spearman bilatérale au risque de 5%.

Sur la table de Spearman à 9 degrés de liberté et à risque de 5% bilatérale, nous avons : (Voir Annexe « Table de la loi Spearman bilatérale »)

$$\rho_{s;0,05} = 0,7$$

Donc :

$$|\rho_s| \text{ observé} = 0,5337 < \rho_{s;0,05} \text{ théorique} = 0,7$$

Le résultat obtenu ne laisse aucune chance à l'hypothèse alternative, c'est-à-dire que l'ordre chronologique de parution et l'étendue V' sont indépendants.

Si nous prenons en considération l'ordre croissant de V', nous aurons, alors, un coefficient de corrélation égal à -0,537, ce n'est que le signe qui change mais l'interprétation sera la même étant donné qu'elle se fait sur la base de la valeur absolue.

Dans ce premier chapitre, nous avons étudié le phénomène de l'étendue suivant deux optiques différentes mais complémentaires. Dans un premier temps, nous nous sommes focalisée sur la longueur du mot, dans un deuxième temps, nous nous sommes intéressée à la question de l'étendue des occurrences et celle des vocables.

Ainsi, dans la première section, celle de la longueur du mot, nous avons étudié les différentes classes de longueur, tout en faisant le rapprochement entre les données observées et la distribution effective des groupes de longueur dans l'œuvre étudiée.

De ce fait, nous avons calculé en premier lieu la moyenne des caractères par mot. Pour un total de 2 039 789 caractères et de 522 421 occurrences, nous avons obtenu une moyenne de 3,90 lettres. Après avoir établi une comparaison entre les moyennes calculées pour les différents textes du corpus, nous nous sommes rendu compte que la moyenne, comme paramètre de position, reste insuffisante pour réaliser une comparaison, nous avons dû, par ailleurs, faire intervenir un paramètre de dispersion, à savoir l'écart type pour mesurer le degré d'éparpillement des observations. Néanmoins, cette démarche s'est avérée trop oblongue, d'autant plus lorsqu'elle est adoptée pour des corpus qui sont formés par un grand nombre de textes. Donc nous nous sommes trouvée dans l'obligation de compléter notre analyse par des tests basés sur les effectifs théoriques et les effectifs réels des mots. Nous avons constaté par la suite que la distribution des lettres dans presque tous les textes du corpus n'est pas le résultat d'un simple hasard, mais au contraire, elle est le produit des effets du genre, du thème et/ou autres.

De la même manière, nous avons étudié les classes de longueur qui allaient des mots composés d'une seule lettre jusqu'à ceux qui sont formés de 11 lettres ou plus. Nous nous sommes aperçue que les mots courts sont les plus abondants, chose qui n'est pas surprenante, du moment que les mots les plus courts (lg2) sont généralement les plus fréquents dans la langue.

Nous avons affiné notre étude par le recours à l'analyse factorielle des correspondances qui nous a donné une vue d'ensemble sur le regroupement des classes de longueur. Nous avons établi que la période de la production n'a pas d'effet sur la distribution des classes de longueur, vu que les textes regroupés en constellations sont produits pendant des périodes distanciées. De ce fait, nous avons cherché d'autres explications en nous basant sur les études effectuées par BRUNET¹⁵⁴ et PARADA-RAMIREZ¹⁵⁵. Ce qui a abouti aux résultats suivants :

- Les textes GENET et ZRIREK attirent les mots courts (lg1, lg2), cependant ils favorisent un emploi fort des mots longs (lg9, lg10 et lg11), ce qui pourrait s'expliquer par le fait que ces textes forment un discours intellectuel avec un style recherché¹⁵⁶. Par conséquent, nous trouvons que ce constat renforce l'idée, avancée par les critiques, selon laquelle les textes GENET et ZRIREK sont bien des essais.
- MILLE, AÎLEN et PARCOURS privilégient les mots qui sont en faveur d'un registre familier, à savoir les mots courts et moyens (lg3, lg5, lg6 et

¹⁵⁴ Etienne BRUNET, *op.cit.*, 1789, p.246.

¹⁵⁵ José Gregorio PARADA RAMIREZ, *op.cit.*, 2013, p. 53

¹⁵⁶ Quoique nous avons signalé l'attraction des deux textes pour la lg1 et lg2, ça n'a aucune influence sur l'interprétation que nous avons suggéré, parce que les mots courts sont les plus fréquents dans la langue, et c'est généralement les mots grammaticaux qui font que la fréquence de cette classe de longueur soit élevée.

- lg7). Par ailleurs, l'utilisation de ces classes de longueur est tout à fait normale du moment qu'il s'agit de textes appartenant au genre romanesque.
- Le texte ELHAKI qui est à cheval entre les mots longs ainsi que les mots courts et moyens est aussi un roman, cependant son attraction pour les mots longs pourrait se justifier par l'utilisation de l'auteur d'un discours soutenu abordant les écritures mystiques de certaines figures intellectuelles prestigieuses qui cheminent le lecteur dans des espaces et des périodes pluriels.
 - Les textes ABOUNOUR (Nouvelle) et FEMME (Essai) marquent une symétrie entre deux groupes de longueur : les mots longs et courts (lg9, lg10, lg11, lg1, lg2) et les mots courts et moyens (lg3, lg5, lg6 et lg7). Le texte LETTRE (récit épistolaire) n'appartient à aucune constellation.

Dans la seconde section, nous nous sommes penchée sur l'analyse de l'étendue du texte. Nous avons commencé tout d'abord par présenter le classement des textes selon l'étendue N, puis nous avons étudié la corrélation entre le nombre d'occurrences de chaque segment et l'ordre chronologique de parution des textes en utilisant le test de corrélation des rangs de Spearman. Nous avons conclu que les deux variables sont dépendantes l'une de l'autre.

En effet, nous avons constaté que le développement de l'étendue est articulé en trois phases. Dans la première, l'étendue des quatre premiers textes évoluent selon une tendance haussière, pour atteindre le pic dans ELHAKI, puis nous assistons à une phase de tendance baissière qui prend fin par ABOUNOUR, avant d'effectuer un second pic dans ZRIREK. Finalement nous assistons à une décroissance relative aux textes FEMME et LETTRES, ceux dont l'étendue est la plus faible de toutes les partitions composant notre corpus.

Le caractère élevé de l'étendue des quatre premiers textes d'El Maleh, peut-être expliqué par son acharnement pour l'écriture. En effet, notre auteur était très sensible à la variable « temps », il se voyait comme une matière vouée à la déliquescence et voulait étaler dans ses textes, avant que la fin ne le surprenne, tous ce qu'il a envie d'exprimer. C'est la raison pour laquelle les premiers écrits d'El Maleh se caractérisent par une étendue élevée comparativement aux autres textes qui sont publiés après.

Cependant, l'étendue élevée de ZRIREK qui marque la fin de la deuxième phase peut être due à des raisons d'ordres thématiques. C'est vrai que les productions d'El Maleh abordent plus ou moins les mêmes thèmes, cependant dans *Le Café bleu*, *Zrerek* nous assistons à une concentration d'idées sur la littérature et l'écriture allégorique. A notre humble avis c'est ce qui peut justifier la hausse de l'étendue au niveau de la deuxième phase.

De la même façon, nous avons étudié l'étendue V, en nous appuyant sur la loi binomiale, pour identifier la nature déficitaire ou excédentaire du vocabulaire des différentes partitions. Nous avons déduit, alors, que nous sommes en présence d'un vocabulaire déficitaire dans l'ensemble du corpus. Néanmoins, nous avons mis en évidence les textes les moins déficitaires, à savoir : ELHAKI, ABOUNOUR, FEMME, PARCOURS. Cela dit que l'auteur aurait utilisé de presque tout son lexique dans ces textes.

Ensuite, nous avons procédé au calcul de l'écart réduit de la distribution des vocables pour tester si cette distribution est régie par une loi ou non. Nous avons prouvé, à travers des calculs complexes mais fiables, que la répartition des vocables dans notre corpus n'est pas le simple jeu du hasard mais elle pourrait être le produit des effets du style et du thème traité.

En d'autres termes, le lexique personnel d'El Maleh (vocabulaire) ne dépend pas de la date de parution de ses textes. Ce qui pourrait laisser entendre que l'auteur diversifie ses choix thématiques. En effet, si nous nous référons au contenu des textes, nous pourrions facilement constater que l'auteur alterne parfois entre certains thèmes. Par exemple dans ses premières productions, il est question de : la mémoire, l'histoire, la culture, l'exode des marocains juifs, le drame palestinien, etc. Cependant, dans GENET, à travers les thèmes de la littérature et l'écriture allégorique, El Maleh analyse son propre style d'écriture tout en le comparant avec d'autres écrivains, notamment, Jean Genet qui se taille la part du lion dans cette analyse. Par ailleurs, dans ABOUNOUR l'auteur redécouvre encore une fois les valeurs culturelles marocaines ancestrales. En arrivant à ZRIREK nous retrouvons la tragédie palestinienne, mais également l'écriture allégorique qui domine presque tout le contenu de ce texte. D'ailleurs, ce dernier évoque des écrivains dont les œuvres se caractérisent par la force et la puissance d'une écriture allégorique. Nous y trouvons également des réflexions sur la langue et la littérature. Ainsi, nous pouvons conclure que les thèmes abordés par El Maleh ne suivent pas une trajectoire continue.

Une fois nous avons étudié le phénomène de la longueur du mot et celui de l'étendue du corpus, nous nous intéresserons, dans le chapitre suivant, à l'étude de la richesse lexicale et l'accroissement du vocabulaire.

CHAPITRE 2

Richesse lexicale et accroissement du vocabulaire

La structure lexicale peut être abordée sous deux aspects différents, l'un statique et l'autre dynamique. Le premier aspect, la richesse lexicale, traite le corpus étudié « *comme un ensemble clos et achevée* »¹⁵⁷. Elle permet de décrire la structure lexicale d'un texte quelconque. Elle consiste à comparer un certain nombre de textes qui peuvent être d'étendues égales ou inégales et de qualifier ceux qui sont riches de ceux qui sont pauvres. Notons qu'il n'y a aucune péjoration dans « pauvre » ni valorisation dans « riche », ces deux notions ne reflètent que l'utilisation ou la non utilisation du vocabulaire contenu dans le lexique d'une langue donnée.

Loin d'une conception statique des textes et qui considère le corpus comme une chose achevée, nous nous intéresserons, par la suite, à l'accroissement du vocabulaire qui met l'accent sur le caractère dynamique du corpus. Il est donc question d'étudier « *la façon dont le vocabulaire d'un texte se constitue quand on suit ce texte du premier mot au point final* »¹⁵⁸. D'une autre manière, l'étude de l'accroissement du vocabulaire détermine l'évolution de ce dernier au fur et à mesure que l'œuvre progresse. Donc, c'est une exploration qui se fait d'un point

¹⁵⁷ Charles MULLER, *Principes et Méthodes de La Statistique Lexicale*, Paris, Classiques Hachette, 1992, p. 116

¹⁵⁸ Charles MULLER, *op.cit.*, 1967, p. 67

de vue chronologique. Elle détermine l'apport du vocabulaire au fil du temps et met en évidence, pour un fragment déterminé du texte, les unités qui n'ont pas été employées antérieurement mais qui apparaissent dans le fragment choisi. Notons que pour effectuer ce type d'investigation, le corpus doit être découpé en tranche et l'évaluation de l'accroissement du vocabulaire se réalise par l'accroissement cumulé de tous les segments composant le corpus.

Nous commencerons par l'étude du corpus sous le premier angle, à savoir la richesse du vocabulaire, puis nous enchaînerons sur le second qui abordera l'accroissement du vocabulaire.

A. Richesse lexicale

La définition de la richesse lexicale a été longuement débattue dans les débuts des études lexicométriques, mais l'on est arrivé à accepter la définition qui voit dans la richesse lexicale un « *lieu de comparaison entre deux ou plusieurs textes en fonction de leur étendue respective et du nombre de vocables relevés dans chacun d'eux* »¹⁵⁹. Autrement dit, il s'agit d'un outil qui permet de décrire la structure lexicale d'une œuvre donnée ou de la comparer à d'autres corpus. La richesse lexicale est donc un fait de structure, elle est indépendante du contenu¹⁶⁰.

On n'a pas attendu les études quantitatives pour qualifier de « riche » ou de « pauvre » le vocabulaire d'un écrivain ou d'une œuvre littéraire, écrit Charles Muller¹⁶¹. En effet, cette appréciation qui reste subjective se fonde sur la présence,

¹⁵⁹ Ménard NATHAN, *Mesure de la richesse lexicale. Théorie et vérifications expérimentales : Etudes stylométriques et sociolinguistiques*, Genève, Slatkine, 1979, p. 16.

¹⁶⁰ Charles MULLER, *op. cit.*, 1967, p.115

¹⁶¹ Ibid

dans le corpus, de quelques vocables jugés rares, ou bien, au contraire, sur l'absence de tels éléments du lexique. Néanmoins, une proportion considérable de vocables rares, dans un texte donné, ne signifie pas nécessairement qu'il s'agit d'un vocabulaire riche. D'où la nécessité de recourir au traitement mathématique de la richesse lexicale pour apporter une solution que l'on estime objective à un problème qui n'a reçu que des réponses impressionnistes et subjective¹⁶².

Avant d'enchaîner sur ce qui suit, une clarification terminologique s'impose : par richesse lexicale, il ne faut pas entendre richesse du lexique, mais plutôt la richesse de ce que Charles Muller appelle le « lexique de situation »¹⁶³ c'est-à-dire du vocabulaire. Donc, richesse lexicale équivaut à richesse du vocabulaire.

Pour mesurer la richesse lexicale de notre corpus, la lexicométrie nous fournit à cet égard plusieurs méthodes qui cherchent, en quelque sorte, à apprécier l'aspect statique de la structure lexicale d'une manière objective. Vu que le nombre des méthodes, proposées par les spécialistes, est assez important, il serait judicieux de choisir et de tester celles qui sont les plus appropriées. De toutes les méthodes mises en œuvre, nous avons opté pour quelques-unes afin de les appliquer à notre corpus.

La première méthode qui reste la plus traditionnelle en statistique lexicale est celle qui évalue les relations qui existent entre la longueur d'un corpus (N), l'étendue de son vocabulaire (V) et l'effectif de ses hapax. C'est une méthode

¹⁶² Ch. BERNET, « *Faits lexicaux, Richesse du vocabulaire. Résultats.* », In : Ph Thoirô, D. Labbé, D. Serant (éds). Paris-Genève, Champion – Slatkine, 1988, p. 2.

¹⁶³ Charles MULLER, *op. cit.*, 1967, pp. 44-46.

basique à laquelle nous faisons appel pour avoir une première impression sur la richesse lexicale de notre corpus. Quant aux autres méthodes, il s'agit en fait de la méthode binomiale de MULLER et de l'indice w d'Etienne Brunet. Notre choix a été mené par deux considérations : premièrement parce que ces techniques ont pu satisfaire un grand nombre de spécialistes reconnus, deuxièmement parce qu'elles ont marqué un moment important de la recherche en lexicométrie.

Passons donc à l'application de ces méthodes à notre corpus. Nous allons commencer, tout d'abord par l'évaluation qualitative de V d'après N , puis enchaîner sur la méthode binomiale de Muller, pour finir avec l'indice w de Brunet.

1. Appréciation qualitative de V d'après N

Rappelons que nous appelons N l'étendue du corpus qui représente le nombre des occurrences, V l'étendue du vocabulaire ou le nombre de vocables et V_1 l'effectif des hapax, c'est-à-dire les mots employés une seule fois.

1.1. Rapport V/N

Tableau II. 17 : Calcul du rapport V/N

TEXTES	V	N	V/N	Rang
PARCOURS	9467	64566	0,147	5
AILEN	10322	75357	0,137	6
MILLE	10109	75802	0,133	7
ELHAKI	13054	109549	0,119	9
GENET	8139	53949	0,151	4
ABOUNOUR	5520	26350	0,209	2

ZRIREK	10622	88200	0,120	8
FEMME	1755	6083	0,289	1
LETTRES	4681	22565	0,207	3

Une lecture succincte du tableau montre que ce ratio est très sensible aux valeurs de l'étendue des textes N. Le quotient est plus élevé dans le texte le plus court, à savoir FEMME. Selon cette formule les textes tendent à garder leurs positions déterminées par l'étendue N de leurs textes.

Cela dit, qu'en sera-t-il du rapport V_1/N ?

1.2. Rapport V_1/N

Tableau II. 18 : Calcul du rapport V_1/N

TEXTES	V	V1	N	V_1/N	Rang
PARCOURS	9467	1551	64566	0,024022	4
AILEN	10322	1914	75357	0,025399	2
MILLE	10109	1714	75802	0,022612	6
ELHAKI	13054	2690	109549	0,024555	3
GENET	8139	496	53949	0,009194	9
ABOUNOUR	5520	517	26350	0,019620	7
ZRIREK	10622	1548	88200	0,017551	8
FEMME	1755	140	6083	0,023015	5
LETTRES	4681	617	22565	0,027343	1
Corpus	28638	11187	522421	0,021414	

En dépit du changement des rangs des différents textes, le fragment LETTRES, marqué par une étendue faible, occupe le premier rang, ELHAKI s'est déplacé du neuvième rang¹⁶⁴ au troisième. Mais, il n'y a aucune logique derrière ce constat, donc ce rapport également reste influencé par l'étendue du texte et ne rend pas compte de la richesse lexicale. Ce constat est déjà affirmé par Etienne Brunet dans ce qui suit :

*« Malheureusement ces quotients, ne reflètent pas seulement les variations de la richesse ou de la pauvreté du vocabulaire. Un facteur indésirable perturbe les résultats : c'est l'influence de l'étendue de l'œuvre : le quotient est plus élevé dans le texte les plus courts (dans un texte de moins de vingt mots, c'est-à-dire au niveau de la phrase, il avoisine 1), il s'abaisse dans les textes longs (jusqu'à 0,001 dans le grand corpus du Trésor de la langue française). »*¹⁶⁵

Ces deux quotient V/N et V_1/N ont le même dénominateur commun N , donc ils varient en fonction de l'étendue N . A priori nous pouvons avancer que l'utilisation de l'étendue, dans le calcul des ratios, perturbe les résultats, d'où la nécessité de mettre en évidence un quotient qui ne soit pas exprimé en fonction de N .

1.3. Rapport V_1/V

Tout en cherchant à éliminer le facteur parasite qui influence les quotients, à savoir l'étendue des textes, P. Guiraud propose quelques formules en vue d'atténuer ces effets. Il s'agit notamment du rapport V_1/V .

164 Tableau n° 15 : Calcul du rapport V/N

165 Etienne BRUNET., *op.cit.*, 1978, p.23.

Tableau II. 19 : Calcul du rapport V_1/V

TEXTES	V	V ₁	V ₁ /V	Rang
PARCOURS	9467	1551	0,163832259	4
AILEN	10322	1914	0,18542918	2
MILLE	10109	1714	0,169551884	3
ELHAKI	13054	2690	0,206067106	1
GENET	8139	496	0,060941148	9
ABOUNOUR	5520	517	0,09365942	7
ZRIREK	10622	1548	0,145735266	5
FEMME	1755	140	0,07977208	8
LETTRES	4681	617	0,131809442	6
CORPUS	28638	11187	0,390634821	

L'utilisation de ce ratio a permis d'affiner les résultats, en effet, GENET occupe le premier rang en terme de richesse lexicale et FEMME le huitième rang, mais ce résultat reste insuffisant puisque V et V₁ sont eux mêmes influencés par N, d'ailleurs E. Brunet le confirme en disant :

« La première formule tend à mesurer la dispersion du vocabulaire en comparant le nombre de mots employés une fois au nombre de vocable. Mais, comme ce rapport varie avec l'étendue du vocabulaire en s'amenuisant quand celle-ci croît, un correctif a été proposé, qui fait varier V₁ et V d'un pas différent (V₁ est porté au cube et V au carré). »¹⁶⁶

¹⁶⁶ Etienne BRUNET, *op.cit.*, 1978, p.23.

Pour réduire l'effet de l'étendue, Pierre Guiraud¹⁶⁷ propose deux nouvelles formules dont l'une aboutit à l'autre :

Si nous sommes en face de deux textes a et b de longueurs inégales, remarquant que $\frac{V_1^3}{V^2} = \text{constante}$, connaissant le rapport $p = N_a / N_b$, nous pouvons prévoir quel sera respectivement le vocabulaire V_b et les hapax V_{1b} du texte b en appliquant les formules suivantes :

$$V_b = \sqrt[2]{p} V_a \text{ et } V_{1b} = \sqrt[3]{p} V_{1a}$$

Nous n'appliquons pas cette formule à notre corpus pour deux raisons, premièrement parce qu'E. Brunet remarque que cette formule est impraticable lorsque le texte étudié est de dimension réduite. Deuxièmement, parce qu'il démontre que la première formule aboutit à une seconde qui est la suivante : $V/\sqrt{N} = \text{constante}$.

Dans ce qui suit, nous appliquerons la deuxième formule de Guiraud à notre corpus.

1.4. Formule de GUIRAUD

En voulant neutraliser l'effet des perturbations stylistiques, P. GUIRAUD, a opté pour un indice qui s'appuie essentiellement sur N et V . Certes cette méthode¹⁶⁸ n'est pas non plus indépendante de l'étendue des textes, mais elle

¹⁶⁷ Pierre GUIRAUD, *Problèmes et méthodes de la statistique linguistique*, p.89, cité par Brunet, 1978, *op.cit.*, p.p. 24-27.

¹⁶⁸ Zoubeïr MOUELHI, « *La Richesse Lexicale dans une Perspective de Lexicométrie Arabe : Étude contrastive de 5 méthodes de mesure : Application au « Imtâ wa-l-muânasa de Tawhîdî »* », in : revue électronique Texte et corpus, n°3, Août 2008, Actes des journées de la linguistique de Corpus, 2007, p.p. 271-284

demeure moins influencée par cette dernière, du moment qu'elle ne met pas en quotient V et N mais plutôt V et la racine carré de N, et ce, dans le but de minimiser l'influence de l'étendue sur la mesure de la richesse lexicale.

La formule proposée par P. GUIRAUD pour calculer la richesse lexicale est très simple :

$$R = \frac{V}{\sqrt{N}}$$

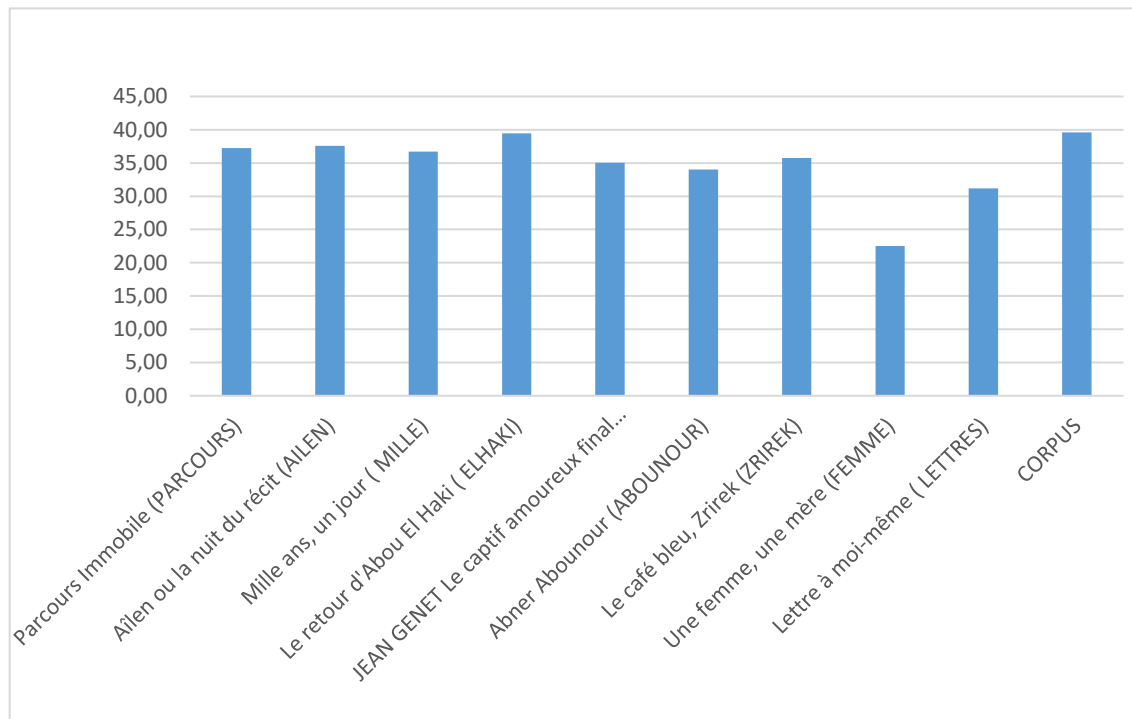
Le tableau, ci-dessous, présente les résultats obtenus suite à l'adoption de la formule sus-cité. Dans l'ordre des colonnes, nous avons mentionné les différents textes classés par ordre chronologique, le nombre de vocable V de chaque texte ainsi que le nombre d'occurrence N, le ratio $\frac{V}{\sqrt{N}}$ qui représente l'indice de la richesse lexicale et enfin, dans la toute dernière colonne, nous avons classé l'indice calculé par ordre décroissant.

Tableau II. 20 : Richesse lexicale selon la formule de GUIRAUD

TEXTES	Nombre de vocable (V)	Nombre d'occurrences (N)	$\frac{V}{\sqrt{N}}$	Rang
PARCOURS	9467	64566	37,26	3
AILEN	10322	75357	37,60	2
MILLE	10109	75802	36,72	4
ELHAKI	13054	109549	39,44	1
GENET	8139	53949	35,04	6
ABOUNOUR	5520	26350	34,01	7
ZRIREK	10622	88200	35,77	5
FEMME	1755	6083	22,50	9
LETTRES	4681	22565	31,16	8
Total	28638	522421	39,62	

D'après ce tableau, nous constatons que le texte, le plus court, FEMME a enregistré une valeur minimale avec un indice de 22,5 tandis que le texte ELHAKI, jugé le plus long, a enregistré une valeur maximale de 39,44. De même, l'application de la formule sur l'ensemble du corpus donne la valeur de 39,62. Cette valeur obtenue reste très proche de celle du texte le plus long à savoir ELHAKI.

Figure II. 4 : Richesse lexicale selon la formule de P. GUIRAUD



Nous résumons à travers l'histogramme présenté ci-dessus, d'une façon visuelle, le classement des textes en nous basant sur la richesse lexicale obtenue selon la formule de P. GUIRAUD.

À la lumière de cet histogramme, nous constatons que l'étendue des textes a une influence très nette sur la richesse lexicale. En effet le texte le plus long est celui qui a l'indice le plus grand et l'indice le plus faible est réservé au texte le plus court.

Quoique cette méthode soit estimée comme étant l'outil le plus simple pour apprécier la richesse lexicale, dans la mesure où elle requiert peu de données et un minimum d'opérations mais, elle enregistre en quelque sorte une certaine déficience du moment que tous les quotients étudiés souffrent de l'effet parasite de l'étendue. Par le mot déficience nous ne voulons pas exprimer un refus catégorique de l'utilisation de cette méthode ; en effet, les ratios que nous avons présentés peuvent servir comme un outil d'exploration¹⁶⁹ qui peut rendre au chercheur en lexicométrie d'énormes services, ne serait-ce que pour choisir quelle autre méthode utiliser en tandem avec elle.

Normalement, nous devrions vérifier par la suite la formule de Yule-Herdan, mais il s'avère, d'après plusieurs chercheurs, entre autres, A. Hefied, que cette formule donne des résultats décevants. En effet, ce dernier stipule qu' : « *elle propose un indice basé sur la fréquence moyenne($\bar{f} = N/V$) : $\sum V_i = \frac{(\bar{f}_i - \bar{f})^2}{v}$, mais les résultats en sont décevants. D'ailleurs elle ne mesure que les hautes fréquences et doit donc être abandonnée au profit de la loi binomiale* »¹⁷⁰.

De ce fait, nous allons suivre le même cheminement et passer à l'application de la loi binomiale de MULLER.

¹⁶⁹ Z. MOUELHI, *op.cit.*, 2007, p. 491

¹⁷⁰ Ali HEFIED, *op. cit.*, p.149

1.5.Méthode binomiale

La méthode binomiale que propose MULLER pour mesurer la richesse lexicale se base sur la distribution des fréquences. En adoptant une approche comparative, cette méthode permet d'apprécier la richesse lexicale de deux textes *a* et *b* et de montrer qu'un texte *a* est plus riche qu'un texte *b* si l'une des conditions suivantes est remplie :

- ✓ $V_a \geq V_b$ quand $N_a \leq N_b$
- ✓ $(N_a - N_b) < (V_a - V_b)$ quand $V_a > V_b$ et $N_a > N_b$
- ✓ $(N_a / V_a) < (N_b / V_b)$ quand $V_a > V_b$ et $N_a > N_b$

Si nous appliquons ces formules à notre corpus, en accouplant les textes qui ont des longueurs voisines, nous obtiendrons : (AÎLEN, MILLE) ; (PARCOURS, GENET) ; (ABOUNOUR, FEMME) ; (ELHAKI, ZRIREK) ; (ABOUNOUR, FEMME). Nous procéderons, par la suite, à la vérification de chaque hypothèse pour chaque binôme :

- ✓ **Première condition : $V_a \geq V_b$ quand $N_a \leq N_b$**

Prenons comme exemple le binôme (AÎLEN, MILLE), nous avons :

$$N_{AÎLEN}=75357 < N_{MILLE}=75802$$

$$V_{AÎLEN}=10322 > V_{MILLE}=10109$$

Donc le vocabulaire de **AÎLEN** est plus riche que celui de **MILLE**. Ce raisonnement n'est pas valable pour les autres binômes, car la différence des étendues entre les textes est importante.

✓ **Deuxième condition :** $(N_a - N_b) < (V_a - V_b)$ quand $V_a > V_b$ et $N_a > N_b$:

Nous allons appliquer cette condition au binôme (PARCOURS, GENET), nous aurons donc :

$$N_{\text{PARCOURS}} = 64566 > N_{\text{GENET}} = 53949$$

$$V_{\text{PARCOURS}} = 9467 > V_{\text{GENET}} = 8139$$

$$N_{\text{PARCOURS}} - N_{\text{GENET}} = 64566 - 53949 = 10617$$

$$V_{\text{PARCOURS}} - V_{\text{GENET}} = 9467 - 8139 = 1328$$

$$N_{\text{PARCOURS}} - N_{\text{GENET}} = 10617 > V_{\text{PARCOURS}} - V_{\text{GENET}} = 1328$$

Cette condition n'est pas vérifiée pour les autres binômes.

✓ **Troisième condition :** $(N_a / V_a) < N_b / V_b$ quand $V_a > V_b$ et $N_a > N_b$

Reprenons le même exemple du binôme (PARCOURS, GENET) :

$$N_{\text{PARCOURS}} = 64566 > N_{\text{GENET}} = 53949$$

$$V_{\text{PARCOURS}} = 9467 > V_{\text{GENET}} = 8139$$

$$\frac{N_{\text{PARCOURS}}}{V_{\text{PARCOURS}}} = \frac{64566}{9467} = 6,82$$

Et

$$\frac{N_{\text{GENET}}}{V_{\text{GENET}}} = \frac{53949}{8139} = 6,63$$

Donc, cette troisième condition n'est pas remplie par le couple (PARCOURS, GENET), de même pour les autres couples.

Généralement, ces conditions sont vérifiées pour les textes qui ont un faible écart en N , cependant une fois l'écart devient important, la vérification de ces trois conditions sera remise en question. Chose qui nous conduit à l'utilisation de la fréquence moyenne du vocabulaire¹⁷¹ $\bar{f}$ qui s'obtient grâce à l'équation suivante $\bar{f} = \frac{N}{V}$ et selon la déduction qui affirme que : si pour deux textes a et b , nous avons $N_a > N_b$, $V_a > V_b$ et $\bar{f}_a < \bar{f}_b$, le texte a est, alors, plus riche que le texte b .

Dans le tableau ci-dessous, nous appliquerons cette formule à notre corpus :

¹⁷¹ Charles MULLER, *op.cit.*, 1968, p.158.

Tableau II. 21 : Vérification des conditions de la fréquences moyennes

TEXTES	N _a	N _b	N _a -N _b	V _a	V _b	V _a -V _b	$\bar{f}_a = N_a/V_a$	$\bar{f}_b = N_b/V_b$	$\bar{f}_a - \bar{f}_b$
(PARCOURS, AILEN)	64566	75357	-10791	9467	10322	-855	6,82	7,30	-0,5
(PARCOURS, MILLE)	64566	75802	-11236	9467	10109	-642	6,82	7,50	-0,7
(PARCOURS, ELHAKI)	64566	109549	-44983	9467	13054	-3587	6,82	8,39	-1,6
(PARCOURS, GENET)	64566	53949	10617	9467	8139	1328	6,82	6,63	0,2
(PARCOURS, ABOUNOUR)	64566	26350	38216	9467	5520	3947	6,82	4,77	2,0
(PARCOURS, ZRIREK)	64566	88200	-23634	9467	10622	-1155	6,82	8,30	-1,5
(PARCOURS, FEMME)	64566	6083	58483	9467	1755	7712	6,82	3,47	3,4
(PARCOURS, LETTRES)	64566	22565	42001	9467	4681	4786	6,82	4,82	2,0
(AILEN, MILLE)	75357	75802	-445	10322	10109	213	7,30	7,50	-0,2
(AILEN, ELHAKI)	75357	109549	-34192	10322	13054	-2732	7,30	8,39	-1,1
(AILEN, GENET)	75357	53949	21408	10322	8139	2183	7,30	6,63	0,7
(AILEN, ABOUNOUR)	75357	26350	49007	10322	5520	4802	7,30	4,77	2,5
(AILEN, ZRIREK)	75357	88200	-12843	10322	10622	-300	7,30	8,30	-1,0
(AILEN, FEMME)	75357	6083	69274	10322	1755	8567	7,30	3,47	3,8
(AILEN, LETTRES)	75357	22565	52792	10322	4681	5641	7,30	4,82	2,5
(MILLE, ELHAKI)	75802	109549	-33747	10109	13054	-2945	7,50	8,39	-0,9
(MILLE, GENET)	75802	53949	21853	10109	8139	1970	7,50	6,63	0,9
(MILLE, ABOUNOUR)	75802	26350	49452	10109	5520	4589	7,50	4,77	2,7
(MILLE, ZRIREK)	75802	88200	-12398	10109	10622	-513	7,50	8,30	-0,8

(MILLE, FEMME)	75802	6083	69719	10109	1755	8354	7,50	3,47	4,0
(MILLE, LETTRES)	75802	22565	53237	10109	4681	5428	7,50	4,82	2,7
(ELHAKI, GENET)	109549	53949	55600	13054	8139	4915	8,39	6,63	1,8
(ELHAKI, ABOUNOUR)	109549	26350	83199	13054	5520	7534	8,39	4,77	3,6
(ELHAKI, ZRIREK)	109549	88200	21349	13054	10622	2432	8,39	8,30	0,1
(ELHAKI, FEMME)	109549	6083	103466	13054	1755	11299	8,39	3,47	4,9
(ELHAKI, LETTRES)	109549	22565	86984	13054	4681	8373	8,39	4,82	3,6
(GENET, ABOUNOUR)	53949	26350	27599	8139	5520	2619	6,63	4,77	1,9
(GENET, ZRIREK)	53949	88200	-34251	8139	10622	-2483	6,63	8,30	-1,7
(GENET, FEMME)	53949	6083	47866	8139	1755	6384	6,63	3,47	3,2
(GENET, LETTRES)	53949	22565	31384	8139	4681	3458	6,63	4,82	1,8
(ABOUNOUR, ZRIREK)	26350	88200	-61850	5520	10622	-5102	4,77	8,30	-3,5
(ABOUNOUR, FEMME)	26350	6083	20267	5520	1755	3765	4,77	3,47	1,3
(ABOUNOUR, LETTRES)	26350	22565	3785	5520	4681	839	4,77	4,82	0,05
(ZRIREK, FEMME)	88200	6083	82117	10622	1755	8867	8,30	3,47	4,8
(ZRIREK, LETTRES)	88200	22565	65635	10622	4681	5941	8,30	4,82	3,5

Nous constatons que pour l'intégralité des textes pris deux à deux, les écarts vont dans le même sens, d'où la relation d'équivalence suivante :

$$N_a > N_b \text{ et } V_a > V_b \text{ implique } \bar{f}_a > \bar{f}_b$$

Il est évident qu'une telle relation n'est pas identique à celle proposée par Muller et par conséquent la comparaison des textes en se référant à la condition évoquée par Muller est remise en cause.

Nous ne pouvons pas utiliser cette méthode faute d'inégalité de longueurs des textes, chose déjà confirmée par Muller qui énonce que : « *D'autre part, la méthode par comparaison n'est d'aucune utilité quand les deux textes sont de longueurs très inégales.* »¹⁷²

Ainsi, en étudiant la corrélation entre N et V nous pouvons arriver à d'autres conclusions. En effet, lorsqu'un segment occupe le même rang en N et V, nous pouvons dire que l'étendue du vocabulaire y est normale par rapport à l'intégralité du corpus. Cependant, s'il existe un décalage entre le classement de V et N nous pouvons déduire que le vocabulaire est exceptionnellement riche si $N_r - V_r$ est négatif, ou bien, il est relativement pauvre quand $N_r - V_r$ est positif. Maintenant, procédons à la vérification des rangs de N, V et $\bar{f}$.

Le tableau suivant présente les résultats obtenus :

¹⁷² Charles MULLER, *op.cit.*, 1968, p.159.

Tableau II. 22 : Corrélation entre N et V

TEXTES	V	Rang V	Nombre d'occurrences (N)	Rang N	Rang N - Rang V	$\bar{f}=N/V$	Rang $\bar{f}$
PARCOURS	9467	5	64566	5	0	6,82	5
AILEN	10322	3	75357	4	1	7,30	4
MILLE	10109	4	75802	3	-1	7,50	3
ELHAKI	13054	1	109549	1	0	8,39	1
GENET	8139	6	53949	6	0	6,63	6
ABOUNOUR	5520	7	26350	7	0	4,77	8
ZRIREK	10622	2	88200	2	0	8,30	2
FEMME	1755	9	6083	9	0	3,47	9
LETTRES	4681	8	22565	8	0	4,82	7
CORPUS	28638		522421			18,24	

Tout d'abord, nous constatons que sept romans présentent une étendue normale du vocabulaire, cette dernière se traduit par la valeur nulle de la différence des rangs $N_r - V_r$, du moment qu'ils ont le même classement en N et en V. Il s'agit de PARCOURS (5-5), ELHAKI (1-1), GENET (6-6), ABOUNOUR (7-7), ZRIREK (2-2), FEMME (9-9), LETTRES (8-8). Tandis que la différence entre N_r et V_r donne un écart positif pour les deux textes restants, ce qui signifie qu'il s'agit de textes pauvres en vocabulaire. Ces textes sont : MILLE et AILEN qui ont un écart de (+1). En nous basant sur l'écart entre le classement de N et V, nous obtenons le classement suivant :

**PARCOURS = ELHAKI = GENET = ABOUNOUR = ZRIREK = FEMME =
LETTRES > AILEN = MILLE**

En nous basant sur la fréquence moyenne des textes, nous obtenons le classement suivant :

**FEMME > ABOUNOUR > LETTRES > GENET > PARCOURS > AILEN >
MILLE > ZRIREK > ELHAKI**

En nous référant au principe selon lequel, plus la fréquence moyenne est faible, plus le texte est riche et quand elle est élevée, cela signifie que le texte est pauvre. Nous pouvons déduire alors que FEMME est le texte le plus riche tandis qu'ELHAKI sera le texte le plus pauvre.

Arrivant à ce niveau, un problème se pose, malgré les quatre formules proposées par MULLER, nous n'arrivons pas à les vérifier pour toutes les combinaisons binaires du corpus. De même, ces méthodes restent incomplètes puisqu'elles suggèrent juste un classement partiel des textes soumis à l'analyse, ce qui nous amène à l'utilisation d'une autre méthode qui puisse surpasser ces limites qui permette d'étendre la comparaison à des textes de longueurs

différentes. Il s'agit du modèle théorique qui se fonde sur la loi binomiale¹⁷³ et qui permet le calcul de V_0' , c'est-à-dire le nombre de mots qui manqueraient dans le texte selon la formule suivante :

$$V_0' = \sum_{i=1}^n q^i V_i = q^1 V_1 + q^2 V_2 + \cdots \dots + q^n V_n$$

Nous appliquerons cette méthode à notre corpus, les résultats seront présentés dans le tableau qui suit. Signalons que nous avons aussi exposé la valeur des hapax et l'écart réduit.

¹⁷³ Charles MULLER, *Initiation aux méthodes de la statistique linguistique*, Paris, Hachette, 1973, p.172.

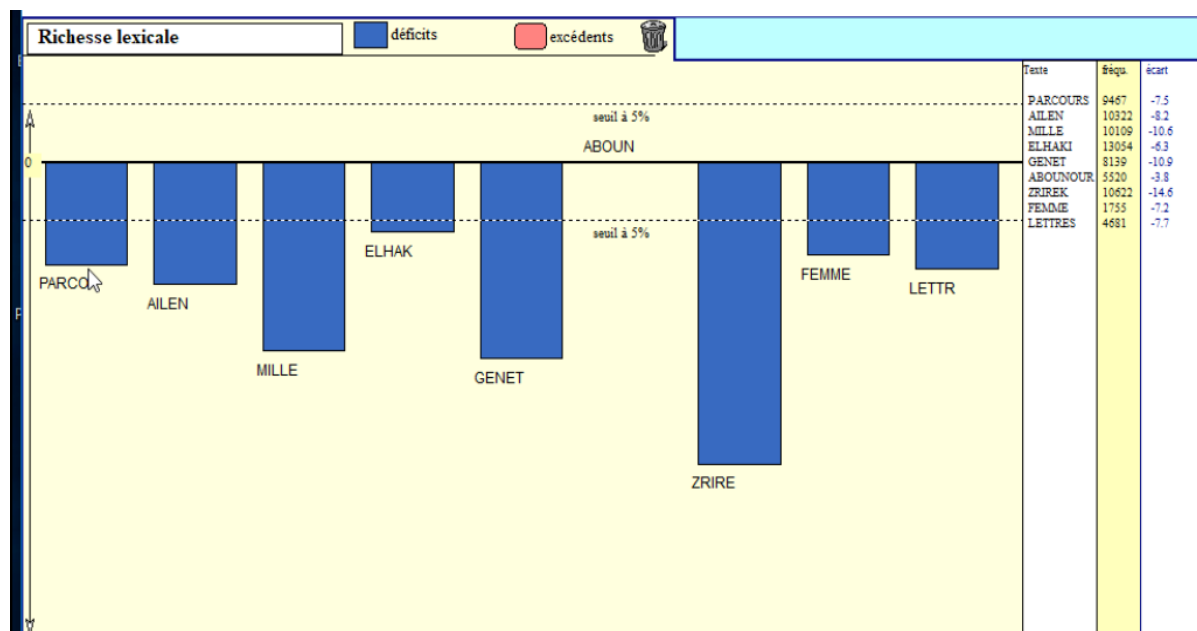
Tableau II. 23 : Richesse lexicale selon la loi binomiale

N	TEXTES	Valeur réelle	Valeur théorique	Écart	Écart réduit	Hapax	Écart réduit	Rang
1	PARCOURS	9467	10226	-759	-7,51	1551	4,84	4
2	AILEN	10322	11192	-870	-8,22	1914	8,08	6
3	MILLE	10109	11230	-1121	-10,58	1714	2,44	7
4	ELHAKI	13054	13798	-744	-6,33	2690	7,99	2
5	GENET	8139	9182	-1043	-10,88	496	-20,48	8
6	ABOUNOUR	5520	5812	-292	-3,83	517	-2,04	1
7	ZRIREK	10622	12242	-1620	-14,64	1548	-8,6	9
8	FEMME	1755	2082	-327	-7,17	140	0,86	3
9	LETTRES	4681	5236	-555	-7,67	617	6,22	5

Au regard de ce tableau, nous pouvons constater que tous les écarts entre l'effectif théorique et l'effectif réel sont négatifs, donc il s'agit d'un lexique déficitaire dans les différents textes du corpus, chose qui traduit le phénomène de la spécialisation lexicale : les mêmes vocables existent dans les mêmes textes, en vertu des contraintes thématiques, stylistiques ou de la situation de discours. De même, nous remarquons que l'écart entre le texte qui mobilise le vocabulaire le plus « riche » ABOUNOUR et celui qui mobilise le vocabulaire le plus « pauvre » ZRIREK est très important, la différence atteint presque 3 écarts types.

La courbe de la richesse lexicale établie par le logiciel Hyperbase, visualise les résultats que nous avons produit dans le tableau précédent :

Figure II. 5 : Richesse lexicale (Vocabulaire déficitaire)



La représentation affiche un dépassement du seuil de 5% dans le sens déficitaire pour l'ensemble des textes, exception faite du texte ABOUNOUR ; par contre dans le sens excédentaire, il n'est même pas atteint.

En classant les neuf textes selon leur richesse, nous aurons :

ABOUNOUR > ELHAKI > FEMME > PARCOURS > LETTRES > AILEN > MILLE > GENET > ZRIREK

Avant de chercher une explication à un tel classement, essayons de le confirmer au moyen d'une autre méthode plus simple et qui repose sur l'examen des valeurs N et V, il s'agit de l'indice w de Brunet. D'ailleurs Muller écrit, à ce propos : « *L'intérêt de cet indice réside d'abord dans la facilité relative d'emploi : certes la recette est sophistiquée mais les ingrédients sont simples puisqu'ils se réduisent à N et V, sans apport extérieur.* »¹⁷⁴. Dans le point suivant, nous irons à la découverte de cette méthode tout en l'appliquant à notre corpus.

1.6. Indice w d'Etienne Brunet

En étudiant la distribution de N et de V, BRUNET a observé que les deux séries croissent mais avec des vitesses différentes. Ce qui lui a permis de suggérer de nouveaux critères pour la mesure de la richesse lexicale. Selon l'indice w d'Etienne Brunet, le texte « b » est plus riche que « a » si :

$$(V_b / V_a) > (N_b / N_a) \text{ quand } V_b > V_a \text{ et } N_b > N_a$$

Si nous vérifions cette formule pour deux textes de notre corpus de recherche, le résultat sera :

¹⁷⁴ Charles MULLER, *op.cit.*, 1968, p.51.

Textes	V	N
LETTRES (a)	4681	22565
ABOUNOUR (b)	5520	26350
V_b/V_a	1,179235206	
N_b/N_a	1,167737647	

Quand : $V_b = 5520 > V_a = 4681$ et $N_b = 26350 > N_a = 22565$

Donc : $V_b / V_a = 1,1792 > N_b / N_a = 1,167737647$

C'est-à-dire le texte *b*, ici ABOUNOUR est plus riche que le texte *a* LETTRES.

Si nous prenons deux textes dont l'étendue n'est pas voisine, nous aurons :

Textes	V	N
PARCOURS (a)	9467	64566
AILEN (b)	10322	75357
V_b/V_a	1,090313721	
N_b/N_a	1,167131307	

Ici, la relation d'équivalence n'est pas vérifiée. En effet, quand :

$V_b = 10322 > V_a = 9467$ et $N_b = 75357 > N_a = 64566$,

$V_b/V_a = 1,090313721 < N_b/N_a = 1,167131307$.

Donc cette formule n'est valable que lorsque les deux textes comparés ont une étendue voisine.

Afin de mettre en œuvre toutes les comparaisons, il est nécessaire de réitérer le calcul à 36 combinaisons ($C_9^2 = 36$). Pour élargir le champ de comparaison, BRUNET propose un nouveau principe qui peut s'étendre à trois textes : a, b et c.

Il s'agit, bel et bien, du principe de transitivité selon lequel :

$$\text{si } a > b \text{ et } b > c \text{ alors } a > c$$

Autrement dit, si a est plus riche que b et b est plus riche que c , alors a est plus riche que c .

A ce niveau, E. BRUNET soulève un autre problème. En effet, les formules, que nous avons déjà présentées, ne donnent de résultats fiables que lorsque les textes sont de longueurs voisines. Ainsi, pour remédier à ce problème, BRUNET s'est inspiré de la formule de GUIRAUD pour réduire la progression de N afin qu'elle suive fidèlement celle de V . Convaincu que la racine carrée est un facteur de réduction « *trop constant et trop rigide pour annuler les distorsions quand les œuvres sont d'étendue trop inégale. Et ce qui est pire, cette distorsion agit dans un sens puis dans l'autre, ce qui ruine toute possibilité de correction* »¹⁷⁵, Etienne Brunet a conçu un indice de mesure de la richesse lexicale, quoiqu'il se base sur les deux données de base N et V , mais incluant un facteur de réduction plus maniable et plus fidèle que la simple racine carrée de N . De même, il a proposé le facteur $1/V$ qui n'est que le réciproque de V pour minimiser au maximum l'influence de l'étendue des textes sur la valeur de la richesse lexicale. Après avoir testé des exposants fractionnaires (α) qui doivent être appliqués à V , il a obtenu une valeur qui se situe aux alentours de 0,172 pour les classements les plus fiables.

¹⁷⁵ Etienne BRUNET, *op.cit.*, 1978, p. 49.

Généralement, l'indice w peut prendre des valeurs allant de 10 à 25. De plus, il progresse dans un sens inverse à celui de la richesse lexicale, c'est-à-dire, lorsque l'indice W est élevé, le vocabulaire sera, alors, pauvre et vice-versa.

Ainsi, pour obtenir l'indice w , trois opérations sont à réaliser. Pour concrétiser ces transformations, nous allons les opérer sur l'un des segments de notre corpus. Prenons l'exemple du texte PARCOURS où $N = 64566$ occurrences et $V = 9467$ vocables.

• **Réduction de V :**

$$V_r = V^{0,172} = 9467^{0,172} = 4,829570709$$

• **Réduction de V_r :**

$$a = \frac{1}{V_r} = \frac{1}{4,829570709} = 0,2070577408$$

• **Réduction de N :**

$$N^a = 64566^{0,2070577408} = 9,907145403$$

L'indice obtenu n'est autre que l'indice w de BRUNET ($w = 9,90$). Son principal avantage est qu'il reste insensible à l'étendue de N . Appliqué à la totalité de notre corpus, l'indice w permet d'obtenir le classement suivant.

Tableau II. 24 : Richesse lexicale selon l'indice w de BRUNET

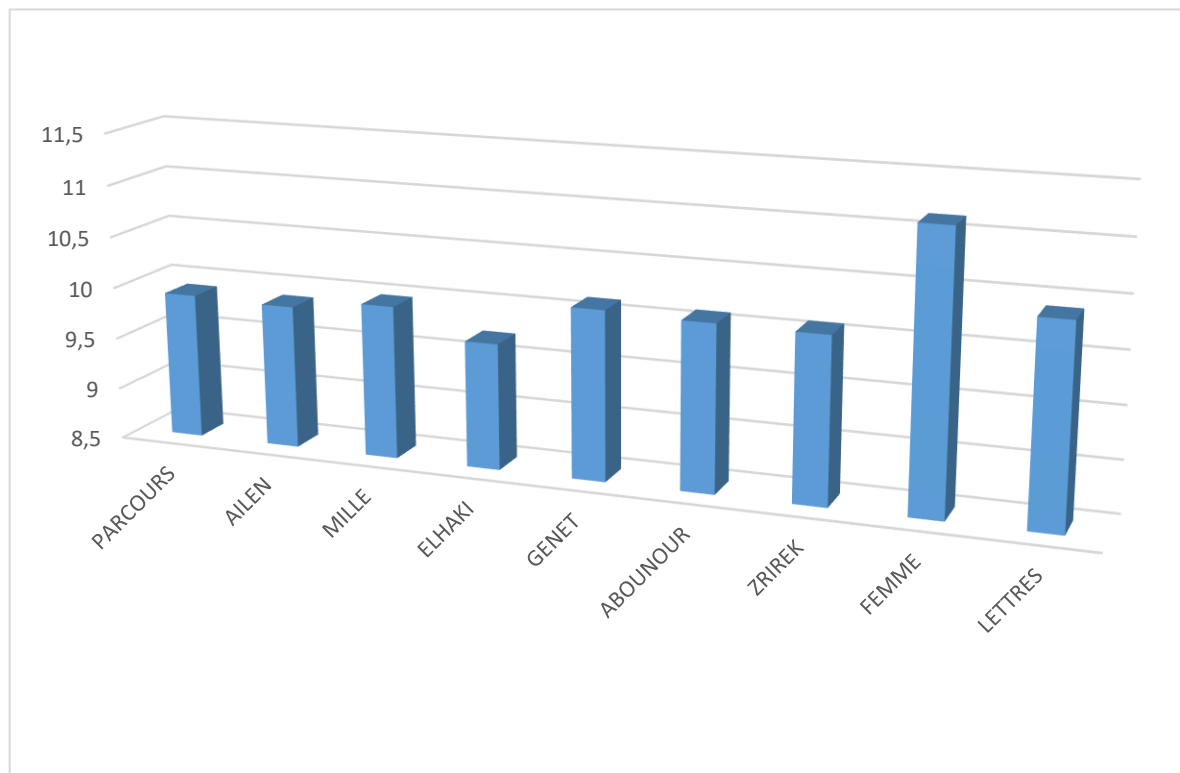
Textes	V	N	V _r	a	W	Rang
PARCOURS	9467	64566	4,829570709	0,207057741	9,907145403	3
AÎLEN	10322	75357	4,901933136	0,204001151	9,884138592	2
MILLE	10109	75802	4,884384111	0,204734103	9,977851684	4
ELHAKI	13054	109549	5,103967504	0,195926012	9,713827364	1
GENET	8139	53949	4,705634815	0,212511178	10,12976496	7
ABOUNOUR	5520	26350	4,401629882	0,22718857	10,10068503	6
ZRIREK	10622	88200	4,926148313	0,202998354	10,09071601	5
FEMME	1755	6083	3,614228987	0,276684185	11,14308779	9
LETTRES	4681	22565	4,278566104	0,233723162	10,41114929	8

Le tableau présente la répartition des neuf textes d'Edmond Amran El Maleh selon l'indice w de Brunet. Comme la richesse lexicale et l'indice w évoluent dans des sens inverses, nous pouvons dire que LE premier rang, qui correspond à l'indice le plus faible, fait référence au texte le plus riche qui est dans notre cas le texte ELHAKI et ainsi de suite :

**ELHAKI > AÎLEN > PARCOURS > MILLE > ZRIREK > ABOUNOUR >
GENET > LETTRES > FEMME**

En visualisant le tableau ci-dessus sous forme d'histogramme, le résultat sera comme suit :

Figure II. 6 : Richesse lexicale selon l'indice w de BRUNET



Nous sommes là en droit de nous interroger sur ce qui ordonne ce classement : le style, le thème, le genre ou d'autres paramètres ?

Dans son *Discours sur l'Autre à travers quatre récits de voyage en Orient*, Véronique Magri stipule que :

« la faible richesse lexicale d'une œuvre révèle la réutilisation des mêmes termes. Il s'agit certes d'une loi du discours puisque'un registre thématique fait l'unité et la cohérence d'un texte et conduit tout énonciateur à puiser dans un certain lexique et à en exclure d'autres »¹⁷⁶

Pourrions-nous donc conclure, à partir de cette citation, que les essais d'El Maleh sont pauvres en vocabulaire parce qu'ils réutilisent les mêmes termes ? ou leur faible richesse est-elle due à leur genre et au style dans lequel ils s'inscrivent ? En tout cas nous avons constaté des régularités entre les essais et l'utilisation des mots longs, ce qui range lesdits textes dans le discours intellectuel et soutenu ; cependant nous trouvons que l'indice w les classe avec les textes les moins riches. Pourrions-nous dire que l'auteur réutilise les mêmes termes ? cause très plausible ne serait-ce que pour les mots *écriture*, *allégorie* et *langue* qui sont très récurrents dans ces textes.

Le classement des essais parmi les textes pauvres infirme également ce qu'avance José Gregorio PARADA-RAMIREZ dans sa thèse à propos de la prépondérance de l'essai quant à la richesse lexicale :

« une parenthèse doit être faite par rapport à l'influence du genre en ce qui concerne la richesse lexicale. En effet, les Essais occupent la place prépondérante, suivis des Nouvelles et des Romans qui, en général, et malgré leur proportion écrasante, tendent à être les textes les plus « pauvres »¹⁷⁷

¹⁷⁶ Véronique MAGRI, *Le Discours sur l'Autre à travers quatre récits de voyage en orient*, Paris, Honoré champion, 1995, p.65.

¹⁷⁷ José Gregorio PARADA RAMIREZ, *op.cit.*, 2013, p. 33.

L'œuvre maléhiennne réfute aussi les conclusions de véronique Magri concernant la relation entre l'étendue et la richesse lexicale : « *plus un texte est long, plus il y a de chance de réemployer les mêmes mots* »¹⁷⁸. Chose qui n'est pas valable pour les textes d'El Maleh, d'ailleurs l'indice w classe le texte le plus long, ELHAKI, en première position de point de vue richesse lexicale.

Enfin l'écriture d'El Maleh va aussi à l'encontre de ce qu'affirme José Gregorio PARADA-RAMIREZ en disant : « *les textes les plus courts jouissent d'une richesse lexicale plus importante* »¹⁷⁹, car dans notre cas les textes les plus courts sont les moins riches.

De toute façon, en analysant le classement obtenu à travers l'indice w, nous constatons que la richesse lexicale de l'œuvre d'El Maleh s'avère très importante au début de sa carrière d'écrivain, puis elle rechute avec le texte GENET pour stagner, avant de marquer une rechute spectaculaire au niveau du texte FEMME. Nous pouvons dire alors que l'œuvre maléhiennne était à son apogée lors des premières productions qui sont tous des romans et qui s'intéressent aux thèmes relatifs à : l'exode des marocains juifs, l'histoire, la mémoire, la culture, etc.

Nous remarquons, également, qu'il existe une certaine convergence entre les classements que nous avons effectués moyennant la première et la troisième formule de la loi binomiale et l'indice w.

¹⁷⁸ Véronique MAGRI, *op.cit.*, 1995, p. 64.

¹⁷⁹ José Gregorio PARADA-RAMIREZ, *op.cit.*, 2013, p. 30.

Pour affiner la comparaison, nous dresserons un tableau qui représente l'ensemble des méthodes adoptées lors de notre exploration de la richesse lexicale, ainsi que le classement obtenu pour chaque démarche.

Tableau II. 25 : Comparaison des méthodes de mesure de la richesse lexicale

TEXTES	W	GUIRAUD ($V/\sqrt{N}$)	Loi binomiale	$\bar{f}$
PARCOURS	3	3	4	5
AILEN	2	2	6	4
MILLE	4	4	7	3
ELHAKI	1	1	2	1
GENET	7	6	8	6
ABOUNOUR	6	7	1	8
ZRIREK	5	5	9	2
FEMME	8	9	3	9
LETTRES	9	8	5	7

Les cases colorées mettent en évidence les convergences entre les quatre méthodes. Le texte ELHAKI occupe la première position en se basant sur l'indice w, la formule de Guiraud et $\bar{f}$. De même, nous remarquons que les deux premiers classements (l'indice W et la formule de Guiraud) *rangent* les textes AÎLEN, PARCOURS, MILLE et ZRIREK respectivement en deuxième, troisième, quatrième et cinquième position. Pour ce qui est du deuxième et quatrième classement (la formule de Guiraud et $\bar{f}$), ils placent les textes GENET et FEMME respectivement en sixième et neuvième position. La loi binomiale se trouve en discordance avec les autres méthodes.

Tout au long de cette section, nous avons essayé d'exposer, d'analyser et d'appliquer à notre corpus, cinq des plus importantes méthodes de mesure de la richesse lexicale. Au terme de ce long examen exploratoire et contrastive, et après avoir présenté la synthèse des résultats fournis par les différentes méthodes et les points de convergence et de divergence des différents classements. Nous pouvons dire que l'indice W de Brunet est la méthode la plus convenable pour réduire l'influence de l'étendue N sur la richesse lexicale. En effet, il suffit de remarquer, par exemple, que les textes ELHAKI et ZRIREK, qui ont les plus grandes étendues de tous les segments de notre corpus de recherche (109549 et 88200 occurrences), se sont vus classés, au niveau de la richesse lexicale, au premier et cinquième rang. D'ailleurs, André Cossette est arrivé au même constat, en effet, il a appliqué l'indice w à des corpus dont l'étendue varie de 671 364 à 37 652 402 (TLF) occurrences et s'est aperçu, finalement, que l'indice w jouit d'une autonomie et d'une stabilité presque parfaite.¹⁸⁰

Nous venons d'examiner l'un des aspects de la structure lexicale des œuvres d'Edmond Amran El Maleh sous un angle statique. Nous aborderons, par la suite, la structure lexicale d'un point de vue dynamique et ce, à travers l'étude de l'accroissement du vocabulaire.

B. Accroissement du vocabulaire

Dans le point précédent, nous nous sommes bornée à déterminer l'effectif de V pour les différentes valeurs de N. Maintenant, avec l'accroissement du vocabulaire, nous nous intéresserons à l'évolution de V en tenant compte du déroulement chronologique de l'œuvre. Il est donc question de relever, pour

¹⁸⁰ André COSSETTE, *La richesse lexicale et sa mesure*, Paris, Champion, 1994, p. 148.

chaque segment du corpus, le nombre de vocables qui y apparaissent pour la première fois, dans le but d'identifier l'apport du vocabulaire au fil du temps. Après, nous explorerons l'ordre chronologique inverse, autrement dit, nous prendrons comme point de départ le dernier texte afin de faire apparaître les thèmes abandonnés ou qui n'apparaîtront plus.

Dans ce sens ch. MULLER¹⁸¹ explique qu'à tout moment de l'acte de discours, le locuteur emploie un certain lexique dans lequel il puise, l'étendue et la structure de ce lexique déterminent l'apport lexical qui est plus grand une fois le lexique est plus étendu et plus faible si les probabilités d'emploi sont inégalement réparties. De même Etienne Brunet, lors de son étude sur le vocabulaire de Proust, stipule que :

*« Proust n'a guère le souci de renouveler son vocabulaire tout au long de la composition de la **Recherche** ou que du moins ce souci s'arrête à **Sodome**. L'accroissement du vocabulaire étant le reflet des variations thématiques, on peut soutenir qu'au-delà de **Sodome** les thèmes proustiens se renouvellent peu et qu'en abordant l'épisode de l'Albertine le lecteur qui a lu ce qui précède se retrouve dans un univers familier »¹⁸²*

De ce fait, la structure du vocabulaire peut être influencée par les choix thématiques qui agissent sur l'apport lexical. Nous essaierons donc de situer les ruptures thématiques qui ont provoqué soit un afflux de nouveaux vocables, soit l'épuisement d'un thème.

Observons, dans un premier temps, l'accroissement chronologique dans notre corpus.

¹⁸¹ Charles MULLER, *op. cit.*, 1977, p.130

¹⁸² Etienne BRUNET, *Le vocabulaire de Proust I, Étude quantitative*, Genève, SLATKINE, 1983, pp. 21-22

1. Accroissement chronologique

Nous examinerons l'accroissement chronologique en nous référant à deux techniques différentes. La première est la formule $y = ax^b$, tandis que la deuxième sera l'indice w de Brunet.

1.1. Accroissement chronologique selon la formule $y = ax^b$

En faisant appel à un réajustement de deux séries parallèles (vocabulaire cumulé et étendue cumulée) nous calculerons l'accroissement du vocabulaire grâce à la fonction $y = ax^b$, où x représente le vocabulaire cumulé et y l'étendue cumulée théorique.

L'écart entre les deux étendues, théorique et réelle, est alors calculé pour chaque texte, puis pondéré par l'étendue de chacun. Notons que le logiciel *Hyperbase* ne fournit pas les valeurs théoriques et l'écart est calculé pour chaque texte de la série.

Pour clarifier d'avantage ce que nous venons de présenter, nous étalons l'exemple de calcul pour les deux premiers textes, PARCOURS et AÎLEN :

Pour PARCOURS, nous avons : $N_{\text{réel}} = 64566$ ainsi que $V_{\text{réel}} = 9467$

Si nous appliquons la formule $y = ax^b$, avec :

$$a = 1,77482274805508 \times 10^{-3}$$

et

$$b = 1,89619888429778$$

Nous aurons : $y = 1,77482274805508 \times 10^{-3} \times 9467^{1,89619888429778}$

$$y = 61495,65898$$

Donc l'écart résultant est de :

$$\text{Écart} = 61495,65898 - 64566 = -3070,341024$$

Et par conséquent l'écart pondéré sera de :

$$\frac{-3070,341024}{9467 - 3070,341024} = -0,4799913573 = -0,48$$

Une fois la méthode de calcul exemplifiée, nous appliquerons la même procédure à tous les textes de notre corpus. Les résultats seront présentés dans le tableau suivant :

Tableau II. 26 : Accroissement du vocabulaire dans l'ordre chronologique

Textes	Acc	Vocab	Vocc.cum	Occu	Occu.cum	Occu.cum	Écart	Δ Écart	Pondéré	Rang
PARCOURS	9467	9467	9467	64566	64566	61495,65898	-3070,341024	-3070,341024	-0,48	6
AÎLEN	5476	10322	14943	75357	139923	146123,2351	6200,235083	9270,576107	1,23	1
MILLE	3714	10109	18657	75802	215725	222597,5933	6872,593315	672,3582323	0,09	4
ELHAKI	4548	13054	23205	109549	325274	336640,1363	11366,13635	4493,543033	0,41	3
GENET	2040	8139	25245	53949	379223	394961,7342	15738,73417	4372,597824	0,81	2
ABOUNOUR	632	5520	25877	26350	405573	413920,9891	8347,989062	-7390,74511	-2,8	8
ZRIREK	1939	10622	27816	88200	493773	474702,584	-19070,416	-27418,40506	-3,11	9
FEMME	150	1755	27966	6083	499856	479568,3368	-20287,66324	-1217,247244	-2	7
LETTRES	672	4681	28638	22565	522421	501654,5235	-20766,47654	-478,8132942	-0,21	5
a	0,001774823				b	1,896198884				
r^2	0,996647585118246				r	0,998322335279666				

Nous donnons l'exemple de AÎLEN pour illustrer le calcul de la variation de l'écart :

$$\Delta \text{Écart} (AÎLEN) = 6200,235083 - (-3070,341024) = 9270,576107$$

La deuxième colonne du tableau expose l'apport de chaque texte en vocable. Nous remarquons, ainsi, une certaine régularité qui se manifeste par une baisse de la valeur en progressant dans l'œuvre. Signalons que Brunet a abouti au même constat lors de ces recherches :

« On remarquera que ces effectifs vont décroissant, car à mesure qu'un corpus se développe, plus faible deviennent les chances d'y trouver des mots inédits. »¹⁸³

Néanmoins, les textes ELHAKI, ZRIREK et LETTRES font dérogation à la règle, du moment qu'ils présentent des valeurs plus grandes que celles des textes qui les précèdent. Ceci est peut-être dû à leurs étendues qui restent supérieures aux autres (pour ELHAKI et ZRIREK), à une rupture thématique dans le corpus, ce qui engendre l'afflux de nouveaux vocables, ou bien à l'introduction « *de mots rares ou de néologismes* »¹⁸⁴. Rappelons ce que nous avons dit au niveau du premier chapitre à propos des classes de longueur, le texte ELHAKI quoiqu'il appartienne au genre romanesque, il est aussi attiré par les mots longs qui sont en faveur d'un style recherché, alors qu'il ne doit héberger que les mots courts et moyens (comme c'est le cas des textes qui le précèdent : PARCOURS, AÎLEN et MILLES qui favorisent l'utilisation des mots courts et moyens). Ce choix stylistique est peut-être la conséquence du thème traité. En effet, dans ELHAKI nous assistons à une mise en valeur de la tradition musulmane de savoir, de culture

¹⁸³ Etienne BRUNET, *op.cit.*, 1981, p. 53.

¹⁸⁴ Ibid., p.53.

et de refus de tyrannie, un thème qui reste plus ou moins absent des trois premiers romans. Les idées d'El Maleh prennent forme à travers le récit épico-lyrique de vie d'individualités marquantes, comme : Amine El Andaloussi, Sofiane Abou El Haki, Ahmed El Ghazouli et dont l'évocation s'accompagne du plaisir de la méditation sur les vertus et les joies de l'écriture.

Le même constat reste valable pour ZRIREK, un essai qui se caractérise par une prépondérance des mots longs et qui étale des réflexions sur les enjeux de l'écriture tout en faisant appel à des intellectuels éminents, comme : le romancier Juan Goytisolo, le poète José-Angel Valente, Jean Genet, khaïr-Eddine et Nissaboury. De même, à travers une série d'évocations relatives au soufisme, l'auteur choisit de présenter des figures appartenant aux trois religions théologiques, tels : Abraham Aboulafia, Ibn Arabi et Juan de la Cruz. Donc c'est le contenu sémantique et thématique de ces deux textes qui a imposé l'apparition de mots nouveaux.

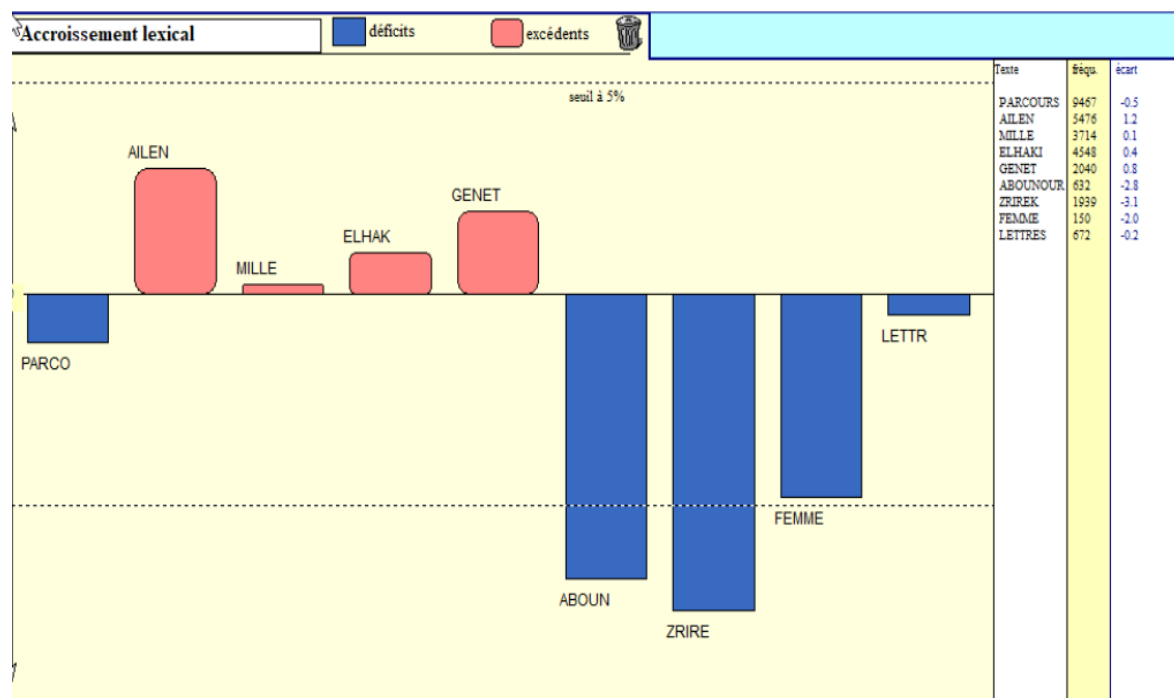
Pour le troisième LETTRES, le renouvellement lexical est dû à la fois au thème et au genre. En effet ledit texte est le premier à aborder l'exil d'El Maleh, mais aussi il est le premier et dernier récit épistolaire de toute la production littéraire de notre auteur.

Cependant, une remarque s'impose, c'est que les données offertes par la deuxième colonne ne donnent pas une appréciation exacte de l'accroissement du vocabulaire dans notre corpus, car elle mesure uniquement l'apport du dernier texte par rapport à celui qui le précède. Il faut bien évidemment recourir aux résultats de pondération ainsi présentés dans l'avant dernière colonne, pour avoir une idée plus rigoureuse sur l'accroissement du vocabulaire. Les écarts pondérés classent AÎLEN en première position du point de vue apport lexical, suivi de

GENET, puis ELHAKI, ensuite MILLE, pour arriver à LETTRES, PARCOURS, ABOUNOUR et enfin ZRIREK. Notons, qu'il n'existe pas une relation d'équivalence entre la notion d'accroissement lexical et celle de richesse lexicale. En effet, un segment peut jouir d'un vocabulaire riche sans avoir apporté de nouveaux vocables par rapport aux textes qui le précèdent. De la même manière, il peut y avoir un texte pauvre en vocabulaire mais qui actualise un nombre important de vocables qui n'existe pas dans ses antécédents.

Nous pouvons dire que l'accroissement lexical est un moyen permettant la mise en exergue des variations stylistiques et thématiques au sein d'un corpus donné. La courbe suivante montre bien cette variation :

Figure II. 7 : Accroissement lexicale



Ce graphique expose l'accroissement lexical sur la base de l'écart réduit, tout en respectant l'ordre de parution des textes. Ainsi, nous constatons qu'AÎLEN vient en première position de point de vue apport lexical avec un écart de (1,2), suivi du texte GENET avec un écart de (0,8), puis ELHAKI avec un écart de (0,4), ensuite MILLE avec un écart de (0,1). L'accroissement lexical est excédentaire dans ces quatre derniers, alors qu'il est déficitaire dans les autres avec des écarts de (-0,2) ; (-0,5) ; (-2) ; (-2,8) et (-3,1) respectivement pour LETTRES, PARCOURS, FEMME, ABOUNOUR et ZRIREK.

Nous remarquons également que le seuil de 5% est dépassé deux fois, mais dans le sens négatif, avec les textes qui ont des écarts de (-2,8) pour ABOUNOUR et (-3,1) pour ZRIREK.

En somme, nous pouvons établir un classement des textes en fonction de leur accroissement, allant du plus fort au plus faible :

Tableau II. 27 : Accroissement lexical

Textes	Rang
PARCOURS	6
AÎLEN	1
MILLE	4
ELHAKI	3
GENET	2
ABOUNOUR	8
ZRIREK	9
FEMME	7
LETTRES	5

Dans le point suivant, nous ferons subir les mêmes données à l'indice W de BRUNET afin de pouvoir comparer les résultats et d'en tirer des conclusions.

1.2. Indice w d'Étienne Brunet

Cet indice nous permet de calculer des effectifs théoriques qui permettent de dégager des différences avec la réalité. Cet effectif s'obtient à partir de la formule suivante :

$$N_{\text{théorique}} = w V^a$$

V étant le vocabulaire cumulé pour chaque texte.

$$a = 0,176 - \frac{0,01}{\log N}$$

$$a = 0,176 - \frac{0,01}{\log 522421}$$

$$a = 0,1742511431$$

Donc, pour N = 522421 et V = 28638, dans notre corpus :

$$w = N^{V^{-a}}$$

$$w = 522421^{28638^{-0,174}}$$

$$w = 9,095390256 = 9,09$$

De ce fait, la valeur w est de 9,09. Cet indice nous permettra de calculer l'étendue théorique cumulée de l'œuvre d'EL MALEH, et cela, en saisissant la valeur observée de V cumulée d'un texte à l'autre.

Présentons la méthode employée lors des calculs pour les deux premiers textes PARCOURS et AÎLEN avant de la généraliser à l'ensemble des textes.

Étape 1 : Calcul du nombre d'occurrences cumulées

- Le nombre d'occurrences cumulées jusqu'au texte PARCOURS est de 64566
- Le nombre d'occurrences cumulées jusqu'au texte AÎLEN est obtenu par l'addition du nombre d'occurrences du texte AÎLEN au nombre d'occurrences cumulées des textes qui le précèdent ; dans notre cas, il s'agit uniquement du texte PARCOURS, donc le calcul est le suivant : $64566 + 75357 = 139923$

Étape 2 : Calcul du nombre de vocables cumulés

- Le nombre de vocables cumulés jusqu'au texte PARCOURS est de 9467
- Le nombre de vocables cumulés jusqu'au texte AÎLEN est obtenu par l'addition du nombre de vocables du texte AÎLEN au nombre de vocables cumulés de tous les textes qui le précèdent. Pareillement à l'étape précédente, il s'agit uniquement du texte PARCOURS, donc le résultat est le suivant : $9467 + 5476 = 14943$

Étape 3 : Détermination de la valeur théorique N

Nous pouvons déduire aisément la valeur théorique N en nous référant aux valeurs précédentes, tout en appliquant la formule suivante :

$$N_{théorique} = w^{v^a}$$

- Ceci dit, pour le texte PARCOURS la valeur théorique N sera :

$$N_{théorique} = 9,09^{9467^{0,00177482}} = 63406,80$$

L'écart cumulé sera, alors, le nombre d'occurrences cumulées moins le nombre d'occurrences cumulées théoriques :

$$\text{Écart cumulé} = 64566 - 63406,80 = 1159,2$$

➤ Pour ce qui est de AÎLEN :

$$N_{théorique} = 9,09^{14943^{0,00177482}} = 159888,80$$

et

$$\text{L'écart cumulé} = 139923 - 159888,8 = -19965,85$$

Nous signalons que l'écart est toujours égal à la différence entre deux écarts cumulés successifs, par exemple :

$$\text{L'écart de AILEN} = \text{Écart cumulé de AILEN} - \text{Écart cumulé de PARCOURS}$$

$$= -19965,85 - 1159,20$$

$$= -21125,05$$

Les résultats obtenus pour l'ensemble des textes sont présentés dans le tableau suivant :

Tableau II. 28 : Accroissement du vocabulaire selon $w = 9,09$

Textes	Acc	Vocab	Voc.cu	Occu	Occ.cum	N théorique	Écart	Δ Écart	Écart Pondéré	Rang
PARCOURS	9467	9467	9467	64566	64566	63406,80	1159,20	1159,20	0,02	9
AÎLEN	5476	10322	19789	75357	139923	293451,34	-153528,34	-154687,54	-2,05	8
MILLE	3714	10109	29898	75802	215725	757555,55	-541830,55	-388302,21	-5,12	7
ELHAKI	4548	13054	42952	109549	325274	1847207,96	-1521933,96	-980103,41	-8,95	6
GENET	2040	8139	51091	53949	379223	2889683,46	-2510460,46	-988526,50	-18,32	5
ABOUNOUR	632	5520	56611	26350	405573	3789410,19	-3383837,19	-873376,73	-33,15	3
ZRIREK	1939	10622	67233	88200	493773	6035543,56	-5541770,56	-2157933,37	-24,47	4
FEMME	150	1755	68988	6083	499856	6479453,35	-5979597,35	-437826,79	-71,98	1
LETTRES	672	4681	73669	22565	522421	7774947,29	-7252526,29	-1272928,93	-56,41	2

La dernière colonne met en exergue le classement des textes par ordre décroissant, d'où le résultat suivant :

**FEMME > LETTRES > ABOUNOUR > ZRIREK > GENET >
ELHAKI > MILLE >> AÎLEN > PARCOURS**

Ce classement ne converge pas avec celui de la richesse lexicale, du moment que tous les textes ont vu leur rang changé.

Mais ce qui est marquant c'est que les textes dont l'étendue est la plus faible et qui sont aussi les derniers à être publiés se voient classer dans la première et la deuxième position. De ce fait FEMME et LETTRES actualisent un nombre de vocables qui n'existent pas dans les textes précédents. Soulignons à ce sujet, une évocation de Guy Dugas selon laquelle les auteurs judéo-maghrébins de notre époque seraient :

*« si antévertis sur leur passé qu'ils ne parviennent parfois plus
...qu'à produire une œuvre répétitive, dramatiquement ankylosée
autour de quelques faits de mémoire »¹⁸⁵.*

De ce fait, El Maleh a voulu que ces deux textes marquent un changement radical dans son œuvre, que ce soit au niveau de la forme qu'au niveau du contenu. D'ailleurs, l'avant dernier texte LETTRES est le seul récit épistolaire qui existe dans tout le corpus, de même, c'est le seul qui parle de l'exil d'El Maleh, de son arrivée à Paris. Tandis que le dernier texte FEMME est le seul qui aborde sa découverte de la ville d'Asilah et sa rencontre avec la mère de son fils spirituel

¹⁸⁵ Guy DUGAS, *La littérature judéo-maghrébine ! d'expression française. Entre Djéba et Cagayous*, Paris, L'Harmattan, 1990, p. 135

Khalil El Ghrib. Ainsi, le classement auquel nous avons abouti confirme parfaitement bien cette diversification dans l'écriture d'El Maleh.

Toutefois, il existe une autre méthode d'exploration de l'accroissement du vocabulaire, c'est ce que Brunet appelle, l'accroissement inverse. C'est une technique qui permet d'étudier le corpus en remontant du dernier texte jusqu'au premier. Observons, donc, ce que donne cet examen.

2. Accroissement inverse

Contrairement à l'accroissement chronologique qui porte sur l'ensemble de l'œuvre suivant son ordre de parution, l'accroissement inverse consiste à remonter dans le temps, en prenant comme point de départ le dernier texte de la série chronologique pour effectuer la même analyse dans l'ordre inverse. Etienne Brunet dit à propos de ces deux techniques :

*« Les deux visées sont en fait complémentaire, la première désigne les thèmes qui apparaissent pour la première fois, la seconde ceux qui n'apparaîtront plus. La première marque le jaillissement d'une inspiration, la seconde son épuisement »*¹⁸⁶.

Donc, il s'agit pour l'accroissement chronologique de détecter la naissance d'un mouvement, tandis que l'accroissement inverse essaiera de mettre en évidence les ruptures chronologiques qui marquent la fin d'un mouvement et l'épuisement de l'inspiration de l'auteur.

¹⁸⁶ Etienne BRUNET, *op.cit.*, 1978, p.81.

Ainsi, pour rendre compte de l'accroissement du vocabulaire dans l'ordre inverse, nous dressons le tableau suivant :

Tableau II. 29 : Accroissement du vocabulaire selon l'ordre inverse

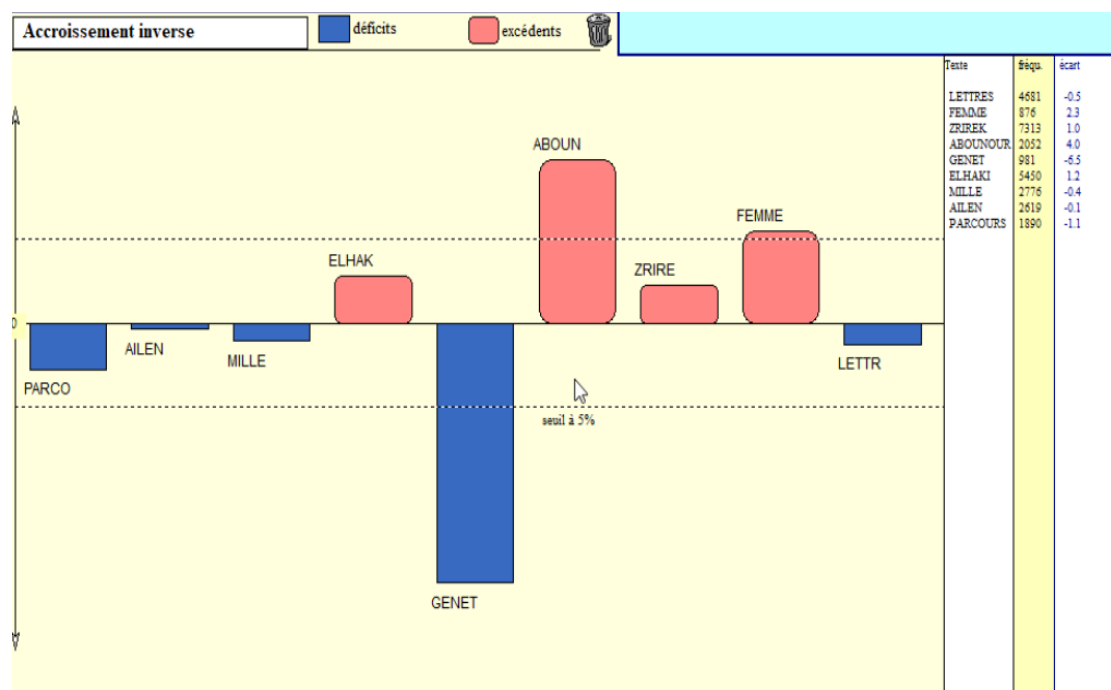
Textes	Acc	Vocab	Vocc.cum	Occu	Occu.cum	Occu.cum	Écart	ΔÉcart	Écart Pondéré	Rang
LETTRES	4681	4681	4681	22565	22565	21391,0254	-1173,9746	-1173,9746	-0,52	3
FEMME	876	1755	5557	6083	28648	28874,38757	226,3875675	1400,362168	2, 3	8
ZRIREK	7313	10662	12870	88200	116848	125409,4257	8561,42567	8335,038103	0, 59	6
ABOUNOUR	2052	5520	14922	26350	143198	162435,6965	19237,69651	10676,27084	4,05	9
GENET	981	8139	15903	53949	197147	181566,8473	-15580,15274	-34817,84925	-6,45	1
ELHAKI	5450	13054	21353	109549	306696	303972,8732	-2723,126811	12857,02593	-1,17	7
MILLE	2776	10109	24129	75802	382498	376405,9977	-6092,002301	-3368,87549	-0,44	4
AÎLEN	2619	10322	26748	75357	457855	450727,7701	-7127,229891	-1035,22759	-0,14	5
PARCOURS	1890	9467	28638	64566	522421	507885,3093	-14535,69074	-7408,460849	-1,15	2
a	0,008164178				b	1,748699436				
r^2	0,9969949827566				r	0,998496360913048				

Nous dégageons, à partir de ce tableau, que la tendance régulière à la baisse est inversée dans deux romans : ZRIREK (7313), ELHAKI (5450). Donc ces deux textes signalent la fin d'un mouvement et l'épuisement de l'inspiration de l'auteur.

Le mouvement dont il s'agit, ici, est surtout d'ordre générique et thématique : le texte ELHAKI marque une rupture avec le genre romanesque, d'ailleurs aucun des cinq textes publiés après n'admettent l'étiquette de roman, du moment que l'auteur a délaissé ce genre en faveur des essais et des nouvelles. ZRIREK est aussi le dernier essai qui aborde l'écriture comme thème principale.

Le classement présenté dans le tableau ci-dessus peut être illustré sous forme d'histogramme :

Figure II. 8 : Accroissement du vocabulaire selon l'ordre inverse



Pareillement à l'accroissement chronologique, nous constatons l'existence de valeurs positives et d'autres négatives. Le seuil de 5% est dépassé dans le sens

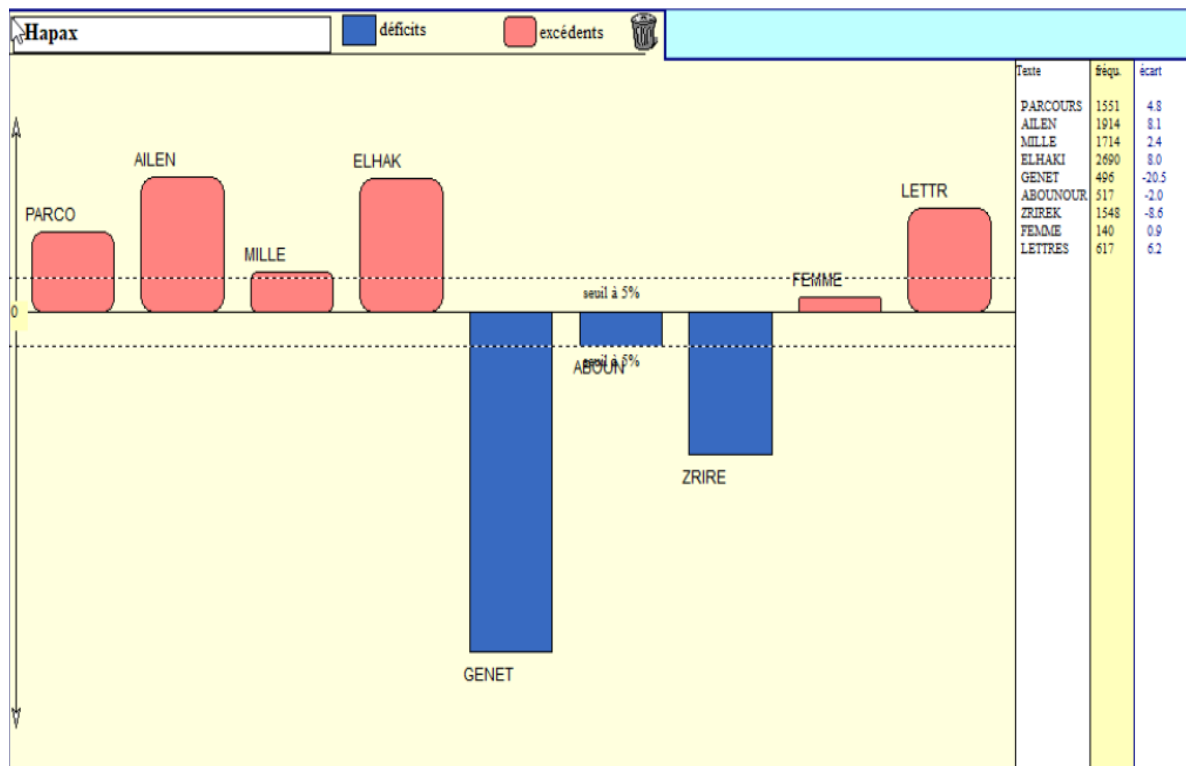
négatif avec GENET, d'ailleurs c'est le premier texte qui traite des thèmes de l'écriture et la littérature). De même, les deux textes ABOUNOUR et FEMME dépassent le seuil de 5% dans le sens positif.

En nous basant sur les écarts réduits, nous pouvons classer les neufs textes comme suit :

**ABOUNOUR > FEMME > ELHAKI > ZRIREK > AÎLEN > MILLE > LETTRES
> PARCOURS > GENET**

Cette étude de l'accroissement inverse permet de dévoiler un éclatement lexical au niveau du texte ABOUNOUR, chose que l'ordre chronologique n'a pas pu relever. Ceci est peut-être dû aux effets du genre, car ce texte est le seul recueil de nouvelles qui existe dans le corpus d'El Maleh. Cette étude a aussi inversé le classement du texte GENET en le plaçant au dernier rang, alors qu'il occupait une position privilégiée dans l'accroissement chronologique.

Figure II. 9 : Distribution des hapax



En nous référant à l'histogramme de la distribution des hapax et à celui de l'accroissement chronologique, nous remarquons l'existence de quelques ressemblances. Les seules différences relevées concernent les textes PARCOURS, FEMME et LETTRES qui ont une valeur positive au niveau des Hapax tandis qu'elle reste négative au niveau de l'accroissement chronologique. Aussi, le texte GENET enregistre une valeur négative au niveau des hapax tandis qu'elle demeure positive dans l'accroissement chronologique.

En somme, pour mieux appréhender la structure du vocabulaire de notre corpus, nous avons traité dans ce chapitre la notion de richesse lexicale et celle de l'accroissement du vocabulaire.

Pour ce qui est de la première, la richesse lexicale, elle traite le corpus sous un angle statique. C'est une notion indépendante du contenu, elle se rapporte à la structure. Elle prend en charge le texte dans sa globalité et le considère comme un ensemble clos et achevé, dans le but de distinguer les textes les plus riches de ceux qui sont les plus pauvres. Tandis que la deuxième notion, à savoir, l'accroissement du vocabulaire, elle vise à renseigner sur la contribution de chaque texte quant au renouvellement lexical et de mettre en évidence les thèmes abandonnés par l'auteur. Elle traite, donc, le corpus sous un angle dynamique.

Abordons la question de la richesse lexicale, nous avons fait recours à plusieurs formules et techniques. Parmi toutes les méthodes qui existent, nous nous sommes permis de choisir celles qui ont pu satisfaire un grand nombre de spécialistes en lexicométrie, à savoir la formule de GUIRAUD, l'indice w de BRUNET et la méthode binomiale de MULLER.

Après exploitation de ces techniques, nous avons pu affirmer que l'indice w de BRUNET reste la formule la plus fiable, du moment qu'elle neutralise ou du moins réduit l'influence parasitaire de l'étendue, pour ne prendre en considération que la distribution de N et de V . Une application de cet indice à notre corpus a permis de classer les textes de celui qui a le vocabulaire le plus riche à celui qui a le vocabulaire le plus faible :

**ELHAKI > AÎLEN > PARCOURS > MILLE > ZRIREK >
ABOUNOUR > GENET > FEMME > LETTRES**

Pour trouver une explication à ce classement nous nous sommes référée à plusieurs recherches, cependant nous avons constaté que nos résultats ne convergent pas avec ce qui a été avancé dans lesdites études.

Il est à signaler que les classements relatifs à la formule de Guiraud et à l'indice w sont semblables dans 56 % des cas. Ceci n'est pas surprenant du moment que la deuxième formule s'inspire de la première. D'ailleurs le concepteur de la technique le dit clairement :

« Pour ce faire nous sommes inspiré de la démarche de M. Guiraud qui propose de réduire la progression de N afin qu'elle accompagne fidèlement celle de $V..$ »¹⁸⁷

Aussi, les deux formules rangent ELHAKI et AÎLEN respectivement dans la première et la deuxième position, chose qui tend à se justifier par la présence d'un grand nombre d'hapax dans ces deux textes.

Après avoir exploré le corpus du point de vue richesse lexicale, nous nous sommes intéressée dans un deuxième temps à l'accroissement du vocabulaire. Cette étape a été réalisée suivant deux techniques contradictoires mais complémentaires. L'une parcourt les œuvres en respectant leur enchainement chronologique de parution, pour nous renseigner sur l'apport lexical de chaque texte, l'autre inverse le processus en explorant le corpus à partir du dernier texte écrit, dans le but de mettre en exergue les thèmes abandonnés.

Nous avons commencé par l'accroissement chronologique. Ainsi, nous avons procédé à la valorisation de l'apport lexical de chaque texte, moyennant la

¹⁸⁷ Charles MULLER, *op.cit.*, 1968, p. 49.

fonction $y = ax^b$. Cette technique nous a permis de tracer la progression du vocabulaire pour déterminer les nouveaux vocables qui apparaissent dans chaque segment. En nous référant à Brunet, nous avons signalé qu'en général les valeurs diminuent au fur et à mesure que nous progressons vers le dernier texte, du moment que les chances de trouver des mots inédits deviennent faibles. Cependant, dans notre corpus nous avons décelé, tantôt un accroissement normal qui se caractérise par une diminution du vocabulaire et c'est d'ailleurs le cas des couples : (PARCOURS, AÎLEN) ; (AÎLEN, MILLE) ; (ELHAKI, GENET) ; (GENET, ABOUNOUR) ; (ZRIREK, FEMME), tantôt un accroissement anormal qui se reflète par une augmentation de l'apport lexical. Cela est clair, surtout pour ELHAKI, ZRIREK et LETTRES. Le renouvellement lexical de ces trois textes est peut-être dû à leurs étendues qui restent supérieures aux autres (pour ELHAKI et ZRIREK), à une rupture thématique dans le corpus, ce qui engendre l'afflux de nouveaux vocables, ou bien à l'introduction « *de mots rares ou de néologismes* »¹⁸⁸. Rappelons ce que nous avons dit au niveau du premier chapitre à propos des classes de longueur, le texte ELHAKI quoi qu'il appartienne au genre romanesque, il est aussi attiré par les mots longs qui sont en faveur d'un style recherché, alors qu'il ne doit héberger que les mots courts et moyens comme c'est le cas des textes qui le précèdent : PARCOURS, AÎLEN et MILLES qui favorisent l'utilisation des mots courts et moyens. Ce choix stylistique est peut-être la résultante du thème traité. En effet, dans ELHAKI nous assistons à une mise en valeur de la tradition musulmane de savoir, de culture et de refus de tyrannie, un thème qui reste plus ou moins absent des trois premiers romans. Les idées d'El Maleh prennent forme à travers le récit épico-lyrique de vie d'individualités marquantes, comme : Amine El Andaloussi, Sofiane Abou El Haki, Ahmed El

¹⁸⁸ Ibid., p.53.

Ghazouli et dont l'évocation s'accompagne du plaisir de la méditation sur les vertus et les joies de l'écriture.

Le même constat reste valable pour ZRIREK, un essai caractérisé par une prépondérance des mots longs et qui étale des réflexions sur les enjeux de l'écriture en faisant appel à des intellectuels éminents, comme : le romancier Juan Goytisolo, le poète José-Angel Valente, Jean Genet, khair-Eddine et Nissaboury. De même, à travers une série d'évocations relatives au soufisme, l'auteur choisit de présenter des figures appartenant aux trois religions théologiques, tels : Abraham Aboulafia, Ibn Arabi et Juan de la Cruz. Donc c'est le contenu sémantique et thématique de ces deux textes qui a ordonné l'apparition de mots nouveaux.

Pour le troisième texte LETTRES, le renouvellement lexical est dû à la fois au thème et au genre. En effet ledit texte est le premier à aborder l'exil d'El Maleh, mais aussi, il est le premier et le dernier récit épistolaire de toute la production littéraire de notre auteur.

Signalons que nous avons établi un classement basé sur le calcul de l'écart réduit. L'illustration graphique de ce dernier a montré que la courbe est en faveur d'un vocabulaire excédentaire de manière significative, à partir du deuxième texte, mais elle décline brusquement à partir de ABOUNOUR. La chute que nous avons observée dans les quatre derniers textes correspond bien à la rupture dans l'écriture de notre auteur et à l'abandon du genre romanesque mais aussi de certains thèmes.

Les premiers romans, caractérisés par un vocabulaire riche avec des apports lexicaux importants, reflètent surtout l'intention manifeste de l'écrivain de vouloir tout exprimer sur l'exode des marocains juifs et de jaillir la spontanéité créatrice de son imagination. Donc, il semble bien que les récits où nous suivons le

destin des juifs, qui sont partis à la recherche de la terre promise, apportent plus de vocabulaire que les parties qui sont consacrées à d'autres thèmes.

Toutefois, nous avons relevé quelques irrégularités. PARCOURS, nous l'avons vu, ne s'inscrivait pas dans la chronologie : il était avec son apport lexical pauvre plus proche des quatre derniers textes, que ceux qui lui sont contemporains. De la même façon, nous avons constaté que le troisième texte ELHAKI, présente un apport lexical assez spectaculaire. Cet apport est significatif lorsque nous examinons le corpus dans son ensemble.

Ensuite, pour mieux apprécier le phénomène de l'accroissement du vocabulaire, nous avons mesuré l'apport lexical à partir du dernier texte jusqu'au premier, d'où le résultat suivant :

**FEMME > LETTRES > ABOUNOUR > ZRIREK > GENET >
ELHAKI > MILLE >> AÎLEN > PARCOURS**

Ce classement ne converge pas avec celui de la richesse lexicale, du moment que tous les textes ont vu leur rang changé. Soulignons, par ailleurs, qu'un segment peut jouir d'un vocabulaire riche sans avoir apporté de nouveaux vocables par rapport aux textes qui le précèdent. De la même manière, il peut y avoir un texte pauvre en vocabulaire actualisant un nombre important de vocables qui n'existent pas dans ses antécédents.

Mais ce qui est marquant dans ce classement, c'est que les textes dont l'étendue est la plus faible et qui sont aussi les derniers à être publiés se voient classer dans la première et la deuxième position. De ce fait FEMME et LETTRES actualisent un nombre de vocables qui n'existent pas dans les textes précédents. Pour trouver une explication à cet état de fait, nous nous sommes inspirée de Guy

Dugas pour qui les auteurs judéo-maghrébins de notre époque ne parviennent qu'à reproduire une œuvre répétitive sans innovation.

Donc, El Maleh a voulu que ces deux textes marquent un changement radical dans son œuvre, que ce soit au niveau de la forme qu'au niveau du contenu. D'ailleurs, l'avant dernier texte LETTRES est le seul récit épistolaire qui existe dans le corpus, de même, c'est le seul qui parle de l'exil d'El Maleh, de son arrivée à Paris. Tandis que le dernier texte FEMME est le seul qui aborde sa découverte de la ville d'Asilah et sa rencontre avec la mère de son fils spirituel Khalil El Ghrib. Ainsi, le classement auquel nous avons abouti confirme parfaitement bien cette diversification dans l'écriture d'El Maleh.

Finalement nous avons testé une autre méthode, à savoir : l'accroissement inverse. C'est une technique qui permet d'étudier le corpus en remontant du dernier texte jusqu'au premier. Cette méthode a montré que la tendance régulière à la baisse est inversée dans deux romans : ZRIREK (7313), ELHAKI (5450). Donc ces deux textes signalent la fin d'un mouvement et l'épuisement de l'inspiration de l'auteur.

Le mouvement dont il s'agit ici est surtout d'ordre générique et thématique : ELHAKI marque une rupture avec le genre romanesque. D'ailleurs aucun des cinq textes publiés après n'admettent l'étiquette de roman, du moment que l'auteur a délaissé ce genre en faveur des essais et des nouvelles. ZRIREK est aussi le dernier essai qui aborde l'écriture comme thème principale.

Toutefois, le classement établi sur la base de l'écart réduit a dévoilé un éclatement lexical au niveau du texte ABOUNOUR, chose que l'ordre chronologique n'a pas pu relever. Ceci est peut-être dû aux effets du genre,

d'ailleurs ce texte est le seul recueil de nouvelles qui existe dans tout le corpus d'El Maleh.

Après avoir étudié le phénomène de la richesse lexicale et celui de l'accroissement du vocabulaire, nous abordons maintenant une autre question qui se rapporte à la fréquence des mots et à la manière avec laquelle ils sont distribués dans le corpus. Il s'agit des catégories grammaticales.

CHAPITRE 3

Catégories grammaticales

Après avoir approché la structure lexicale des œuvres d'El Maleh du point de vue de la « richesse lexicale » et de « l'accroissement du vocabulaire », nous nous intéresserons dans ce qui suit aux éléments qui les composent, à savoir les parties du discours, ou bien, les catégories grammaticales.

De nombreux spécialistes en lexicométrie ont pu affirmer qu'il y a une corrélation entre la distribution des catégories grammaticales et les caractéristiques des œuvres littéraires, citons, entre autres, E. BRUNET qui stipule que les catégories grammaticales¹⁸⁹ restent un indice révélateur de la distinction des genres, des styles et des écrivains. Il affirme, aussi, que le choix d'une telle ou telle catégorie s'opère d'une manière instinctive et spontanée, d'ailleurs, il serait difficile pour un écrivain de prétendre¹⁹⁰ que le choix se fait d'une manière intentionnelle. Aussi, GUIRAUD a prouvé que la distribution des parties du discours dans les œuvres littéraires est profondément influencée¹⁹¹ par le genre et

¹⁸⁹ Etienne BRUNET, *op.cit.*, 1983, p. 836

¹⁹⁰ Etienne BRUNET, *op.cit.*, 1981, p. 295

¹⁹¹ Pierre GUIRAUD, *op.cit.*, pp. 34-53

l'époque où elle s'inscrit.

L'étude de la distribution des catégories grammaticales permet donc de saisir le choix intuitif et authentique qu'opère l'écrivain lors de l'acte scriptural et, ainsi, apporter de nouveaux éléments de diagnostic, pour comprendre au mieux les particularités de l'écriture maléhiennne.

A cet effet, nous découvrirons, tout d'abord, comment les catégories grammaticales se répartissent dans notre corpus, afin de mener une comparaison externe et une autre interne. Cette dernière se base essentiellement sur le calcul des effectifs théoriques et des écarts réduits. Ensuite, nous aurons recours à deux autres tests statistiques pour dégager les lois qui régissent la distribution des catégories grammaticales dans les œuvres d'El Maleh. Il s'agit du test de Pearson et de l'analyse factorielle des correspondances. Le premier test sert à exprimer l'écart existant entre le réel et le théorique en probabilité pour détecter les divergences et les convergences qui existent entre les textes. Tandis que le deuxième vise à faire ressortir les affinités qui s'établissent entre les catégories grammaticales dans les différentes parties du corpus.

A. Distribution des catégories grammaticales

Le logiciel *Hyperbase* répartit les textes selon dix catégories fondamentales, à savoir : verbes, substantifs, adjectifs, déterminants, pronoms, numéraux, interjections, prépositions, adverbes et conjonctions. Le tableau ci-dessous met en évidence cette répartition.

Tableau II. 30 : Répartition des textes selon les catégories grammaticales

	PARC	AÎL	MIL	ELHA	GEN	ABOU	ZRIR	FEM	LET	TOT
Verbe	7727	10090	10409	15025	6710	3506	10966	743	3122	68298
Sub	12756	17074	16826	22949	11864	5739	19172	1276	4297	111953
Adj	4414	5345	4848	7477	3552	1795	6078	387	1434	35330
Num	214	258	216	244	268	99	323	20	87	1729
Pron	6202	8166	9117	13780	5922	2973	9985	766	3233	60144
Adv	3048	3329	3387	4978	2625	1303	4278	350	1415	24713
Déter	5959	7429	7606	10601	5639	2570	8902	616	1897	51219
Prépo	7727	9923	10190	14891	7110	3701	12398	884	3147	69971
Conj	1553	1632	2309	3203	2088	816	3552	207	935	16295
Interj	29	80	26	85	21	16	22	0	8	287

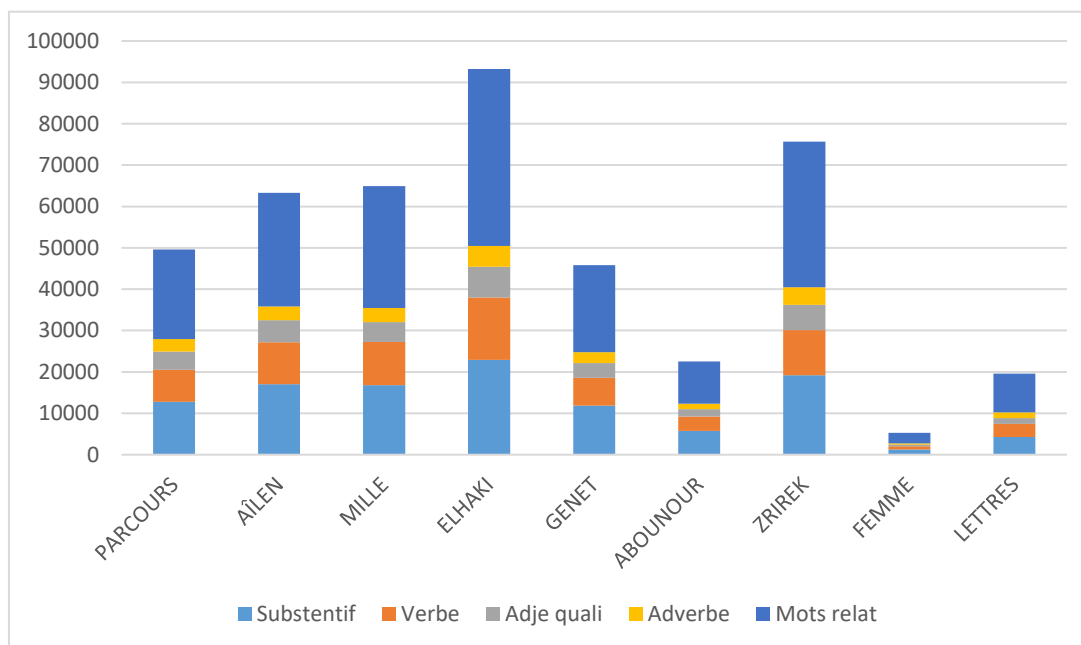
La version lemmatisée du corpus compte 522421 occurrences. Après avoir éliminé la ponctuation et les symboles, ce chiffre a été révisé à la baisse pour atteindre le seuil de 439939 occurrences. Pour faciliter l'exploitation des données présentées dans le tableau ci-dessus, nous avons regroupé dans la classe des mots de relation, tous les mots grammaticaux ainsi que les interjections. Le tableau suivant présente cette nouvelle répartition :

Tableau II. 31 : Répartition des textes selon les catégories grammaticales

TEXTES	Substantif	Verbe	Adjectif	Adverbe	Mot relationnel	TOTAL
PARCOURS	12756	7727	4414	3048	21684	49629
AÎLEN	17074	10090	5345	3329	27488	63326
MILLE	16826	10409	4848	3387	29464	64934
ELHAKI	22949	15025	7477	4978	42804	93233
GENET	11864	6710	3552	2625	21048	45799
ABOUNOUR	5739	3506	1795	1303	10175	22518
ZRIREK	19172	10966	6078	4278	35182	75676
FEMME	1276	743	387	350	2493	5249
LETTRES	4297	3122	1434	1415	9307	19575
TOTAL	111953	68298	35330	24713	199645	439939

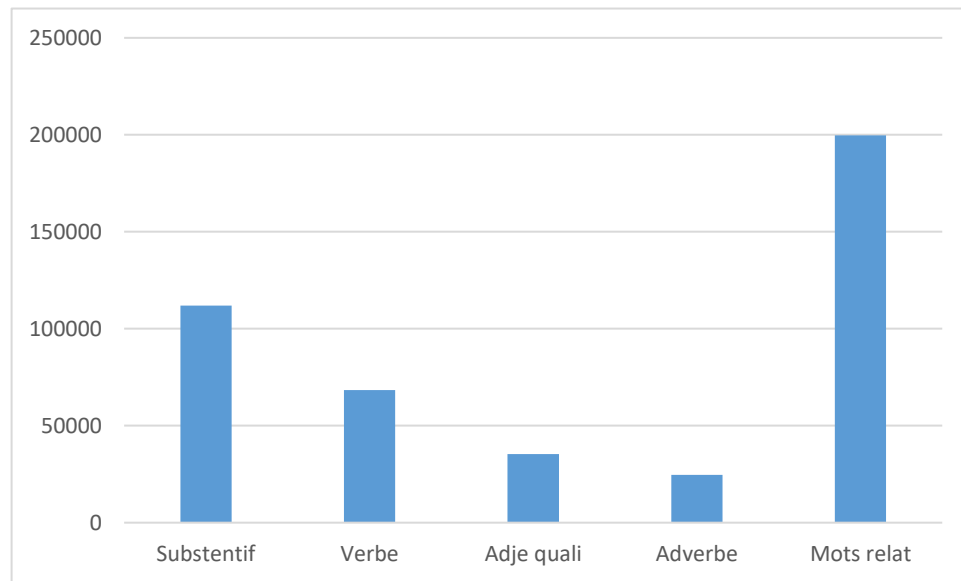
Ce tableau peut être, également, représenté sous forme d'un histogramme :

Figure II. 10 : Répartition des catégories grammaticales dans les textes



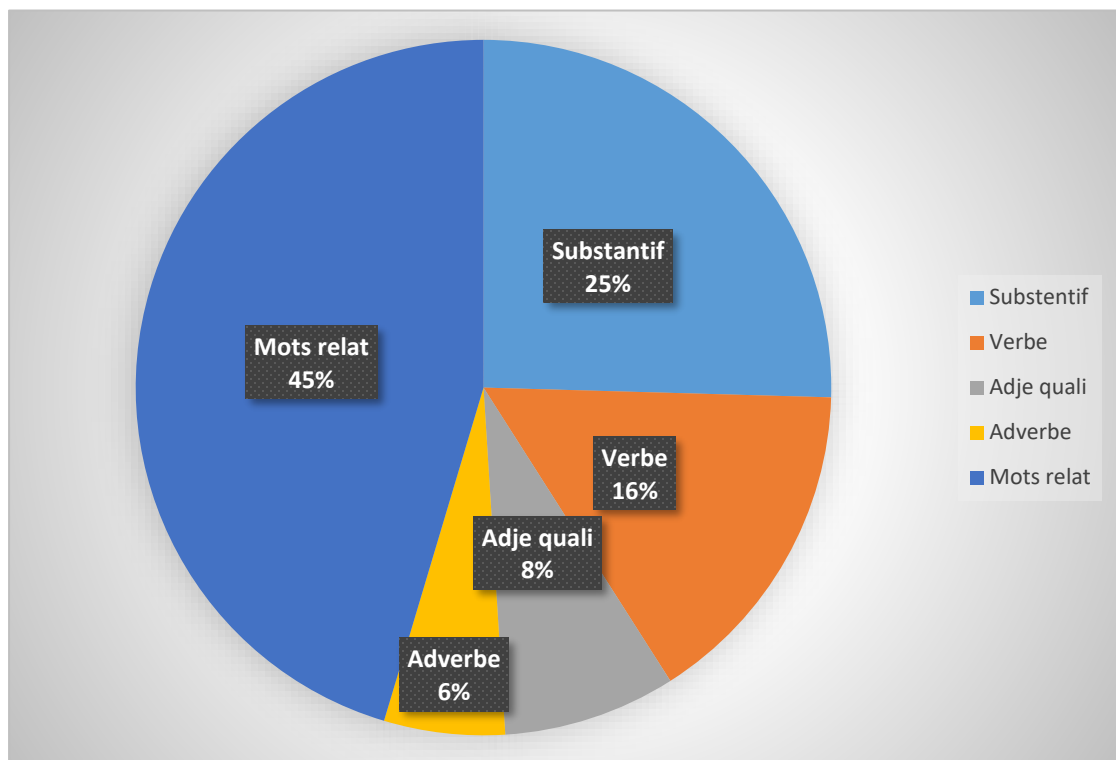
Ce graphique visualise la distribution des catégories grammaticales au sein de chaque texte. Avant de se livrer aux commentaires, nous proposons d'illustrer sous forme d'historgramme le classement hiérarchique des différentes catégories grammaticales de notre corpus.

Figure II. 11 : Répartition des catégories grammaticales dans le corpus



Nous pouvons traduire ces résultats en pourcentages dans un diagramme circulaire (Camembert) :

Figure II. 12 : Répartition des catégories grammaticales dans le corpus



Une lecture rapide de ces graphiques révèle une prépondérance des mots de relation avec un taux de 45%, suivi des substantifs avec un taux 25%. Les verbes se mettent à la troisième position avec un pourcentage de 16%. Les adjectifs et les adverbes se rangent, respectivement, à la quatrième et cinquième position, avec un pourcentage de 8% et de 6%.

Nous avons déjà signalé au début de ce chapitre que l'utilisation des catégories grammaticales varie d'un écrivain à un autre. Chez El Maleh, nous avons pu dévoiler qu'il a une tendance à favoriser les mots de relation sur les autres catégories, ceci ne constitue pas une particularité, mais une tendance générale en français. Cela étant dit, nous pouvons comparer notre écrivain avec d'autres auteurs.

B. Comparaison externe

L'utilisation des catégories grammaticales est une sorte d'emprunte propre à chaque écrivain. Le rôle de la statistique lexicale est de déceler les caractéristiques de cette emprunte scripturale. En adoptant la même catégorisation grammaticale, BRUNET¹⁹² a fait une étude détaillée sur les vocabulaires de plusieurs écrivains, dans le but de détecter le penchant de chacun d'eux à utiliser ou à négliger une catégorie grammaticale quelconque. Il a pu, donc, constater que parmi les écrivains du XIX^{ème} siècle, CHATEAUBRIAND préférait la catégorie nominale, ZOLA chérissait les participes et les verbes, PROUST avait un penchant pour les mots de relation, HUGO optait pour l'utilisation des substantifs et des verbes et enfin GIRAUDOUX qui favorisait les adjectifs.

Concernant Edmond Amrane El Maleh, nous constatons qu'il rejoint PROUST en préférant l'utilisation des mots de relation et d'ailleurs, c'est l'une des caractéristiques de la langue française. De même, il s'approche d'HUGO pour ce qui est de l'utilisation des substantifs et des verbes.

Après avoir comparé les effectifs réellement observés dans notre corpus de recherche avec d'autres écrivains, nous nous intéresserons, dans le point suivant, au calcul des effectifs théoriques des différentes catégories grammaticales, dans le but de construire un modèle théorique des catégories.

C. Effectifs théoriques des catégories grammaticales

Comme nous l'avons déjà explicité, nous procéderons au calcul des effectifs théoriques. Ces derniers nous serviront pour effectuer une comparaison

¹⁹² Etienne BRUNET, *op.cit.*, 1988, pp. 171-172

interne entre les différents textes par le biais de l'écart réduit. La comparaison, que nous allons mener, peut avoir deux issues différentes. En effet, si l'écart réduit est significatif, cela peut nous renseigner sur le style de l'écrivain, par contre, s'il n'est pas significatif, cela affirme l'inscription de l'écrivain dans le fonctionnement normal de la langue.

Le calcul des effectifs théoriques nécessite la connaissance de la probabilité « p » qui correspond à la part des catégories grammaticales d'un texte par rapport à l'intégralité du corpus.

Tableau II. 32 : Effectifs des catégories par textes & leurs probabilités

Textes	Effectifs	probab.p	Probab.q
PARCOURS	49629	0,112808821	0,887191179
AÎLEN	63326	0,143942683	0,856057317
MILLE	64934	0,147597735	0,852402265
ELHAKI	93233	0,211922562	0,788077438
GENET	45799	0,104103069	0,895896931
ABOUNOUR	22518	0,051184369	0,948815631
ZRIREK	75676	0,172014757	0,827985243
FEMME	5249	0,0119312	0,9880688
LETTRES	19575	0,044494805	0,955505195
TOTAL	439939	1	0

En nous basant sur les données présentées dans le tableau ci-dessus, nous pouvons calculer l'effectif théorique de chaque catégorie moyennant la formule suivante :

Effectifs théorique = Effectif total d'une catégorie $\times$ Probabilité « P » du texte

Sachant que :

$$P = \frac{\text{Effectif total des catégories grammaticales d'un texte donné}}{\text{Effectif total des catégories grammaticales du corpus}}$$

Nous allons illustrer cette formule en calculant l'effectif théorique des substantifs dans le premier texte de notre corpus (PARCOURS) :

*Effectifs théoriques*_{substantifs}

$$= \text{Effectifs total des substantifs} \times \text{Probabilité « P » du texte}$$

$$= 111953 \times 0,112808821222942$$

$$= 12629,28596$$

Cependant, l'effectif réel des substantifs dans le texte PARCOURS est de 12756 occurrences, donc :

$$\text{L'écart absolu} = 12756 - 12629,2859623721$$

$$= 126,7140376279$$

Soit un excédent de 126,71 par rapport au modèle.

Ainsi, en appliquant la même démarche à l'ensemble des catégories grammaticales, nous obtenons le tableau des effectifs théoriques suivant :

Tableau II. 33 : Répartition des textes selon les effectifs théoriques des catégories grammaticales

Textes	Substantif	Verbe	Adjectif	Adverbe	Mots de relation
PARCOURS	12629,29	7704,62	3985,54	2787,84	22521,72
AÎLEN	16114,82	9831,00	5085,49	3557,26	28737,44
MILLE	16524,01	10080,63	5214,63	3647,58	29467,15
ELHAKI	23725,37	14473,89	7487,22	5237,24	42309,28
GENET	11654,65	7110,03	3677,96	2572,70	20783,66
ABOUNOUR	5730,24	3495,79	1808,34	1264,92	10218,70
ZRIREK	19257,57	11748,26	6077,28	4251,00	34341,89
FEMME	1335,73	814,88	421,53	294,86	2382,00
LETTRES	4981,33	3038,91	1572,00	1099,60	8883,17
TOTAL	111953	68298	35330	24713	199645

Après avoir calculé les effectifs théoriques, nous passerons au calcul des écarts réduits, en appliquant la formule suivante :

$$\text{Ecart réduit} = \frac{\text{Effectif réel} - \text{Effectif théorique}}{\sqrt{\text{Effectif théorique} \times q}}$$

Dans les tableaux qui suivent, nous confrontons, les effectifs réels avec ceux théoriques, pour déboucher sur l'écart absolu qui est obtenu en faisant la différence entre les deux effectifs cités auparavant. Aussi, en rapportant cet écart absolu rapporté à la racine carrée du produit de l'effectif théorique et la probabilité d'échec q , nous obtenons l'écart réduit z .

Tableau II. 34 : Répartition des textes selon les effectifs théoriques des substantifs

TEXTES	Effectif réel	Effectif théorique	Écart absolu	Écart réduit (z)
PARCOURS	12756	12629,286	126,714	1,197
AÎLEN	17074	16114,815	959,185	8,167
MILLE	16826	16524,009	301,991	2,545
ELHAKI	22949	23725,367	-776,367	-5,678
GENET	11864	11654,651	209,349	2,049
ABOUNOUR	5739	5730,244	8,756	0,119
ZRIREK	19172	19257,568	-85,568	-0,678
FEMME	1276	1335,734	-59,734	-1,644
LETTRES	4297	4981,327	-684,327	-9,919
TOTAL	111953	111953		

Tableau II. 35 : Répartition des textes selon les effectifs théoriques des verbes

TEXTES	Effectif réel	Effectif théorique	Écart absolu	Écart réduit (z)
PARCOURS	7727	7704,617	22,383	0,271
AÎLEN	10090	9830,997	259,003	2,823
MILLE	10409	10080,630	328,370	3,542
ELHAKI	15025	14473,887	551,113	5,160
GENET	6710	7110,031	-400,031	-5,012
ABOUNOUR	3506	3495,790	10,210	0,177
ZRIREK	10966	11748,264	-782,264	-7,931
FEMME	743	814,877	-71,877	-2,533
LETTRES	3122	3038,906	83,094	1,542
TOTAL	68298	68298		

Tableau II. 36 : Répartition des textes selon les effectifs théoriques des adjectifs

Textes	Effectif réel	Effectif théorique	Écart absolu	Écart réduit (z)
PARCOURS	4414	3985,536	428,464	7,205
AÎLEN	5345	5085,495	259,505	3,933
MILLE	4848	5214,628	-366,628	-5,499
ELHAKI	7477	7487,224	-10,224	-0,133
GENET	3552	3677,961	-125,961	-2,194
ABOUNOUR	1795	1808,344	-13,344	-0,322
ZRIREK	6078	6077,281	0,719	0,010
FEMME	387	421,529	-34,529	-1,692
LETTRES	1434	1572,001	-138,001	-3,561
TOTAL	35330	35330		

Tableau II. 37 : Répartition des textes selon les effectifs théoriques des adverbes

Textes	Effectif réel	Effectif théorique	Écart absolu	Écart réduit (z)
PARCOURS	3048	2787,844	260,156	5,231
AÎLEN	3329	3557,256	-228,256	-4,136
MILLE	3387	3647,583	-260,583	-4,673
ELHAKI	4978	5237,242	-259,242	-4,035
GENET	2625	2572,699	52,301	1,089
ABOUNOUR	1303	1264,919	38,081	1,099
ZRIREK	4278	4251,001	26,999	0,455
FEMME	350	294,856	55,144	3,231
LETTRES	1415	1099,600	315,400	9,73
TOTAL	24713	24713		

Tableau II. 38 : Répartition des textes selon les effectifs théoriques des mots de relation

Textes	Effectif réel	Effectif théorique	Écart absolu	Écart réduit (z)
PARCOURS	21684	22521,717	-837,717	-5,926
AÎLEN	27488	28737,437	-1249,437	-7,966
MILLE	29464	29467,150	-3,150	-0,019
ELHAKI	42804	42309,280	494,720	2,709
GENET	21048	20783,657	264,343	1,937
ABOUNOUR	10175	10218,703	-43,703	-0,44
ZRIREK	35182	34341,886	840,114	4,98
FEMME	2493	2382,004	110,996	2,288
LETTRES	9307	8883,165	423,835	4,497
TOTAL	199645	199645		

Lorsque la valeur absolue d'un écart réduit est supérieure ou égale à 2, dans le sens positif ou négatif, le test se considère comme significatif. Nous constatons donc que, parmi les 45 valeurs obtenues ci-dessus (5 catégories grammaticales × 9 textes), 28 sont significatives (62,22%) et seulement 17 qui ne le sont pas.

En affectant la valeur 1 au test significatif et la valeur 0 à celui qui ne l'est pas, nous aurons le tableau suivant :

Tableau II. 39 : Traduction binaire de la significativité du test

Textes	Substantif	Verbe	Adjectif	Adverbe	Mots relation	Total
PARCOURS	0	0	1	1	1	3
AÎLEN	1	1	1	1	1	5
MILLE	1	1	1	1	0	4
ELHAKI	1	1	0	1	1	4
GENET	1	1	1	0	0	3
ABOUNOUR	0	0	0	0	0	0
ZRIREK	0	1	0	0	1	2
FEMME	0	1	0	1	1	3
LETTRES	1	0	1	1	1	4
TOTAL	5	6	5	6	6	28

La lecture de ce tableau, nous montre que le texte AÎLEN est le seul à avoir tous les tests significatifs, contrairement à ABOUNOUR qui n'en a aucun. Pour ce qui est des textes restants, MILLE, ELHAKI et LETTRES en ont quatre ; PARCOURS, GENET et FEMME en ont trois. Finalement le texte ZRIREK présente un seul test significatif.

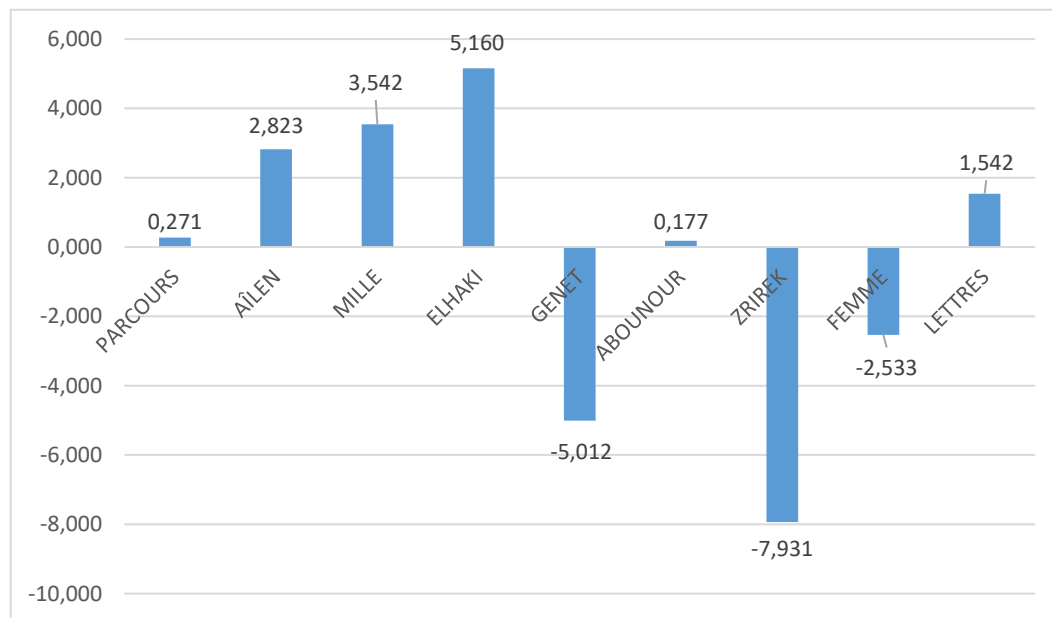
D'après le même tableau, si nous vérifions le nombre des écarts réduits significatifs au niveau des catégories grammaticales, nous constatons alors que la catégorie qui a le plus d'écarts réduits significatifs est celle des verbes, adverbess et mots de relation, avec 6 écarts supérieurs à 2 pour chacune des trois catégories, suivie de celles des substantifs et des adjectifs avec cinq tests significatifs.

Dans les graphiques qui vont suivre, nous découvrirons les valeurs de l'écart réduit en parcourant l'ensemble des textes de notre corpus de recherche.

Tout d'abord, nous commencerons par les catégories qui ont l'écart le plus significatif, à savoir celles des verbes, adverbess et mots de relation.

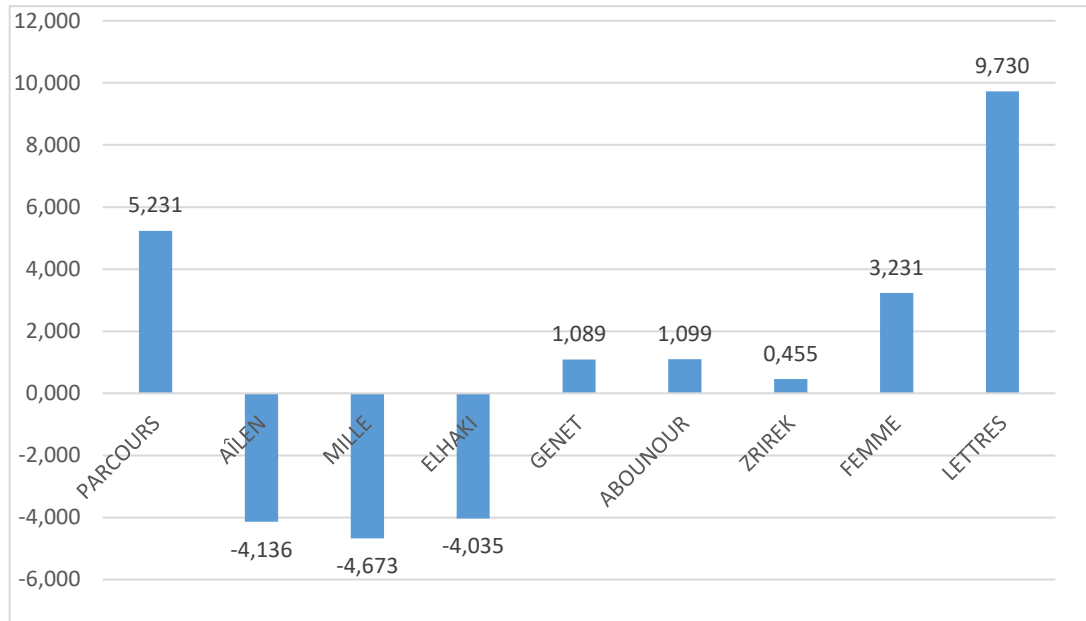
Pour ce qui est des verbes, nous constatons qu'ils ont enregistré trois écarts positifs, ce qui nous renseigne sur l'emploi excédentaire des verbes dans AÎLEN, MILLE et ELHAKI. Par contre, il y a trois autres écarts qui ont des valeurs négatives et qui nous informent sur le caractère déficitaire de cette catégorie dans les textes GENET, ZRIREK et FEMME.

Figure II. 13 : Répartition des textes selon l'écart réduit des verbes



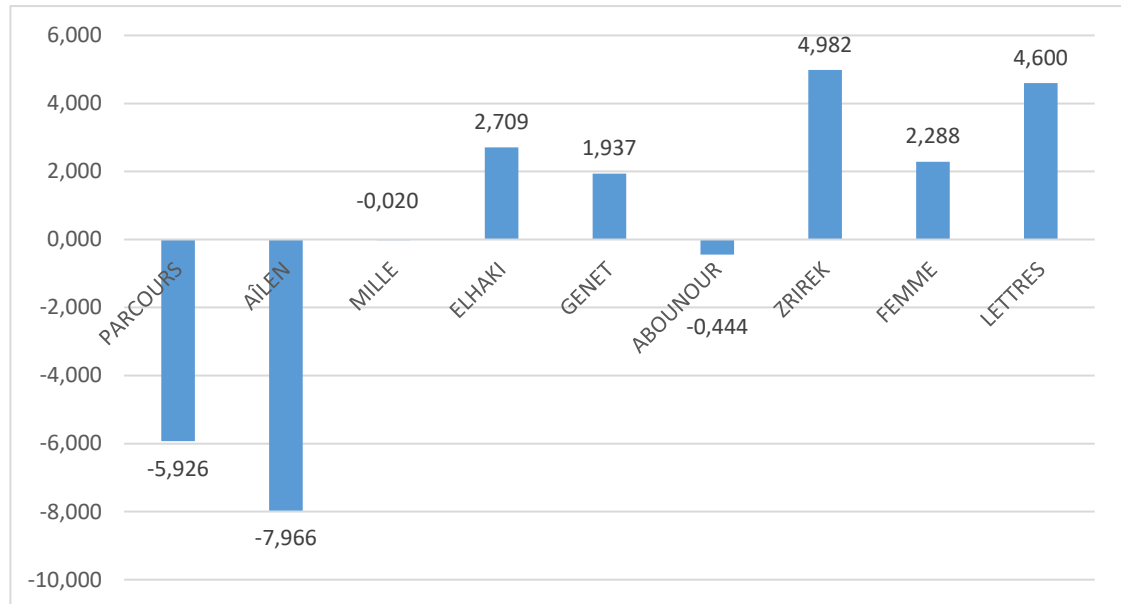
Pour les adverbess, nous relevons que trois textes ont un écart réduit positif PARCOURS, FEMME et LETTRES, ce qui augure de l'emploi excédentaire des adverbess dans ces textes. Réciproquement, nous avons recensé trois autres textes avec des écarts négatifs, à savoir : AÎLEN, MILLE et ELHAKI.

Figure II. 14 : Répartition des textes selon l'écart réduit des adverbes

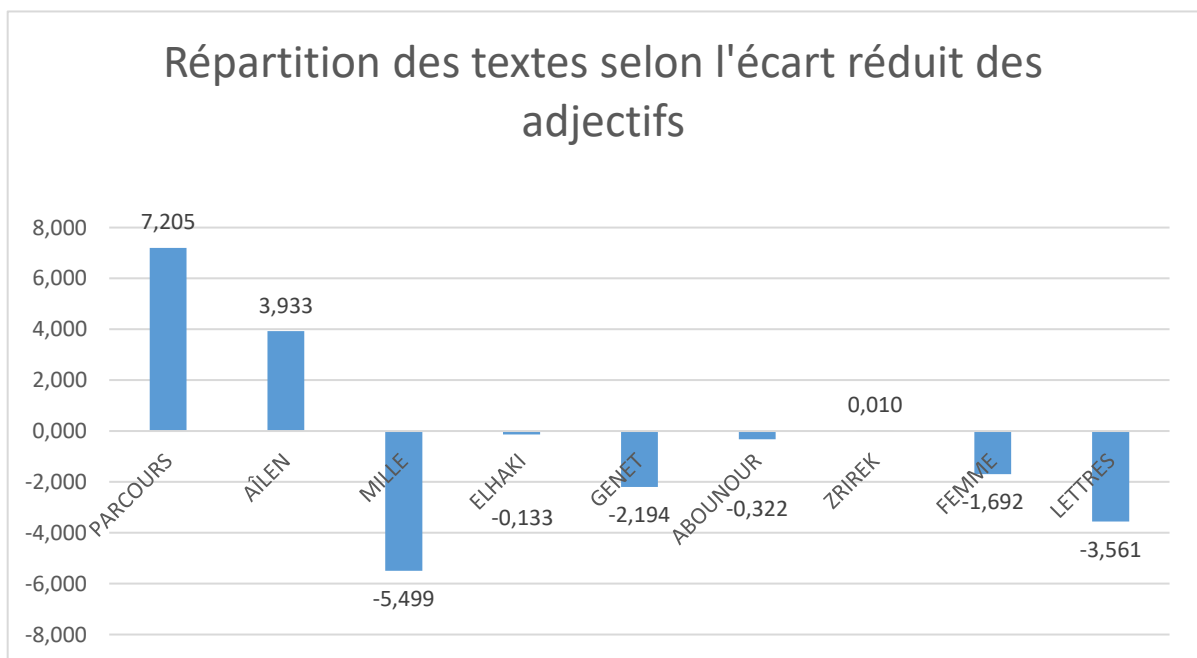


Concernant les mots de relation, nous remarquons que quatre textes, ELHAKI, FEMME, ZRIREK et LETTRES ont un emploi excédentaire des mots de relation, tandis que PARCOURS et AÏLEN présentent un usage déficitaire.

Figure II. 15 : Répartition des textes selon l'écart réduit des mots de relation

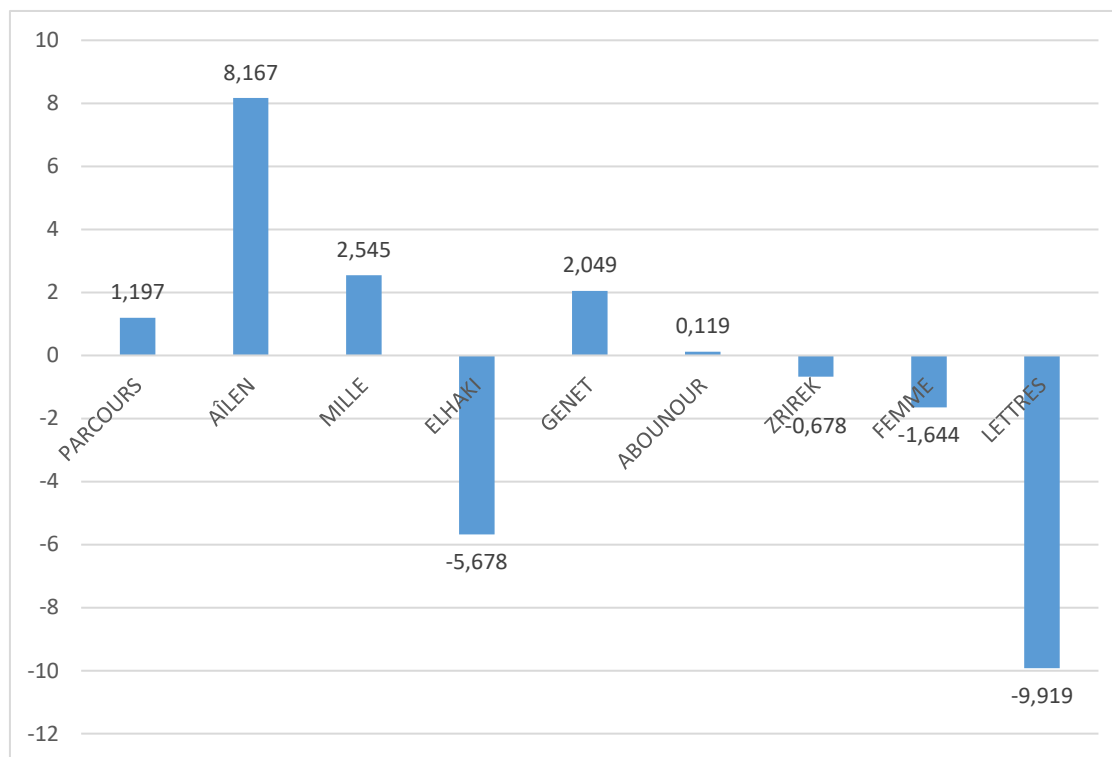


S'agissant des adjectifs, les textes PARCOURS et AÏLEN se caractérisent par un emploi excédentaire de cette catégorie, tandis que MILLE, GENET et LETTRES ont un emploi déficitaire.



En fin de compte, nous clôturons avec la catégorie la moins représentée, à savoir les substantifs. Nous voyons alors que les textes AÎLEN, MILLE et GENET présentent tous un écart positif, ce qui augure de l'emploi excédentaire de cette catégorie, cependant ELHAKI et LETTRES ont des écarts négatifs et donc un emploi déficitaire.

Figure II. 16 : Répartition des textes selon l'écart réduit des substantifs



Nous pouvons dresser un tableau récapitulatif illustrant les résultats obtenus, pour nous constituer une idée claire et nette sur la distribution des catégories grammaticales dans le corpus d'Edmond Amran El Maleh.

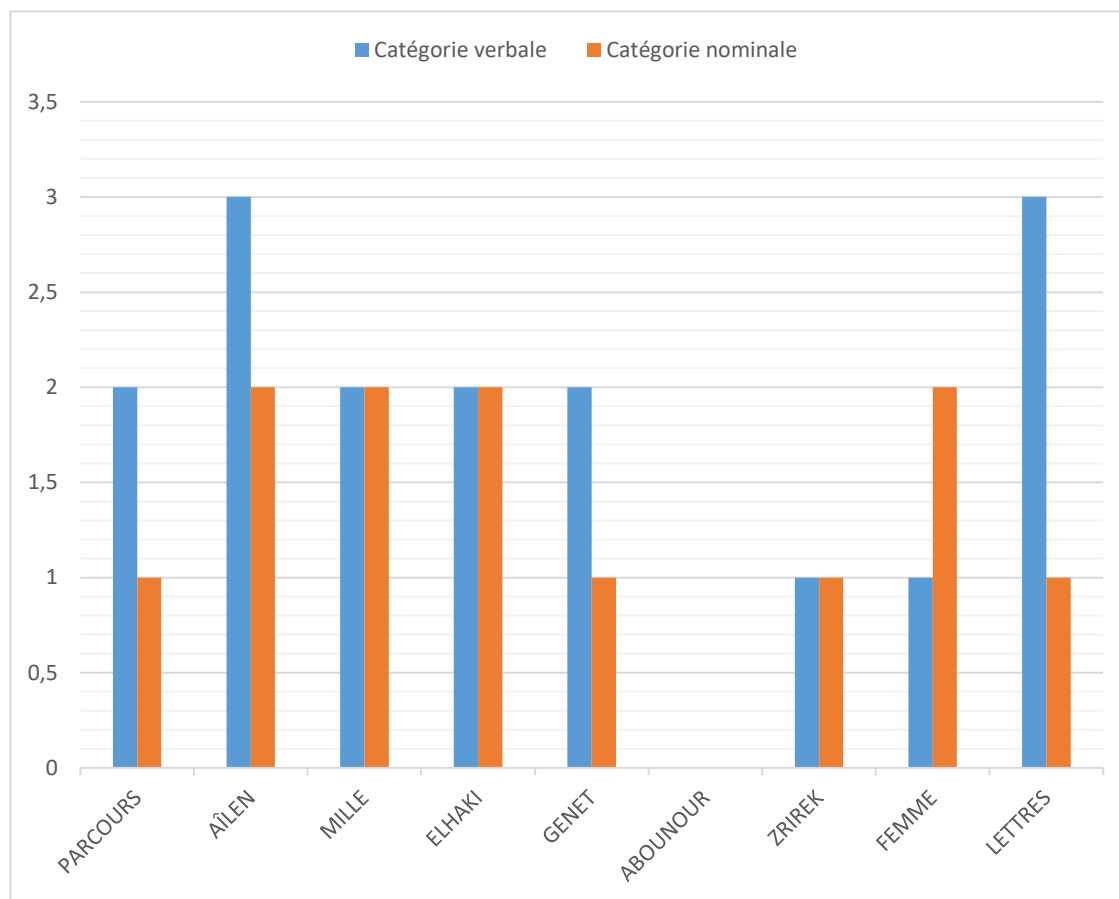
Rappelons que nous affectons la valeur 1 au test significatif et la valeur 0 à celui qui ne l'est pas :

Tableau II. 40 : Répartition des textes et catégories grammaticales selon l'écart réduit

TEXTES	Substantif	Adjectif	Mots de relation	Catégorie nominale	Verbe	Adverbe	Catégorie verbale	Total
PARCOURS	0	1	1	1	0	1	2	3
AÎLEN	1	1	1	2	1	1	3	5
MILLE	1	1	0	2	1	1	2	4
ELHAKI	1	0	1	2	1	1	2	4
GENET	1	1	0	1	1	0	2	3
ABOUNOU	0	0	0	0	0	0	0	0
ZRIREK	0	0	1	1	1	0	1	2
FEMME	0	0	1	2	1	1	1	3
LETTRES	1	1	1	1	0	1	3	4
TOTAL	5	5	6	12	6	6	16	28

Pour mieux appréhender les données affichées sur ce tableau, nous proposons de le convertir en une illustration graphique, tout en accouplant les verbes et les adverbes sous le nom générique « catégorie verbale » et en attribuant aux substantifs, adjectifs et mots de relation l'étiquette de « catégorie nominale ».

Figure II. 17 : Confrontation des textes selon les catégories verbales et nominale



Nous notons l'existence majoritaire d'une corrélation négative entre les deux catégories. En effet, soit les deux pôles évoluent de manière égalitaire soit la catégorie verbale prime sur celle nominale, seul le texte FEMME fait dérogation à la règle.

Autrement dit, dans les textes PARCOURS, AÏLEN, GENET et LETTRES l'utilisation de la catégorie verbale dépasse de loin la catégorie nominale. Pour FEMME c'est l'inverse qui se produit. C'est à dire, le texte fait appel à la catégorie nominale beaucoup plus celle verbale.

Néanmoins, nous constatons une certaine équivalence entre l'emploi des deux catégories dans les textes MILLE, ELHAKI et ZRIREK.

Cependant, nous ne pouvons pas émettre des jugements à ce niveau, nous préférons approfondir l'analyse tout en ayant recours à un autre test pour dire si la distribution des catégories grammaticales dans le corpus relève du hasard ou bien elle est due à d'autres facteurs qui peuvent être d'ordre : stylistiques, thématiques, chronologiques, etc. Il s'agit bien évidemment du test de PEARSON.

D. Test de Pearson

Selon Ch. MULLER, le χ^2 ou test de Pearson¹⁹³ permet d'apprécier en probabilité l'écart constaté entre l'observation réelle et le modèle théorique, quel que soit le nombre de variables. Chose qui nous permettra de voir si cet écart engendré est dû au hasard ou bien, il est le produit d'un autre facteur.

Nous expliciterons la démarche de calcul en appliquant le test de χ^2 à la catégorie des substantifs dans le texte PARCOURS. Nous voulons donc savoir si la répartition de cette catégorie grammaticale dans le premier texte de notre corpus est juste un fait de hasard ou pas. Alors pour répondre à cette interrogation, nous formulons les deux hypothèses suivantes :

$$\begin{cases} H_0: \text{les substantifs se répartissent au hasard dans le texte} \\ H_1: \text{la répartition des substantifs n'est pas due au hasard} \end{cases}$$

¹⁹³ Charles MULLER, *op.cit.*, 1968, p. 95

Posons la métrique générale permettant de mesurer la distance de khi-deux :

$$D^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(N_{ij} - n_{tij})^2}{n_{tij}}$$

Avec : N_{ij} = la fréquence réelle

n_{tij} = La fréquence théorique ou fréquence calculée

Du moment que nous n'avons que les variables lignes, la formule deviendra :

$$D^2 = \sum_{i=1}^m \frac{(N_i - n_{ti})^2}{n_{ti}}$$

Et le nombre des degrés de liberté et de khi-deux sera :

$$\vartheta = k - 1$$

$$= 2 - 1$$

$$= 1$$

D'où la règle de décision suivante :

$$\begin{cases} \text{Si } d^2 > \chi^2_{(k-1)(m-1);\alpha} & \text{on rejette } H_0 \\ \text{Si } d^2 \leq \chi^2_{(k-1)(m-1);\alpha} & \text{on ne peut rejeter } H_0 \end{cases}$$

Le tableau ci-après, présente l'application du test de khi-deux aux substantifs dans le texte PARCOURS :

Tableau II. 41 : Test de Khi2 appliqué à la classe des substantifs dans PARCOURS

PARCOURS	Réel	Théorique	Ecart	Ecart ²	KHI-DEUX
Substantif	12756	12629,29	126,71	16056,45	1,27
Reste	36873	36999,71	-126,71	16056,45	0,43
Total	49629	49629	0		1,71

D'après le tableau, nous avons :

$$d^2 = 1.71 \leq \chi^2_{(2-1);0,05} = 3,841$$

Donc nous ne pouvons rejeter H_0 et par conséquent, la distribution des substantifs dans PARCOURS est un fait du hasard et ne se rapporte à aucun facteur autre que celui aléatoire.

Nous présenterons, ci-dessous, le tableau des khi-deux calculés pour le texte PARCOURS en fonction de toutes les catégories grammaticales.

Tableau II. 42 : Test de Khi2 appliqué à toutes les catégories grammaticales dans PARCOURS

PARCOURS	Réel	Théorique	Ecart	Ecart ²	KHI-DEUX
SUBSTENTIF	12756	12629,29	126,71	16056,45	1,27
RESTE	36873	36999,71	-126,71	16056,45	0,43
TOTAL	49629	49629	0		1,71
VERBE	7727	7704,62	22,38	501,00	0,07
RESTE	41902	41924,38	-22,38	501,00	0,01
TOTAL	49629	49629	0		0,08
Adjectif	4414	3495,79	918,21	843109,57	241,18
RESTE	45215	45643,46	-428,46	183581,70	4,02
TOTAL	49629	49139,25436	489,7456375		245,20
Adverbe	3048	2787,84	260,16	67680,94	24,28
RESTE	46581	46841,16	-260,16	67680,94	1,44
TOTAL	49629	49629	0		25,72
Mots de relation	21684	22521,72	-837,72	701769,96	31,16
RESTE	27945	27107,28	837,72	701769,96	25,89
TOTAL	49629	49629	0		57,05

Nous généraliserons ces calculs à l'ensemble des textes composant notre corpus de recherche selon la distribution des catégories grammaticales. Les résultats seront présentés dans le tableau suivant :

Tableau II. 43 : Test de Khi2 appliqué à toutes les catégories grammaticales dans tous les textes

TEXTES	Substantif	Verbe	Adjectif	Adverbe	Mots de relation	Total
PARCOURS	1,71	0,08	245,20	25,72	57,05	329,75
AÎLEN	76,58	8,08	14,40	15,52	99,46	214,03
MILLE	7,40	12,66	25,78	19,72	0,00	65,57
ELHAKI	34,08	24,84	0,02	13,60	10,59	83,12
GENET	10,59	26,64	4,69	1,13	6,16	49,21
ABOUNOUR	0,02	0,04	0,11	1,21	0,34	1,72
ZRIREK	0,51	61,66	0,00	0,18	37,63	99,98
FEMME	3,58	7,51	3,08	10,93	9,47	34,56
LETTRES	126,10	2,69	13,17	95,85	37,02	274,84
Total	260,57	144,19	306,44	183,86	257,71	1152,77

D'après la table de la loi de khi-deux à 1 degré de liberté, si nous prenons le fractile d'ordre α dont la probabilité est égale à 5%, nous aurons : $\chi^2_{(2-1);0,05} = 3,841$. Donc, la comparaison de toutes les valeurs de khi-deux présentées dans le tableau ci-haut, nous montre que les 2/3 des tests sont significatifs. Si nous dressons un classement en fonction des valeurs des khi2 de chaque catégorie grammaticale contenue dans chaque texte de notre corpus, nous voyons, les mots de relation, à la tête de ce classement, avec 7 valeurs significatives. Les verbes et les adverbes, quant à eux, se partagent la deuxième place avec 6 valeurs significatives. En position ex æquo, nous trouvons, également, les adjectifs et les substantifs qui se mettent au troisième rang, avec 5 valeurs significatives pour chaque catégorie.

Pour une meilleure interprétation, nous procéderons à l'analyse des données du tableau en trois directions :

Premièrement, une lecture verticale, nous montre que :

- Pour les substantifs, nous rejetons H_0 en faveur de H_1 dans AÎLEN, MILLE, ELHAKI, GENET et LETTRES, ce qui nous permet de dire que la distribution des substantifs n'est pas due au hasard, elle trouve son explication dans le style de l'écrivain. Cependant dans les textes restant le test n'est pas significatif.
- Pour les verbes, le test est significatif pour AÎLEN, MILLE, GENET, ZRIREK et FEMME, ce qui permet de rejeter l'hypothèse nulle, contrairement aux textes qui restent.
- Pour les adjectifs, le test est significatif pour PARCOURS, AÎLEN, MILLE, GENET et LETTRES, donc l'hypothèse nulle est rejetée. Cependant pour les autres textes, le test de khi-deux n'est pas significatif.
- Pour les adverbes, le test est significatif pour PARCOURS, AÎLEN, MILLE, ELHAKI, FEMME et LETTRES et ne l'est pas pour les autres textes. Donc, nous pouvons rejeter l'hypothèse nulle pour les textes précédemment cités. Cependant pour le reste, le test de khi-deux n'est pas significatif.
- Pour les mots de relation, le test est significatif pour PARCOURS, AÎLEN, ELHAKI, GENET, ZRIREK, FEMME et LETTRES, ce qui nous permet de rejeter l'hypothèse nulle. Cependant pour le reste le test de khi-deux n'est pas significatif.

Si nous nous intéressons à la somme des khi-deux de chaque colonne, nous serons alors à 8 degrés de liberté $\chi^2_{(9-1);0,05} = 15,507$. A cet effet, nous constatons que le $\chi^2_{calculé(reel)} > \chi^2_{théorique}$ (lu sur la table de la loi de Khi-deux) pour l'ensemble des catégories grammaticales. Ce qui traduit le rejet de l'hypothèse nulle en faveur de l'hypothèse alternative, selon laquelle la distribution des catégories grammaticales dans l'ensemble du corpus est un effet de style et non pas du hasard.

Deuxièmement, une lecture horizontale nous renseigne sur la distribution des catégories grammaticales dans chaque texte pris isolément. Parcourant, maintenant, les textes un à un pour voir ce qu'il en est :

- Pour PARCOURS, l'emploi des adjectifs, adverbess et des mots de relation n'est pas le résultat d'un simple jeu de hasard. Cependant la distribution des substantifs et des verbes est un phénomène aléatoire.
- Pour AÎLEN, la distribution de la totalité des catégories grammaticales trouve son explication dans le style de l'écrivain.
- Pour MILLE, exception faite des mots de relation, la distribution de la totalité des catégories grammaticales n'est pas due au hasard mais elle est expliquée par le style.
- Pour ELHAKI, c'est le style qui entre en jeu dans la distribution de la totalité des catégories grammaticales (exception faite des adjectifs).

- Pour GENET, la distribution de la totalité des catégories grammaticales (exception faite des adverbes) n'est pas due au hasard mais elle est expliquée par le style.
- Pour ABOUNOUR, la distribution de l'intégralité des catégories grammaticales ne peut être expliquée par le style, elle est due au hasard.
- Pour ZRIREK, la répartition des verbes et mots de relation est expliquée par le style, cependant la distribution des autres catégories grammaticales ne peut être expliquée que par le hasard.
- Pour FEMME, la distribution des verbes, adverbes et mots de relation est expliquée par le style, pour les autres c'est un fait de hasard.
- Pour LETTRES, la distribution de la totalité des catégories grammaticales (exception faite des verbes) n'est pas due au hasard mais elle est expliquée par le style.

Si nous prenons en considération la somme des khi-deux de chaque ligne, c'est-à-dire pour chaque texte pris isolément, nous serons alors à 4 degrés de liberté $\chi^2_{(5-1);0,05} = 9,4877$, Encore une fois, nous constatons que le $\chi^2_{calculé} > \chi^2_{théorique}$ pour l'ensemble des textes, exception faite de ABOUNOUR. Cela signifie que la distribution des catégories grammaticales dans un même texte a subi l'effet du style de l'écrivain. Seulement le texte ABOUNOUR qui a dérogé à cette règle.

Finalement, nous allons étudier la totalité des moyennes des khi-deux qui est de 1152,77 que ce soit pour les lignes ou les colonnes. Dans ce cas, c'est l'intégralité des variables qui entrent en jeu, donc le degré de liberté va être changé. En effet, ce dernier se traduira par le produit obtenu pour $n = 9$ et $k = 5$:

$$\vartheta = (k - 1)(n - 1) = (9 - 1)(5 - 1) = 32$$

D'après la table de la loi de khi-deux à 32 degrés de liberté, si nous prenons la fractile d'ordre α dont la probabilité est égale à 5%, nous aurons :

$$\chi^2_{(9-1)(5-1);0,05} = 46,194$$

Dans ce cas, le khi-deux calculé qui est de 1152,77 sera supérieur à celui théorique et par conséquent, nous rejetons H_0 en faveur de H_1 .

Quand le degré de liberté est supérieur à 30, nous pouvons interpréter le khi2 calculé à travers une autre méthode tout en aboutissant au même résultat. Il s'agit de la transformation du khi2 en écart réduit à travers l'approximation de Fisher selon laquelle :

$$\text{Soit } X \sim \chi^2_{(\vartheta)} \text{ , si } \vartheta > 30 \text{ ; alors } \sqrt{2X} - \sqrt{2\vartheta - 1} \approx N(0,1)$$

$$\text{Donc } Z = \sqrt{2 \times 1152,77} - \sqrt{2 \times 32 - 1} = 40,07878505$$

D'après la table de la loi normale centrée et réduite nous avons : $Z_{\alpha=Z_{0,05}} = 1,645$

et par conséquent $Z = 40,07878505 > Z_{\alpha=Z_{0,05}} = 1,645$

Donc nous rejetons H_0 en faveur de H_1 .

Le résultat obtenu est très significatif. Ainsi, cette étude menée sur la distribution des catégories grammaticales, dans tous les segments de notre corpus, affirme que cette dernière est fortement influencée par le style de l'écrivain. Donc elle n'est pas du tout un fait du hasard.

A la fin de cette étude, nous exploiterons les mêmes données dans une analyse plus synthétique, à savoir l'analyse factorielle des correspondances.

E. Analyse factorielle des correspondances

L'analyse factorielle des correspondances¹⁹⁴ traite un tableau de contingence croisant deux variables non métriques. Elle vise à rendre visible les ressemblances ou les dissemblances qui caractérisent les variables étudiées, en mesurant la distance qui existe entre chaque profil de distribution, à savoir : les modalités marquées au niveau des lignes d'une part, les modalités notées au niveau des colonnes d'autre part, puis les modalités associant lignes et colonnes simultanément.

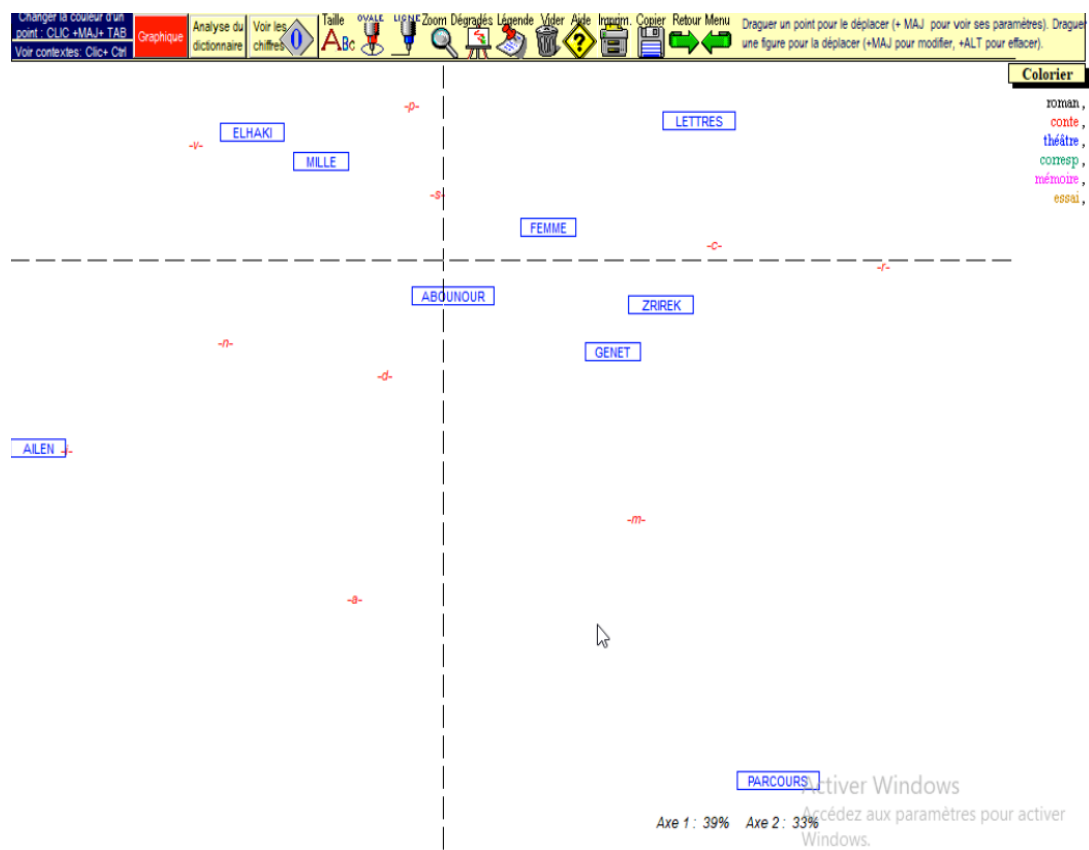
Notons que la formule de calcul des distances se base sur la métrique du khi-deux. Si deux profils se confondent, cette distance est nulle ($\chi^2 = 0$) ; et sa valeur croît au fur et à mesure que les deux profils comparés se différencient ; elle atteint son maximum quand ils sont symétriquement opposés. C'est précisément cette matrice de distances entre les modalités qui est représentée graphiquement sur un plan. Ce dernier est déterminé statistiquement par un système d'axes, appelés axes factoriels, de telle façon que l'essentiel des proximités et des éloignements entre les points du tableau se voit sur les plans.

¹⁹⁴ Caumont DANIEL, Ivanaj SILVESTER, *Analyse des données – Méthodes et outils statistiques, applications à la prise de décision, cas d'entreprises détaillés*, Paris, Dunod, 2017, p.p.93-96

L'information totale, correspondante aux données contenues dans l'intégralité du tableau, se décompose en autant de facteurs que le plus petit des nombres de lignes et de colonnes. Pourtant, tous les facteurs ne contiennent pas autant d'information. En effet, les premiers sont ceux qui en récupèrent le maximum, plus nous en calculons de nouveaux, moins ils seront représentatifs.

Sur cette base et en appliquant l'analyse factorielle des correspondances à la répartition des catégories grammaticales selon les textes, le logiciel *Hyperbase* nous fournit la carte suivante :

Figure II. 18 : AFC de la répartition des catégories grammaticales



Nous voyons, tout d'abord, sur ce graphique, que le premier facteur représente 39 % de l'information globale, tandis que le deuxième en représente 33 %. De ce fait, le premier plan factoriel formé des deux premiers axes principaux restitue 72 % de l'inertie totale, c'est-à-dire, de l'information globale contenue dans les variables (les textes et les catégories grammaticales).

Avant de commencer l'interprétation, nous voulons bien préciser la signification des abréviations contenues dans le graphique : Verbe(v), substantif(n), Adjectif(a), Numéral(m), pronom(p), adverbe (r), déterminant (d), préposition(s), Conjonction (c), Interjection(i).

Si nous prenons en considération la répartition des catégories grammaticales, nous avons le premier axe qui oppose les verbes, les pronoms, les prépositions, les conjonctions aux substantifs, déterminants, adjectifs, numéraux, adverbes, interjections.

Nous pouvons remarquer avec Véronique Magri-Mourgue¹⁹⁵ que cette opposition substantif /verbe peut s'expliquer par la binarité espace /temps. L'univers espace est rendu en langue française par le substantif, tandis que l'univers temps, il est restitué par le verbe.

Si nous prenons en considération la répartition des textes du corpus selon le premier axe, nous constatons que le texte AÎLEN favorise un emploi fort des interjections, suivi des substantifs, des déterminants et des adjectifs. Les textes PARCOURS, ZRIREK et GENET forment une constellation avec les adverbes, les adjectifs numéraux, les déterminants et les adjectifs. Du côté opposé, les textes

¹⁹⁵ Véronique Magri MOURGUE, *Le Voyage à Pas Comptés, Pour une Poétique du Récit de Voyage au XIX^e Siècle*, Paris, champion, 2009, pp.55- 85.

LETTRES, FEMME, ELHAKI, MILLE évoluent suivant une forme géométrique ovale avec les verbes, les pronoms, les prépositions et les conjonctions. Toutefois, ELHAKI est marqué par l'intensité de l'utilisation des verbes, MILLE également mais avec une intensité moindre.

L'axe vertical, oppose les essais FEMME, ZRIREK, GENET, qui attirent les conjonctions, les adverbes et les adjectifs numéraux ; aux romans MILLE, ELHAKI, AILEN qui se rangent à leur tour du côté des verbes, des pronoms, des prépositions, des substantifs, des déterminants, des interjections et des adjectifs.

Cependant, le texte ABOUNOUR prend une position particulière proche du centre de gravité (l'origine du repère orthogonale), tout en étant proche des pronoms, des adjectifs, des déterminants et des prépositions.

Après avoir approché la structure lexicale des œuvres d'El Maleh du point de vue « richesse lexicale » et « accroissement du vocabulaire », nous nous sommes intéressée dans le troisième chapitre aux parties du discours, ou bien, aux catégories grammaticales.

De nombreux spécialistes en lexicométrie ont pu affirmer qu'il y a une corrélation positive entre la distribution des catégories grammaticales et les caractéristiques des œuvres littéraires. Citons, entre autres, E. BRUNET qui stipule que les catégories grammaticales¹⁹⁶ restent un indice révélateur de la distinction des genres, des styles et des écrivains.

L'étude de la distribution des parties du discours permet donc de saisir le choix intuitif et authentique qu'opère l'écrivain lors de l'acte scriptural et ainsi, apporter de nouveaux éléments de diagnostic pour comprendre au mieux les particularités de l'écriture maléhienne.

Rappelons que le logiciel *Hyperbase* répartit les textes selon dix catégories fondamentales, à savoir : verbes, substantifs, adjectifs, déterminants, pronoms, numéraux, interjections, prépositions, adverbes et conjonctions. Cependant pour faciliter l'exploitation des données, nous avons regroupé dans la classe des mots de relation, tous les mots grammaticaux ainsi que les interjections. Donc, les catégories grammaticales retenues sont les suivantes : le substantif, le verbe, l'adjectif, l'adverbe et les mots de relation.

Après avoir vérifié la répartition de ces catégories grammaticales à travers les neuf romans qui forment notre corpus, nous avons trouvé que les mots de relation sont les plus fréquents avec un taux de 45%, suivi des substantifs avec un taux de

¹⁹⁶ Etienne BRUNET, *op.cit.*, 1983, p. 836

25%. Les verbes se mettent à la troisième position avec un pourcentage de 16%. Les adjectifs et les adverbes se rangent, respectivement, à la quatrième et cinquième position, avec un pourcentage de 8% et de 6%. Donc, notre écrivain a tendance à favoriser les mots de relation sur les autres catégories. Cependant, ces résultats ne fournissent que des informations internes à l'œuvre, mais, comme dit A. Hefied : « *En statistique linguistique le mot-clef a toujours été celui de la comparaison* »¹⁹⁷. De ce fait, en comparant El Maleh avec d'autres écrivains, nous avons constaté qu'il rejoint PROUST en préférant l'utilisation des mots de relation, et d'ailleurs, c'est l'une des caractéristiques de la langue française. De même, il s'approche d'HUGO pour ce qui est de l'utilisation des substantifs et des verbes. Néanmoins la distribution des catégories grammaticales à laquelle nous avons abouti n'est pas la même dans tous les textes. Cette disparité pourrait trouver son explication dans le style, le genre ou le thème abordé.

Ensuite, nous avons appliqué un ensemble de tests à notre corpus afin de nous renseigner sur le style de l'écrivain en fonction des écarts jugés significatifs. Nous avons alors calculé les écarts entre les effectifs de chaque catégorie en utilisant les valeurs des écarts réduits. Cette démarche a mis en évidence la formation de deux axes, celui des verbes qui attire les adverbes et celui des substantifs qui aime les adjectifs.

De même, nous avons eu recours à deux autres tests statistiques pour dégager les lois qui régissent la distribution des catégories grammaticales dans les œuvres d'El Maleh. Il s'agit du test de Pearson et de l'analyse factorielle des correspondances. Le premier sert à exprimer l'écart existant entre le réel et le théorique en probabilité pour détecter les divergences et les convergences qui

¹⁹⁷ Ali HEFIED, *op. cit.*, p. 234

existent entre les textes. Tandis que le deuxième vise à faire ressortir les affinités qui s'établissent entre les catégories grammaticales dans les différentes parties de notre corpus. Ainsi, le test de Pearson a montré que la distribution des catégories grammaticales, dans tous les segments, est fortement influencée par le style de l'écrivain. Donc elle n'est pas du tout un fait du hasard. Ceci dit, l'écriture maléhiennne est effectivement bien forgée et bien pensée.

Dans l'analyse factorielle des correspondances, nous avons constaté qu'il n'y a pas de frontières claires entre les genres et la préférence d'une catégorie grammaticale bien précise, par exemple nous pouvons trouver des essais qui privilégient les adverbes, les adjectifs numéraux, les déterminants et les adjectifs, c'est le cas d'ailleurs de ZRIREK et GENET ; comme nous pouvons détecter des romans qui favorisent l'emploi des mêmes catégories (PARCOURS).

Cependant et pour la deuxième fois nous avons relevé une opposition entre deux pôles, à savoir la catégorie verbale et nominale. Soulignons avec Véronique Magri-Mourgue¹⁹⁸ que cette opposition substantif /verbe peut s'expliquer par la binarité espace/temps. L'univers espace est rendu en langue française par le substantif, tandis que l'univers temps est restitué par le verbe. Ce constat rejoint ce que nous avons avancé au niveau de la première partie. En effet, l'objectif même de l'écriture maléhiennne est de restituer la mémoire séculaire de sa terre natale (espace) à une époque révolue (temps). Ajoutons que l'utilisation des verbes est aussi motivée par le mouvement et le déplacement. En effet, voyage et errance structurent l'ensemble des récits d'El Maleh par le biais de ses protagonistes qui nous offrent une sorte de voyage à la fois physique et mental. Ils parcourent

¹⁹⁸ Magri-Mourgue VERONIQUE, *op.cit.*, 2009, pp. 55- 85.

l'espace et *le temps*, nous communiquant ainsi une représentation des événements vus à travers un homme éclairé.

Soulignons que cette bipolarité ne témoigne pas d'un signe d'originalité pour l'écriture maléhiennne du moment qu'elle a été observée chez d'autres écrivains, comme le signale E. Brunet. Ce qui inscrit la production d'Edmond Amran El Maleh dans le cadre général de la littérature française.

Au terme de cette deuxième partie, une synthèse des différents aspects de la structure du lexique, soulevés dans les trois chapitres, s'impose. Dès le premier chapitre nous avons souligné que le choix qu'opère le scripteur du point de vue longueur des mots n'est nullement automatique, car genre, époque ou thème et bien d'autres facteurs pourraient influencer l'emploi d'un tel ou tel mot. Ainsi, dans la première section, celle de la longueur des mots, nous avons étudié les différentes classes de longueur, tout en faisant appel à différentes méthodes de calcul, en commençant par la moyenne, en passant par l'écart type et en terminant par des tests basés sur les effectifs théoriques et les effectifs réels des mots. Nous avons constaté finalement que la distribution des lettres dans presque tous les textes du corpus n'est pas le résultat d'un simple hasard, mais au contraire, elle est le produit des effets du genre, du thème et/ou autres. De la même manière, nous avons soulevé la prépondérance des mots courts par rapport aux autres classes. Néanmoins, c'est un constat qui reste normal vu que les mots les plus courts (lg2) sont généralement les plus fréquents dans la langue.

Après plusieurs étapes de calcul, nous avons affiné notre étude par le recours à l'analyse factorielle des correspondances qui nous a donné une vue d'ensemble sur le regroupement des classes de longueur. Nous avons établi que la période de la production n'a pas d'effet sur la distribution des classes de longueur, vu que les textes regroupés en constellations sont produits pendant des périodes distancées. De même, nous avons constaté des régularités entre les romans et l'utilisation des mots courts et moyens qui sont en faveur d'un registre familier, et entre les essais et la tendance de l'auteur à utiliser des mots longs, ce qui renvoient souvent au style soutenu. Par ailleurs, nous avons soulevé quelques exceptions. Par exemple ELHAKI qui est un roman marque une symétrie entre deux groupes de longueur : les mots longs (lg9, lg10, lg11) et les mots courts et moyens (lg3, lg5,

lg6 et lg7). Ceci pourrait se justifier par l'utilisation de l'auteur d'un discours soutenu abordant les écritures mystiques de certaines figures intellectuelles prestigieuses qui cheminent le lecteur dans des espaces et des périodes pluriels.

Dans la seconde section, nous nous sommes penchée sur l'analyse de l'étendue des occurrences N. Nous avons alors constaté que l'étendue des quatre premiers textes évolue selon une tendance haussière, ce qui pourrait être expliqué par l'acharnement d'El Maleh pour l'écriture. D'ailleurs, notre écrivain était très sensible à sa limitation par le « temps » et il se voyait comme une matière vouée à la déliquescence. De ce fait, il voulait étaler dans ses textes, avant que la fin ne le surprenne, tout ce qu'il a envie d'exprimer. Cependant, nous assistons par la suite à une décadence de l'étendue avant d'effectuer un second pic dans ZRIREK. Ce pic peut être justifié par des facteurs d'ordre thématiques. En effet, quoique la majorité des textes d'El Maleh abordent plus ou moins les mêmes thèmes, mais dans *Le Café bleu*, *Zrerek* nous assistons à une concentration d'idées sur la littérature et l'écriture allégorique. À notre humble avis c'est ce qui pourrait justifier la hausse de l'étendu au niveau de ce texte. Finalement nous repérons une décroissance relative aux textes FEMME et LETTRES, ceux dont l'étendue est la plus faible de toutes les partitions composant notre corpus.

En étudiant l'étendue V, nous avons déduit que le vocabulaire d'El Maleh est déficitaire dans l'ensemble des textes. Néanmoins, nous avons mis en évidence les segments les moins déficitaires, à savoir : ELHAKI, ANOUNOUR, FEMME, PARCOURS. Cela dit que l'auteur aurait usé de presque tout son lexique dans ces textes.

En appliquant le coefficient de corrélation des rangs à notre corpus, pour comparer l'ordre chronologique de parution des textes avec l'étendue des

occurrences N et celle des vocables V, nous nous sommes rendu compte que le lexique personnel d'El Maleh ne dépend pas de la date de parution des textes. Ce qui pourrait laisser entendre que l'auteur diversifie ses choix thématiques. En effet, si nous nous référons au contenu des textes, nous pourrions facilement constater que l'auteur alterne parfois entre certains thèmes. Par exemple dans ses premières productions, il est question de : la mémoire, l'histoire, la culture, l'exode des marocains juifs, le drame palestinien, etc. Cependant, dans GENET, à travers les thèmes de la littérature et l'écriture allégorique, El Maleh analyse son propre style d'écriture tout en le comparant avec d'autres écrivains. Par ailleurs, dans ABOUNOUR l'auteur redécouvre encore une fois les valeurs culturelles marocaines ancestrales. En arrivant à ZRIREK nous retrouvons la tragédie palestinienne, mais également l'écriture allégorique qui domine presque tout le contenu de ce texte. D'ailleurs, ce dernier a rassemblé les écrivains dont les œuvres se caractérisent par la force et la puissance d'une écriture allégorique. Nous y trouvons également des réflexions sur la langue et la littérature. Ainsi, nous pouvons conclure que les thèmes abordés par El Maleh sont incessamment revisités.

Le second chapitre s'est attardé sur la richesse lexicale et l'accroissement du vocabulaire. Abordons le premier volet, nous avons testé plusieurs formules et techniques, finalement nous avons pu affirmer que l'indice W de BRUNET reste la formule la plus fiable, du moment qu'elle neutralise ou du moins réduit l'influence parasitaire de l'étendue. Une application de cet indice à notre corpus a permis de classer les textes de celui qui a le vocabulaire le plus riche à celui qui a le vocabulaire le plus faible :

**ELHAKI > AÎLEN > PARCOURS > MILLE > ZRIREK >
ABOUNOUR > GENET > FEMME > LETTRES**

Pour trouver une explication à ce classement nous nous sommes référée à plusieurs études, en l'occurrence celles de : Véronique Magri-Mourgues et José Gregorio PARADA-RAMIREZ. Cependant nous avons constaté que nos résultats ne convergent pas avec ce qui a été avancé par ces chercheurs

Après avoir exploré le corpus du point de vue richesse lexicale, nous nous sommes intéressée dans un deuxième temps à l'accroissement du vocabulaire. Cette étape a été réalisée suivant deux techniques contradictoires mais complémentaires. La première vise l'examen des textes selon leur enchaînement chronologique de parution, tandis que l'autre inverse le processus en explorant le corpus à partir du dernier texte écrit.

Ainsi, dans la première étape nous avons procédé à la valorisation de l'apport lexical de chaque partition. En nous référant à Brunet nous avons signalé qu'en général les valeurs diminuent au fur et à mesure que nous progressons vers le dernier texte vu que les chances de trouver des mots inédits deviennent faibles. Cependant, dans notre corpus nous avons décelé, tantôt un accroissement normal qui se caractérise par une diminution du vocabulaire, tantôt un accroissement anormal qui se reflète par une augmentation de l'apport lexical, C'est le cas d'ailleurs des textes ELHAKI, ZRIREK et LETTRES. Le renouvellement lexical de ces trois textes pourrait être dû à leurs étendues qui restent supérieures aux autres (pour ELHAKI et ZRIREK) , à une rupture thématique dans le corpus, ce qui engendre l'afflux de nouveaux vocables, ou bien à l'introduction « de mots rares ou de néologismes »¹⁹⁹. Dans le cas d'ELHAKI, l'apport lexical peut être expliqué par les thèmes traités. D'ailleurs, dans ce texte nous assistons à une mise en valeur de la tradition musulmane de savoir, de culture et de refus de tyrannie,

¹⁹⁹ Etienne BRUNET, *op.cit.*, 1981, p. 53.

un thème qui reste plus ou moins absent des trois premiers romans. Les idées d'El Maleh prennent forme à travers le récit épico-lyrique de vie d'individualités marquantes, comme : Amine El Andaloussi, Sofiane Abou El Haki, Ahmed El Ghazouli et dont l'évocation s'accompagne du plaisir de la méditation sur les vertus et les joies de l'écriture.

Le même constat reste valable pour ZRIREK, un essai caractérisé par une prépondérance des mots longs et qui étale des réflexions sur les enjeux de l'écriture en faisant appel à des intellectuels éminents, comme : le romancier Juan Goytisolo, le poète José-Angel Valente, Jean Genet, Khaïr-Eddine et Nissaboury. De même, à travers une série d'évocations relatives au soufisme, l'auteur choisit de présenter des figures appartenant aux trois religions théologiques, tels : Abraham Aboulafia, Ibn Arabi et Juan de la Cruz. Donc c'est le contenu sémantique et thématique de ces deux textes qui a ordonné l'apparition de mots nouveaux.

Pour le troisième texte LETTRES, le renouvellement lexical est dû à la fois au thème et au genre, car ce texte est le premier à aborder l'exil d'El Maleh, mais aussi il est le premier et le dernier récit épistolaire de toute la production littéraire de notre auteur.

Signalons que nous avons établi un autre classement basé sur le calcul de l'écart réduit. L'illustration graphique de ce dernier a montré que les premiers romans se caractérisent par un vocabulaire riche avec des apports lexicaux importants, ce qui reflète surtout l'intention manifeste de l'écrivain de vouloir tout exprimer sur l'exode des marocains juifs et de jaillir la spontanéité créatrice de son imagination. Cependant nous avons repéré une chute au niveau des quatre derniers textes ce qui correspond bien à la rupture dans l'écriture de notre auteur et à l'abandon du genre romanesque mais aussi de certains thèmes. Donc, il semble

que les récits où nous suivons le destin des juifs, qui sont partis à la recherche de la terre promise, apportent plus de vocabulaire que les parties qui sont consacrées à d'autres thèmes.

Ensuite, pour mieux apprécier le phénomène de l'accroissement du vocabulaire, nous avons mesuré l'apport lexical à partir du dernier texte jusqu'au premier. Le résultat était frappant parce que les textes FEMME et LETTRES qui ont l'étendue la plus faible et qui sont aussi les derniers à être publiés étaient classés dans la première et la deuxième position. Cela signifie qu'ils actualisent un nombre de vocables qui n'existent pas dans les textes précédents. Pour trouver une explication à cet état de fait, nous nous sommes inspirée de Guy Dugas²⁰⁰ pour qui les auteurs judéo-maghrébins de notre époque n'arrivent qu'à reproduire une œuvre répétitive, dramatiquement hantée par quelques faits de mémoire.

Cela étant dit, nous trouvons qu'El Maleh a voulu opérer un changement radical dans son œuvre que ce soit au niveau de la forme qu'au niveau du contenu. D'ailleurs, l'avant dernier texte LETTRES est le seul récit épistolaire qui existe dans tout le corpus, de même, c'est le seul qui parle de l'exil d'El Maleh, de son arrivée à Paris. Tandis que le dernier texte FEMME est le seul qui aborde sa découverte de la ville d'Asilah et sa rencontre avec la mère de son fils spirituel Khalil El Ghrib. De ce fait, le classement auquel nous avons abouti confirme parfaitement bien cette diversification dans l'écriture maléhienne.

Nous avons achevé ce chapitre par l'application d'une autre méthode, à savoir : l'accroissement inverse. Ce dernier a montré que la tendance régulière à la baisse est inversée dans deux romans : ZRIREK (7313), ELHAKI (5450). Donc

²⁰⁰ Guy DUGAS, *op. cit.*, 1990, p. 135

ces deux textes signalent la fin d'un mouvement et l'épuisement de l'inspiration de l'auteur. Le mouvement dont il s'agit ici est surtout d'ordre générique et thématique : le texte ELHAKI marque une rupture avec le genre romanesque. D'ailleurs aucun des cinq textes publiés après n'admettent l'étiquette de roman, du moment que l'auteur a délaissé ce genre en faveur des essais et des nouvelles. ZRIREK est aussi le dernier essai qui aborde l'écriture comme thème principale.

Toutefois, le classement établi sur la base de l'écart réduit a dévoilé un éclatement lexical au niveau du texte ABOUNOUR, chose que l'ordre chronologique n'a pas pu relever. Ceci pourrait être dû aux effets du genre, d'ailleurs ce texte est le seul recueil de nouvelles qui existe dans tout le corpus d'El Maleh.

Arrivant au troisième chapitre, nous nous sommes intéressée aux parties du discours. L'étude que nous avons réalisée a montré la tendance d'El Maleh à favoriser les mots de relation sur les autres catégories. Néanmoins la distribution à laquelle nous avons abouti n'est pas la même dans tous les textes. Cette disparité pourrait trouver son explication dans le style, le genre ou le thème abordé.

Ensuite, nous avons appliqué un ensemble de tests à notre corpus afin de nous renseigner sur le style de l'écrivain. Entre autres nous citons le test de Pearson et l'analyse factorielle des correspondances. Ainsi, le premier a montré que la distribution des catégories grammaticales, dans tous les segments, est fortement influencée par le style de l'écrivain. Donc elle n'est pas du tout un fait du hasard. Ceci dit, l'écriture maléhienne est effectivement bien forgée et bien pensée.

Dans l'analyse factorielle des correspondances, nous avons constaté qu'il n'y a pas de frontières claires entre les genres et la préférence d'une catégorie grammaticale bien précise, par exemple nous pouvons trouver des essais qui

privilégient les adverbes, les adjectifs numéraux, les déterminants et les adjectifs, c'est le cas d'ailleurs de ZRIREK et GENET ; comme nous pouvons détecter des romans qui favorisent l'emploi des mêmes catégories (PARCOURS).

Cependant nous avons relevé une opposition entre deux pôles, à savoir la catégorie verbale et nominale. Cette opposition²⁰¹ substantif /verbe peut s'expliquer par la binarité espace/temps. L'univers espace est rendu en langue française par le substantif, tandis que l'univers temps est restitué par le verbe. Ce constat rejoint ce que nous avons avancé au niveau de la première partie. En effet, l'objectif même de l'écriture maléhiennne est de restituer la mémoire séculaire de sa terre natale (espace) à une époque révolue (temps). Ajoutons que l'utilisation des verbes est aussi motivée par le mouvement et le déplacement. En effet, voyage et errance structurent l'ensemble des récits d'El Maleh par le biais de ses protagonistes qui nous offrent une sorte de voyage à la fois physique et mentale. Ils parcourent l'espace et le temps, nous communiquant ainsi une représentation des événements vus à travers un homme éclairé.

²⁰¹ Véronique Magri MOURGUE, *op.cit.*, 2009, pp.55- 85.

TROISIEME PARTIE

Contenu lexical

Si la seconde partie a été consacrée à la structure du lexique, troisième se fixe comme objectif de compléter notre étude précédente en abordant le contenu lexical. Pour ce faire, nous présenterons trois chapitres où nous étudierons d'abord le phénomène de la connexion lexicale ; ensuite nous mettrons en lumière les spécificités du vocabulaire étudié, pour finir avec les manifestations du plurilinguisme dans l'œuvre d'Edmond Amran El Maleh.

Le premier chapitre se penchera sur la mesure de la distance qui sépare les textes les uns des autres pour se prononcer sur leur proximité ou leur éloignement. Plusieurs méthodes seront mises en œuvre : d'abord nous utiliserons la méthode Jaccard bien qu'elle soit très sensible aux différences des étendues des textes comparés, puis pour remédier aux insuffisances de la première méthode, nous aurons recours à la loi binomiale. Nous nous intéresserons aussi aux représentations graphiques, à l'analyse factorielle des correspondances et à l'analyse arborée pour rendre compte des rapprochements et des éloignements existants entre les textes composant notre corpus.

Le deuxième chapitre abordera les spécificités lexicales d'Edmond Amrane El Maleh en s'attardant sur l'étude du vocabulaire excédentaire et celui déficitaire. En d'autres termes, nous mettrons en lumière les mots-clés qui sont les plus employés et qui se regroupent sous les mêmes champs sémantiques pour relever les thèmes privilégiés de l'auteur.

Pour ce qui est du troisième chapitre, il sera réservé à l'étude du plurilinguisme dans l'œuvre d'El Maleh. Ainsi, notre objectif sera de faire surgir des phénomènes linguistiques divers et variés. Il s'agit, en effet, de l'emprunt, du calque, de l'alternance codique et de l'interférence. De ce fait, la mission que nous nous sommes fixée est de décrire et comprendre le mode de fonctionnement des phénomènes linguistiques sus-

cités, pour déceler leur contribution dans la mise en évidence de la mémoire, l'histoire et la culture marocaine.

CHAPITRE PREMIER

Connexion lexicale

Elaborée et définie par Charles Muller en 1967, la connexion lexicale étudie les liens existants entre deux textes et la distance lexicale qui les sépare pour comparer leur vocabulaire²⁰². Elle permet de savoir dans quelle mesure les textes étudiés présentent des ressemblances au niveau de leur structure lexicale. Ainsi, elle s'est très utile pour étudier les caractéristiques d'une œuvre, le style d'un écrivain ou le vocabulaire d'une époque donnée, etc.

Notons que la mesure de la connexion lexicale a même constitué un champ fertile pour effectuer des travaux lexicométriques sur la paternité d'une œuvre. D'ailleurs, le principe²⁰³ de base pour l'attribution d'un texte à un auteur consiste à le comparer à d'autres textes parfaitement authentifiés. En effet, si deux textes ont été rédigés par un même auteur, nous pouvons supposer qu'ils partageront un même vocabulaire, des catégories grammaticales et des structures syntaxiques communes, à l'insu même de celui qui les a produits et malgré leurs éventuelles différences thématiques. Par contre, deux autres textes écrits par des auteurs différents

²⁰² Signalons que pierre Guiraud était le premier a annoncé dans son ouvrage *Problèmes et méthodes de la Statistique Linguistique* qu'on pourrait établir un tableau de corrélations lexicales entre plusieurs textes en les prenant deux à deux, pour identifier les mots qu'elles ont en commun et ceux qui sont propre à chacun ; mais vu la lourdeur de la démarche, il a fini par l'abandonné du moment qu'il ne disposait pas à l'époque de moyens de calcul. P.GUIRAUD, 1959, *op.cit.*, p. 129

²⁰³ Sylvie MELLET, *Lemmatisation et encodage grammatical permettent-ils de reconnaître l'auteur d'un texte*, Nice, CNAS, Bases, corpus et langue, UMR 6039, 2002, pp.1-5

présenteront des structures linguistiques spécifiques à chacun d’eux, ce qui se traduira par leur éloignement.

Le calcul de la connexion lexicale entre deux textes a le plus souvent reposé sur une comparaison du vocabulaire actualisé dans les textes étudiés. Les variables prises en considération sont la part du vocabulaire qui est commune aux deux textes et les parts qui sont spécifiques à chacun d’eux. Le calcul, simple dans son principe, est renouvelé autant de fois qu’il y a de paires de textes à comparer et peut donc devenir assez lourd, de ce fait, le recours à l’outil informatique devient indispensable.

Notons que la mesure de la connexion lexicale peut s’appliquer soit à des ensembles de formes, soit à des ensemble de lemmes. Dans le premier cas, le vocabulaire d’un texte sera évalué en fonction du nombre de formes qui le composent ; dans le second, il sera évalué en fonction du nombre de vocables employés.

Enfin, pour une meilleure interprétation, les résultats sont représentés graphiquement grâce à l’analyse factorielle des correspondances, cette dernière permet de dégager les principaux facteurs explicatifs des proximités ou des éloignements. L’analyse arborée²⁰⁴ qui met en valeur la hiérarchie classificatoire des textes est aussi utilisée notamment pour présenter la distribution des hautes et des basses fréquences comme possible critère de différenciation générique.

L’enjeu de ce chapitre sera d’étudier la distance des différents textes du corpus et bien sûr déduire l’existence ou non d’une connexion entre les neuf œuvres qui font l’objet de notre étude. Pour ce, nous testerons différentes méthodes utilisées par les

²⁰⁴ Barthélemy J.P. & Luong X. « *Sur la topologie d'un arbre phylogénétique : aspects théoriques, algorithmiques et applications à l'analyse de données textuelles* », *Mathématiques et Sciences Humaines* n° 100, 1987, pp. 57-80.

analyses statistiques des textes littéraires. Seront ainsi expérimentées les fonctions implémentées dans le logiciel *Hyperbase* : la mesure de la " distance intertextuelle " qui repose sur la méthode Jaccard, mais aussi l'évaluation de la " connexion lexicale " que l'on doit à Ch. Muller.

A. Méthode Jaccard

La méthode Jaccard se base sur deux notions principales, à savoir : le vocabulaire commun et le vocabulaire exclusif. Autrement dit, elle se borne à établir le rapport entre le vocabulaire commun à deux textes et celui qui n'existe que dans l'un des deux pour se prononcer sur leur connexion lexicale. Notons que ladite méthode ne s'intéresse pas aux fréquences des mots, elle ne considère que leur présence ou leur absence dans les textes considérés.

Pour appliquer la méthode Jaccard à notre corpus, nous devons prendre en compte tous les appariements de textes deux à deux. Nous avons neuf textes, ce qui porte alors le nombre de combinaison à :

$$C_9^2 = \frac{A_9^2}{2!} = \frac{9 \times 8}{2 \times 1} = 36$$

Dans les points suivants et tout en respectant le principe de la méthode Jaccard, nous mesurerons la connexion lexicale pour les 36 binômes sous différents angles et selon plusieurs formules de calcul.

1. Vocabulaire commun à deux textes

La notion de vocabulaire commun est l'une des variables qui permettent d'apprécier le rapprochement ou l'éloignement de deux textes donnés. En effet, deux textes sont considérés proches s'ils partagent un vocabulaire commun très important, en revanche, ils sont éloignés s'ils présentent un vocabulaire commun très faible.

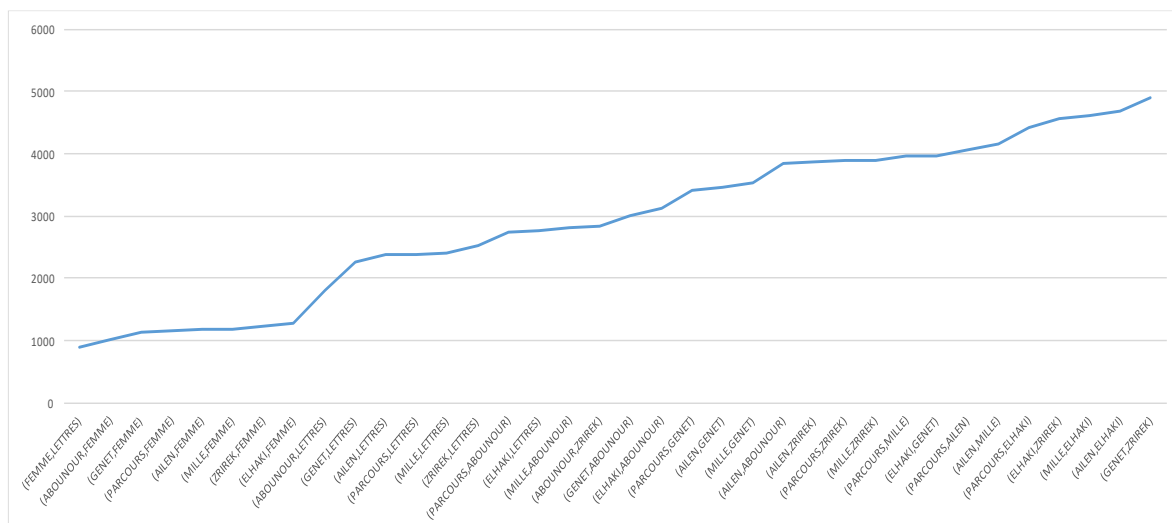
En faisant appel au logiciel *Hyperbase*, ce dernier nous offre le tableau ci-dessous qui met en exergue les valeurs communes à chaque couple de texte.

Tableau III. 1 : Effectifs des formes communes dans le corpus

	PARC	AÎLEN	MILLE	ELHAKI	GENET	ABOUN	ZRIREK	FEMME	LETTRE
PARC	0	4051	3958	4417	3403	2739	3883	1167	2379
AÎLEN	0	0	4161	4682	3464	3842	3862	1187	2370
MILLE	0	0	0	4616	3542	2811	3893	1189	2400
ELHAKI	0	0	0	0	3973	3120	4568	1275	2761
GENET	0	0	0	0	0	3008	4902	1125	2263
ABOU	0	0	0	0	0	0	2830	1023	1805
ZRIREK	0	0	0	0	0	0	0	1235	2535
FEMME	0	0	0	0	0	0	0	0	905
LETTRE	0	0	0	0	0	0	0	0	0

Le tableau affiche le relevé des mots qui sont actualisés dans les deux textes de chaque binôme. Le vocabulaire commun d'un texte et lui-même est représenté par la valeur nulle. Cependant, pour une meilleure interprétation des données, nous procéderons au classement des résultats de manière hiérarchique, ce qui donne la courbe suivante :

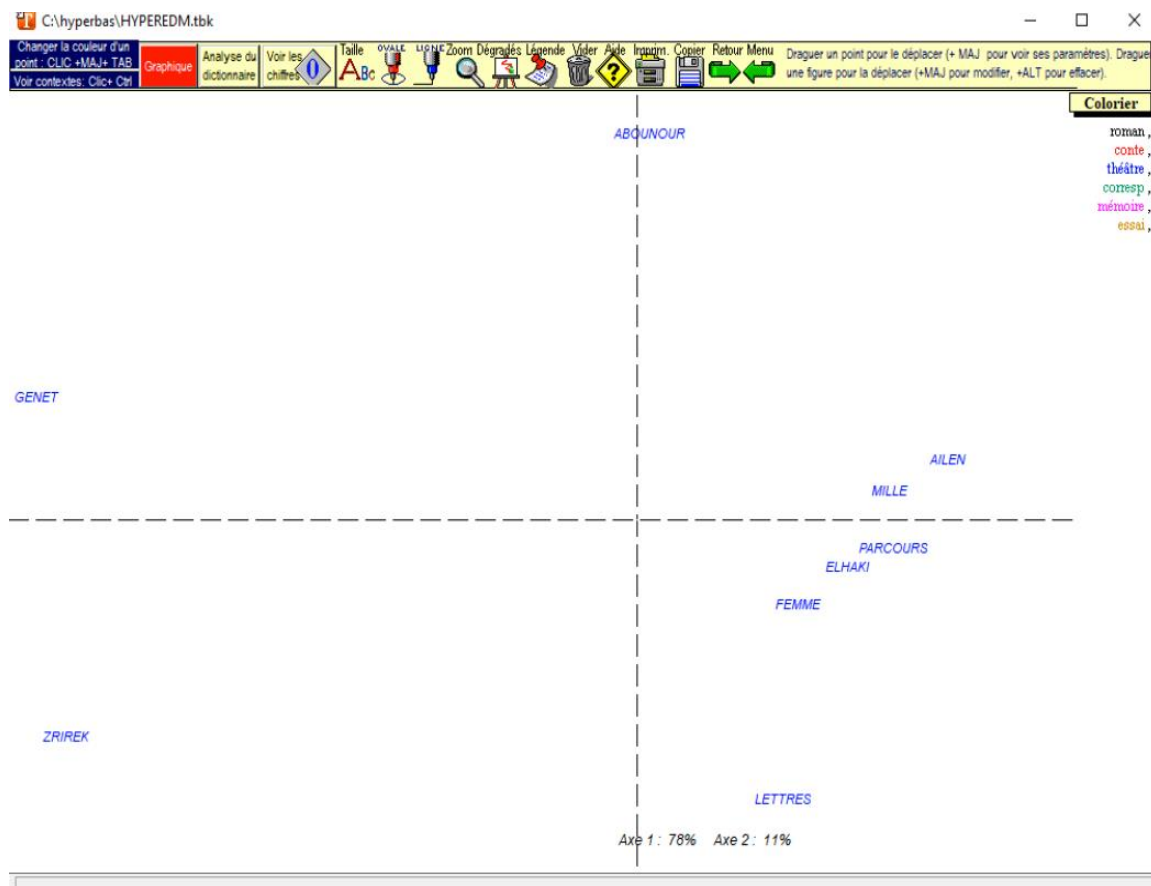
Figure III. 1 : Effectifs des formes communes dans le corpus



Nous avons procédé à un tri par ordre croissant du vocabulaire commun des textes pris deux à deux, puis nous l'avons traduit en un graphique mettant en évidence un classement hiérarchique. La valeur planchée est de 905 pour le couple (FEMME, LETTRES) et celle maximale est de 4902 pour le couple (GENET, ZRIREK). De ce fait, les deux textes les plus proches sont GENET et ZRIREK alors que les plus éloignés sont FEMME et LETTRES. La moyenne arithmétique des valeurs est de l'ordre de 2926,22. Nous remarquons alors que 50% des couples ont une valeur inférieure à cette moyenne, cependant les 50% restants ont une valeur qui dépasse la moyenne.

Pour une matrice de 36 binômes le tableau reste difficile à interpréter. C'est la raison pour laquelle nous fut recours à l'analyse factorielle des correspondances. Cette dernière nous permettra de produire une cartographie mettant en œuvre les différentes distances qui séparent les textes, chose qui rendra l'interprétation plus aisée :

Figure III. 2 : Analyse factorielle de la distance lexicale (lemmes)



Nous constatons que les deux axes factoriels récupèrent 89% de l'inertie totale dont 78% de l'information récupérée revenant au premier axe. A première vue, nous remarquons qu'il y a plusieurs couples qui se rapprochent et d'autres qui s'éloignent. Sur le plan vertical, nous avons à droite deux couples de textes qui paraissent être proches, à savoir (AILEN, MILLE) et (ELHAKI, PARCOURS). À gauche aucun texte ne semble être proche l'un de l'autre. Si nous prenons en considération le deuxième axe factoriel, horizontal, les textes qui s'attirent que ce soit en haut de l'axe ou en bas, sont également (AILEN, MILLE) et (ELHAKI, PARCOURS). Ce

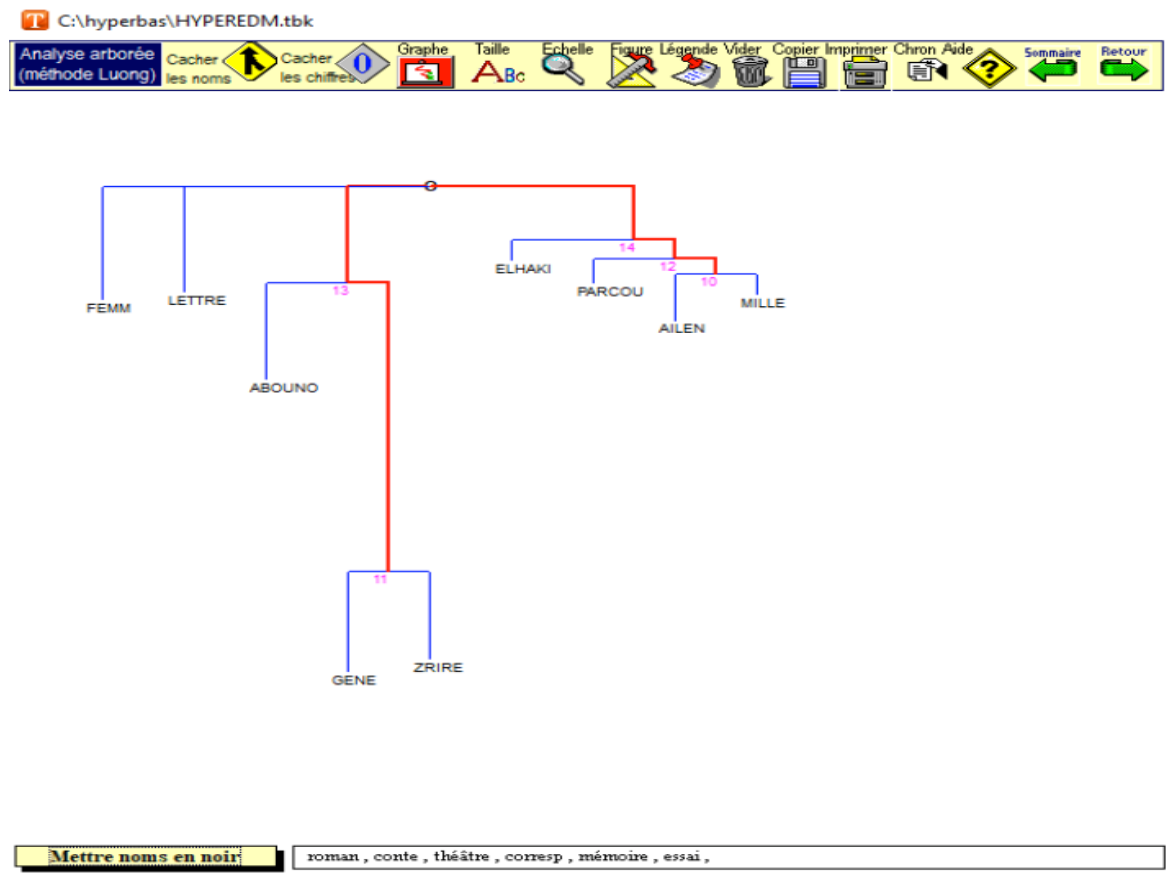
rapprochement, confirmé sur tous les plans peut être dû à la thématique traitée dans chacun des textes formant les couples susmentionnés. Dans (AILEN, MILLE) les thèmes de l'exode juive, de la guerre et du colonialisme s'avèrent être très présent. Le couple (ELHAKI, PARCOURS) traite également les thèmes de l'exode et le militantisme marocain.

Aussi pouvons-nous facilement constater que les deux couples forment une constellation à droite de l'axe vertical. Ceci n'est pas étonnant, du moment que la thématique de l'exode des juifs marocains leur est commune. Aussi si nous prenons en considération la chronologie nous remarquerons que les quatre textes se succèdent. De même, le texte ELHAKI qui est le dernier à être rédigé, constitue une récapitulation des sujets traités dans les trois autres romans.

Outre l'analyse factorielle des correspondances, le logiciel Hyperbase permet également de représenter la distance lexicale selon la technique de classification de Xuan Luong 118, appelée analyse arborée. Cette dernière offre une possibilité d'interprétation plus confortable.

Ainsi, nous présentons dans la figure ci-dessous l'analyse arborée de la distance lexicale des neuf textes calculés sur V :

Figure III. 3 : Analyse arborée de la distance lexicale (lemmes)



Le graphique ci-dessus met en lumière une représentation rectangulaire et hiérarchique, la lecture que nous effectuerons se fera sur le plan vertical et de gauche à droite.

La première proximité que l'on peut noter est celle des textes (GENET, ZRIREK) avec ABOUNOUR. Juste à côté nous avons un deuxième nœud qui regroupe plusieurs textes. Encore une fois les couples (AILEN, MILLE) et (ELHAKI, PARCOURS) se montrent très proches, ceci peut être dû aux raisons que nous avons déjà citées au niveau de l'analyse factorielle. Une autre remarque émerge, c'est la filiation des trois romans AILEN, MILLE et PARCOURS à ELHAKI. En effet, ce

texte qui chapote les trois premiers regroupe les thèmes majeurs évoqués dans ces textes.

Toutefois cette analyse demeure incomplète du moment qu'elle n'est pas fondée sur d'autres examens neutralisant quelques facteurs perturbateurs, tels que la différence de taille dans un même couple ainsi que les inégalités des étendues d'un couple à l'autre. « *L'idéal serait de pouvoir exprimer, pour un couple de textes, le degré de connexion lexicale par un indice numérique, ce qui permettrait des classements et des appréciations utiles.* »²⁰⁵

Donc, pour compléter l'analyse et exprimer la connexion lexicale de manière plus détaillée, nous nous intéresserons à la distance intertextuelle.

2. Distance intertextuelle

Pour mesurer la distance intertextuelle, les textes sont comparés deux à deux. Généralement la distance absolue entre deux textes A et B est obtenue par l'addition du vocabulaire des deux textes moins leur intersection, soit : $(A \cup B) - (A \cap B)$

Notons que la mesure de la distance intertextuelle pourra nous renseigner sur la proximité²⁰⁶ des textes à travers la comparaison de la distribution de tous les mots de chaque couple. Ainsi, si la majorité du vocabulaire actualisé dans le couple est commun aux deux textes, la distance sera faible. Par contre, si les deux textes comparés ne partagent pas beaucoup de vocables et que chacun est caractérisé par un

²⁰⁵ Charles MULLER, *op.cit.*, 1967, p.169.

²⁰⁶ Véronique Magri MOURGUE, « *Distance intertextuelle et connexion lexicale : outils de catégorisation générique ou stylistique ?* », In : International Conference on Statistical Analysis of Textual Data, n°10, Paris, 2010, pp. 334-340.

vocabulaire spécifique qui lui est propre la distance sera élevée. En utilisant le logiciel *Hyperbase* nous obtenons le tableau suivant :

Tableau III. 2 : Confrontation des textes deux à deux selon la distance globale

	PARCOURS	AÎLEN	MILLE	ELHAKI	GENET	ABOUNOUR	ZRIREK	FEMME	LETTRE
PARCOURS	437	448	448	438	489	490	478	538	510
AÎLEN	448	428	441	429	497	485	501	536	524
MILLE	448	441	420	421	474	482	483	531	510
ELHAKI	438	429	421	420	459	463	450	513	470
GENET	489	497	474	459	296	413	297	545	513
ABOUNOUR	490	485	482	463	413	412	492	549	547
ZRIREK	478	501	483	450	297	492	296	520	495
FEMME	538	536	531	513	545	549	520	512	583
LETTRE	538	524	510	470	513	547	495	583	469

Tout d’abord, nous remarquons que le tableau est symétrique par rapport à la diagonale, ceci est tout à fait normal du moment que la distance calculée pour le couple **ab** est la même que celle du couple **ba**. En principe la diagonale du tableau ne devrait comporter que des zéros du moment que la distance entre un texte et lui-même est nulle. Or les valeurs nulles influencent négativement les calculs multidimensionnels, c’est la raison pour laquelle E. Brunet a choisi de remplacer les zéros par des valeurs arbitraires qui ne sont que la valeur minimale de la ligne ou de la colonne.

Rappelons que les calculs réalisés par le logiciel Hyperbase pour obtenir le tableau susmentionné portent sur les lemmes. La distance calculée, pour chaque

combinaison, prend en considération l'étendue de l'un et de l'autre vocabulaire en vertu de la formule :

$$d = \left(\frac{a - ab}{a} \right) + \left(\frac{b - ab}{b} \right)$$

Avec **ab** est la partie commune aux vocabulaires **a** et **b** ainsi que **(a – ab)** et **(b – ab)** correspondent respectivement à leurs parties privatives. Chaque élément présente la somme des deux quotients $\left(\frac{a-ab}{a} \text{ et } \frac{b-ab}{b} \right)$, c'est-à-dire qu'il expose le rapport du vocabulaire exclusif au vocabulaire globale. Nous signalons que les valeurs obtenues sont comprises entre 0 et 2, mais pour des raisons de simplicité, ces valeurs sont exprimées en centaine, par exemple 0,437 a été remplacé par 437.

A partir des résultats affichés dans le tableau, nous pouvons déduire que le couple (GENET, ZRIREK) est le plus proche du moment qu'il a enregistré la valeur la plus basse, à savoir 297, cela veut dire que les deux textes partagent un nombre important de mots communs. Par contre le couple (ABOUNOUR, FEMME) est le plus éloigné avec une distance de 549. Ce qui signifie que les deux textes sont caractérisés par un vocabulaire privatif spécifique à chacun d'eux.

Pour compléter l'examen de la distance intertextuelle, qui a abouti à dégager des liens de parenté lexicale entre certains textes, nous passons maintenant à l'indice de connexion lexicale qui n'est que son complémentaire.

3. Indice de connexion lexicale

L'exploration de la connexion lexicale de deux textes permet de voir si les textes examinés partagent le même contenu lexical ou non. Selon Etienne Brunet²⁰⁷, la connexion lexicale est supposée forte lorsque les deux textes sont proches l'un de l'autre et partagent un grand nombre de mots, si ce n'est pas le cas nous pouvons alors examiner l'indépendance du premier à l'égard du second et inversement l'indépendance du second par rapport au premier.

La connexion et l'indépendance se marquent par des indices qui ne sont que le rapport entre les vocables communs et les autres.

Soient :

- $V(A)$ le vocabulaire du texte A
- $V(B)$ le vocabulaire du texte B
- $V(A \cap B)$ est le vocabulaire commun aux textes A et B, c'est-à-dire l'intersection des deux textes.
- $V(A \cup B)$ est la somme des mots non communs, c'est-à-dire le nombre de vocables qui font partie de A ou de B et non aux deux.

L'indice de connexion, alors, est le rapport entre le vocabulaire commun et le lexique exclusif.

²⁰⁷ Etienne BRUNET, *op.cit.*, 1978, p. 371

$$\text{Indice de connexion} = \frac{\text{Vocabulaire commun}}{\text{Vocabulaire commun}} = \frac{V(A \cap B)}{V(A \cup B)}$$

- $V(A \Delta B) = V(A \cup B) - V(B)$ représente le vocabulaire exclusif du texte A
- $V(B \Delta A) = V(A \cup B) - V(A)$ représente le vocabulaire exclusif du texte B

L'indice d'indépendance des deux textes est donc mesuré par la proportion des mots exclusifs dans le vocabulaire de chacun :

$$\text{Indice d'indépendance de A} = \frac{\text{Vocabulaire exclusif de A}}{\text{Vocabulaire de A}} = \frac{V(A \Delta B)}{V(A)}$$

et

$$\text{Indice d'indépendance de B} = \frac{\text{Vocabulaire exclusif de B}}{\text{Vocabulaire de A}} = \frac{V(B \Delta A)}{V(B)}$$

Nous présenterons, tout d'abord, l'exemple de calcul pour le premier texte PARCOURS avec les autres textes avant de le généraliser à l'ensemble du corpus.

Tableau III. 3 : Exemple de la connexion lexicale du texte Parcours avec les autres textes du corpus

PARC/A	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de B
AÎLEN	7081	7643	4051	6622	3030	3592	0,61	0,43	0,47
MILLE	7081	7270	3958	6435	3123	3312	0,62	0,44	0,46
ELHAKI	7081	8855	4417	7102	2664	4438	0,62	0,38	0,50
GENET	7081	6288	3403	6563	3678	2885	0,52	0,52	0,46
ABOUNOU	7081	4331	2739	5934	4342	1592	0,46	0,61	0,37
ZRIREK	7081	7840	3883	7155	3198	3957	0,54	0,45	0,50
FEMME	7081	1539	1167	6286	5914	372	0,19	0,84	0,24
LETTRE	7081	3697	2379	6020	4702	1318	0,40	0,66	0,36

Si nous extrayons les données concernant les deux textes Parcours et Mille, nous aurons :

Référence	V
PARCOURS (Texte A)	7081
MILLE (Texte B)	7643
Vocabulaire commun $V(A \cap B)$	4051
Vocabulaire de l'ensemble A et B $V(A \cup B)$	6662
Vocabulaire exclusif de A $V(A \Delta B)$	3030
Vocabulaire exclusif de B $V(B \Delta A)$	3592

De ce fait, l'indice de connexion est le suivant :

$$\text{Indice de connexion} = \frac{\text{Vocabulaire commun}}{\text{Vocabulaire de l'ensemble}} = \frac{V(A \cap B)}{V(A \cup B)} = \frac{4051}{6662} = 0,61$$

Si nous appliquons les formules de calcul de l'indépendance des deux textes nous aurons :

$$\text{Indice d'indépendance de A} = \frac{\text{Vocabulaire exclusif de A}}{\text{Vocabulaire de A}} = \frac{V(A \Delta B)}{V(A)} = \frac{3030}{7081} = 0,43$$

$$\text{Indice d'indépendance de B} = \frac{\text{Vocabulaire exclusif de B}}{\text{Vocabulaire de A}} = \frac{V(B \Delta A)}{V(B)} = \frac{3592}{7643} = 0,47$$

Une fois la démarche de calcul expliquée, nous pouvons à ce moment la généraliser à l'ensemble des neufs textes et présenter le tableau de l'indice de connexion :

Tableau III. 4 : Connexion lexicale des binômes

PARC	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de B
AÎLEN	7081	7643	4051	6622	3030	3592	0,61	0,43	0,47
MILLE	7081	7270	3958	6435	3123	3312	0,62	0,44	0,46
ELHAKI	7081	8855	4417	7102	2664	4438	0,62	0,38	0,50
GENET	7081	6288	3403	6563	3678	2885	0,52	0,52	0,46
ABOU	7081	4331	2739	5934	4342	1592	0,46	0,61	0,37
ZRIREK	7081	7840	3883	7155	3198	3957	0,54	0,45	0,50

FEMME	7081	1539	1167	6286	5914	372	0,19	0,84	0,24
LETTRE	7081	3697	2379	6020	4702	1318	0,40	0,66	0,36
AÎLEN	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de B
MILLE	7643	7270	4161	6591	3482	3109	0,63	0,46	0,43
ELHAKI	7643	8855	4682	7134	2961	4173	0,66	0,39	0,47
GENET	7643	6288	3464	7003	4179	2824	0,49	0,55	0,45
ABOU	7643	4331	3842	4290	3801	489	0,90	0,50	0,11
ZRIREK	7643	7840	3862	7759	3781	3978	0,50	0,49	0,51
FEMME	7643	1539	1187	6808	6456	352	0,17	0,84	0,23
LETTRE	7643	3697	2370	6600	5273	1327	0,36	0,69	0,36
MILLE	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de A
ELHAKI	7270	8855	4616	6893	2654	4239	0,67	0,37	0,48
GENET	7270	6288	3542	6474	3728	2746	0,55	0,51	0,44
ABOU	7270	4331	2811	5979	4459	1520	0,47	0,61	0,35
ZRIREK	7270	7840	3893	7324	3377	3947	0,53	0,46	0,50
FEMME	7270	1539	1189	6431	6081	350	0,18	0,84	0,23
LETTRE	7270	3697	2400	6167	4870	1297	0,39	0,67	0,35
ELHAKI	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de A
GENET	8855	6288	3973	7197	4882	2315	0,55	0,55	0,37
ABOU	8855	4331	3120	6946	5735	1211	0,45	0,65	0,28
ZRIREK	8855	7840	4568	7559	4287	3272	0,60	0,48	0,42

FEMME	8855	1539	1275	7844	7580	264	0,16	0,86	0,17
LETTRE	8855	3697	2761	7030	6094	936	0,39	0,69	0,25
GENET	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de A
ABOU	6288	4331	3008	4603	3280	1323	0,65	0,52	0,31
ZRIREK	6288	7840	4902	4324	1386	2938	1,13	0,22	0,37
FEMME	6288	1539	1125	5577	5163	414	0,20	0,82	0,27
LETTRE	6288	3697	2263	5459	4025	1434	0,41	0,64	0,39
ABOU	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de A
ZRIREK	4331	7840	2830	6511	1501	5010	0,43	0,35	0,64
FEMME	4331	1539	1023	3824	3308	516	0,27	0,76	0,34
LETTRE	4331	3697	1805	4418	2526	1892	0,41	0,58	0,51
ZRIREK	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de A
FEMME	7840	1539	1235	6909	6605	304	0,18	0,84	0,20
LETTRE	7840	3697	2535	6467	5305	1162	0,39	0,68	0,31
FEMME	Voc A	Voc B	Voc commun	Voc exclusif de A et B	Voc exclusif en A	Voc exclusif en B	Connexion	Indépendance de A	Indépendance de A
LETTRE	1539	3697	905	3426	634	2792	0,26	0,41	0,76

Tableau III. 5 : Connexion lexicale (indice brut)

	PARC	AÎL	MILL	ELHA	GENE	ABOUN	ZRIREK	FEMME	LETTRE
PARCOURS	0	0,61	0,62	0,62	0,52	0,46	0,54	0,19	0,4
AÎLEN	0	0	0,63	0,66	0,49	0,9	0,5	0,17	0,36
MILLE	0	0	0	0,67	0,55	0,47	0,53	0,18	0,39
ELHAKI	0	0	0	0	0,55	0,65	0,48	0,86	0,69
GENET	0	0	0	0	0	0,65	1,13	0,2	0,41
ABOUNOUR	0	0	0	0	0	0	0,43	0,27	0,41
ZRIREK	0	0	0	0	0	0	0	0,18	0,39
FEMME	0	0	0	0	0	0	0	0	0,26
LETTRE	0	0	0	0	0	0	0	0	0

Pour notre corpus, les deux textes qui ont un indice de connexion fort sont (GENET et ZRIREK : 1,13) cela veut dire qu'ils ont une part commune plus large que celle spécifique à chacun. Alors que les textes les plus éloignés sont ZRIREK et FEMME, ces derniers ont un indice faible de 0,18. Il est question ici de deux textes qui abordent des sujets différents, donc l'intersection entre les éléments de cet ensemble sera étroite.

Notons que la majorité des binômes qui ont intégré le texte FEMME ont enregistré un indice de connexion trop faible. En effet, presque tous les textes d'El Maleh traitent plus ou moins les mêmes thèmes (le colonialisme, la guerre, l'écriture allégorique, l'exode juive, etc.), cependant l'écrivain, vers la fin de sa vie, avait le souci de sombrer dans la redondance, c'est la raison pour laquelle il a entamé l'écriture du texte FEMME qui reste plus ou moins différent des autres textes surtout sur le plan thématique. Donc l'éloignement du texte FEMME des autres est dû au besoin de l'écrivain de créer un changement dans son œuvre.

Signalons que BRUNET²⁰⁸ a mentionné dans son ouvrage, *Le Vocabulaire de Jean Giraudoux, Structure et évolution*, que pour mesurer avec précision la distance lexicale, il ne faut pas s'intéresser uniquement à la présence ou l'absence d'un tel ou tel vocable, mais il faut prendre en considération la fréquence des mots.

De ce fait, pour pouvoir se prononcer sur l'éloignement ou le rapprochement des textes, nous devons exploiter d'autres méthodes qui intègrent la fréquence du mot comme variable. Nous essayerons bien évidemment la loi binomiale.

B. Essai de la loi binomiale

Dans le présent point, nous testerons une autre approche, quoiqu'elle soit rarement appliquée vu la lourdeur de ses calculs, il s'agit en fait de la loi binomiale. Précisons que le recours à cette méthode est dicté par les limites de la méthode Jaccard qui ignore la fréquence des mots et leur distribution. D'ailleurs, Brunet²⁰⁹ stipule à ce sujet que les fréquences élevées perdent tout poids dans le calcul, puisqu'elles se trouvent nécessairement dans la partie commune du vocabulaire.

Pour tenir compte de la fréquence, Charles Muller a proposé une approche qui se base sur la loi binomiale. En effet, cette méthode permet de mesurer avec précision la distance²¹⁰ qui sépare le vocabulaire de deux textes. Les mesures ne vont pas se réduire à une simple confrontation binaire qui constate la présence ou l'absence d'un vocable particulier, mais elles cherchent si cette présence est insistante ou fugitive, essentielle ou marginale.

²⁰⁸ Etienne BRUNET, *op.cit.*, 1978, p. 375

²⁰⁹ Etienne BRUNET, *op.cit.*, 2006, p. 52

²¹⁰ Charles MULLER, *op.cit.*, 1976, p. 375

Après avoir réuni les deux textes à comparer, il faut leur attribuer les fréquences selon leur distribution, c'est-à-dire, prendre en considération l'effectif des mots employés 1 fois, 2 fois, n fois par texte. Cependant, puisque le logiciel *Hyperbase* ne permet pas une telle fonction, le travail devrait se faire manuellement, ce qui rend la tâche trop longue et trop délicate.

Une fois la liste des effectifs réels est établie, le calcul binomial permet alors de prévoir comment théoriquement devrait se répartir ces effectifs dans les deux textes. Ainsi, pour chaque binôme sera constitué un tableau d'effectif à double entrée. L'axe des ordonnées regroupera les mots selon leur fréquence dans le premier texte, cependant l'axe des abscisses présentera la fréquence des mots dans le second. La première colonne regroupe les vocables manquant dans (A) mais qui existent dans (B). Tandis que les effectifs des vocables qui sont dans le premier texte (A) et non dans le second (B) sont présentés dans la première ligne. Entre les deux se trouvent les vocables communs qui sont placés sur une diagonale descendante.

Après avoir procédé aux calculs nécessaires, la somme des χ^2 donne la mesure de la distance entre les deux textes, compte tenu des degrés de liberté, qui peuvent varier d'un binôme à l'autre.

S'agissant de notre corpus, il englobe neuf textes engendrant 36 combinaisons. De ce fait, nous avons opté pour l'application de la méthode en question à deux couples, l'un constitué de deux textes d'étendues voisines (AILEN, MILLE), l'autre est formé par deux textes dont les étendues sont différentes (EL HAKI, LITTRES).

1. Application de la loi binomiale aux textes EL HAKI et LETTRES

L'application de la loi binomiale au binôme EL HAKI et LETTRES exige, d'une part, l'élaboration du tableau des effectifs réels ainsi que celui des effectifs

théoriques, d'autre part, la confrontation des deux tableaux afin de déboucher sur un écart. Le tableau suivant illustre les effectifs réels des textes ELHAKI et LETTRES.

Tableau III. 6 : Effectifs réels des textes EL HAKI et LETTRES

	0	1	2	3	4	5	6	7	8	9
0	0	1294	125	30	11	4	3	2	0	1
1	6161	606	103	39	9	3	1	2	1	
2	1752	333	85	20	8	2	1	1		
3	763	210	80	27	6	0	3			
4	403	143	51	12	8	4				
5	221	99	43	20	9					
6	150	91	41	22						
7	86	50	17							
8	64	44								
9	44									

L'élaboration du tableau des effectifs théoriques nécessite tout d'abord le calcul des probabilités p et sa complémentaire q :

$$p = N_a / (N_a + N_b) = 22565 / (22565 + 109549) = 0,170799$$

Avec :

N_a est le nombre d'occurrences du texte LETTRES

N_b est l'ensemble des occurrences du texte ELHAKI

$$q = N_b / (N_a + N_b) = 109549 / (22565 + 109549) = 1 - p = 0,829201.$$

Nous appliquerons la loi binomiale pour calculer les effectifs théoriques pour chaque classe de fréquence :

$$(p + q)^n = p^n + C_n^{n-1} p^{n-1} q + C_n^{n-2} p^{n-2} q^2 + \dots + C_n^2 p^2 q^{n-2} + C_n^1 p^1 q^{n-1} + q^n$$

Considérons les extensions binomiales suivantes pour n allant de 1 jusqu'à 9 :

$$(p + q)^0 = 1$$

$$(p + q)^1 = p^1 + q^1$$

$$(p + q)^2 = p^2 + 2p q + q^2$$

$$(p + q)^3 = p^3 + 3p^2 q + 3p^1 q^2 + q^3$$

$$(p + q)^4 = p^4 + 4p^3 q + 6p^2 q^2 + 4p^1 q^3 + q^4$$

$$(p + q)^5 = p^5 + 5p^4 q + 10p^3 q^2 + 10p^2 q^3 + 5p^1 q^4 + q^5$$

Prenons la fréquence 1 : nous additionnons les effectifs de la diagonale correspondant à la fréquence 1, nous aurons $1294 + 6161 = 7455$ puis nous multiplions la valeur obtenue par $p = 0,170799$ pour avoir l'effectif théorique du texte LETTRES qui est de $7455 \times p^1 = 7455 \times 0,170799 = 1273,30654 \approx 1273$, ensuite par $q = 1 - p = 0,829201$ pour trouver l'effectif théorique d'EL HAKI qui est de $7455 \times q^1 = 7455 \times 0,829201 = 6181,69346 \approx 6182$;

Pour la fréquence 2 : nous additionnons les effectifs de la diagonale correspondante à la fréquence 2, nous aurons $125 + 606 + 1752 = 2483$ puis nous multiplions la valeur obtenu par $p^2 = 0,029172$ pour avoir l'effectif théorique du texte LETTRES qui est de $2483 \times p^2 = 2483 \times 0,029172 = 72,44 \approx 72$ ensuite par $q = 1 - p = 0,829201$ pour trouver l'effectif théorique d'EL HAKI qui est de $2483 \times q^2 = 2483 \times 0,687574 = 1707,25 \approx 1707$; et reste l'effectif théorique médian $2483 \times p q = 2483 \times 2 \times 0,170799 \times 0,829201 = 703,3182 \approx 703$

Pour la fréquence 3 : nous calculons les effectifs de la diagonale afférente à la fréquence 3, ce qui donne $30 + 103 + 333 + 763 = 1229$, puis nous déterminons les quatre termes du polynôme $p^3, 3p^2q, 3p^1q^2$ et q^3 , ce qui donne : $1229 \times p^3 = 1429 \times 0,170799^3 = 1229 \times 0,004983 = 6,12 \approx 6$, $1229 \times 3 \times p^2 \times q = 1229 \times 0,170799^2 \times 0,829201 = 89,19 \approx 89$, $1229 \times 3 \times p \times q^2 = 1229 \times 0,170799 \times 0,829201^2 = 432,99 \approx 433$, $1229 \times q^3 = 1429 \times 0,829201^3 = 1229 \times 0,570136 = 700,70 \approx 701$.

Nous reproduisons, ainsi, la même démarche de Charles MULLER, telle qu'elle est décrite dans son ouvrage « *Initiation à la Statistique linguistique* », pour l'ensemble des fréquences jusqu'à la neuvième.

Tableau III. 7 : Effectifs théoriques des textes LETTRES et EL HAKI

	0	1	2	3	4	5	6	7	8	9
0	0	1273	72	6	0	0	0	0	0	0
1	6182	703	89	12	0	0	0	0	0	
2	1707	433	90	16	0	0	0	0		
3	701	291	79	19	0	0	0			
4	354	193	71	19	0	0				
5	187	137	55	20	6					
6	111	90	49	20						
7	62	69	41							
8	42	50								
9	27									

Après avoir présenté les tableaux des effectifs réels et théoriques, il convient maintenant de mettre en relief le tableau des écarts qui sert de base pour la détermination des valeurs de khi-deux.

Tableau III. 8 : Ecart entre effectifs théoriques et réels des textes EL HAKI et LETTRES

	0	1	2	3	4	5	6	7	8	9
0	0	21	53	24	11	4	3	2	0	1
1	-21	-97	14	27	9	3	1	2	1	
2	45	-100	-5	4	8	2	1	1		
3	62	-81	1	8	6	0	3			
4	49	-50	-20	-7	8	4				
5	34	-38	-12	0	3					
6	39	1	-8	2						
7	24	-19	-24							
8	22	- 6								
9	17									

Le tableau III.8 décèle la présence d'écarts positifs sur la première ligne qui revient au texte LETTRES : 21, 53, 24, 11, 4, 3, 2, 0, 1. Il en est de même pour la première colonne qui est attribuée au texte ELHAKI ainsi que pour la partie située au-dessus de la diagonale principale. Cependant, cette dernière présente des écarts négatifs dans l'ensemble -97, -5 et même sur la partie située en bas de la diagonale. Ce constat met en avant le phénomène de la spécialisation lexicale, du moment que chaque texte peut faire appel aux exigences du thème, du genre ou autres.

Sur la base des écarts, nous pourrions alors calculer le carré des écarts comme le montre le tableau ci-dessous :

Tableau III. 9 : Écarts au carré des textes EL HAKI et LETTRES

	0	1	2	3	4	5	6	7	8	9
0	0	441	2809	576	121	16	9	4	0	1
1	441	9409	196	729	81	9	1	4	1	
2	2025	10000	25	16	64	4	1	1		
3	3844	6561	1	64	36	0	9			
4	2401	2500	400	49	64	16				
5	1156	1444	144	0	9					
6	1521	1	64	4						
7	576	361	576							
8	484	36								
9	289									

Reste maintenant à présenter, ci-dessous, le tableau qui mettra en exergue les valeurs des Khi-deux.

Tableau III. 10 : Khi-deux des textes EL HAKI et LETTRES

LETT ELHAKI	0	1	2	3	4	5	6	7	8	9	Total
0	0	0,3464	39,013	96	0	0	0	0	0	0	135,36
1	0,07133	13,384	2,2022	60,75	0	0	0	0	0		76,407
2	1,18629	23,094	0,2777	1	0	0	0	0			25,558
3	5,48359	22,546	0,0126	3,3684	0	0	0				31,411
4	6,78248	12,953	5,6338	2,5789	0	0					27,948
5	6,18181	10,540	2,6181	0	1,5						20,840
6	13,7027	0,0111	1,3061	0,2							15,219
7	9,29032	5,2318	14,048								28,570
8	11,5238	0,72									12,243
9	10,7037										10,703
Total	64,9260	88,828	65,113	163,89	1,5	0	0	0	0	0	384,26

A première vue, nous découvrons que les valeurs des khi-deux sont de plus en plus fortes en se rapprochant de la première ligne réservée au vocabulaire exclusif de **LETTRES** ou de la première colonne dédiée au vocabulaire exclusif de **EL HAKI**. Pourtant, la somme des khi-deux de la première ligne est largement supérieure à celle de la première colonne, ce qui veut dire que le texte **LETTRES** possède plus de mots qui lui sont propres que le texte **EL HAKI**.

Reste à signaler que le calcul des khi-deux a permis d'avoir une valeur totale de 384,265 pour 33 degrés de liberté et comme le degré de liberté est strictement supérieur à 30, la loi de khi-deux converge vers une loi normale centrée et réduite. Cette convergence est appelée l'approximation de Fisher qui est donnée par la formule suivante :

Soit : $X \sim \chi^2(n)$, si $n > 30$

Alors : $\sqrt{2X} - \sqrt{2n - 1} \sim N(0,1)$

Donc : $Z = \sqrt{2 \times 384,265} - \sqrt{2 \times 33 - 1} = 19,98 \sim N(0,1)$

La valeur de l'écart réduit ainsi calculé s'avère très significative au risque de 5%, ce qui exclut l'effet du hasard et justifie, par voie de conséquence, le phénomène de spécialisation lexicale. En outre, nous ne pouvons éliminer l'influence de l'étendue des textes.

2. Application de la loi binomiale aux textes MILLE et AÏLEN

Pour compléter l'analyse, nous appliquerons les mêmes tests au binôme **MILLE** et **AÏLEN** dont les étendues sont voisines. Dans ce cadre, nous sommes amenée à élaborer le tableau des effectifs réels des deux textes ainsi que celui des effectifs théoriques.

Tableau III. 11 : Effectifs réels des textes MILLE et ÀILEN

	0	1	2	3	4	5	6	7	8	9
0	0	3735	776	244	93	41	18	6	12	5
1	3806	1048	396	182	80	43	26	13	5	
2	859	403	235	149	68	34	27	14		
3	272	186	117	91	57	35	34			
4	108	85	71	61	52	30				
5	50	41	44	45	22					
6	23	22	27	28						
7	14	22	16							
8	15	10								
9	7									

Suivant la même démarche appliquée au premier binôme (LETTRE/EL HAKI), nous reproduirons les calculs pour l'ensemble des fréquences jusqu'à la neuvième.

Tableau III. 12 : Effectifs théoriques des textes MILLE et AILEN

	0	1	2	3	4	5	6	7	8	9
0	0	3782	675	166	51	17	6	0	0	0
1	3759	1341	495	202	82	34	11	8	0	
2	667	492	301	164	84	34	27	12		
3	163	200	163	111	56	55	28			
4	50	81	83	55	68	42				
5	16	33	33	54	42					
6	0	11	27	28						
7	0	8	12							
8	0	0								
9	0									

Une fois les tableaux des effectifs réels et théoriques présentés, il reste à les confronter pour déboucher sur le tableau des écarts qui nous permettra de déterminer les valeurs des khi-deux.

**Tableau III. 13 : Écarts entre effectifs théoriques et réels des textes MILLE et
ÀILEN**

	0	1	2	3	4	5	6	7	8	9
0	0	-47	101	78	42	24	12	6	12	5
1	-47	-293	-99	-20	-2	9	13	5	5	
2	192	-89	-66	-15	-16	0	0	2		
3	109	-14	-46	-20	1	-20	6			
4	58	4	-12	6	-16	-12				
5	34	8	11	-9	-20					
6	23	11	0	0						
7	14	14	4							
8	15	10								
9	7									

A première vue, nous décelons que la diagonale principale présente des écarts négatifs dans l'ensemble. Cependant, elle divise le tableau en deux parties, la première située en haut est marquée par des écarts positifs, contrairement à la partie située en bas qui présente majoritairement des écarts négatifs. Nous remarquons également que la première ligne qui revient au texte **MILLE** est caractérisée dans l'ensemble par des écarts positifs : 101, 78, 42, 24, 12, 6, 12, 5. Il en est de même pour la première colonne qui est attribuée au texte **ÀILEN**. Ces remarques mettent en exergue le phénomène de la spécialisation lexicale qui peut être expliquée par des raisons thématiques, chronologique ou autres.

Sur la base du tableau ci-dessus, nous pouvons alors calculer le carré des écarts :

Tableau III. 14 : Écarts au carré des textes MILLE et ÀILEN

	0	1	2	3	4	5	6	7	8	9
0	0	2209	10201	6084	1764	576	144	36	144	25
1	2209	85849	9801	400	4	81	225	25	25	
2	36864	7921	4356	225	256	0	0	4		
3	11881	196	2116	400	1	400	36			
4	3364	16	144	36	256	144				
5	1156	64	121	81	400					
6	529	121	0	0						
7	196	196	16							
8	225	100								
9	49									

À partir de cette étape, il nous sera aisé de produire le tableau portant sur les valeurs des Khi-deux.

Tableau III. 15 : Khi-deux des textes MILLE et ÀILEN

MILLE	0	1	2	3	4	5	6	7	8	9	Total
ÀILEN											
0	0	0,5840825	15,112593	36,650602	34,588235	33,882353	24	0	0	0	144,81787
1	0,5876563	64,0186428	19,8	1,980198	0,0487805	2,3823529	20,454545	3,125	0		112,39718
2	55,268366	16,0995935	14,471761	1,3719512	3,047619	0	0	0,3333333			90,592624
3	72,889571	0,98	12,981595	3,6036036	0,0178571	7,2727273	1,2857143				99,031068
4	67,28	0,19753086	1,7349398	0,6545455	3,7647059	3,4285714					77,060293
5	72,25	1,93939394	3,6666667	1,5	9,5238095						88,87987
6	0	11	0	0							11
7	0	24,5	1,3333333								25,833333
8	0	0									0
9	0										0
Total	268,27559	119,319244	69,100888	45,760901	50,991007	46,966005	45,74026	3,4583333	0	0	649,61223

Nous constatons que les valeurs des khi-deux sont de plus en plus fortes en se rapprochant de la première ligne ou de la première colonne. Par ailleurs, ce qui est remarquable, c'est que la somme des khi-deux de la première colonne vaut presque deux fois la somme des khi-deux de la première ligne, ce qui veut dire que le texte **AÏLEN** possède plus de mots qui lui sont propre que le texte **MILLE**. Ce résultat converge avec celui constaté dans la richesse lexicale et de l'accroissement lexical.

Notons que le total des khi-deux s'élève à 649,61223 pour 41 degrés de liberté et puisque le degré de liberté est strictement supérieur à 30, la loi de khi-deux converge vers une loi normale centrée et réduite, donc :

$$Z = \sqrt{2 \times 649,61223} - \sqrt{2 \times 41 - 1} = 27,04 \sim N(0,1)$$

Ainsi, nous trouvons que la valeur de l'écart réduit est très significative au risque de 5%, ce qui exclut l'effet du hasard et justifie le phénomène de la spécialisation lexicale.

En définitive, l'étude des deux couples moyennant la loi binomiale a rendu compte d'un décalage entre la distribution théorique et celle réelle, ce qui ne permet pas de généraliser des conclusions. De plus, nous ne pouvons pas exploiter cette loi sur toutes les partitions du corpus en raison de l'effet des étendues auxquelles les mesures probabilistes sont très sensibles.

Somme toute, l'étude des rapprochements entre les textes de notre corpus a été effectuée, tout d'abord, en utilisant la méthode Jaccard qui ne considère que la présence ou l'absence des mots dans les textes considérés, abstraction faite de leur fréquence. Ensuite, nous avons eu recours à une autre méthode qui se base sur la loi binomiale mais qui tient compte de la fréquence des mots.

S'agissant de la méthode Jaccard, nous avons commencé, tout d'abord, par calculer le vocabulaire commun à chaque binôme pour avoir une idée sur l'éloignement ou le rapprochement des textes étudiés. Pour une bonne interprétation des résultats, nous avons utilisé l'analyse factorielle et celle arborée, ce qui nous a permis de dresser une liste de couples qui s'attirent soit parce qu'ils partagent le même contenu lexical, le même genre ou bien les mêmes sujets. Il s'agit bien évidemment des couples (GENET, ZRIREK) ; (AILEN, MILLE) et (ELHAKI, PARCOURS). Notons que ces rapprochements convergent avec les résultats obtenus dans l'analyse de la longueur des mots. En effet, cette dernière a montré une attraction des textes GENET et ZRIREK pour les mots longs, ce qui est en faveur de l'avis des critiques qui classent ces deux textes du côté des essais. La même remarque reste valable pour les binômes (AILEN, MILLE) et (ELHAKI, PARCOURS) qui forment tous une constellation dans l'AFC des longueurs des mots en attirant les mots courts et moyens, ce qui renforce l'idée qu'il s'agit bien évidemment de romans. Déjà ces textes se regroupent aussi dans l'analyse factorielle et celle arborée de la distance lexicale.

Ce rapprochement, confirmé sur tous les plans peut être dû, aussi, à la thématique traitée dans chacun des textes formant les couples susmentionnés. Déjà dans les binômes (AILEN, MILLE) et (ELHAKI, PARCOURS) nous trouvons les thèmes de l'exode juif, de la guerre, du colonialisme, du militantisme marocain, etc. Aussi, si nous prenons en considération leur chronologie de parution, nous remarquerons que les quatre textes se succèdent. En plus, le texte ELHAKI qui est le dernier à être publié, constitue

une récapitulation des sujets traités dans les trois autres romans. Soulignons également que les textes GENET et ZRIREK se partagent les thèmes de la littérature, des écritures mystiques et de l'allégorie.

Pour compléter l'analyse et exprimer la connexion lexicale de manière plus détaillée, nous nous sommes intéressée à la distance intertextuelle qui pourra nous renseigner sur la proximité des textes à travers la comparaison de la distribution de tous les mots de chaque couple. Également dans cette étape, la parenté des textes GENET et ZRIREK était prouvée, cependant, les textes qui restent les plus éloignés sans ABOUNOUR et FEMME. Ce qui signifie que les deux textes sont caractérisés par un vocabulaire privatif spécifique à chacun d'eux.

Nous avons mis en œuvre un autre indice qui permet de mesurer l'indépendance lexicale de chaque texte par rapport à l'autre, à savoir : l'indice de connexion lexicale. Ce dernier a montré une forte connexion entre GENET et ZRIREK ; cela veut dire qu'ils ont une part commune plus large que celle spécifique à chacun. Par ailleurs, tous les binômes qui ont intégré le texte FEMME ont enregistré un indice de connexion trop faible. Ceci reste normal du moment que l'écrivain, vers la fin de sa vie, avait le souci de sombrer dans une sempiternelle répétition, c'est la raison pour laquelle il a entamé l'écriture du texte FEMME qui reste plus ou moins différent des autres textes surtout sur le plan thématique. Donc l'éloignement du texte FEMME des autres est dû au besoin de l'écrivain de créer un changement dans ses écrits.

En plus de la méthode Jaccard, nous avons testé une autre approche empruntée à Ch. MULLER, à savoir : la loi binomiale. En effet, cette dernière permet de présenter théoriquement la répartition des effectifs dans les deux textes considérés en vue de la comparer à une autre réel. Néanmoins, les résultats obtenus ne pourraient être fiables que pour des couples de textes dont les étendues sont voisines.

Tout compte fait, ce chapitre constitue une première appréciation des rapprochements et des éloignements qui existent entre les différentes partitions du corpus. Cette analyse sera complétée par l'étude des spécificités internes qui mettront en valeur les penchants thématiques de chaque texte et par conséquent, nous renseignerons sur la charge sémantique que l'auteur veut transmettre à travers son écriture.

CHAPITRE 2

Vocabulaire spécifique du corpus

Dans le présent chapitre, nous nous focaliserons sur la spécificité de l'écriture d'Edmond Amran El Maleh suivant deux volets. Le premier se chargera de comparer le corpus à un autre externe, en l'occurrence, *le Trésor de la Langue Française informatisé (TLFi)*, tandis que le deuxième examinera les variations de chaque texte par rapport au corpus. Partant, l'objectif de cette étude est de dégager les thèmes privilégiés par l'auteur. Ce sera donc une analyse sémantique.

Notons que le logiciel *Hyperbase* permet de générer une liste automatique regroupant toutes les occurrences du vocabulaire spécifique classées par ordre décroissant de l'écart réduit et une autre lemmatisée. Cependant, nous avons opté pour les formes. Ce choix trouve sa justification dans le fait qu'un auteur pourrait jouer consciemment ou inconsciemment sur certains aspects de la langue, comme : la flexion grammaticale, les temps verbaux, les noms propres, etc.

Notre étude ne se limitera pas à un fait bien précis, mais elle essaiera souvent d'analyser les champs sémantiques pouvant aider à isoler les thèmes saillants, que ce soit dans le corpus pris intégralement ou dans chaque texte pris isolément.

A. Vocabulaire spécifique excédentaire et déficitaire : Étude exogène

Admettons que la fréquence n'a de valeur que si elle est comparée à une autre théorique qui sera la norme et la base de calcul. S'agissant de notre étude, ce sont les

données du TLFi qui seront considérées comme norme et qui mettrons en relief les excédents et les déficits du vocabulaire d'El Maleh.

A priori, notre étude sera facilitée par le logiciel Hyperbase qui effectue les calculs nécessaires en sollicitant le bouton SPÉCIFICITÉS. En effet, la méthode²¹¹ consiste à calculer l'écart réduit de chacun des vocables du texte par rapport à la sous fréquence théorique, puis elle les classe.

Soulignons que le TLFi a paru entre 1971 et 1994, de ce fait certains de nos résultats seraient relatifs, à défaut d'une mise à jour récente du TLFi, d'une part, d'autre part, en raison de l'évolution de la langue. Une autre remarque s'impose, c'est qu'il existe aussi quelques aspects qui peuvent influencer la comparaison, comme le registre de langue, la taille du corpus traité ainsi que la fréquence du mot en question.

Cependant, pour une meilleure lecture des résultats, la liste du vocabulaire spécifique sera répartie en quatre catégories grammaticales, ce qui aboutit aux données suivantes : 255 substantifs (61,89%), 70 verbes (16,99%), 76 adjectifs (18,44%) et 11 adverbes (2,67%).

Ainsi, nous nous pencherons sur l'analyse exogène du vocabulaire spécifique de l'œuvre d'Edmond Arman El Maleh. Mais soulignons d'emblée que nous nous sommes abstenue de dresser la liste du vocabulaire excédentaire et déficitaire, justement pour éviter l'encombrement des pages. Cependant, les spécificités de l'écriture maléhienne seront mise en valeur en dressant les champs sémantiques qui peuvent mettre en relief les préférences thématiques de l'auteur, sans pour autant oublier de prendre en considération les mots ayant les écarts réduits les plus significatifs.

²¹¹ Charles MULLER., *op.cit.*, 1968, p. 204

1. Appartenances linguistiques et ethniques

En agglomérant le vocabulaire excédentaire des appartenances linguistiques et ethniques nous obtenons le tableau suivant :

Tableau III. 16 : Appartenances linguistiques et ethniques (Substantifs)

Écart réduit	TLF	Corpus	Mots
177.56	42	106	marocains
53.64	1543	147	arabe
41.45	2052	134	juif
40.86	371	54	hébreu
35.89	756	69	juive
28.57	2452	105	Juifs
15.83	14932	195	français
13.22	3368	66	Origine
10.97	5763	81	française
8.62	5992	43	étranger

Le mot qui se trouve en tête de liste du vocabulaire spécifique excédentaire n'étonnera aucun lecteur de l'œuvre soumise à l'analyse. Il s'agit en effet du mot **marocains** avec un z de + 177,56 suivi du mot **arabe** en deuxième position avec un écart réduit de + 53,64 et du mot **juif** en troisième position avec un z de + 41,45.

Une autre remarque s'impose, c'est que les substantifs *français(e)*²¹² et **arabe** sont particulièrement au singulier alors que le substantif **juif** se présente à la fois au

²¹² Pour le mot *français* qui porte toujours le (s), qu'il soit au pluriel ou au singulier, nous avons consulté le logiciel *Hyperbase* pour lever l'ambiguïté, nous avons constaté finalement que le substantif *français* est particulièrement au singulier.

pluriel et au singulier. La même remarque reste valable pour le substantif **marocain** qui est souvent au pluriel. De ce fait, nous pourrions croire qu'en évoquant les français et les arabes c'est surtout l'individu qui prédomine, par contre en parlant des **marocains** et des **juifs** c'est surtout des communautés qu'il s'agit.

Cet intérêt pour ces deux communautés n'est-il pas justifié par l'origine ethnique de l'écrivain lui-même, ou bien c'est la disparition chaotique des juifs marocains qui est à l'origine de ce constat, une disparition que l'auteur qualifie au niveau de ces romans de « *fracture sismique* » ou d'« *incheqq* ». D'ailleurs, tout au long de son œuvre, El Maleh nous offre des textes imbus de nostalgie et de regret pour un monde perdu. Il pleure l'harmonie qui régnait entre les juifs marocains et les différentes catégories sociales, ethnologique et religieuses composant la communauté marocaine à une époque donnée.

Cependant, si la thématique de l'exode de la communauté juive est si importante pour El Maleh, pourquoi le mot **juif** occupe-t-il la troisième position après **marocain** et **arabe** ?

Si nous empruntons le raisonnement de Pierre Guiraud, selon lequel :

*« la fréquence des mots chez un auteur est non seulement un des éléments les plus caractéristiques de sa vision et de son style, mais celui par lequel il agit le plus puissamment sur la sensibilité du lecteur et sur la langue. »*²¹³

²¹³ P. GUIRAUD, *op. cit.*, 1954, p. 97

Nous pouvons déduire, alors, que la fréquence serait très significative pour relever les caractéristiques de l'écriture chez un auteur quelconque.

Suivant la même logique, si nous prenons en considération la fréquence des mots, nous pouvons constater que le substantif **juif** est repris 308 fois (qu'il soit au masculin, au féminin ou au pluriel), suivi du mot **français** avec une fréquence de 276 et du mot **arabe** qui a été repris 147 fois, finalement nous trouvons le mot **marocains** avec une fréquence de 106 fois. Nous sommes là devant un paradoxe, le premier classement, celui des écarts réduits met le mot **marocains** au sommet de la liste²¹⁴, cependant le deuxième classement qui se base essentiellement sur les fréquences, place le mot **juif** en tête de liste. Donc le lecteur est amené inconsciemment à garder en mémoire le mot le plus fréquent à savoir le substantif **juif**.

En consultant la liste des adjectifs nous trouvons que le mot **marocains** affiche également l'indice le plus élevé, mais en vérifiant le contexte de l'adjectif **marocain(s)** il est dans 74 % des cas placés à côté du substantif **juifs** en tant qu'adjectif qualificatif. De ce fait nous pouvons postuler que la communauté juive marocaine occupe une place prépondérante dans le raisonnement d'El Maleh.

²¹⁴ Notons que les écarts réduits sont fortement influencés par la fréquence des mots au niveau de TLFi

Tableau III. 17 : Appartenances linguistiques et ethniques (Adjectifs)

Écart	TLF	Corpus	Mots
177.56	42	106	Marocains
53.64	1543	147	Arabe
41.45	2052	134	juif
28.57	2452	105	juifs
15.83	14932	195	français
10.97	5763	81	française
3.89	756	69	juive

Une dernière remarque relative au tableau n° III.16, c'est que l'écart réduit du substantif *origine* dépasse de loin celui du substantif *étranger*, ceci ne peut que corroborer les constats auxquels nous avons abouti, à savoir l'attachement de l'écrivain à ces racines judéo-marocaines.

2. Littérature

En examinant la liste du vocabulaire spécifique excédentaire nous trouvons également des mots qui relèvent du champ sémantique de la littérature (tableau n° III.18). A leur tête, il y a les substantifs *écriture* et *allégorie*. Alors examinant ces deux concepts chacun dans un volet d'analyse à part :

Tableau III. 18 : Littérature

Écart	TLF	Corpus	Mots
116.39	2942	433	écriture
90.83	172	80	allégorie
76.58	2291	254	texte
54.12	225	55	métaphore
47.24	5211	250	récit
35.46	9192	267	langue
33.09	5620	190	lecture
25.49	22762	357	histoire
21.4	23318	321	mots
20.21	16605	247	écrit
19.8	4902	114	page
16.35	5074	100	poésie
15.43	5896	105	langage
14.41	21358	235	livre
13.01	5055	84	pages
12.03	12122	142	œuvre
12.01	3339	61	recherche
11.14	5351	78	poète
10.87	4050	64	littérature
10.86	4351	67	personnage
8.6	12855	122	cours
8.46	23380	190	savoir
6.03	35801	235	mot
5.4	8362	70	auteur
4.99	20917	141	lettres

2.1. Écriture

Ce n'est pas surprenant que le mot *écriture* occupe une place très importante dans le vocabulaire spécifique de notre corpus, d'ailleurs c'est l'essence même de l'entreprise maléhienne, mais que serait l'objectif de son écriture ? Quel est le message qu'il désire véhiculer à travers son acte scriptural ?

En effet, accablé de chagrin et d'amertume, notre écrivain va puiser dans un passé douloureux et définitivement clos pour écrire non pas sa propre vie, mais « *écrire la vie* »²¹⁵. Toutefois, il ne choisit pas d'écrire n'importe quelle vie, mais celle de sa terre-mère, le Maroc.

Cette nécessité pressante de vouloir écrire « la vie du Maroc » trouve son origine non pas uniquement dans la volonté de transmettre un héritage fabuleux mais dans le désir d'El Maleh de vouloir vivre et revivre un certain passé par l'intermédiaire du récit : « *je le voyais vivre et je vivais de le voir vivre* »²¹⁶ révèle le narrateur à propos du parcours du personnage principal dans l'épilogue de *Le retour d'Abou El Haki*. D'où l'utilisation du vocabulaire contenu dans le tableau n° III.19.

Tableau III. 19 : Vie

Écart	TLF	Corpus	Mots
26.93	6479	173	naissance
12.67	95018	682	vie
12.58	6157	93	vivant

²¹⁵ Nous empruntons cette expression à : Ghyslaine Hadouch, *Entre Mémoire et Oubli, L'écriture de l'Histoire*, in : Anouar Ben Msila. *Edmond Amran El Maleh, Art, culture et écriture* : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013, p. 272.

²¹⁶ Edmond Amran EL MALEH, *Le retour d'Abou El Haki*, Grenoble, La Pensée Sauvage, 1988. p. 279.

C'est donc sans doute pour revivre et restituer une vie archéologique, jalonnée de joies et de souffrances (tableaux n° III.20 et n° III.21), qu'El Maleh prend la plume et transcrit une histoire millénaire, l'histoire de la présence juive sur la terre marocaine qui est marquée jusqu'à présent par les synagogues, les cimetières juifs ou encore les mellahs qui continuent à hanter le paysage mémorial marocain.

Tableau III. 20 : Joie

Écart	TLF	Corpus	Mots
27.84	4270	140	chant
13.58	5556	92	folie
13.50	6876	105	fête
5.35	7479	64	spectacle
4.14	16922	111	rire
3.06	13194	82	sourire

Tableau III. 21 : Souffrance

Écart	TLF	Corpus	Mots
17.50	1779	57	choc
14.80	56794	487	mort
12.59	2579	54	cimetière
12.39	3920	69	ténèbres
12.15	4102	70	souffrance
9.41	4068	58	angoisse
8.28	24159	193	guerre
7.57	9739	93	morts
7.49	6095	66	inquiétude
6.96	7865	76	cri
5.39	10111	81	tombe
4.96	7043	59	solitude

2.2. Allégorie

La reconstitution d'une vie dans ses multiples voix, dans ses différentes facettes serait presque un projet impossible. Néanmoins notre auteur avait besoin tout de même de s'aventurer et relever le défi quoiqu'il fût conscient de la complexité de l'entreprise qu'il était en train de mener :

« comment pouvait-il naïvement parler d'avancer, de parcours tracé sur un espace miraculeusement unifié, dompté, chimère, délices d'une séduction subtile, il se parait de pouvoirs magiques, il tenait en sa main le temps

dont sa vie s'imbibait comme une éponge et le temps de l'écriture, l'un et l'autre travaillant à se perdre, affrontés dans un combat sanglant fatal »²¹⁷.

Parmi les stratégies qui lui ont servi pour atteindre son objectif, il y a le procédé de l'allégorie. Ce dernier est formé de trois composantes qui se compénètrent et se correspondent mutuellement, à savoir : l'image, le monde et le langage. Ainsi, tout au long de son œuvre, il a manié ses trois composantes en mobilisant des champs sémantiques divers et variés, par exemple, nous trouvons des substantifs relatifs au langage (tableau n° III.22) :

Tableau III. 22 : Langage

Écart	TLF	Corpus	Mots
39.66	15544	398	parole
20.91	9261	175	mémoire
20.75	50550	535	voix
6.98	10617	95	Conversation
5.26	16803	120	paroles

De même, pour décrire le monde, l'auteur avait recours au vocabulaire contenu dans les tableaux n° III.23, III.24 et III.25 :

²¹⁷ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, Paris, André Dimanche, 2000, p. 159

Tableau III. 23 : Nature

Écart	TLF	Corpus	Mots
39.27	5584	220	espace
36.24	2172	122	écho
31.37	16710	344	lumière
23.63	4849	131	poussière
21.45	1210	55	remparts
19.70	2994	85	paysage
17.99	3140	81	aube
17.78	7431	135	univers
17.12	4942	102	sable
15.37	1961	54	éclats
14.37	5196	92	désert
13.05	5616	90	horizon
12.65	41225	354	ciel
12.65	2885	58	océan
12.22	7419	103	éclat
11.89	21003	208	mer
11.41	4023	66	infini
10.67	7640	96	sol
10.60	48707	372	terre
10.19	17148	165	vent
10.16	6455	83	fleur
10.14	4161	62	vagues
9.22	3833	55	terres
9.11	4317	59	matière

8.18	7438	80	eaux
7.70	24663	190	feu
7.29	5580	61	pierres
5.84	5608	54	source
5.36	6858	60	courant
5.23	18756	131	ombre
4.94	15541	110	beauté
4.89	12179	90	jardin
4.30	15952	107	pierre
4.29	27026	167	soleil
3.06	8400	56	champ

Tableau III. 24 : Parties du corps

Écart	TLF	Corpus	Mots
40.35	2364	141	vision
36.11	21444	447	regard
25.39	26224	390	visage
21.41	36960	438	corps
18.56	7914	145	chair
17.57	3178	80	visages
13.99	6314	102	émotion
11.67	6995	96	souffle
9.09	60566	418	main
9.08	33764	261	sens
8.20	25925	203	sang
8.17	21937	178	oeil
7.32	18917	151	face

6.47	15419	122	figure
6.30	6689	64	ventre
5.83	7960	70	peau
4.96	8793	70	doigts
4.78	24372	158	pieds
4.37	9070	68	oreille
4.19	20312	130	pied
3.00	61526	323	tête

Tableau III. 25 : Lieux et habitations

Écart	TLF	Corpus	Mots
28,59	2752	112	terrasse
21,30	2947	90	miroir
17,85	3653	88	cit��
17,28	29308	327	ville
12,79	4039	72	seuil
12,74	7212	104	portes
11,66	9407	117	lieux
10,29	26314	228	pays
10,09	6156	80	rues
9,98	3417	54	boutique
8,75	34920	264	lieu
8,47	8942	93	murs
8,26	5850	68	bureau
7,88	12179	112	bord
7,88	8216	84	entr��e
5,64	7380	65	demeure

4,76	43825	261	maison
4,61	8816	68	mur
3,90	25365	154	rue
3,44	8701	60	salon

Mais l'essence même de l'allégorie, réside surtout dans sa représentation picturale, d'où l'utilisation du champ sémantique relatif au tableau n° III.26 :

Tableau III. 26 : Peinture

Écart	TLF	Corpus	Mots
25.80	9016	203	image
17.67	4213	95	images
13.25	4132	75	toile
3.37	8951	61	couleur
3.10	15780	96	art

En abordant la peinture, au niveau de la première partie, nous avons évoqué que littérature et peinture chez El Maleh se complètent et se perfectionnent. Pour Anouar Ben Msila dans son article « *l'écrivain, le peintre et l'allégoricien* »²¹⁸, l'articulation de la peinture et de l'écriture dans l'œuvre d'Edmond Amran El Maleh est procurée par le recours à l'allégorie, laquelle assure la complémentarité entre le pictural (visuel) et le scriptural (verbal) mais aussi entre le texte et le monde.

De ce fait, l'allégorie fait voir le monde et, du coup, ouvre la voie vers un chemin de clairvoyance et de reconnaissance de l'autre. C'est sous cet angle que M.-C. Dufour

²¹⁸ Anouar BEN MSILA, « *L'écrivain, le peintre et l'allégoricien* », in : Anouar Ben Msila. Edmond Amran El Maleh, Art, culture et écriture : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013, p. 273

El Maleh écrit en parlant de l'allégorie : « *Elle est...la forme générale de la puissance d'appréhension du monde, de nous-mêmes dans notre rapport au monde* »²¹⁹. Elle recule devant toute pensée absolument abstraite pour tisser des liens avec une approche du sens humaine et sensible, d'où l'immanence de l'émotionnel et du passionnel, dans l'œuvre soumise à l'analyse, à travers les termes contenus dans le tableau suivant :

Tableau III. 27 : Passion

Écart	TLF	Corpus	Mots
24.84	11069	223	désir
3.20	13174	83	passion

Ainsi, de l'écriture allégorique El Maleh dégage une représentation picturale consubstantielle à la réalité parce qu' « *il n'écrit pas par thème ou notion, mais par figure, car est figuratif ce qui possède un correspondant dans le monde naturel* ». Dans cette optique, le procédé de l'allégorie rétablit l'empire narratif maléhien qui s'exprime en se dissimulant dans des figures qui représentent le monde et crée le reflet d'un réel circonscrit. Ainsi, l'allégorie fait du monde et du texte/écriture deux pôles qui se génèrent mutuellement et s'interpénètrent l'une dans l'autre. Elle fait ressurgir la gravité de l'exil, la cruauté de l'exode, la fatalité de l'oubli. De même, elle récupère l'histoire du Maroc, revisite sa diversité culturelle, dessine le parcours tortueux d'un homme asthmatique et d'un militant politique, ce sont là autant de manifestations de l'écriture allégorique d'El Maleh qui est assimilée à une passerelle cruciale et primordiale pour pénétrer dans l'univers de l'œuvre maléhienne et en tracer des voies de lecture, sinon d'écriture.

²¹⁹ Marie-Cécile DUFOUR, *L'écriture allégorique*, Grenoble, La Pensée sauvage, 1999, p.67

3. Art culinaire

Nous enchaînons avec un autre groupement du vocabulaire spécifique, celui de l'art culinaire :

Tableau III. 28 : Art culinaire

Écart	TLF	Corpus	Mots
35.47	2152	119	thé
33.05	738	63	saveur
13.98	6811	107	café
10.71	4420	67	fruit
9.54	6854	83	verre
5.48	8288	70	pain

Les mots relatifs à l'art culinaire articulent les textes d'El Maleh de manière régulière. Ceci reflète l'intérêt que l'auteur accorde à ce domaine. Cependant, nous n'allons pas tarder sur ce champ sémantique, parce qu'il fera l'objet d'une analyse détaillée au niveau du troisième chapitre.

4. Faits historiques et politiques dans l'œuvre d'El Maleh : entre réel et fictif

Dans un monde moderne accablé par la rapidité des événements, le passé nous interpelle et nous incite à revisiter les étapes parcourues de notre existence, c'est ce qu'El Maleh nous offre à travers ces récits qui font œuvre de mémoire et assurent la transmission des événements passés, tout en se situant dans le sillage de l'histoire. En effet, son œuvre raconte une histoire à travers laquelle elle prend sa revanche sur l'Histoire, elle éclaire plusieurs faits historiques importants tels : la guerre mondiale, le

protectorat, l'indépendance, la guerre du Liban, la guerre civile d'Espagne, le communisme, la création de l'état d'Israël qui est corollaire à la destruction de la Palestine, etc. D'où l'émergence des substantifs contenus dans le tableau n° III.29 :

Tableau III. 29 : Politique

Écart	TLF	Corpus	Mots
15.89	6132	110	violence
15.44	6255	109	révolution
14.07	2222	54	exil
13.23	10401	136	politique (44)
9.78	16661	158	parti (105)
7.34	22457	173	ordre
5.25	6795	59	fureur
3.38	17399	107	liberté

Cependant, El Maleh n'introduit pas, dans son texte, des repères réels qui peuvent renvoyer le lecteur directement à un événement donné. Cela nous pousse à se demander comment l'écrivain se joue de son lecteur pour présenter un fait réel tout en le dissimulant dans l'imaginaire des conteurs ?

En effet, qu'il s'agisse de la colonisation française, du protectorat, de l'exil du roi, de la résistance nationale, de l'exode des juifs marocains ou encore des politiques ambiguës qui voulaient réduire au silence un peuple affamé de liberté, El Maleh nous suggère une approche toute particulière pour éclaircir des pans de l'Histoire marocaine souvent occultés et méconnus. En effet, lorsque ces événements réels sont introduits dans les récits maléhiens, ils sont enrobés de brumes de fiction, ils ne s'inscrivent plus dans l'« historique » mais ils appartiennent plutôt à l'univers du narrateur. Ceci est

confirmé par El Maleh lui-même dans ses entretiens avec Marie Redonnet, notamment à propos de *Parcours Immobile* :

*« Dans mon livre, la présence française, l'exil du roi, le Glaoui, je l'évoque sous la forme jubilatoire, dans une espèce de délire. Il apparaît qu'une façon de se détendre de l'importance des événements est de les traiter de façon ludique. Ce qui est important, ce n'est plus alors tellement l'événement mais la façon de l'avoir vécu dans l'écriture ».*²²⁰

Ainsi, l'auteur crée une nouvelle histoire dépourvue de toute contrainte d'objectivité et dans laquelle il se soustrait à ce lourd poids de ce que l'on a coutume d'appeler « historicité ». Il se permet de se laisser porter par les souffles de la jubilation, comme il nous l'explique plus haut, en dissimulant dans la narration du fait historique le regard critique d'un homme éclairé. Donc, à travers le prisme de son imagination, l'auteur prend une bonne distance par rapport aux événements qu'il introduit dans son œuvre pour les revivre à sa manière, tout en mêlant la fiction aux faits réels. D'où l'utilisation du champ sémantique du tableau n° III.30 :

²²⁰ Marie REDONNET, 2005, *op.cit.*, p.10.

Tableau III. 30 : Fiction

Écart	TLF	Corpus	Mots
55.29	318	67	mythe
30.44	1099	72	symbole
29.50	1191	73	pouvoirs
28.41	3195	121	signes
15.40	5028	95	mystère
14.59	10053	142	puissance
7.60	5765	64	aventure
7.31	14342	122	voyage
6.08	5716	56	sagesse
5.84	26136	179	pouvoir
5.67	13362	103	secret
5.44	6047	55	héros

Nonobstant, il ne faut pas s'écarter du but principal de la narration maléhiennne et s'égarer dans les méandres poétiques des mots, il faut surtout que lecteur averti décèle et reconnaisse la part historique des événements que l'auteur dissimule dans les replis des oublis et des souvenirs.

5. Quête de l'identité

Quoique le terme d'identité, « *ce mot abominable de consonance policière* », ou encore « *cette identité fliquée* », ne plaise pas à Edmond Amran El Maleh, il ressort de l'examen minutieux de ses récits que l'auteur inscrit son œuvre dans la littérature de quête, une quête de l'identité. Dans ses récits, la recherche de l'identité se trouve métaphorisée par le motif du voyage. Ce qui peut expliquer l'émergence des verbes de

déplacement et des substantifs relatifs au voyage dans la liste du vocabulaire spécifique excédentaire (tableaux n° III.31 et n° III.32) :

Tableau III. 31 : Verbes de déplacement

Écart	TLF	Corpus	Mots
26.22	19389	329	venait
22.51	20657	307	allait
15.19	4685	90	venaient
12.96	24382	243	passé
12.74	4148	73	venus
12.01	3339	61	recherche
10.41	5195	73	arrivait
9.78	16661	158	parti
8.43	8434	89	court
6.18	7803	71	passait
5.55	12531	97	partir
4.31	28688	176	vient
4.17	15244	102	venu
4.11	11782	82	marche

Tableau III. 32 : Substantifs relatifs au voyage

Écart	TLF	Corpus	Mots
69.76	491	105	parcours
28.54	5258	161	destination
19.79	15110	229	mouvement
14.58	8897	131	passage
14.11	4858	87	distance
6.47	18369	140	chemin
6.46	15599	123	retour
7.31	14342	122	voyage

Ainsi, voyage et errance structurent l'ensemble de ses récits par le biais de son protagoniste qui nous offre une sorte de voyage à la fois physique et mental. Il parcourt l'espace et le temps, nous communiquant ainsi une représentation des événements vus à travers un homme en dissension avec les décisions prises par le groupe religieux auquel il appartient et, en même temps, vivement préoccupé par le devenir de la communauté dont il est issu. Cette double position fait de l'auteur/protagoniste un observateur scrutateur des événements et, en même temps, un témoin inquisiteur de leur réalisation. Aussi, le parcours ou le chemin emprunté par le sujet est particulier, dans la mesure où il le ramène au point de départ. Ce voyage/retour se manifeste sur deux plans : d'un côté, le retour physique du protagoniste à son pays d'origine. D'un autre, le retour à l'Histoire à travers la remembrance et l'évocation des événements passés. Ce voyage/retour constitue une technique ajustée au besoin de l'auteur de mettre en exergue son attachement à son pays natal : Le Maroc, la terre qui a apporté protection et sécurité aux communautés juives aux moments les plus vulnérables de leur existence au nord du

détroit de Gibraltar²²¹, mais aussi une manière de rappeler que le remède à l'errance n'est possible que dans le retour aux origines.

L'itinéraire parcouru dans les œuvres d'El Maleh ne prend pas uniquement la forme d'un voyage effectif ou un périple, il est aussi une sorte de voyage intérieur et subjectif, puisque le voyage signifie non seulement le déplacement d'un espace extérieur vers un autre, mais aussi une descente vers soi, une sorte de parcours existentiel qui mène le sujet au tréfonds de son être, à la recherche d'un « paysage unique » que chaque homme possède au fond de lui-même. D'ailleurs El Maleh le confirme en disant :

*« on porte en soi une fois pour toutes un paysage, un paysage unique.
Hors de quoi il n'est nulle vie vivante. »²²²*

La parole qui murmure au sein des récits d'El Maleh et qui écrase le silence constitue un voyage métaphysique et cérébral permettant à l'auteur de se débarrasser des chaînes qui font de lui un prisonnier d'un seul et unique espace, et de s'ouvrir à sa propre histoire ainsi qu'à celle des autres.

6. La mémoire dans l'œuvre d'El Maleh

Nous avons signalé au niveau de la première partie que les récits d'El Maleh non jamais cessé de dérouter par une structure non linéaire qui se développe le plus souvent sans ponctuation et sans connecteurs logiques avec notamment une pluralité de narrateurs, de temporalités et de langues. Ce qui permet à ces récits de véhiculer un

²²¹ Nous faisons allusions à la Reconquista espagnole et les décrets de la « limpieza de sangre » (pureté du sang) ayant pour but la purification ethnique, morale et religieuse de l'Espagne chrétienne des communautés musulmanes et juives, mais aussi au génocide nazi.

²²² Edmond Amran EL MALEH, *op. cit.*, 2000, p. 54

message, de ne pas sombrer dans la prolixité, de ne pas se transformer en bruit, c'est bien la mémoire.

En fait, en parcourant l'œuvre d'El Maleh, le lecteur peut remarquer facilement que la mémoire tient une place privilégiée dans la configuration même du récit, au point que souvenirs et écriture se dissolvent pour former un tout solide et homogène. Cet auteur pleinement consommé et consumé par la recherche du passé, nous fait circuler dans les labyrinthes de la mémoire pour reconstituer un héritage culturel spécifique à sa terre-mère, ceci justifie également l'actualisation de substantifs relatifs aux champs sémantiques de la communauté, de la nature, de l'art culinaire... bref des substantifs qui servent à décrire une réalité quelconque.

Dans son ouvrage *la mémoire, l'histoire et l'oubli*, Paul Ricœur nous explique toute la complexité de l'acte de « faire mémoire »²²³ :

*« lui aussi a son ambition, sa revendication, sa prétention, celle de représenter en fidélité le passé. Or la phénoménologie de la mémoire, dès l'époque de Platon et d'Aristote, a proposé une clé d'interprétation du phénomène mnémonique, à savoir le pouvoir de la mémoire de rendre présente une chose absente survenue auparavant. Présence, absence, antériorité, représentation forment ainsi la toute première chaîne conceptuelle du discours de la mémoire. L'ambition de fidélité de la mémoire précéderait ainsi l'ambition de vérité de l'histoire dont il resterait à faire la théorie distincte. »*²²⁴

Ainsi, nous pourrions dire en ce qui concerne l'écriture d'El Maleh qu'il s'agit d'un acte de « faire mémoire ». Quoiqu'il refuse d'attribuer le caractère autobiographique à ses récits et déclare que son ambition n'est pas d'ordre historique, il

²²³ Paul RICOEUR, *La mémoire, L'Histoire, L'Oubli*, Paris, Editions du seuil, collection « l'ordre philosophique », 2000, p. 26.

²²⁴ Ibid., p.296.

serait difficile de ne pas déceler dans ses textes une certaine volonté d'amener le lecteur à *revivre le passé*. En effet, l'œuvre d'El Maleh reste imbue d'indices qui infirment ses dires et qui permettent au récit d'atteindre le but ultime qui le justifie, celui de reconstituer ce qui s'est réellement passé.

Déjà, en consultant la liste du vocabulaire spécifique, nous remarquons que la majorité des verbes relevés sont conjugués soit à l'imparfait soit au présent de narration, chose qui est tout à fait normale dans une œuvre qui cherche principalement à décrire le vécu d'une communauté à une époque révolue. Les faits relatés prennent ainsi plus de relief et la valeur durative de l'imparfait accentue l'effet dramatique des événements racontés. De même, l'utilisation du présent de narration donne l'illusion que des faits passés appartiennent au présent, comme s'ils se déroulaient ici et maintenant. Il crée alors une impression d'actualité et rend l'action plus vivante et plus proche de nous.

Soulignons que le recours à la mémoire explique, aussi, l'apparition des verbes qui font appel aux facultés de l'homme, tel que le montre le tableau suivant :

Tableau III. 33 : Verbes de réflexion

Écart	TLF	Corpus	Mots
30.69	12002	277	savait
20.72	21654	299	sait
20.21	16605	247	écrit
17.75	1850	59	souvient
16.02	9188	143	parlait
11.61	2945	55	nommer
10.97	3064	54	écoutait
10.46	11932	129	lire
9.56	15209	146	écrire
8.68	4997	63	connaissait
8.46	23380	190	savoir
8.31	21409	176	parle
8.28	28238	218	disait
7.62	38410	270	parler
7.36	12157	108	voyait

Par ailleurs, cette mémoire n'est pas toujours heureuse, elle est affectée d'une blessure profonde infligée par l'annihilation d'une existence séculaire de juifs marocains. Elle est affectée également par les épisodes malheureux qui ont jalonné l'histoire marocaine, tels le colonialisme et même quelques événements marquant de l'histoire internationale, comme « la Shoah » et les massacres perpétrés à l'encontre du peuple palestinien.

Nous sommes donc en droit de se demander pourquoi ce recours à la mémoire ? C'est bien évidemment parce que la mémoire est l'unique arme possible contre toutes formes d'usurpation et de falsification de l'histoire. Reconstituer une mémoire devenue

imaginaire, c'est reconstruire la **vérité**, c'est essayer de redonner sens et de réinsuffler la vie à ce qui a disparu.

Tableau III. 34 : Vérité

Écart	TLF	Corpus	Mots
24.54	10993	220	présence
13.47	5322	89	réalité
4.37	20922	135	vérité

D'ailleurs, derrière la résurrection du passé qui se lit en filigrane dans les récits d'El Maleh, il y a une quête de **vérité** et cette obsession pour la vérité provient d'« *un choc émotionnel* » ou la blessure secrète qui suscite en lui l'envie d'écrire :

« J'écris à partir d'un choc émotionnel qui déclenche l'écriture....., le choc est de sentir qu'il y a une sorte d'oubli qui se fait concernant les événements historiques du Maroc, que les personnages disparaissent..... »²²⁵.

Donc, ce recours à la mémoire traduit un engagement de la part d'El Maleh, celui de vouloir sauver des événements importants des méandres de **l'oubli**, d'où l'utilisation des substantifs contenus dans le tableau n° III.35 :

²²⁵ Marie REDONNET, *op.cit.*, 2005, p. 126.

Tableau III. 35 : Oubli

Écart	TLF	Corpus	Mots
20.09	2908	85	oubli
19.43	2632	78	néant
17.20	7193	129	absence
16.62	25515	290	silence

D'ailleurs, au fil de ses récits, l'auteur exprime un certain plaisir à avoir du recul sur son passé et sur celui de sa communauté afin de comprendre où et quand s'était faite « *la ligne de fracture* »²²⁶. Il refuse qu'on ensevelisse la communauté juive marocaine dans le passé comme si elle n'avait jamais existé. Ainsi, il poursuit son écriture mémorieuse et s'éternise dans son cheminement non seulement pour que les marocains n'oublient pas leur passé commun, mais aussi pour empêcher sa propre mémoire de défaillir.

7. Communauté

Le tableau ci-dessous montre que le monde dans l'œuvre d'El Maleh est caractérisé par le communautarisme : communauté (+25,27), gens (+9,63), amis (+6,93), êtres (5,01). L'excès de ces vocables ne signifie-il pas que l'auteur a un esprit de solidarité ?

Nous remarquons également que le mot homme(s) est excédentaire qu'il soit au singulier ou au pluriel. Cela ne voudrait-il pas dire que l'emploi de ce vocable fait allusion au rôle de l'homme dans le militantisme comme c'est le cas de la majorité des protagonistes d'El Maleh ? De même, la faiblesse de l'écart du vocable *homme* au

²²⁶ Edmond Amran EL MALEH, *op. cit.*, 2002, p.16

singulier par rapport à *hommes* au pluriel n'est-il pas dicté par le fait que le narrateur le remplace souvent par le nom propre qui le désigne ?

Tableau III. 36 : Communauté

Écart	TLF	Corpus	Mots
25.27	949	56	communauté
9.63	43337	326	gens
6.93	22318	168	amis
6.15	60410	369	hommes
5.44	23695	161	enfants
5.16	16082	115	jeunes
5.01	6538	56	êtres
4.82	42893	257	personne
4.53	6631	54	enfance
4.42	127370	671	homme
4.07	22575	141	famille
3.73	49649	276	mère

Une autre remarque s'impose c'est que la femme n'est excédentaire dans l'œuvre d'El Maleh que lorsqu'elle remplit le rôle de mère. Ceci est peut-être dû à l'attachement enfantin de l'auteur à sa maman. D'ailleurs, à maintes reprises, il ne cesse d'évoquer les

crises d'asthme dont il était victime et le soin particulier que sa mère lui procure. Ce qui est amusant, c'est que l'évocation du mot *mère* est souvent jumelé à des plats ou à des repas, même lors des crises d'asthme l'auteur souvent rappelle que la maman du protagoniste a toujours eu le soin de lui apporter un verre de thé et du pain beurré une fois que la crise est passée.

Tableau III. 37 : Famille

Écart	TLF	Corpus	Mots
-3,2	34812	115	femme
-4,02	62783	212	père
-10,31	50752	71	fille

Le tableau n° III.37 est surprenant en affichant un déficit des mots *femmes, fille et père*. Cependant, ce déficit n'est-il pas corrigé par l'usage excédentaire du mot *famille* qui affiche un écart réduit de + 4,06. Peut-être aux yeux de l'auteur le rôle de ses trois personnes n'est important que dans le cadre de la *famille* qui symbolise la solidarité, la procréation et la continuité de l'espèce. Ceci n'est-il pas confirmé par l'emploi excédentaire des mots : *vie, naissance, vivant* (tableau n° III.19) ? mais aussi du substantif *mère* qui incarne elle également la procréation ?

8. La notion de temps

En analysant le temps chez Giraudoux, Etienne Brunet stipule que :

«(le temps) n'est pas celui de l'« horloge interne », le temps vécu, le temps imprécis qui ne distingue que les instants et les moments. Le temps de Giraudoux se mesure au cadran (+6) solaire. Il suit le soleil dans sa course : aube, aurore, matin, midi, soir, minuit, nuit, jour, journée, tous ces aspects du jour dépassant de loin le seuil significatif. Il accompagne la terre dans le rythme plus large de ses mouvements saisonniers : semaine, mois, an, printemps, saison (...) les divisions les plus courtes et les plus précises appartiennent aussi au vocabulaire positif : heure, minute, seconde (...) Giraudoux ne rumine pas le temps, il butine et le goûte, sans arrière-pensée, sans inquiétude et sans parcimonie, sachant bien que le temps est le fonds qui manque le moins »²²⁷

Tableau III. 38 : Temps

Écart	TLF	Corpus	Mots
19.80	23000	302	instant
15.15	105797	798	temps
14.47	2773	63	date
12.96	24382	243	passé
12.77	7726	109	époque
12.37	47595	391	nuit
10.63	4872	71	siècles
10.20	13934	142	années
8.90	91550	586	jour
5.53	8701	73	veille

²²⁷ Etienne BRUNET, 1978, **op.cit.**, p.523

Pareillement à Giroudoux, la notion du temps est présente dans l'œuvre d'El Maleh de la division la plus courte jusqu'à la division la plus longue. Pour le cadran solaire, nous trouvons les substantifs *jour* et *nuît* qui incarnent la lumière et l'obscurité ainsi que la continuité. Pour les mouvements saisonniers nous relevons les mots suivants : *années, siècles, époque*. S'agissant des divisions courtes nous ne rencontrons que le mot *instant*.

Notons que notre auteur était très sensible à sa limitation par le temps et il se voyait comme une matière vouée à la disparition et à la déliquescence. D'ailleurs, en lisant ses premiers romans, nous avons l'impression qu'il entreprend une course contre les aiguilles d'une montre pour étaler, avant que la fin ne le surprenne, l'expérience d'une vie, d'un périple riche et sinueux par ses pérégrinations, par son militantisme pour la liberté du pays et l'affirmation de la dignité du citoyen marocain après l'indépendance. Par l'écriture, il brise les soucis que suscite l'idée de mort et tend vers l'éternité et, en même temps, il se réjouissait que sa production allât l'orienter vers la purification des souillures que l'exil lui a causées, en se réfugiant dans l'évocation lyrique de son enracinement immémorial dans la terre marocaine.

B. Analyse du vocabulaire spécifique de chaque texte : étude endogène

Si le point précédent nous a offert l'opportunité de dégager les préférences thématiques de tous le corpus comparé au TLFi, le présent point s'intéressera à l'étude de chaque texte par rapport à l'ensemble de l'œuvre d'El Maleh et ce dans le but de déceler le ou les thèmes qui pourraient en émerger. Signalons d'erechef que le logiciel *Hyperbase* effectue les calculs nécessaires et présente automatiquement le vocabulaire spécifique à chaque partition.

Il n'est pas sans intérêt de signaler qu'il serait difficile d'approcher l'œuvre d'El Maleh en dissociant ses textes les uns des autres. Cependant, pour braquer la lumière

sur l'ensemble des caractéristiques et des reliefs de ces partitions, rien ne nous empêche de regrouper les textes qui partagent la même thématique et qui hébergent les mêmes champs sémantiques.

Dans chacun des points qui suivent nous commencerons, tout d'abord, par dresser les listes du vocabulaire spécifique de chaque texte, avant de nous pencher sur leur analyse.

1. Parcours Immobile / Aïlen ou la nuit du récit / Mille ans, un jour / Le retour d'Abou El Haki

Commençons par les premières productions d'El Maleh qui sont toutes dominées par une question centrale, à savoir : la disparition de la communauté juive marocaine et les retombés de cette tragédie sur le devenir judéo-arabe. En effet, cette blessure profonde pousse El Maleh à s'interroger sur la raison qui a bousculé le sort de toute une communauté vivant en toute sérénité sur sa terre-mère. La recherche d'une réponse à cette question libère la méditation d'El Maleh qui, en revisitant son passé, étale tous les ingrédients d'une société vivant en harmonie. Il essaye, en fait, de réhabiliter la mémoire sociale et individuelle d'un peuple, fondamentalement interculturelle.

Lesdits textes partagent les champs sémantiques de la littérature, du militantisme, de la politique, de la religion juive, etc. Cependant dans chaque texte il y a un champ sémantique qui prime sur les autres.

Tableau III. 39 : Vocabulaire spécifique PARCOURS

Ecart	Corpus	PARC	Mots
26.5	283	205	aïssa
19.8	2052	556	était
19.8	158	118	parti
18.0	92	81	josua
12.2	1351	322	avait
11.4	59	44	communiste
10.5	44	35	camarades
10.5	26	26	houmrani
10.1	31	28	haroun
9.8	23	23	jacquet
9.0	50	33	casablanca
8.7	23	21	staline
7.8	5897	925	des
7.7	30	22	groupe
7.7	23	19	moscou
7.7	23	19	communistes
7.2	258	75	étaient
6.9	141	49	famille
6.8	71	32	ligne
6.8	357	91	histoire
6.8	28	19	camarade
6.7	1639	296	comme
6.4	22	16	militants

Ecart	Corpus	PARC	Mots
6.4	13	12	responsabilités
6.3	23	16	militante
6.2	307	78	allait
6.2	16	13	biographe
6.2	134	44	juif
6.1	136	44	politique
5.9	99	35	juste
5.9	20	14	heureusement
5.9	17	13	hram
5.9	17	13	clandestinité
5.8	66	27	immobile
5.6	86	31	simple
5.6	277	68	savait
5.6	19	13	secrétaire
5.6	122	39	peur
5.5	406	90	pouvait
5.4	195	52	français
5.4	10	9	oueld
5.3	6877	997	les
5.3	682	133	vie
5.3	15	11	responsable
5.2	31	16	militant
5.2	24	14	policrière

Ecart	Corpus	PARC	Mots
5.2	112	35	tard
5.1	86	29	soir
5.1	11	9	hmida
5.0	369	80	hommes
5.0	14	10	inspecteur
5.0	14	10	congrès
4.9	342	75	grand
4.9	337	74	jamais
4.9	2709	420	pas
4.8	21	12	flics
4.8	18	11	quarante
4.8	18	11	dirigeants
4.8	18	11	asthme
4.8	154	41	rue
4.8	15	10	safi
4.7	414	85	très
4.7	25	13	raconte
4.7	14567	1983	la
4.6	67	23	marocain
4.6	321	69	mots
4.6	22	12	mensonge
4.6	13	9	appareil
4.5	394	80	autres
4.5	341	72	peu
4.5	339	71	quelques

Ecart	Corpus	PARC	Mots
4.5	218	51	disait
4.5	11430	1571	le
4.4	90	27	libre
4.4	2636	401	ne
4.4	11	8	clandestin
4.3	58	20	angoisse
4.3	214	49	petit
4.3	156	39	chambre
4.2	26	12	midi
4.2	22	11	récitant
4.2	22	11	policier
4.2	126	33	surtout
4.2	12	8	ouvrière
4.2	12	8	lénine
4.1	39	15	école
4.1	16	9	flic
4.1	157	38	avaient
4.1	10	7	bandeau
4.0	59	19	honte
4.0	49	17	marge
4.0	373	73	monde
4.0	24	11	familiale
4.0	20	10	embark
4.0	13	8	technique
3.9	81	23	française

Ecart	Corpus	PARC	Mots
3.9	51	17	révolutionnaire
3.9	42	15	sage
3.9	42	15	partis
3.9	37	14	folle
3.9	29	12	tranquille
3.9	29	12	inlassablement
3.9	14	8	période
3.8	7275	1003	il
3.8	72	21	certitude
3.8	43	15	trouvait
3.8	329	65	venait
3.8	30	12	marocains
3.8	26	11	villa
3.8	26	11	chargé
3.8	208	45	mer
3.8	138	33	nouvelle
3.8	121	30	part
3.8	11	7	tracts
3.8	11	7	comprenait
3.8	108	28	voyait
3.7	90	24	venaient
3.7	73	21	venus
3.7	68	20	bureau
3.7	4984	701	dans
3.7	19	9	londres

Ecart	Corpus	PARC	Mots
3.7	1295	205	sans
3.7	12	7	fraternité
3.7	12	7	bateaux
3.7	105	27	parcours
3.6	9	6	moulin
3.6	9	6	marginalité
3.6	586	102	jour
3.6	41	14	récits
3.6	36	13	torture
3.6	28	11	carte
3.6	207	44	ailleurs
3.6	161	36	enfants
3.6	143	33	parlait
3.6	1234	194	où
3.5	958	154	être
3.5	3403	486	du
3.5	33	12	essaouira
3.5	24	10	national
3.5	20	9	action
3.5	13	7	tribus
3.5	13	7	lyautey
3.4	70	19	peau
3.4	671	112	homme
3.4	64	18	spectacle
3.4	63	18	date

Ecart	Corpus	PARC	Mots
3.4	39	13	cependant
3.4	268	52	grande
3.4	26	10	maroc
3.4	21	9	colonel
3.4	126	29	beaucoup
3.4	102	25	émotion
3.4	10	6	vacances
3.4	10	6	roches
3.4	10	6	crise
3.4	10	6	aîné
3.3	60	17	papier
3.3	56	16	dehors
3.3	56	16	creux
3.3	46	14	venue
3.3	46	14	samedi
3.3	40	13	millions
3.3	314	59	ça
3.3	31	11	écrivait
3.3	27	10	engagement
3.3	213	43	toujours
3.3	18	8	vigilance
3.3	14	7	responsables
3.3	11	6	postales
3.3	11	6	piano
3.3	11	6	nés

Ecart	Corpus	PARC	Mots
3.3	11	6	biographie
3.3	11	6	aïoua
3.2	42	13	fès
3.2	28	10	cas
3.2	15	7	voisins
3.2	15	7	rif
3.2	15	7	coton
3.2	15	7	asilah
3.2	108	25	importe
3.1	798	127	temps
3.1	54	15	coups
3.1	43	13	indépendance
3.1	34	11	nationale
3.1	33	11	important
3.1	33	11	franchir
3.1	317	58	yeux
3.1	24	9	coupée
3.1	20	8	soupçonner
3.1	20	8	fascination
3.1	20	8	espagnols
3.1	16	7	répondre
3.1	16	7	passaient
3.1	16	7	choeur
3.1	12	6	réunion
3.1	12	6	répression

Ecart	Corpus	PARC	Mots
3.1	12	6	docteur
3.1	12	6	découvrait
3.1	116	26	france
3.1	109	25	époque
3.0	9	5	tante
3.0	9	5	recueillait
3.0	9	5	pâque
3.0	9	5	importante

Ecart	Corpus	PARC	Mots
3.0	9	5	fermait
3.0	9	5	calotte
3.0	71	18	passait
3.0	30	10	ennemi
3.0	257	48	personne
3.0	25	9	saura
3.0	167	34	soleil
3.0	117	26	lieux

Dans *Parcours Immobile* l’auteur dresse le parcours de Aïssa le militant communiste qui lutte pour l’indépendance de son pays. Ce personnage découvre finalement qu’il soutient un parti qui ne sert que les propres intérêts de ses dirigeants qui sont en quête perpétuelle de pouvoir et de privilèges.

Ainsi, à travers les protagonistes de *Parcours Immobile*, l’auteur démystifie quelques idéologies, notamment communiste et sioniste qui, toutes les deux, exploitent les sentiments profonds des hommes, à savoir : le désir de liberté et d’identité. Pour faire passer toutes ses idées aux lecteurs, El Maleh mobilise un vocabulaire varié relevant des champs sémantiques de la politique (*parti, communiste, camarades, Staline, Moscou, communistes, histoire, camarade, responsabilités, politique, responsable, responsables, policière, inspecteur, flics, ouvrière, Lénine, policier, flic, colonel, Maroc, espagnols, dirigeants, soupçonner, France, ennemi*) et du militantisme (*militants, militante, clandestinité, clandestin, hommes, militant, révolutionnaire, national, bureau, nationale, répression, indépendance, réunion, engagement, vigilance, coups, tracts*).

Notons qu'à l'origine même de *Parcours immobile*, il y eut bien l'idée de présenter un témoignage sur l'expérience politique, mais ce projet fut vite abandonné par El Maleh :

« Au cours des années qui ont suivi la date de 1965, [...], la question s'est posée de savoir si je n'allais pas aborder une analyse critique et historique du Parti communiste marocain. De 1945 à 1959, date à laquelle je démissionnai et cessai toute activité politique, j'assumai des responsabilités à la direction du PCM en qualité de membre du bureau politique et du secrétariat, ayant en charge le plus souvent les questions relevant de la presse. [...] sur cette toile de fond, le parcours d'un jeune Juif marocain, issu d'une famille de négociants [...]. J'abandonnai assez vite ce projet d'un livre traitant de cette période et de cette expérience d'une activité militante au sein du PCM.²²⁸ »

De ce fait, l'idée de témoignage qui, nécessairement, revêtirait la forme d'une autobiographie, était abandonnée pour ne pas sombrer dans « *la sempiternelle autocritique de triste mémoire. J'ai voulu rompre catégoriquement avec cet héritage*²²⁹ ».

Cependant, quoique l'auteur refuse d'attribuer le caractère autobiographique à son premier roman, *Parcours immobile*, de nombreux indices évoquent la vie même d'El Maleh : appartenance à la communauté juive marocaine, membres de la famille, état de santé, études, amitiés, engagement politique, etc. Même les références historiques et spatiales coïncident (enfance et adolescence à Safi et à Essaouira, activités politiques à Casablanca), ce qui explique l'écart positif de certaines villes marocaines comme : *Essaouira, Asilah, Safi et Casablanca*.

²²⁸ Edmond Amran EL MALEH, « Préface » de l'auteur à l'occasion de la réédition de *Parcours immobile* en 2000, pp. 9-10.

²²⁹ Ibid., p. 11

Tableau III. 40 : Vocabulaire spécifique AILEN

Écart	Corpus	AILEN	Mots	Écart	Corpus	AILEN	Mots
15.4	55	55	jawad	5.5	9	9	fumé
13.5	60	53	baba	5.5	9	9	dikr
13.1	41	41	krimo	5.5	9	9	azzedine
12.9	40	40	mounir	5.5	88	34	calme
10.1	26	26	boujmaa	5.5	20	14	saïd
9.8	7275	1338	il	5.4	354	90	ciel
9.1	535	157	voix	5.4	2427	446	sur
8.5	26	23	aïlen	5.4	18	13	akbar
8.5	24	22	titi	5.4	165	51	vent
7.9	31	24	cahier	5.3	43	21	vas
7.8	25	21	mica	5.2	31	17	balles
7.6	43	28	gosses	5.2	14	11	bar
7.6	1351	296	avait	5.1	12	10	douar
7.3	112	48	terrasse	5.1	10	9	abderrahim
7.1	314	94	ça	5.0	624	136	rien
7.1	14	14	pepe	5.0	23	14	acier
7.0	118	48	allah	5.0	208	58	mer
6.2	5897	1019	des	4.9	358	87	ont
6.1	195	62	rouge	4.9	140	43	vieille
5.9	6877	1166	les	4.9	126	40	immense
5.9	18	14	pan	4.9	112	37	bord
5.8	10	10	téhéran	4.8	85	30	aimait
5.8	10	10	chicot	4.8	691	146	ils

Écart	Corpus	AILEN	Mots
4.7	243	63	passé
4.7	107	35	café
4.6	758	156	tu
4.6	20	12	mardochée
4.6	17	11	plastique
4.6	17	11	majdoub
4.6	17	11	abdelkrim
4.5	249	63	jours
4.5	12	9	têtu
4.4	118	36	haute
4.4	10	8	ramadan
4.4	10	8	haïk
4.3	36	16	dix
4.3	32	15	mouches
4.3	22	12	horloge
4.3	13	9	rive
4.2	60	22	salon
4.2	60	22	petites
4.2	60	22	maisons
4.2	37	16	hélas
4.2	33	15	prince
4.2	326	75	gens
4.2	317	74	coeur
4.2	11	8	simon
4.2	11	8	selham

Écart	Corpus	AILEN	Mots
4.2	11	8	haillons
4.2	11	8	dansait
4.1	9	7	parfumée
4.1	9	7	omar
4.1	9	7	mitsvah
4.1	9	7	meurtries
4.1	9	7	coiffeur
4.1	14	9	jouent
4.1	103	31	petits
4.0	51	19	poitrine
4.0	51	19	chiens
4.0	327	74	ville
4.0	28	13	fauteuil
4.0	24	12	fines
4.0	137	38	voulait
3.9	29	13	voyageurs
3.9	277	64	savait
3.9	25	12	matelas
3.9	25	12	ecoute
3.9	203	50	mains
3.9	115	33	femmes
3.8	76	24	cri
3.8	62	21	vagues
3.8	46	17	samedi
3.8	34	14	élégance

Écart	Corpus	AILEN	Mots
3.8	323	72	tête
3.8	30	13	siège
3.8	26	12	heureuse
3.8	19	10	noces
3.8	143	38	noir
3.8	13	8	veste
3.8	13	8	marabout
3.8	10	7	whisky
3.8	10	7	sardine
3.8	10	7	restés
3.7	60	20	salle
3.7	43	16	ya
3.7	20	10	terrasses
3.7	111	31	rire
3.6	70	22	pain
3.6	53	18	revenait
3.6	40	15	voiture
3.6	24	11	fermées
3.6	17	9	ordinaires
3.6	11	7	gosse
3.5	81	24	heures
3.5	77	23	regardait
3.5	49	17	colère
3.5	33	13	lait
3.5	25	11	inspiré

Écart	Corpus	AILEN	Mots
3.5	25	11	aveugles
3.5	2255	383	son
3.5	1930	332	lui
3.5	173	42	petite
3.5	1548	271	ses
3.5	121	32	assis
3.4	93	26	morts
3.4	88	25	noire
3.4	43	15	énorme
3.4	418	85	main
3.4	30	12	etat
3.4	26	11	oh
3.4	26	11	décor
3.4	19	9	mosquée
3.3	90	25	horizon
3.3	71	21	blancheur
3.3	61	19	pierres
3.3	53	17	vents
3.3	35	13	crête
3.3	31	12	indifférence
3.3	31	12	incertain
3.3	23	10	laine
3.3	16	8	riad
3.2	83	23	matin
3.2	73	21	venus

Écart	Corpus	AILEN	Mots
3.2	45	15	arrête
3.2	41	14	vague
3.2	344	71	lumière
3.2	32	12	délices
3.2	20	9	exigence
3.2	17	8	lunettes
3.2	17	8	arrêta
3.2	13	7	réussite
3.2	13	7	quartiers
3.2	122	31	oui
3.2	10	6	secrètes
3.2	10	6	revenus
3.2	10	6	pst
3.2	10	6	impasse
3.2	10	6	bouteilles
3.2	10	6	asséché
3.1	51	16	plis
3.1	51	16	douceur
3.1	46	15	allaient
3.1	33	12	grains
3.1	29	11	enveloppait
3.1	1849	312	sa

Écart	Corpus	AILEN	Mots
3.1	135	33	blanche
3.1	105	27	fête
3.0	92	24	sûr
3.0	52	16	pensait
3.0	52	16	idées
3.0	317	65	yeux
3.0	307	63	allait
3.0	30	11	sereine
3.0	30	11	lisse
3.0	30	11	anglais
3.0	26	10	souriait
3.0	26	10	faraji
3.0	22	9	cuir
3.0	18	8	rentrer
3.0	18	8	couteau
3.0	18	8	couscous
3.0	18	8	carton
3.0	11	6	roseau
3.0	11	6	moteur
3.0	11	6	glissait
3.0	11	6	chinois

Aïlen ou la nuit du récit abrite également un vocabulaire relatif au domaine militaire (*balles, meurtries, cri, colère, morts, couteau*), du moment que l'œuvre dresse une critique du système socio-politique marocain, mais cette fois-ci après l'indépendance. En effet, le texte met en évidence, à maintes reprises, les répressions militaires perpétuées contre les manifestants, notamment les affrontements sanglants de 1965 à Casablanca et ceux de 1981. Cependant les champs sémantiques prépondérants sont ceux qui se rapportent à la description de la misère sociale des marocains qui est dû essentiellement à la cruauté et le cynisme des détenteurs de pouvoir (arme, savoir, argent), mais également aux méfaits des arrivistes qui exploitent les paysans, les ouvriers, etc. La description que fait l'auteur convoque les champs sémantiques de la nature (*ciel, vent, fumé, mer, rive, horizon, pierres, vents, lumière, etc*), des lieux (*terrasse, douar, salon, maison, ville, etc*), de la religion (*allah, dikr, ramadan, mosquée*), des aliments (*sardine, pain, lait, grains, café, couscous, etc*), mais pour décrire le luxe dont jouit les arrivistes et les détenteurs de pouvoirs, nous trouvons des termes comme : *riad, délices, réussite, fête, rire, bouteilles, whisky, etc.*

Tableau III. 41 : Vocabulaire spécifique Mille

Écart	Corpus	MILLE	Mots	Écart	Corpus	MILLE	Mots
32.8	351	347	nessim	6.2	258	76	étaient
17.9	96	86	yeshuaa	6.2	11	11	yeshouaa
14.1	47	47	majid	6.2	11	11	legaye
11.1	2052	479	était	6.0	327	89	ville
10.7	29	29	kiji	5.8	29	18	synagogue
9.6	24	24	elhachmi	5.8	12	11	mule
9.5	1351	324	avait	5.8	10	10	pinhas
9.0	24	23	antoinette	5.7	77	32	regardait
8.7	31	26	commandant	5.7	326	87	gens
8.2	218	80	disait	5.6	317	84	yeux
8.2	18	18	hamad	5.5	9	9	sosie
8.0	24	21	louis	5.5	9	9	enchanteur
7.6	73	38	hamza	5.5	212	62	père
7.6	16	16	kaddish	5.5	157	50	avaient
7.6	16	16	isabelita	5.4	29	17	ange
7.5	6877	1214	les	5.4	11	10	thuya
7.5	1930	399	lui	5.4	11	10	telaviv
7.3	15	15	teddy	5.3	323	83	tête
6.9	7275	1264	il	5.3	105	37	juifs
6.8	13	13	nsismat	5.2	19	13	cousin
6.8	13	13	ari	5.1	10	9	zahra
6.5	22	17	ruben	5.1	10	9	khôl
6.4	33	21	rabbi	5.0	23	14	brûlé
6.3	1548	316	ses	5.0	183	53	enfant

Écart	Corpus	MILLE	Mots
4.9	62	25	sentait
4.9	47	21	avançait
4.9	13	10	mahia
4.8	16	11	youssef
4.8	16	11	jérusalem
4.7	9	8	hitler
4.7	9	8	asfi
4.7	9	8	amizmiz
4.7	108	35	voyait
4.6	54	22	boutique
4.6	390	91	visage
4.5	691	144	ils
4.5	314	76	ça
4.4	74	26	douleur
4.4	329	78	venait
4.4	15	10	cars
4.4	10	8	chleuh
4.3	55	21	quelqu'
4.2	243	60	devant
4.2	137	39	voulait
4.1	71	24	filles
4.1	27	13	beyrouth
4.1	14	9	vitre
4.0	80	26	visages
4.0	35	15	nez

Écart	Corpus	MILLE	Mots
4.0	31	14	the
4.0	18	10	observait
4.0	12	8	prier
4.0	12	8	joseph
3.9	88	27	regarde
3.9	70	23	doigts
3.9	69	23	lèvres
3.9	56	20	communauté
3.9	25	12	prières
3.9	161	42	enfants
3.9	15	9	sarah
3.9	120	34	paroles
3.8	67	22	avons
3.8	66	22	noirs
3.8	13	8	pèlerinage
3.8	114	32	table
3.8	10	7	etranger
3.7	355	77	maintenant
3.7	3443	576	se
3.7	31	13	donnait
3.7	23	11	montait
3.7	20	10	liban
3.7	134	36	faisait
3.6	36	14	ouvrait
3.6	249	57	jours

Écart	Corpus	MILLE	Mots
3.6	233	54	puis
3.6	173	43	petite
3.6	14	8	prénom
3.5	77	23	prière
3.5	50	17	foule
3.5	141	36	famille
3.4	9	6	palpait
3.4	9	6	cousines
3.4	9	6	arrivés
3.4	491	98	quand
3.4	47	16	horreur
3.4	30	12	appelait
3.4	2427	411	sur
3.4	22	10	rabbins
3.4	22	10	osait
3.4	19	9	juives
3.4	19	9	ignorait
3.4	19	9	faisaient
3.4	107	29	vu
3.3	95	26	milieu
3.3	91	25	bois
3.3	4984	803	dans
3.3	44	15	cherchait
3.3	406	83	pouvait
3.3	35	13	moha

Écart	Corpus	MILLE	Mots
3.3	35	13	miel
3.3	35	13	haine
3.3	342	72	grand
3.3	31	12	odeur
3.3	16	8	roulait
3.3	114	30	tenait
3.2	68	20	bureau
3.2	45	15	larmes
3.2	32	12	habillé
3.2	24	10	paupières
3.2	13	7	caïd
3.2	13	7	accouraient
3.2	10	6	accord
3.1	85	23	aimait
3.1	42	14	commençait
3.1	38	13	lève
3.1	38	13	affaires
3.1	29	11	juin
3.1	29	11	atlas
3.1	21	9	vin
3.1	14	7	savante
3.0	5897	934	des
3.0	43	14	pendant
3.0	34	12	souk
3.0	30	11	djellaba

Écart	Corpus	MILLE	Mots
3.0	277	58	savait
3.0	26	10	fqih
3.0	203	45	mains
3.0	18	8	of

Écart	Corpus	MILLE	Mots
3.0	11	6	dames
3.0	11	6	chaise
3.0	11	6	cérémonial

Mille ans, un jour reprend le drame de la communauté juive marocaine, cependant cette fois-ci, l’auteur jumelle cette thématique à la tragédie du peuple palestinien, il met le lecteur face à un double paradoxe ; en effet, tout au long de cet œuvre, il se demande comment une communauté qui a vécu sur la terre marocaine depuis plus de mille ans, a pu, en un jour, quitter son pays d’origine ? Parallèlement, il se demande comment une communauté qui a été victime de la shoah arrive-t-elle, un jour, à être le catalyseur du drame palestinien ?

Encore une fois ce constat active le désir d’El Maleh pour trouver une réponse à son interrogation, il se met donc, à travers son protagoniste principal Nessim, à revisiter le passé pour mettre en scène les liens intimes qui unissaient les juifs et la population arabo-berbère du Maroc. Pour ce faire, il fait appel à un vocabulaire, divers et varié, qui sert à peindre des scènes de la vie quotidienne. Ainsi, nous avons noté au niveau du vocabulaire spécifique excédentaire de *Mille ans, un jour*, l’existence des champs sémantiques de la communauté (*gens, foule, famille, communauté, enfants, fille, enfant, dames*), du corps humain (*visage, tête, doigts, lèvres, mains, paupières, visages, nez*) et de la religion (*prier, prières, pèlerinage, prière, fqih*). La pluralité ethnique du Maroc est aussi mise en évidence, en effet l’auteur fait allusion à la culture juive à travers les mots suivants : *synagogue, juifs, mahia, rabbins, juives, kaddish, rabbi*. La culture berbère est pareillement présente par l’utilisation de mots comme : *chleuh et atlas*. Même les choix onomastiques sont très révélateurs de la

diversité ethnique des protagonistes, nous trouvons ainsi des prénoms d'origine juive (*joseph, sarah, youssef, pinhas*), des prénoms arabes (*zahra, elhachmi, hamza, hamad*) et un prénom berbère (*moha*).

Les verbes dominants dans ce texte sont particulièrement des verbes d'action tels : *montait, donnait, venait, ouvrait, avançait, faisait, cherchait, roulait*. Notons que tous les verbes relevés dans le vocabulaire excédentaire de *Mille ans, un jour* sont conjugués à l'imparfait, ceci montre que l'auteur cherche à faire une description des événements dans le passé.

Soulignons que la majorité des souvenirs d'El Maleh se déroulent à Marrakech, Amizmiz, Safi, Essaouira, ce qui explique l'apparition de deux villes marocaines dans la liste du vocabulaire excédentaire, à savoir : *Amizmiz et asfi* qui, toutes les deux, ont enregistré un écart positif de + 4,7.

Nous avons signalé plus haut que l'auteur relie, dans *Mille ans, un jour*, la question des juifs marocains au drame palestinien. D'ailleurs, Nessim, son protagoniste principal, est tourmenté par les massacres de Sabra et Chatila qui sont, pour lui, liés à la disparition des juifs marocains qui ont quitté leur terre ancestrale pour rejoindre d'état d'Israël.

De même, il évoque très souvent la guerre civile du Liban, ceci justifie l'écart positif des noms des villes suivantes : *Jérusalem, Tel-Aviv, Beyrouth, Liban*, mais aussi l'utilisation du champ sémantique de la souffrance : *larmes, horreur, douleur, brûlé*.

Tableau III. 42 : Vocabulaire spécifique d’El HAKI

Ecart	Corpus	ELHAKI	Mots	Ecart	Corpus	ELHAKI	Mots
24.1	149	149	sofian	6.6	73	41	al
18.3	758	367	tu	6.5	85	45	vois
18.1	97	95	amine	6.1	16	15	poissons
15.0	65	65	bourhil	6.0	20	17	jellaba
14.9	64	64	borghès	6.0	13	13	hanoumat
14.4	70	67	ahmed	6.0	13	13	abdelaaziz
13.8	1385	501	je	5.9	38	25	haj
12.5	47	47	ismaël	5.6	246	90	moi
12.3	46	46	nezha	5.6	1307	358	ces
12.3	46	46	agar	5.5	11	11	zaïm
12.0	228	127	te	5.5	11	11	qaraouyine
9.7	30	30	andaloussi	5.5	11	11	merouani
9.1	99	62	as	5.5	11	11	hindi
8.6	183	90	sais	5.5	11	11	cigare
8.3	23	23	aboubaker	5.5	11	11	adolescente
8.1	22	22	ghazouli	5.3	15	13	serge
7.9	21	21	kadhafi	5.2	10	10	felouque
7.9	21	21	abderrehman	5.1	24	17	abou
7.7	445	163	me	5.1	167	64	ma
7.7	213	94	toi	4.7	133	52	vieux
7.2	34	27	moulay	4.6	76	34	tes
7.0	90	49	jardin	4.6	40	22	souviens
7.0	17	17	finie	4.5	71	32	dis
6.6	887	269	nous	4.5	23	15	sauterelles

Ecart	Corpus	ELHAKI	Mots	Ecart	Corpus	ELHAKI	Mots
4.5	16	12	université	4.0	16	11	tigre
4.4	42	22	chef	4.0	159	55	fut
4.4	34	19	entends	4.0	158	55	suis
4.4	31	18	monsieur	3.9	57	25	écoute
4.4	26	16	faraji	3.9	418	121	main
4.4	12	10	verras	3.9	24	14	adressait
4.3	35	19	amours	3.9	14	10	lalla
4.3	24	15	vendredi	3.8	10	8	pardonne
4.3	17	12	nasser	3.8	10	8	étudiants
4.2	79	33	étais	3.7	83	32	verre
4.2	22	14	esther	3.7	43	20	ya
4.2	22	14	caire	3.7	34	17	vais
4.2	119	45	thé	3.7	20	12	connais
4.1	9	8	ustad	3.7	1548	383	ses
4.1	9	8	chérie	3.7	15	10	kif
4.1	7274	1666	un	3.6	94	35	VRAI
4.1	45	22	aimé	3.6	88	33	cité
4.1	42	21	fès	3.6	71	28	cher
4.1	28	16	puits	3.6	63	26	fit
4.1	18	12	pouvais	3.6	41	19	sommeil
4.1	11	9	nil	3.6	38	18	longuement
4.1	11	9	caporal	3.6	13	9	surpris
4.0	74	31	coloniale	3.6	13	9	chine
4.0	298	92	notre	3.5	9	7	gardant
4.0	16	11	voleur	3.5	9	7	approchait

Ecart	Corpus	ELHAKI	Mots	Ecart	Corpus	ELHAKI	Mots
3.5	9	7	antoine	3.3	12	8	marchant
3.5	9	7	agitation	3.2	220	66	présence
3.5	391	111	nuit	3.2	18	10	battre
3.5	33	16	vint	3.2	116	39	es
3.5	33	16	mêlés	3.2	10	7	volupté
3.5	33	16	accueillir	3.2	10	7	ravir
3.5	299	88	sait	3.2	10	7	miséricorde
3.5	261	78	mon	3.2	10	7	humiliation
3.5	16	10	oasis	3.1	78	28	garde
3.4	228	69	pays	3.1	40	17	religion
3.4	14	9	indigène	3.1	13	8	naguère
3.3	71	27	respect	3.1	13	8	contrées
3.3	71	27	étranger	3.1	13	8	chèvres
3.3	58	23	reconnaître	3.1	123	40	retour
3.3	58	23	laissant	3.0	28	13	nouveaux
3.3	58	23	jeunesse	3.0	25	12	revenu
3.3	47	20	rouges	3.0	25	12	paraissait
3.3	32	15	armée	3.0	22	11	fumée
3.3	29	14	fût	3.0	19	10	métropole
3.3	23	12	reprenait	3.0	19	10	iblis
3.3	23	12	mouvements	3.0	19	10	caravane
3.3	20	11	faiblesse	3.0	128	41	âme
3.3	12	8	obscur	3.0	120	39	paroles
3.3	12	8	maréchal				

Le retour d'Abou El Haki s'inscrit dans la filiation de ces précédents, déjà l'œuvre commence par un « il » qui est donné comme s'il était connu du lecteur, ce qui laisse croire que la première phrase n'est que la suite d'une histoire qui a déjà commencé ailleurs. Ainsi, *Le Retour d'Abou El Haki* englobe, à son tour, presque tous les thèmes relevés dans les trois premiers textes : l'écriture, la politique, le colonialisme, l'enfance, l'effondrement stalino-communiste, l'Intifada, Beyrouth, etc.

Cependant, nous assistons à une mise en valeur de **la culture musulmane** ainsi que ses principes de tolérance et de refus de tyrannie qui contrastent avec la cruauté de certaines idéologies actuelles. Les représentants de ces valeurs, dans *Le Retour d'Abou El Haki*, sont des figures prestigieuses qui cheminent les lecteurs dans des espaces et des périodes pluriels et qui marquent le rôle de l'islam dans l'épanouissement des relations entre les diverses ethnies formant la société marocaine.

Le savoir est aussi mis en exergue, nous trouvons ainsi dans le vocabulaire excédentaire des mots comme : *université, qaraouyne, ustad, étudiants*, mais aussi des noms de lieux qui incarnent la renaissance spirituelle et religieuse dans le monde arabe : *Fès, Le Caire, l'Andalousie*.

Déjà les protagonistes d'El Maleh eux aussi s'affirment à travers un savoir séculaire. Nous trouvons par exemple la figure noble et originale du « qari tager²³⁰ » (érudit et négociant) qui est centrale dans *Le Retour d'Abou El Haki*. C'est un type de personnage qui s'inscrit dans un courant où le savoir n'est pas une fin en soi mais une quête spirituelle jamais achevée. Notons que dans le monde juif et musulman, cette

²³⁰ Incarnée à travers les personnages de Moulay Aboubaker Lasri dont la poésie possède une force d'enchantement originelle, ou bien Amine al-Andaloussi que l'auteur vénère comme un grand maître.

figure²³¹ a joué un rôle très important, dans la culture andalouse et maghrébine, à une époque où la circulation de l'information et de la science était souvent liée à celles des hommes et des biens.

2. Le café bleu, ZrIREK / Jean Genet, Le Captif amoureux et autres essais / Abner Abounour

Tableau III. 43 : Vocabulaire spécifique ZRIREK

Ecart	Corpus	ZRIREK	Mots	Ecart	Corpus	ZRIREK	Mots
17.1	433	216	écriture	9.4	39	33	angel
14.9	277	149	genet	9.3	53	39	mystique
13.4	3746	937	est	9.3	254	104	texte
12.8	7039	1581	et	9.2	267	107	langue
12.2	256	124	jean	9.0	55	39	walter
11.5	77	57	benjamin	8.5	80	47	allégorie
11.1	105	67	langage	8.5	2844	652	que
11.0	142	80	oeuvre	8.4	130	63	expérience
10.8	247	112	écrit	8.3	20	20	eddine
10.5	190	93	lecture	8.3	20	20	dufour
10.1	72	50	poétique	8.1	22	21	maleh
9.8	37	33	josé	8.0	24	22	arabi
9.7	79	51	juan	7.7	72	41	littéraire
9.6	121	66	pensée	7.7	64	38	littérature
9.5	100	58	poésie	7.7	109	53	amoureux

²³¹ Zafrani Haïm, Lieux de rencontre et de dialogue, in : Horizons Maghrébins - Le droit à la mémoire, N°27, 1994. Présences d'Edmond Amran El Maleh. p.p. 13-25

Ecart	Corpus	ZRIREK	Mots	Ecart	Corpus	ZRIREK	Mots
7.6	34	26	valente	5.6	70	32	auteur
7.6	22	20	khair	5.6	556	146	fait
7.5	17	17	san	5.6	41	23	création
7.5	17	17	barzakh	5.6	30	19	totalité
7.5	102	50	captif	5.6	10	10	ifergan
7.4	50	32	saurait	5.6	10	10	cruz
7.4	38	27	écrivain	5.5	23	16	kabbale
7.3	16	16	allégorique	5.5	136	50	serait
6.8	89	43	réalité	5.4	66	30	origine
6.7	78	39	poète	5.4	5885	1149	une
6.7	14	14	giordano	5.4	170	58	propre
6.6	3678	774	qui	5.3	47	24	langues
6.6	261	88	sens	5.2	65	29	proust
6.4	95	43	mystère	5.2	42	22	ibn
6.4	235	80	livre	5.2	27	17	révélation
6.2	2704	580	qu'	5.2	17	13	del
6.2	12	12	bruno	5.2	1118	256	ou
6.1	20	16	los	5.2	11	10	pertinence
6.0	65	32	ceci	5.1	95	37	images
6.0	38	23	interrogatio	5.1	73	31	pouvoirs
6.0	35	22	critique	5.1	487	127	mort
5.9	93	40	vivant	5.1	40	21	révèle
5.9	11	11	shahrâzâd	5.1	25	16	essence
5.8	267	84	ainsi	5.0	41	21	marocaine
5.8	135	51	vérité	5.0	2327	484	cette

Ecart	Corpus	ZRIREK	Mots	Ecart	Corpus	ZRIREK	Mots
5.0	18	13	irréductible	4.5	49	22	trouve
4.9	5614	1087	en	4.5	21	13	pratique
4.9	24	15	poème	4.5	134	44	celui
4.9	203	63	image	4.5	11	9	éditions
4.8	67	28	marocain	4.4	50	22	territoire
4.8	489	124	toute	4.4	47	21	palestiniens
4.8	30	17	maternelle	4.4	47	21	goytisololo
4.8	17	12	aliénation	4.4	44	20	tragique
4.8	135	46	univers	4.4	43	20	destructeur
4.8	123	43	unique	4.4	34	17	référence
4.7	34	18	palestinienne	4.4	1571	332	elle
4.7	28	16	symbolique	4.4	14	10	tentative
4.7	264	75	lieu	4.4	14	10	scientifique
4.7	2339	481	on	4.4	14	10	narrateur
4.7	15	11	soufisme	4.4	14	10	como
4.7	102	37	question	4.3	52	22	nature
4.6	88	33	laquelle	4.3	35	17	entreprise
4.6	8211	1543	à	4.3	31	16	témoignage
4.6	48	22	culture	4.3	22	13	dévoile
4.6	23	14	paradoxe	4.3	17	11	idéologie
4.6	227	66	faut	4.3	114	38	page
4.6	18	12	crâne	4.2	72	27	abord
4.6	18	12	aboulafia	4.2	49	21	agit
4.5	9	8	pluralité	4.2	32	16	évidence
4.5	9	8	auteurs	4.2	29	15	singularité

Ecart	Corpus	ZRIREK	Mots	Ecart	Corpus	ZRIREK	Mots
4.2	21563	3867	de	3.9	11	8	soufis
4.2	10	8	explorer	3.9	11	8	perspective
4.1	30	15	paravents	3.8	9	7	lo
4.1	18	11	sindbad	3.8	9	7	fondement
4.1	135	42	soi	3.8	9	7	croyance
4.1	131	41	tel	3.8	86	29	telle
4.0	38	17	logique	3.8	74	26	prend
4.0	34	16	lisant	3.8	398	97	parole
4.0	28	14	détruire	3.8	36	16	parlant
4.0	220	61	espace	3.8	3403	658	du
4.0	16	10	frappant	3.8	29	14	sceau
4.0	1461	304	même	3.8	23	12	feddayin
4.0	112	36	pourrait	3.8	17	10	duquel
3.9	979	212	y	3.8	106	34	tradition
3.9	89	30	autant	3.7	544	125	dont
3.9	53	21	expression	3.7	48	19	titre
3.9	42	18	précisément	3.7	48	19	destruction
3.9	395	97	non	3.7	30	14	acte
3.9	25	13	oeuvres	3.7	27	13	afin
3.9	25	13	constellation	3.7	24	12	ruskin
3.9	22	12	kabbalistes	3.7	24	12	réel
3.9	19	11	domaine	3.7	12	8	cécile
3.9	122	38	figure	3.7	12	8	affirme
3.9	11	8	zrirek	3.7	109	34	révolution
3.9	11	8	spécifique	3.6	35	15	rapport

Ecart	Corpus	ZRIREK	Mots	Ecart	Corpus	ZRIREK	Mots
3.6	31	14	funambule	3.5	121	36	forme
3.6	18	10	créatrice	3.4	93	29	penser
3.6	141	41	soit	3.4	60	21	tenter
3.6	116	35	tient	3.4	48	18	existence
3.6	107	33	delà	3.4	33	14	ressemblanc
3.6	10	7	impossibilité	3.4	26	12	aléatoire
3.6	10	7	disparition	3.4	23	11	circonstance
3.5	75	25	exemple	3.4	20	10	ouvrage
3.5	55	20	possible	3.4	20	10	impose
3.5	54	20	modernité	3.4	175	47	mémoire
3.5	40	16	signification	3.4	146	41	écrire
3.5	39	16	sinon	3.4	120	35	travail
3.5	325	79	el	3.4	11	7	una
3.5	32	14	pauvreté	3.4	11	7	mellila
3.5	29	13	imaginaire	3.4	11	7	mas
3.5	25	12	origines	3.4	11	7	interprétatio
3.5	22	11	élément	3.4	11	7	assimilation
3.5	22	11	apparente	3.4	11	7	ampleur
3.5	19	10	para	3.4	11	7	affirmation
3.5	19	10	meurt	3.3	753	161	autre
3.5	19	10	complexité	3.3	42	16	méditation
3.5	162	45	sorte	3.3	24	11	transcendan
3.5	13	8	verbe	3.3	24	11	couple
3.5	13	8	réduire	3.3	17	9	plénitude
3.5	13	8	consacré	3.3	17	9	hébraïque

Ecart	Corpus	ZRIREK	Mots	Ecart	Corpus	ZRIREK	Mots
3.3	17	9	dévoilement	3.2	9	6	fragmentaire
3.3	17	9	débat	3.2	67	22	mythe
3.3	17	9	advient	3.2	47	17	rupture
3.3	14	8	procès	3.2	47	17	proximité
3.3	14	8	panthers	3.2	32	13	voudrait
3.3	14	8	lezama	3.2	32	13	textes
3.3	14	8	idéologique				
3.3	14	8	esthétique				
3.3	126	36	esprit				
3.3	108	32	lettre				
3.2	9	6	rationnel				

Le café bleu, Zrerek affiche le nombre le plus élevé de mots spécifiques par rapport au corpus étudié, à savoir 247 occurrences. Nous retrouvons dans ce texte **la tragédie palestinienne** avec les termes : *Destruction, mort, tragique, destructeur, détruire, meurt, aliénation, palestinienne, palestiniens, pauvreté.*

Cependant, *Zrerek* présente un intérêt tout particulier pour **l'écriture** à travers l'usage des mots suivants : *langage, écriture, œuvre, écrit, lecture, poétique, pensée, poésie, texte, langue, allégorie, expérience, littéraire, littérature, écrivain, allégorique, poète, livre, critique, penser, auteur, création, langues, poème, page, verbe, méditation, mémoire, écrire, ouvrage, interprétation, textes, idéologique, éditions, esprit, esthétique, narrateur, idéologie, explorer, scientifique, logique, expression, œuvres, titre, exemple, signification, débat, procès, lisant, domaine, figure, image, créatrice, référence, auteurs, culture, lettre.*

Ce texte étale un ensemble de réflexions sur l'écriture, celle d'El Maleh et celle des autres : le romancier Juan Goytisolo, le poète José-Angel Valente, Jean Genet, Khaïr-Eddine et Nissaboury. En effet, dans *Le café bleu*, El Maleh a rassemblé les auteurs dont les œuvres se caractérisent par la force et la puissance d'une écriture allégorique. Il dresse ainsi des témoignages sur les enjeux de l'écriture, sur ce qu'elle représente pour celui qui en vit spirituellement.

A travers une série d'évocations, l'auteur puise dans un passé lointain pour renvoyer *l'écriture allégorique* vers ses propres origines, vers l'écriture des mystiques de l'ancienne tradition. Ainsi, El Maleh choisit de présenter des figures appartenant aux trois religions théologiques, tels : Abraham Aboulafia, Ibn Arabi et Juan de la Cruz. L'évocation de ces figures fait surgir dans le vocabulaire excédentaire des mots comme : *soufisme, mystique, symbolique, origine, réalité, pouvoirs, essence, univers, question, croyance, fondement, parole, imaginaire, origines, kabbale, kabbalistes complexité, existence, rationnel, soufis, sens, mythe, mystère*.

Tableau III. 44 : Vocabulaire spécifique GENET

Ecart	Corpus	GENET	Mots	Ecart	Corpus	GENET	Mots
16.3	277	127	genet	5.5	33	16	abdallah
13.0	256	102	jean	5.4	34	16	lisant
9.7	27	24	cervantès	5.4	133	36	trois
9.3	3746	563	est	5.3	9	8	gérone
8.1	31	22	dispute	5.3	9	8	eglise
8.1	190	60	lecture	5.3	43	18	fausse
8.0	254	71	texte	5.1	85	26	apparence
7.7	43	25	barcelone	5.1	261	55	sens
7.1	433	95	écriture	5.0	30	14	paravents
7.0	47	24	goytisoló	4.9	79	24	juan
6.9	55	26	métaphore	4.9	31	14	funambule
6.8	7039	903	et	4.9	14567	1683	la
6.7	29	18	moïse	4.8	24	12	ruskin
6.5	65	27	proust	4.8	109	29	révolution
6.5	102	35	captif	4.6	48	17	vingt
6.4	19	14	nahmanide	4.6	34	14	palestinienne
6.4	109	36	amoureux	4.5	80	23	allégorie
6.3	235	58	livre	4.5	27	12	afin
6.1	121	37	forme	4.5	23	11	feddayin
6.0	47	21	palestiniens	4.4	89	24	réalité
6.0	43	20	destructeur	4.4	20	10	graine
6.0	220	54	espace	4.4	1118	162	ou
5.8	14	11	nahman	4.3	2844	364	que
5.6	10	9	makbara	4.2	21	10	rembrandt

Ecart	Corpus	GENET	Mots	Ecart	Corpus	GENET	Mots
4.2	100	25	poésie	3.7	64	17	littérature
4.1	72	20	abord	3.7	26	10	sésame
4.1	31	12	algérie	3.7	229	42	mouvement
4.1	22	10	détruit	3.7	1478	196	mais
4.0	487	79	mort	3.6	72	18	poétique
4.0	3403	422	du	3.6	48	14	destruction
4.0	23	10	caractère	3.6	23	9	adressant
4.0	141	31	soit	3.6	218	40	jusqu'
4.0	1385	190	je	3.6	18	8	marcel
3.9	9	6	jose	3.5	29	10	condition
3.9	77	20	humaine	3.5	19	8	canon
3.9	70	19	auteur	3.5	15	7	déflagration
3.9	65	18	ceci	3.5	146	29	écrire
3.9	59	17	solitude	3.5	11	6	vases
3.9	271	49	parce	3.5	11	6	interruption
3.9	247	46	écrit	3.4	40	12	voudrais
3.9	114	26	page	3.3	90	20	laisser
3.9	107	25	delà	3.3	59	15	matière
3.8	85	21	façon	3.3	31	10	israël
3.8	3678	449	qui	3.2	9	5	présentée
3.8	17	8	messianisme	3.2	9	5	abjection
3.8	17	8	levinas	3.2	81	18	fil
3.8	17	8	blanchot	3.2	74	17	don
3.8	135	29	soi	3.2	62	15	ben
3.7	69	18	commencement	3.2	556	81	fait

Ecart	Corpus	GENET	Mots	Ecart	Corpus	GENET	Mots
3.2	49	13	vivants	3.1	18	7	pressent
3.2	38	11	hôtel	3.1	18	7	convention
3.2	33	10	ressemblance	3.1	18	7	artiste
3.2	32	10	lecteur	3.1	14	6	sexualité
3.2	242	41	ai	3.1	14	6	panthers
3.2	13	6	leçon	3.1	10	5	traduit
3.2	126	25	esprit	3.1	10	5	blin
3.1	34	10	métamorphose	3.0	35	10	refus
3.1	29	9	reprendre	3.0	24	8	réel
3.1	29	9	palmiers	3.0	24	8	faudra
3.1	29	9	formes	3.0	24	8	couple
3.1	28	9	blessure	3.0	19	7	lien

Jean Genet, Le Captif amoureux et autres essais, est également un groupement de textes traitant des écrivains dont l'évocation s'accompagne du plaisir de la méditation sur les vertus de l'écriture et de la littérature. Ainsi, nous pouvons repérer dans la liste du vocabulaire spécifique des mots qui renvoient au champ sémantique de **la littérature**, comme : *livre, lisant, allégorie, lecture, texte, écriture, littérature, métaphore, poésie, auteur, écrit, page, poétique, écrire, traduit, lecteur, esprit, leçon*.

Cet ouvrage révèle surtout le penchant d'El Maleh pour l'imitation du style d'autres écrivains tels : André Malraux, Paul Valéry, Abraham Aboulafia, mais c'est Jean Genet qui se taille la part du lion dans cette analyse. En fait, El Maleh conçoit l'écriture de Genet comme un double de sa propre écriture, il essaye de décrire la similarité qu'il perçoit entre lui et l'auteur du *Captif amoureux* en nous offrons une

série de réflexions portant sur cette ressemblance. Mais ce qui est intéressant, c'est que cette technique représente pour El Maleh la meilleure façon de décrire son propre style. Notons que l'analyse de l'écriture de Genet se base essentiellement sur son dernier livre qui est consacré aux feddayins en Palestine, c'est la raison pour laquelle nous trouvons dans le vocabulaire spécifique excédentaire des termes comme : *révolution, palestiniens, palestinienne, destruction, israël, destructeur, feddayin, makbara, détruit, mort, blessure, déflagration, canon, abjection, refus, humaine, réalité.*

Tableau III. 45 : Vocabulaire spécifique ABOUNOUR

Ecart	Corpus	ABOU	Mots
17.1	45	44	isso
12.9	28	27	abner
8.7	15	14	abdeslam
7.9	11	11	satrape
7.5	10	10	taksiste
7.1	11	10	imzoghen
6.0	13	9	elie
6.0	13	9	brahim
5.6	16	9	chroniqueur
4.8	59	14	matière
4.8	133	22	trois
4.8	13	7	monstrueux
4.7	88	17	cit��
4.6	10	6	bec
4.4	29	9	palmiers
4.2	32	9	d��
4.2	13	6	balustrade
4.0	15	6	observe
4.0	15	6	douze
3.9	16	6	h��te
3.9	10	5	hussein
3.8	7274	437	un
3.8	23	7	c��l��bres

Ecart	Corpus	ABOU	Mots
3.7	1307	96	ces
3.6	34	8	train
3.6	162	20	sorte
3.5	233	25	porte
3.5	167	20	place
3.5	14	5	conteur
3.4	9	4	v��randa
3.4	9	4	magiques
3.4	9	4	d��crets
3.4	21	6	tas
3.4	15	5	dossier
3.4	102	14	sable
3.3	23	6	ind��finissabl
3.3	195	21	rouge
3.3	16	5	��glise
3.3	16	5	advenir
3.2	84	12	pages
3.2	75	11	exemple
3.2	17	5	��tage
3.2	17	5	contraireme
3.2	134	16	heure
3.2	10	4	voitures
3.2	10	4	jnoun

Ecart	Corpus	ABOU	Mots	Ecart	Corpus	ABOU	Mots
3.2	10	4	gaz	3.1	11	4	dormir
3.1	78	11	dessus	3.0	729	55	bien
3.1	44	8	port	3.0	68	10	mur
3.1	11	4	yamine	3.0	12	4	cherkaoui
3.1	11	4	tranchant	3.0	12	4	carrés
3.1	11	4	reine	3.0	12	4	advenu

Abner Abounour : Il s'agit d'un ensemble de micro textes qui s'inscrivent dans la même logique des autres écrits d'El Maleh, celle d'offrir au lecteur, à travers l'écriture, « *des lambeaux de récits arrachés au silence*²³² ». La toile de fond de ces récits est la société marocaine. Tout au long de ce texte, l'auteur étale des figures représentatives telles : Elie le plombier, El Hchiouch qui conduit dans sa barque les touristes vers l'île aux abords d'Essaouira, Brahim qui tenait un bistro-restaurant, Antonia l'Andalouse « *la seule première sage-femme nasrania chrétienne* »²³³, Issou Imzoghen , croyant invétéré, qui se réjouit d'être considéré comme un chérif et dénommé comme de hauts dignitaires « *ya moulay* »²³⁴, etc. ces protagonistes, hommes ou femmes, jeunes ou âgés, incarnent des valeurs éthiques, ils ont le sens de l'équité, de la dignité et de l'humain. En défilant ces figures, El Maleh se présente comme le peintre d'une société riche et variée. Le tableau que dresse l'auteur fait surgir, au niveau du vocabulaire spécifique excédentaire, des noms de métier

²³² Edmond Amran EL MALEH, **Abner, Abounour**, Thionville/Casablanca, La Pensée sauvage/Le Fennec, 1995, p. 14

²³³ Ibid., p. 85

²³⁴ Ibid., p. 61

(taksiste, chroniqueur, conteur), des noms d'objets (*porte, place, véranda, mur, balustrade, voitures, train, matière*), etc.

Également dans ce texte les résonnances et les correspondances avec d'autres œuvres littéraires et philosophiques fonctionnent comme principe générateur du récit. En effet, les auteurs de prédilection qu'on rencontre dans la majorité des textes d'El Maleh sont aussi présents, soit de manière implicite ou explicite : Walter Benjamin, Juan Goytisolo, Jose Angel Valente, Carlos Fuentes, etc.

3. Lettres à moi-même / Une femme, une mère

Nous avons signalé au niveau de la deuxième partie que Guy Dugas a sévèrement critiqué une tendance qui caractérise la littérature judéo-maghrébine d'expression française et qu'il appelait « *l'écriture mémorieuse* »²³⁵. Selon Dugas, les auteurs judéo-maghrébins de notre époque seraient :

« si antévertis sur leur passé qu'ils ne parviennent parfois plus
...qu'à produire une œuvre répétitive, dramatiquement ankylosée autour de
quelques faits de mémoire »²³⁶.

Quant à El Maleh, il a publié deux œuvres dans le crépuscule de sa vie, à savoir : *Une femme, une mère* et *Lettres à moi-même* pour échapper justement à cette ankylose et éviter de sombrer dans la reproduction d'une œuvre répétitive, sans innovation. Notons que ces deux livres ont enregistré le vocabulaire spécifique le plus

²³⁵ Guy DUGAS, *op. cit.*, 1990, p. 62

²³⁶ Ibid., p. 135

faible par rapport aux autres partitions du corpus, mais ceci reste normal du moment que leurs étendues sont les plus modestes.

Tableau III. 46 : Vocabulaire spécifique de FEMME

Ecart	Corpus	FEMME	Mots
6.3	16	7	khalil
5.8	2327	62	cette
5.0	1571	43	elle
4.3	13	4	jaune
4.0	54	6	cimetière
4.0	292	13	vers
3.9	20	4	chaux
3.7	71	6	filles
3.7	25	4	chanson
3.6	261	11	maison
3.5	264	11	lieu
3.5	148	8	plaisir
3.4	118	7	hasard

Ecart	Corpus	FEMME	Mots
3.3	35	4	photo
3.3	130	7	selon
3.2	7274	114	un
3.2	69	5	juive
3.2	308	11	quelque
3.2	18	3	hauteurs
3.1	20	3	assise
3.1	182	8	long
3.0	79	5	certain
3.0	23	3	inscrite
3.0	22	3	arrêt
3.0	22	3	arbres

Une femme, une mère : Nous ne pouvons pas constituer des champs sémantiques riches et variés à travers le vocabulaire spécifique de *Une femme, une mère*, vu son nombre restreint. Cependant nous pouvons consulter quelques mots qui ont enregistrés un écart réduit considérable, comme : maison et cimetière qui constituent tous les deux des points dans une constellation de lieux de mémoire et de création dans ce texte et qui touchent directement à la mémoire judéo-marocaine

Certes la visite du cimetière est l'un des faits de mémoire les plus évoqués dans les textes d'El Maleh²³⁷. Néanmoins, elle est loin de constituer une rigidité dans l'œuvre maléhiennne, parce que cette visite est liée d'une manière dialectique à la maison de Lalla Saïdia, la mère de son fils spirituel, le peintre Khalil El Ghrib. De même elle est employée dans le contexte de sa découverte de la ville d'Asilah.

Contrairement au cimetière marin, la maison de Lalla Saïdia fait son apparition pour la première fois dans *Une femme, une mère*. El Maleh précise que cette maison, située au cœur de la médina, est l'ancienne demeure d'une famille juive, plus précisément d'un hazan²³⁸. Soulignons que cette insistance sur l'emplacement de la maison dans la médina s'oppose à l'idée fixe de certains écrivains et historiens qui décrivent des juifs enfermés dans des mellahs comme s'ils s'agissaient de ghettos européens. D'ailleurs, El Maleh insiste sur le fait que sa ville natale Safi ne disposait pas de Mellah : « *Le mellah n'existait pas il n'en aura jamais l'expérience* » (2000 : 212).

De même, l'auteur ne cachait pas la jouissance mémorielle qu'il ressent quand il rend visite à madame Laghrib :

« Il ne pouvait manquer de trouver fort émouvant qu'elle prête le son de sa voix à une chanson juive dont elle restituait d'une voix incertaine quelques bribes qui, au-delà de sa signification somme toute banale, réveillait, en cette demeure d'un hazan, rendue à elle-même pour un instant,

²³⁷ El Maleh évoque sa découverte du cimetière juif d'Asilah pour la première fois dans son premier livre *Parcours Immobile*. Par la suite ce cimetière ancien, perché sur la côte, et laissé à l'abandon après le décès du dernier juif d'Asilah, est évoqué au moins six fois dans l'œuvre maléhiennne.

²³⁸ Chantre à la synagogue.

des échos lointains, le frémissement d'une vie qui ne s'était pas encore effacée de ces lieux d'exception »²³⁹

Ainsi, nous avons l'impression que l'auteur trouve cette rencontre fort émouvante au point qu'il arrive à toucher le passé zaïlachi à travers la voix et les paroles de Lalla Saïdia. Cependant, nous remarquons, dans la citation plus haut, que l'auteur qualifie la maison de la mère de « lieu d'exception », plus tard dans le même texte cette qualité est aussi attribuée au cimetière juif :

« Son regard s'échappait, courait par-delà les remparts vers ce cimetière juif...lieu d'exception où par hasard s'est nouée une alliance indéfectible »²⁴⁰

Ce lieu d'exception sera le catalyseur d'une dialectique de vie et de mort dans la prose d'El Maleh. En effet, l'appellation du cimetière juif en hébreu, « beit ha-olam » maison d'éternité ou « beit ha-haïm » maison de vie, trouve des échos dans ledit texte :

« Un cimetière vivant, un cimetière mort, car plus personne n'y sera enterré »²⁴¹

²³⁹ Une femme, pp. 6-7

²⁴⁰ Ibid., p. 7

²⁴¹ Ibid., p. 8

Ainsi, le cimetière reste « vivant » parce qu'il continue à témoigner de la présence juive. Mais comme il n'y a personne pour venir prier sur les tombes et personne n'y sera enterré, le cimetière devient mort, au sens figuratif.

Nous relevons également le vocable *Chaux* dans la liste du vocabulaire spécifique de *Un femme, une mère*. D'ailleurs, ceci n'est pas surprenant puisque la ville visitée par El Maleh, Asilah, est représentée physiquement par la chaux (l'un des supports privilégiés par Khalil El Ghrib) et symboliquement par l'hospitalité et le savoir-faire. Aussi, l'évocation de cette matière réveille un art du passé, au sein duquel les juifs occupaient une place distinctive. D'ailleurs, El Maleh lui-même rappelle l'une des méthodes de préparation de cette matière :

« On y ajoute la fameuse nila, ce bleu qui s'incorpore merveilleusement et parfois aussi, chez les juifs surtout, la moghra, une poudre minérale de couleur rouge ou orange »²⁴².

En somme, ledit texte est articulé par trois éléments : la découverte de la ville d'Asilah, le témoignage de la présence juive éveillé par une musulmane et le témoignage plus direct mais aussi puissant du lieu en soi. Choses qui attribuent aux lieux évoqués une dimension exceptionnelle précisément parce qu'ils détiennent la mémoire juive, la mémoire musulmane et la mémoire personnelle de l'écrivain.

²⁴² Ibid., p. 9

Tableau III. 47 : Vocabulaire spécifique LETTRES

Ecart	Corpus	LETTRES	Mots	Ecart	Corpus	LETTRES	Mots
32.8	1408	377	vous	4.6	30	9	formule
12.0	79	38	avez	4.6	17	7	institution
11.2	31	24	collège	4.6	11	6	possibilité
10.7	1385	152	je	4.6	11	6	crs
10.0	145	41	votre	4.5	18	7	romanesque
8.5	12	12	mona	4.5	146	20	écrire
7.8	445	60	me	4.4	50	11	roman
7.2	5614	357	en	4.4	13	6	aimerais
7.0	73	21	êtes	4.2	9	5	philosophie
6.8	729	75	bien	4.2	183	22	sais
6.6	2844	200	que	4.1	38	9	personnages
6.3	1478	118	mais	4.1	111	16	souvent
5.7	100	20	vos	4.0	71	12	cher
5.4	15	8	mai	4.0	17	6	agents
5.3	242	31	ai	4.0	11	5	elias
5.1	23	9	philosophe	4.0	11	5	canetti
5.1	12	7	élèves	3.9	53	10	demande
5.0	9	6	professeurs	3.9	18	6	enseignemen t
5.0	9	6	aviez	3.7	29	7	ayant
4.9	40	11	classe	3.7	21	6	secrets
4.8	79	15	certain	3.6	73	11	arrivait
4.7	35	10	intérêt	3.5	9	4	étiez
4.6	3441	206	ce				

Ecart	Corpus	LETTRES	Mots	Ecart	Corpus	LETTRES	Mots
3.5	88	12	sentiment	3.3	1376	85	tout
3.5	8211	418	à	3.3	11	4	séculaire
3.5	74	11	situation	3.3	11	4	masse
3.5	63	10	fit	3.3	11	4	dise
3.5	2636	151	ne	3.3	11	4	admirablem ent
3.5	24	6	toutefois	3.2	39	7	divin
3.5	24	6	joué	3.2	20	5	sortes
3.4	887	60	nous	3.2	20	5	écheveau
3.4	79	11	manière	3.2	142	15	années
3.4	7039	362	et	3.2	141	15	lettres
3.4	44	8	entrer	3.2	1295	80	sans
3.4	298	26	notre	3.2	12	4	réunion
3.4	2339	135	on	3.1	598	42	dire
3.4	200	20	doute	3.1	30	6	masque
3.4	144	16	enfin	3.1	21	5	garder
3.4	10	4	divan	3.1	21	5	certaines
3.3	94	12	amitié	3.1	103	12	secret
3.3	26	6	corresponda nce	3.0	54	8	exil
3.3	18	5	allez	3.0	43	7	vif
3.3	149	16	trop	3.0	32	6	lecteur
3.3	148	16	plaisir				

Lettres à moi-même : Contrairement aux autres textes d'El Maleh qui parlent du Maroc ou du Moyen Orient, *Lettres à moi-même* s'intéresse principalement à l'arrivée de l'auteur en France en 1965. Ceci ne veut pas dire que le Maroc est absent

de ce texte mais il reste dans l'arrière-plan des lettres en tant que pays natal de l'auteur. En effet, le texte revient souvent sur des événements relatifs au Maroc, comme l'enlèvement de Ben Barka en 1965, le retour de Mohammed V en 1955, la révolte de 1968, la répression sanglante de la manifestation estudiantine à Casablanca en 1965, etc. le Maroc est aussi présent à travers la remémoration de souvenirs personnels, par exemple, lorsque l'auteur décrit sa première entrée comme professeur dans une salle de classe parisienne, il songe à un moment similaire au Maroc. Déjà nous repérons dans la liste du vocabulaire spécifique de nombreux vocables qui renvoient à l'enseignement : *Collège, élèves, professeurs, enseignement, formule, institution, écrire, situation, manière, classe, lecteur, correspondance, lettres, philosophie, doute, philosophe, romanesque, roman, personnages, sentiment.*

Notons que l'auteur éprouve un intérêt tout particulier pour les études et le savoir en général. D'ailleurs il fait allusion à cette idée à maintes reprises dans ces textes. Prenons une citation de *Mille ans, un jour* très explicite dans ce sens :

« yabni, khas teqra ! il te faut apprendre, étudier, pour nous
les juifs, c'est notre seule richesse. »²⁴³

Cette phrase résume l'image qu'El Maleh cherche à donner au savoir : « notre seule richesse ». D'ailleurs si nous revenons à la liste du vocabulaire excédentaire globale nous trouvons que le mot savoir est apparu dans le corpus 190 fois avec un écart réduit de 8,46. Ceci nous rappelle également ce que nous avons dit à propos des juifs au niveau de la première partie concernant le fait qu'ils excellaient dans les langues et même dans les métiers qu'ils exerçaient.

²⁴³ *Mille ans un jour*, p. 144.

Avant d'achever cette analyse, nous précisons que l'une des questions récurrentes qu'El Maleh cherche à élucider dans *Lettres à moi-même* est la nature de son départ en France. D'ailleurs, en lisant ledit texte nous décelons une insistance frappante sur la nature volontaire de ce voyage tout en refusant de lui attribuer le caractère d'un exil. Mais cette idée n'est-il pas un peu forcé si nous nous fions à liste du vocabulaire spécifique ? En effet même si l'auteur résiste à ce vocable pour décrire son départ du Maroc, nous repérons dans ladite liste le mot *exil* avec un écart réduit de + 3, ce qui nous laisse perplexe face à l'idée évoquée par El Maleh.

Au terme de ce deuxième chapitre, consacré à l'étude des spécificités lexicales de l'œuvre d'Edmond Amran El Maleh, nous avons abordé le vocabulaire caractéristique suivant deux logiques. La première de nature exogène a mis l'accent sur les spécificités externes du corpus global, en considérant le T.L.F comme norme et base de calcul. La deuxième de nature endogène a pris en charge les spécificités de chaque texte par rapport au corpus entier.

Ainsi, nous avons découvert un attachement fort de l'auteur à ces ***racines judéo-marocaines***. En effet, la fréquence élevée du substantif ***juif***, que ce soit au singulier ou au pluriel, montre que cette communauté occupe une place prépondérante dans le raisonnement d'El Maleh. Nous le trouvons également excédentaire dans la liste des adjectifs. Ce constat est justifié également par l'écart réduit élevé du vocable ***marocains***.

L'intérêt d'El Maleh pour les origines revêt une signification beaucoup plus profonde, c'est à vrai dire une condamnation de l'exil, un appât qui a séduit des milliers de ces coreligionnaires marocains qui ont troqués leur pays, leur histoire, leur liberté, contre un ghetto qui n'est pas les leur.

Nous remarquons également que l’auteur a un intérêt très profond pour le thème de la *littérature*, parce que dans l’optique d’El Maleh seule la littérature peut restituer une certaine vérité de l’histoire, c’est elle seule qui peut transmettre un héritage fabuleux et rétablir une vie archéologique dans *ses joies et ses souffrances*. L’évocation des vertus de la littérature le ramène naturellement à s’exprimer sur la lecture et l’écriture, d’ailleurs l’indice le plus important que nous avons relevé dans le champ lexical de la littérature est celui du mot *écriture*. Ceci n’est pas étonnant du moment que l’écriture est l’essence même de l’entreprise maléhiennne qui cherche la reconstitution d’une vie dans ses multiples voix, dans ses différentes facettes. Parmi les stratégies qui lui ont servi pour atteindre son objectif, il y a le procédé de *l’allégorie*²⁴⁴. Ce dernier est formé de trois composantes qui se compénètrent et se correspondent mutuellement, à savoir : l’image, le monde et le langage.

Ainsi, tout au long de son œuvre, il a manié ses trois composantes en mobilisant des champs sémantiques divers et variés, par exemple, nous trouvons des substantifs relatifs au *langage*. De même, pour décrire le monde, l’auteur avait recours aux champs sémantiques de *la nature, des parties du corps, des lieux et des habitations*, etc. Mais l’essence même de l’allégorie, réside surtout dans sa représentation picturale, d’où l’utilisation de substantifs relatifs à *la peinture*. Ainsi, le recours à l’écriture allégorique assure une certaine complémentarité entre le pictural et le scriptural mais aussi entre le texte et le monde. De plus, elle fait ressurgir la gravité de l’exil, la cruauté de l’exode, la fatalité de l’oubli. De même, elle récupère l’histoire du Maroc, revisite sa diversité culturelle, dessine le parcours tortueux d’un homme asthmatique et d’un militant politique.

²⁴⁴ Rappelons que le vocable *allégorie* a aussi enregistré un écart réduit très important au niveau du vocabulaire spécifique.

Aussi, à travers le thème de **la mémoire**, l'auteur assure la transmission des événements passés, tout en se situant dans le sillage de **l'histoire**. Cependant, lorsque des événements réels sont introduits dans les récits maléhiens, ils sont enrobés de brumes de fiction, ils ne s'inscrivent plus dans l'« historique » mais ils appartiennent plutôt à l'univers du narrateur. Ainsi, en usant des champs sémantiques de **la politique, de la guerre, du militantisme** mais aussi de **la fiction**, il évoque des événements nationaux et internationaux comme : la guerre mondiale, le protectorat français, l'exil du roi, la résistance nationale, l'indépendance du Maroc, la guerre du Liban, la guerre civile de l'Espagne, le communisme, la création de l'état d'Israël qui est corollaire à la destruction de la Palestine.

Déjà en lisant l'œuvre d'El Maleh, il est facile de déceler chez l'auteur une certaine volonté de vouloir amener le lecteur à **revivre le passé**. D'ailleurs, en consultant la liste du vocabulaire spécifique, nous remarquons que la majorité des verbes relevés sont conjugués soit à l'imparfait soit au présent de narration, chose qui est tout à fait normale dans une œuvre qui cherche principalement à décrire le vécu d'une communauté à une époque révolue. Les faits relatés prennent ainsi plus de relief et la valeur durative de l'imparfait accentue l'effet dramatique des événements racontés. De même, l'utilisation du présent de narration donne l'illusion que des faits passés appartiennent au présent, comme s'ils se déroulaient ici et maintenant. Il crée alors une impression d'actualité et rend l'action plus vivante et plus proche de nous.

Soulignons que le recours à la mémoire explique, aussi, l'apparition des verbes qui font appel aux **facultés de l'homme**, tel : *savoir, reconnaître, se souvenir, connaître, lire, écrire, etc.*

Ce recours à la mémoire est justifié par la volonté de l'écrivain de vouloir restituer un passé devenu imaginaire. De ce fait, derrière la résurrection du passé qui se lit en

filigrane dans les récits d'El Maleh, il y a une quête de *vérité*, il y a le désir de sauver des événements importants des méandres de *l'oubli*, surtout ceux qui se rapportent à la communauté marocaine juive.

De même, il ressort de l'examen minutieux du corpus étudié que l'auteur inscrit son œuvre dans la littérature de quête, une *quête d'identité*. Dans ses récits, la recherche de l'identité se trouve métaphorisée par le motif du voyage. Ce qui peut expliquer l'émergence, dans la liste du vocabulaire spécifique excédentaire, des *verbes de déplacement* et des substantifs relatifs au *voyage*.

Au cours de ladite analyse, nous avons signalé également que le vocable *homme* qui symbolise aux yeux de l'auteur le patriotisme et le militantisme se présente avec un écart réduit élevé, contrairement à la *femme* qui n'est excédentaire dans l'œuvre d'El Maleh que lorsqu'elle remplit le rôle de mère. Ceci est peut-être dû à l'attachement enfantin de l'auteur à sa mère. Dans le même ordre d'idées, nous avons décelé un déficit des mots *femmes, fille et père*. Par ailleurs, nous trouvons que ce déficit est corrigé par l'usage excédentaire du mot *famille* qui affiche un écart réduit élevé. Peut-être aux yeux de l'auteur le rôle de ses trois personnes n'est important que dans le cadre de la *famille* qui symbolise la solidarité, la procréation et la continuité de l'espèce.

La notion du temps est aussi présente dans l'œuvre d'El Maleh de la division la plus courte jusqu'à la division la plus longue, à travers des substantifs comme : *instant, jour, nuit, années, siècles, époque*.

Nous avons aussi souligné que l'art culinaire articulent les textes d'El Maleh de manière régulière. Ce qui reflète l'intérêt que l'auteur accorde à ce domaine.

S'agissant de l'étude endogène nous avons regroupé les textes qui hébergent plus ou moins les mêmes thèmes. Ainsi, nous avons obtenu trois blocs. Le premier qui est constitué par les premières productions d'El Maleh est surtout dominé par une question centrale, à savoir : la disparition de la communauté juive marocaine et les retombés de cette tragédie sur le devenir judéo-arabe. Lesdits textes ont soulevé une prépondérance des thèmes de : la littérature, le militantisme, la politique, la religion juive, la misère sociale, la pluralité ethnique, le savoir, etc.

Dans le deuxième bloc nous retrouvons la tragédie palestinienne avec les champs sémantiques de ***la mort*** et ***la souffrance***. Cependant la réflexion sur ***l'écriture*** et ***la littérature*** est la plus développée ; en cela, El Maleh se rapproche d'écrivains qu'il convoque d'ailleurs abondamment dans son texte, comme : Borges, E. Jabès, W. Benjamin, M. Blanchot, M. Proust, E. Canetti, etc. Le troisième bloc s'écarte légèrement des autres textes en abordant par exemple des thèmes comme l'enseignement et l'exil.

CHAPITRE 3

Plurilinguisme dans l'œuvre d'EDM

Dans les chapitres précédents, nous avons approché l'écriture maléhiennne au point de vue de la langue française. Cependant, dans notre corpus, il ne s'agit pas d'une seule langue, mais plutôt d'une pluralité de langues. D'ailleurs, l'auteur n'hésite pas à introduire, dans son œuvre, des mots qui ne relèvent pas du français et qui infusent au texte une charge imaginative et une sonorité affective propre au Maroc. En effet, le texte d'El Maleh devient quelques fois une mosaïque, où plusieurs langues se mêlent, il forme un enchevêtrement harmonieux entre l'arabe classique, l'arabe dialectal, l'amazigh et le judéo-arabe..., langues qui entretiennent dans l'œuvre d'El Maleh un rapport interne équilibré et cohérent. Notons que de telles situations, celle de pluralité linguistique, présupposent des interactions forcément culturelles.

Donc, il serait aberrant d'approcher le lexique d'El Maleh sans prendre en considération cette dimension plurilingue qui caractérise son écriture. D'autant plus, ce choix nous serait d'une grande importance, du moment qu'il nous servira, de plus en plus, à nous rapprocher de notre but ultime, celui de répondre à la problématique que nous nous sommes fixée au début de ce travail de recherche, à savoir : « ***comment Edmond Amran El Maleh a pu faire de son œuvre, à travers la langue d'écriture et les unités lexicales employées, un outil pour la redécouverte de la mémoire, de l'histoire et de la culture marocaine.*** »

Par ailleurs, le plurilinguisme suggéré par El Maleh nous met face à une situation de contact de langues²⁴⁵ qui, à son tour fait surgir des phénomènes linguistiques divers et variés. Il s'agit, en effet, de l'emprunt, du calque, de l'alternance codique et de l'interférence. De ce fait, la mission qui nous incombe, ici, est de décrire et comprendre le mode de fonctionnement des phénomènes linguistiques sus-cités et, de déceler leur contribution dans la mise en évidence de la mémoire, l'histoire et la culture marocaine.

Pour ce faire, nous commencerons la première section par approcher le phénomène de l'emprunt tout en nous intéressant à ses différentes manifestations. Dans la deuxième section, il sera question du calque linguistique, où nous identifierons ses différents types ainsi que ses spécificités par rapport à l'écriture d'El Maleh. Dans la troisième section, nous tenterons d'approcher le phénomène de l'alternance codique telle qu'il apparaît dans le contexte étudié, et ce par la mise en valeur des différentes langues qui participent à sa réalisation. Finalement, dans la quatrième section, il s'agira de présenter les manifestations de l'interférence phonétique dans l'œuvre d'El Maleh. Tous ces phénomènes linguistiques seront abordés dans l'optique de comprendre leur rôle dans la transmission d'une charge culturelle aussi originelle que la culture marocaine.

A. Emprunts

Multiples sont les linguistes qui se sont intéressés au phénomène de l'emprunt, d'où la profusion des définitions accordées à ce processus. Parmi l'une des plus

²⁴⁵ Notons que la situation de contact de langues dont nous parlons, ici, est loin d'être ordinaire, parce qu'elle ne relève pas du domaine de la vie quotidienne, mais plutôt d'une situation où la langue relève de la création littéraire.

anciennes définitions, qui fait encore référence, il y a celle de Louis Deroy²⁴⁶ pour qui : « *l'emprunt est une forme d'expression qu'une communauté linguistique reçoit d'une autre communauté*²⁴⁷ ». Il existe d'autres définitions plus complexes et plus détaillées mais qui ont tendance à intégrer d'autres variables, par exemples, il y a celle de Hamers²⁴⁸ qui met l'accent dans sa définition sur le critère de distinction entre deux phénomènes à savoir : l'emprunt et l'interférence.

Pour ce qui est de notre étude, nous nous limitons à la conception de Louis Deroy car elle est, nous semble-t-il, pertinente dans la mesure où elle résume d'une manière générale ce en quoi consiste le phénomène de l'emprunt sans, pour autant, insister sur les opérations qui le sous-tendent.

Ainsi, dans le cadre de ce travail, la notion d'emprunt renverra à tout mot appartenant à une langue différente de celle de l'énoncé. Autrement dit, l'analyse de l'emprunt sera centrée sur son acception la plus simple, la plus étroite, qui renvoie au « *seul emprunt de mot ou emprunt lexical [...], le plus fréquent, le plus apparent, le plus largement connu*²⁴⁹ ». Ceci, pour ne pas chevaucher sur les autres manifestations linguistiques que nous aborderons *infra*, à savoir, le calque, l'alternance codique et l'interférence. Nous étudierons ces phénomènes successivement, chacun dans un volet d'analyse à part.

De ce fait, nous nous pencherons sur l'étude des emprunts tels qu'ils se réalisent dans les textes d'El Maleh. Pour cette fin, nous avons constitué un inventaire pour

²⁴⁶ Notons que Deroy, lui même, s'est inspiré de Vittore Pisani en traduisant sa propre conception de l'emprunt.

²⁴⁷ Louis DEROY, *L'emprunt linguistique*, Paris, Les Belles Lettres, 1956, p. 18

²⁴⁸ Josiane F. HAMERS, « Emprunt », In Marie Louis MOREAU, *Sociolinguistique : Concepts de base*, Bruxelles, Mardaga, 1997, p. 136

²⁴⁹ Louis DEROY, *op.cit.*, 1956, p. 21

relever leurs origines linguistiques et leurs principaux champs sémantiques. Par ailleurs, nous examinerons la nature de leurs emplois dans la langue preneuse afin de dresser une typologie qui retrace les contours de ce phénomène dans l'écriture littéraire d'El Maleh.

1. Origines linguistiques et catégories grammaticales des emprunts

L'analyse de l'emprunt suppose qu'on soit en mesure de le reconnaître, d'où l'utilité d'examiner son origine linguistique. Après avoir soumis l'ensemble des emprunts relevés aux différentes opérations de sélection et de classement, nous avons pu tirer les données suivantes :

Figure III. 4 : Origines linguistiques des emprunts

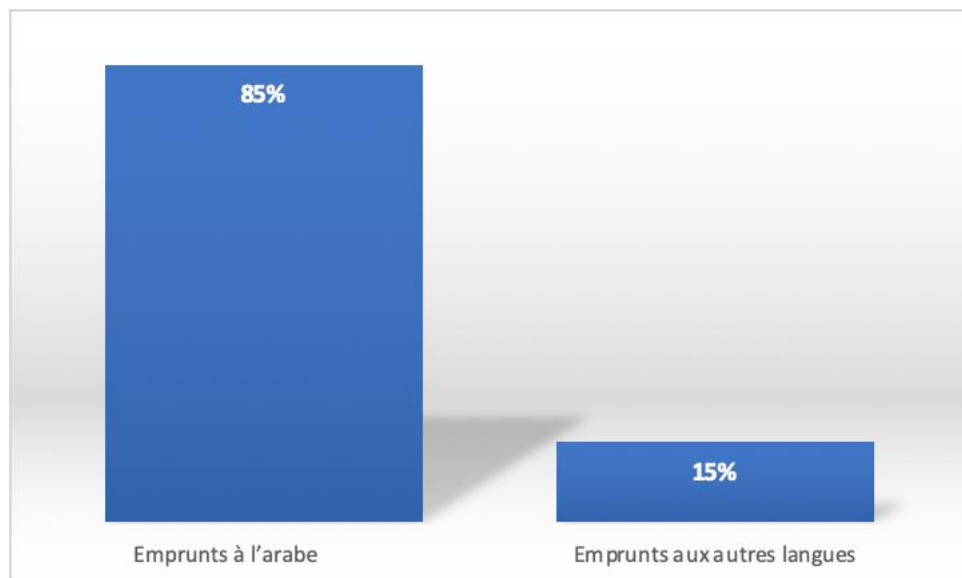
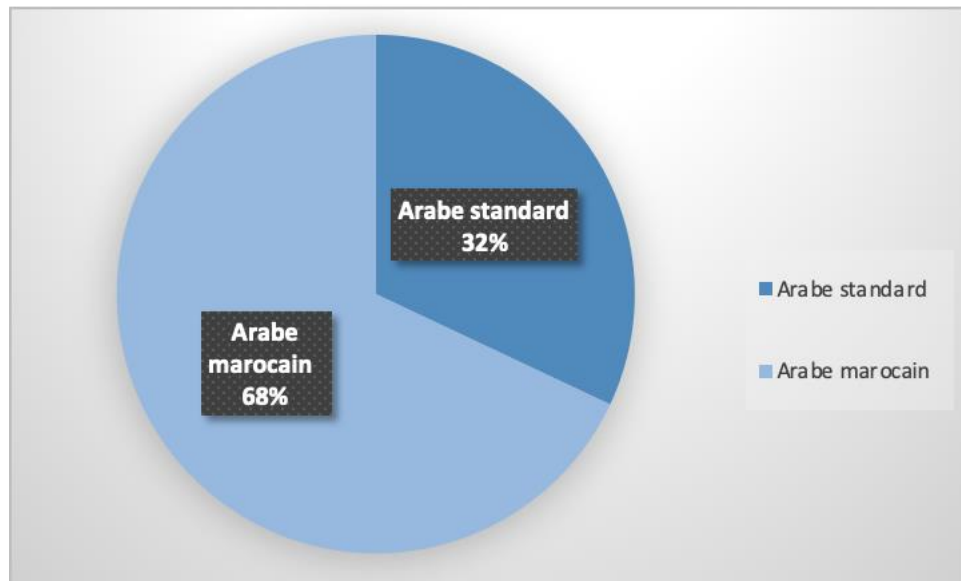


Figure III. 5 : Répartition des emprunts à l'arabe



Le premier graphe montre que la majorité des emprunts proviennent de l'arabe avec un taux de 85%. Le deuxième graphe présente la répartition de ces emprunts à l'arabe et montre que 68 % d'entre eux sont issus de l'arabe marocain alors que les 32% restants découlent de l'arabe 'standard'.

Ainsi, nous pouvons déduire que la langue arabe est la première source linguistique de l'emprunt dans l'écriture d'El Maleh. Par ailleurs, nous relevons également des emplois très fréquents de l'hébreu, chose qui demeure tout à fait normale, du moment que l'auteur est un juif natif de la langue arabe et du judéo-marocain. Nous notons aussi la présence d'emprunts à l'espagnol et à l'amazighe, lesdits emprunts désignent généralement des entités facilement reconnues par le lecteur marocain, comme : *señorita*, *argane*, etc.

1.1. Origines linguistiques de l'emprunt

➤ Emprunts à l'arabe marocain

Les emprunts relevant de l'arabe marocain sont beaucoup plus nombreux que ceux empruntés à l'arabe standard. Ils présentent surtout des réalités de la vie quotidienne (*néfar, khaïma, hammam*), des croyances populaires (*jnouns, chouafa, meskoun*) et des rites culturels (*moussem, hmadcha, zaouïa*) qui caractérisent la société marocaine traditionnelle.

➤ Emprunts à l'arabe standard :

Les emprunts en provenance de l'arabe standard renvoient surtout au domaine liturgique. Ils transposent, généralement, des réalités religieuses (*haj, nabi, dikr, hadith*) et décrivent des aspects de la culture arabo-musulmane (*msid, fqih, hram*) ; mais pas uniquement, puisqu'on relève des mots qui appartiennent au domaine politique comme : « *indimaj, istiqlal* », ou bien « *incheqaq* » que le narrateur traduit par « fracture » pour évoquer le départ des juifs marocains vers l'Etat d'Israël.

➤ Emprunts à l'hébreu :

La majorité des mots hébreux renvoient, au même titre que l'arabe standard, à des concepts liturgiques comme : *pessah, bar-mitsvah, rouah, sofah*, etc. Cependant, il y a des emprunts provenant du judéo-marocain, tels que : *zabador, zdioua, bzahd allah, saykh, skhina, mellah*, etc. Nous avons noté également l'existence de quelques mots de la haketia²⁵⁰, comme « *téfélím* » qui est l'équivalent de la bar-mitsvah. Soulignons que l'utilisation de la haketia dans le corpus d'El Maleh a une

²⁵⁰ « La Haketia est le nom du dialecte judéo-espagnol parlé par les megorachims, les juifs séfarades installés au Maroc à la suite de l'expulsion des juifs d'Espagne en 1492. » <https://fr.wikipedia.org/wiki/Haketia>

connotation particulière puisqu'elle retrace l'itinéraire des juifs expulsés d'Espagne.

➤ **Emprunts à l'espagnol**

L'espagnol pourvoient des mots qui soulignent quelques marques de la présence espagnole au Maroc, notamment au nord du pays, avec l'utilisation d'emprunts comme : *patio, montera, nada, suerte, leche*, etc.

➤ **Emprunts à l'amazighe**

L'amazighe est aussi présent, quoique à moindre degré, mais il présente des entités de la tradition locale amazighe, notamment certains produits connus pour leurs vertus alimentaires et leurs qualités nutritives ou cosmétiques, comme : argan.

➤ **Emprunts à l'anglais :**

Quant à l'anglais, il est rarement utilisé, en plus, il est cantonné à une certaine élite ayant reçu un enseignement supérieur à l'extérieur du Maroc (ex : *wharf, riots*).

1.2. Catégories grammaticales de l'emprunt :

S'agissant des catégories grammaticales des emprunts, nous remarquons à partir des données relevées que la quasi-totalité des emprunts sont des syntagmes nominaux. Ainsi, nous distinguons ceux qui sont simples, comme : un *oued*, un *derb*, la *halqa*, le *macadam*, etc, et ceux qui sont composés comme : *logha el oum, wali ummi, ould essouk, talb maachou, jben el bola*, etc.

Quant aux adjectifs, qui sont moins abondants, ils recouvrent deux catégories : la première concerne les adjectifs qui sont issus directement de l'arabe marocain et qui gardent les marques adjectivales de leur langue d'origine, comme : *souiri, zailachi*, etc. La deuxième recouvre les adjectifs qui sont formés par le biais de la dérivation à partir de noms provenant de l'arabe, c'est le cas, par exemple, de *kiféen* dans

l'expression « firmament kiféen » qui est formé sur *kif* avec le suffixe *-éen*, ou encore *istqlalien*, qui est dérivé à partir d'un emprunt à l'arabe standard *istiqlal* auquel est rajouté le suffixe *-ien*.

Pour ce qui est de la fréquence des syntagmes verbaux, nous pouvons dire qu'elle est trop faible. Néanmoins nous avons détecté quelques verbes comme : *tkelem*, *qol*, etc.

Quant aux règles d'insertions des emprunts arabes dans le texte d'El Maleh, aucune règle n'est appliquée d'une manière systématique, ainsi le terme « *les smaïïia* » adopte l'article français « les » mais garde la terminaison *-ïa* qui marque le masculin pluriel en arabe marocain sur le plan morphologique ; il en est de même pour « *la khamia* ». Cependant, dans les emprunts suivants : « *el ghech, el falas, el bla, el haq, etc* », c'est l'article arabe « *el* » qui est conservé.

Après avoir présenté les origines linguistiques des emprunts inventoriés et leurs catégories grammaticales, nous étalerons, dans les point suivants, les différents types d'emprunts employés dans le texte d'El Maleh, ainsi que les champs sémantiques évoqués par leurs emplois.

2. Emprunts de luxe

Généralement, l'usage de l'emprunt lexical trouve sa première justification dans la volonté de combler un vide dans la langue emprunteuse, en vue de désigner de nouvelles réalités socioculturelles, religieuses, scientifiques, etc. Il constitue à ce titre un emprunt nécessaire ou dénotatif²⁵¹, équivalent d'un néologisme ou d'une création

²⁵¹ Zakia IRAQUI-SINACEUR, « Histoire et emprunt linguistique », in : *Trames de langues : usages et métissages linguistiques dans l'histoire du Maghreb* / Institut de recherche sur le Maghreb contemporain sous la direction de Jocelyne Dakhlia, Paris, Maisonneuve et Larose, 2004, p. 509

de mot. Mais, au-delà d'emprunter par nécessité, d'autres motivations peuvent aussi légitimer le recours à l'emprunt, celles que justifient des « raisons de cœur »²⁵² en créant l'usage de ce que les linguistes nomment « Emprunt de luxe »²⁵³.

Bien que les puristes considèrent ce type d'emprunt comme inutile et superflu puisqu'il vient doubler et parfois même concurrencer un mot de langue encore bien vivant. Mais ça sera faux de dire que pareil emprunt ne répond à aucune nécessité. Pour Louis Deroy²⁵⁴, même s'il n'y a pas de besoin matériel, il existe un autre besoin d'ordre affectif qui revêt les formes les plus variées et présente des nuances divergentes à l'extrême. Certains cas sont très près de l'utilité matérielle, d'autres, au contraire, en sont aussi éloignés que possible. Toutefois, les raisons les plus déraisonnables peuvent avoir un profond intérêt psychologique et parfois des conséquences linguistiques très importantes. Alors, pour Edmond Amrane El Maleh, quelles seraient les raisons de cœur qui justifient son recours à ce type d'emprunt ?

Pour répondre à cette interrogation nous nous proposons tout d'abord de relever tous les emprunts figurant dans notre corpus, chacun avec son équivalent dans le français standard, afin de dévoiler l'intérêt de leur emploi ainsi que les valeurs connotatives auxquelles ils se rapportent.

Soulignons que l'examen de l'emprunt de luxe nous a fait songer au « doublet » tel qu'il a été mis en exergue par Benzakour²⁵⁵, qui, elle-même, s'est inspirée de

²⁵² Louis DEROY, *op.cit.*, 1956, pp. 171-187

²⁵³ Ibid., p. 172

²⁵⁴ Ibid.

²⁵⁵ Fouzia BENZAKOUR, « Le français au Maroc. Le problème des doublets : entre dénotation et connotation », in : *Contacts de langues et identités culturelles, perspectives lexicographiques*, Actes des quatrième journées scientifiques du réseau « Étude du français en francophonie », LATIN, D. et POIRIER, C., Québec: Presses de l'Université Laval, Québec. Canada, 2000b, p. 314.

Lafage²⁵⁶ pour conclure que le doublet est une paire de mots, l'un appartient au français et l'autre est un emprunt lexical ; les deux renvoyant à un même sens.

Le tableau ci-après met en exergue l'ensemble des emprunts de luxe qui, en fait, représente 15% de l'ensemble des emprunts lexicaux répertoriés.

Tableau III. 48 : Emprunts de "luxe"

Emprunts de Luxe	Equivalents en français proposés par l'auteur
Ichraq	Illumination
Berk	Éclair
Aafrit	Démon
Qalam	Roseau
Smaq	Encre
Zgharret	Zouyous
Nazaha	Pique-nique
Thara	Circoncision
Hajjam	Coiffeur
Derb	Quartier
Stouan	Couloir
Saïq	Nettoyage
Merkeb	Voilier
Serra	Nombril
Oum	Mère
Ummî	Illettré
Daher	Apparent

²⁵⁶ Suzanne LAFAGE, *Français écrit et parlé en pays éwé (Sud-Togo)*, SELAF (Coll. sociolinguistique, n° 3), 1985, p. 155.

Baten	Doublure intime de l'être
Jenn	Diable
Deban	Mouches
Maalem	Patron
Kemara	Sale gueule
Beiia	Serment d'allégeance
Serraq-zit	Cafard
Hraoua	Gourdin
Rezq	Chance
Mounafiquin	Hypocrites
Majdoub	Amant inspiré
Bab	Porte
Cheffar	Voleur
Quetaa	Coupeur de route
Mahrach	Pain d'orge
Metrouz	Broderie
Qasida	Poésie/Récitatif
Qsida	Épopée
Douiria	Petit appartement
Tenfiha	Tabac à priser
Tehan	Rate
Smaïtia	Babouchiers
Hamla	Eaux en torrents
Kfen	Linceul
Bisitat	Visites
Boussat	Baisers

Coutché	Cocher
Aar Aar	Bois de thuya
Flann	Un tel
Touiza	Corvée
Kettane	Toile
Hlou	Doux
Sbnia	Foulard
Incheqaq	Déchirure/Ligne de fracture
Sanida	Sucre en poudre
Zman	Dans le temps
Achoubs	Plantes médicinales
Mi'raj	Ascention vers le ciel
Maqamat	Stances
Batels	Barques de pêche
Jedba	Danse
Idrab	Grève
Zbel	Ordure
Chouafa	Cartomancienne
Moulheddine	Athées
Sahar	Sorcier
Aachoubi	Herboriste
Aatar	Épicier
Ahlan	Bienvenue
Stehane	Terrasses
Haja	Chose
Souk	Marché

Aasseb	Nerfs
Dokha	Vertiges
Zerda	Festin
Aajib	Formidable
Maalem	Patron
Khamia	Rideau
Jawoui	Alun
Cadi	Juge
Adib	Lettre
Ouarda	Rose
Khrief	Petit agneau
El ghech	La triche
Babour	Samovar
Douaia	Fiole d'encre
Rohé	Mon âme
Mnam	Rêve
Mlakh	Ange
Aouitqa	Jeune fille
Hmar	Âne/Ignorant
Dbaa	Imbécile
Meskin	Pauvre
Qaraqouz	Fantoches

Les emprunts de luxe mentionnés dans le tableau ci-dessus proviennent, généralement, de l'arabe (marocain et standard). Il n'y a aucune règle quant à leur intégration dans le texte, ils sont soit antéposés à leurs équivalents français soit

postposés.

Cependant, l'usage de ce type d'emprunt ne serait pas le fait d'un pur hasard. L'écrivain qui manie avec aisance la langue française se sert de ces emprunts pour s'approprier la langue française et la faire sienne en l'adaptant aux colorations de sa culture d'origine. A titre d'exemple, le mot *hajjam*, traduit littéralement en français par « coiffeur », fait plonger le lecteur marocain dans un passé lointain où le *hajjam* exerçait toute une panoplie de pratiques qui n'ont rien à voir avec la fonction du « coiffeur » tel qu'il est défini en langue française. En effet, dans Le Petit Robert, le mot "coiffeur" est un nom masculin attesté en 1669 et défini comme étant "le spécialiste de la coiffure" ; il peut être "coiffeur pour hommes" : celui qui "coiffe et fait la barbe", appelé aussi "barbier" dans certaines régions. Alors, qu'en est-il de la signification du mot *hajjam* ?

Le Dictionnaire Colin d'Arabe Dialectal Marocain²⁵⁷ propose pour le mot *hajjam* plusieurs fonctions, dont nous citons les plus fréquentes dans le référentiel marocain traditionnel : couper les cheveux et faire la barbe ; extraire les dents ; tirer le sang mauvais ; circoncire les enfants ; ordonner des cérémonies, etc. De ce fait, le mot *hajjam* est associé dans l'imaginaire populaire marocain à un homme à dons multiples. Alors le vocable français équivalent « coiffeur » suggère-t-il de telles significations ?

D'autres connotations puisées dans le référentiel socioculturel marocain se profilent à travers l'emploi de l'emprunt *Touiza*, auquel l'auteur a donné l'équivalent de « corvée ». Le dictionnaire Larousse indique que le mot corvée recouvre trois

²⁵⁷ Zakia IRAQUI-SINACEUR, *Dictionnaire Colin d'Arabe Dialectal Marocain*, 8 volumes, Rabat : Al Manahil, Ministère des affaires culturelles, 1994, vol. 2, p. 288.

acceptions principales :

- 1) Travail collectif gratuit qui était dû au seigneur ou au roi par le paysan.
- 2) Travail effectué à tour de rôle par certains membres d'un groupe, d'une communauté, d'un corps de troupe ; équipe chargée de ce travail.
- 3) Tâche, obligation pénible ou fastidieuse : La corvée du ménage.

Alors qu'en est-il de son concept dialectal ? La **Touiza** désigne toute un système d'entraide et de coopération fait à titre volontaire et sans profit. Elle répond à des besoins économiques, éthiques et relationnels très importants. Elle est utilisée notamment par les gens qui n'ont pas de moyens financiers pour faire face aux frais de la main-d'œuvre. Elle constitue, à cet effet, l'une des formes les plus expressives de la solidarité sociale.

Généralement la **Touiza** est une pratique connue sur l'ensemble du territoire maghrébin. D'ailleurs ce vocable fait partie de ce que les linguistes appellent « *statalisme*²⁵⁸ » ; il s'agit, en fait, d'un terme qui représente un fait linguistique dont les limites de l'aire de distribution géographique coïncident avec au moins une partie importante d'une frontière politique. Quant à son étymologie, le mot **Touiza**, **Twiza** ou **Twizi**²⁵⁹ est dérivé de la racine berbère **wiz** ou **Iwaz** qui signifie « aider », mais également le col d'une montagne d'accès difficile et périlleux (il est aisé de faire le rapprochement entre la difficulté d'emprunter un passage périlleux dans la montagne et l'entraide comme moyen pour surmonter les difficultés).

²⁵⁸ http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8359112, consulté le 15 septembre 2019

²⁵⁹ Mostefa MIMOUNI, « *La Twiza : entraide d'hier et d'aujourd'hui* », In : colloque franco-algérien « Transmission, mémoire et traumatisme », université de Strasbourg, 9-10 mai 2003 : en ligne sur <http://www.parole-sans-frontiere.org/spip.php?article108#nb1>, consulté le 02 septembre 2019

Qu'elle soit individuelle ou collective, la **Touiza** n'est pas le propre des hommes uniquement. Elle concerne également les femmes dans les travaux que la communauté leur délègue, comme : les funérailles, les mariages, les cérémonies religieuses ; même les enfants sont impliqués. Ainsi, à chaque fois que le travail dépasse les capacités et les moyens d'un individu ou d'un groupe d'individus, il y aura recours à la **Touiza**.

Notons que le travail dans le cadre d'une **Touiza** ne se fait pas sans musique et chants portant sur l'entraide. Parce que ceci crée une ambiance de fête qui ne permet pas à la monotonie de s'installer, surtout quand le travail est trop répétitif (dessablement de palmeraie, creusement des puits, cueillette des olives, etc). De ce fait, la **Touiza** est une situation de travail où « la joie et le plaisir d'être ensemble » doivent prendre le dessus sur la fatigue.

Alors, est-ce que le mot « *corvée* » englobe tous ces valeurs de complémentarité, de fraternité et de solidarité sociale ? Est-ce qu'il peut transmettre toute la charge affective et émotionnelle que revêt le travail dans le cadre d'une **Touiza** ?

Partant, il semble que Blaise Pascal avait raison quand il a dit : « le cœur a ses raisons que la raison ne connaît point ». En effet, quoique contesté par les puristes, les raisons de cœur qui ont poussé l'écrivain à utiliser ce type d'emprunts surpassent de loin les besoins matériels et légitiment ainsi l'usage de ces emprunts tant taxés de l'étiquette de "luxe".

3. Emprunts naturalisés

« L'emprunt est un intrus. Il n'est pas reçu d'emblée dans la langue emprunteuse à l'égal des mots indigènes. Il s'insinue peu à peu, se travestit, se fait familier, laisse oublier son origine étrangère...»

LOUIS DEROY²⁶⁰

A côté des emprunts de luxe, nous avons recensé d'autres emprunts insérés dans le corpus, sans traduction. Lesdits emprunts sont au nombre de 316 mots, mais qui incluent des mots rendus plus ou moins familiers via la langue des pieds-noirs, ou bien à travers la littérature magrébine d'expression française et qui ont finis par être introduits dans les dictionnaires du français. Ils sont, de ce fait, des emprunts naturalisés.

Ainsi, nous jugeons utile de relever ces emprunts, attestés dans les dictionnaires de langue française et de fouiller un peu dans leur histoire, afin de détecter leur langue d'origine avant leur installation dans la langue d'accueil.

Tableau III. 49 : Emprunts naturalisés

Emprunts	Date de parution	Langue et / ou mot d'origine
Arganier	1907	dér. (argan + -ier)
Babouche	1542	Persan pāpuš / turc pāpuş
Bakchich	1846	Turc. Bakšiš / persan bahšīš
Baraka	1903	Arabe. Baraka
Bazar	1546	Persan. Bâzâr

²⁶⁰ Louis DEROY, *op.cit.*, 1956, p.215.

Bled	fin XIX°	Arabe. blad
Burnous	1735	(barnusse 1556) ar. Bournus
Cadi	(cady 1351)	Arabe. qâdi
Caftan	1537	turc qaftān. persan ḥaftān
Caïd	1694	(caïte 1310) ar. Qâid
Casbah	1830	(casouba 1813) ar. qaçaba, qaçba
Cheikh	1838	(scheikh 1631) ar. Chaikh
Chérif	1552	(1528 sérif) ar. Charif
Coran	1657	Arabe / al-Qur'an
Coranique	1877	dér. (coran + -ique)
Couffin	1478	Arabe / quffat
Dahir	1929	Arabe
Djellaba	1849	(1743 jilleba) ar. mar. djallāba
Djebel	1870	Arabe
Douar	1617	(rare avant XIX° .S.) ar. magh.
Fellah	1661	Arabe
Fedayin	1956	Arabe
Goumier	1842	ar. qaum dér. (goum + -ier)
Hadith	1697	Arabe / ḥadīth
Hadj	(hadji 1743)	Arabe
Haïk	1830 (heyque)	Arabe / ḥā'ik
Hammam	1655	arabo-turc hammām
Haschish	1847	(aschy 1556) ar. ḥachīch
Henné	1541	Arabe / ḥinna
Imam	1697	Arabe / imām
Islam	1697	Arabe

Khôl, kohol	kohl 1787 (1646)	Arabe / Kuhl
Kif	1853	Arabe / kif
Marabout	1628 (morabuth)	ar. Murābiṭ port. Marabuto
Médina	1732	Arabe / madiīna
Muezzin	1823	turc. müezzīn arabe.mu'addīn
Oued	1849	(ouadi 1807) ar. Wadi
Ouléma	1765	Arabe / oulamâ
Pacha	1626	(bacha 1457) turc
Ramadan	1896	Arabe / ramadān
Rassoul	-	Arabe / ghāssoul
Riad (ryad)	-	Arabe marocain / riad
Saroual (sarouel)	1887	Arabe / sirwal
Soufi	1834	Arabe / Şūfī
Souk	1848	Arabe / Sūq
Sourate	1842	(surate 1715) ar. Sūrat
Sultan	1540	Arabe / Sulṭān
Tajine	1938	Arabe
Tarbouche	1869	Arabe / Tarbūch
Umma	-	Arabe / umma
Vizir	1457	Turc vezir Arabe wazīr
Youyou	1888	(onomat.)
Zaouïa	1843	Arabe / zāwiya
Zellige	1918	(zelis 1849) ar. magh. zallīz

Ainsi, sur les 316 emprunts non traduits, nous relevons 54 mots ayant chacun une entrée dans les dictionnaires usuels de la langue française comme, Le Petit Robert et

Le Petit Larousse illustré²⁶¹. Ces emprunts, avant d'être assimilés par la langue réceptrice, ont parcouru graduellement plusieurs types d'intégration (phonétique, graphique, morphologique, sémantique, etc.) et ont fini par être bien adaptés au système linguistique de la langue emprunteuse. Cependant, nous pouvons enregistrer au niveau du corpus d'El Maleh quelques fluctuations vis-à-vis de l'orthographe. En effet, quelques emprunts gardent les empreintes de leur langue d'origine, comme : *casbah* qui présente une double graphie *casbah* et *kassabah*, ou bien *calame* que l'auteur note quelques fois *qalam*.

Ces variations graphiques des mots empruntés n'ont pas une grande importance sur le plan sémantique, du moment qu'elles servent uniquement à faire la distinction entre un emprunt et un xénisme²⁶².

S'agissant du volet qui nous intéresse le plus, à savoir le « sens », ces emprunts sont facilement assimilables par le lecteur francophone, du moment qu'ils sont déjà répertoriés dans les dictionnaires du français. D'ailleurs, certains mots évoqués dans le tableau remontent à des dates lointaines, comme : Sultan (1540), Henné (1541), Douar (1617). Les moins anciens d'entre eux remontent à la période du Protectorat : Zellige (1918), Dahir (1929).

²⁶¹ Nous avons également fouillé dans des dictionnaires en ligne comme : *Le Centre nationale des ressources textuelles lexicales*, <https://www.cnrtl.fr/portail/> et la *Base de données lexicographiques panfrancophone*, http://www.bdlp.org/recherche.asp?base=*&AC=oui&AL=oui&AN=oui&BE=oui&BU=oui&CA=oui&CE=oui&CB=oui&CI=oui&FR=oui&LO=oui&MD=oui&MA=oui&IM=oui&NC=oui&QU=oui&RE=oui&RW=oui&SU=oui&TC=oui

²⁶² Le xénisme est un emprunt à une langue étrangère et qui ne subit aucune altération due à la langue hôte (TLFI). Néanmoins, lors du passage de mots arabes vers la langue française, il est difficile de parler de xénisme pur et dur, étant donné l'altération de certains phonèmes arabes par le système phonologique et graphique du français.

4. Emprunts de nécessité

Outre les emprunts de luxe et les emprunts naturalisés, nous avons noté l'existence de mots qui ne sont jamais glosés ou traduits et qui ne figurent pas dans les dictionnaires français. En effet, ces mots, provenant de diverses langues en présence dans le paysage linguistique marocain, servent à véhiculer des réalités culturelles spécifiques que le français est parfois inapte à transmettre. Ils sont à ce titre des emprunts de nécessité.

Généralement, dans l'œuvre d'El Maleh, le recours à ce type d'emprunt répond aux besoins d'exprimer des valeurs identitaires et des croyances populaires typiquement marocaines. À titre d'exemple, le mot « *dada* », cité dans *Le Retour d'Abou El Haki*, remonte à un passé très lointain où des fillettes issues de milieux très pauvres ont été vendues par leurs parents à des familles bourgeoises. Réduite à l'esclavage, ces petites filles leur servent de tout : ménagère, nourrice, même plus, concubine du propriétaire. Ainsi, le mot *dada* est associé dans l'imaginaire marocain à une femme à dons multiples : chanteuse, danseuse, conteuse, celle qui veille au bon déroulement des tâches ménagères, celle qui procure joie et plaisir à tous les membres de la famille tandis que personne ne s'intéresse à ses besoins ni à ses désirs. Alors, existe-il un équivalent en français qui suggère de telles significations ?

Parfois, l'emprunt de nécessité, ou le mot « intraduisible », manifeste chez l'auteur une conscience linguistique aiguë à l'égard du transfert d'un contenu culturel donné dans une langue qui n'est pas la sienne. Un passage de *Parcours immobile*, qui inclut lui-même plusieurs emprunts non traduits, illustre parfaitement cet état de fait:

« Dans mon pays il y a des hommes qui parcourent les espaces de jour de nuit le même pas les conduit à la rencontre d'eux-mêmes voyez-les se découper immuables sous les cieux changeants chevelure crinière ample djellaba ou tchamir

*de bouts d'étoffes rapiécées un couffin un bâton le visage de l'énigme ils marchent avec la régularité d'un astre ils ont aboli les routes les cars les trains les autos les lieux de naissance les routes familières détruit la prison d'un nom **on les appelle Hadaoua** mais cela ne veut rien dire de leur pas sans empreinte ils gommant les plis de l'événement.»²⁶³*

« **Hadaoua** », mot intraduisible, suggère chez l'auteur toute une attitude poétique consistant à récupérer le sens d'un mot auquel la langue française n'a pas offert d'équivalent sémantique. Prenons une autre citation où El Maleh souligne explicitement l'intraduisibilité²⁶⁴ d'un mot :

*« Mouna! C'est lui qu'elle regarde maintenant. Le regard d'un enfant adressé à un vieil homme. **Aamou**, ton vieil oncle, cette appellation particulière, intraduisible en fait, faite de tendresse et de respect indicible²⁶⁵ ».*

Cette phrase souligne ce potentiel d'intransmissibilité contenu dans chaque langue, mais qui demeure chez El Maleh lié à une invitation à accepter la part de mystère contenue en toute chose. En effet, quoique l'auteur ait essayé de traduire littéralement le mot **Aamou** pour faire comprendre au lecteur non arabophone la signification de cet emprunt, il avoue juste après, implicitement, que la traduction proposée n'est pas réussie, parce qu'elle ne communique pas la tendresse et le respect contenu dans cette appellation si particulière.

De ce fait, tout en ayant conscience de la particularité de sa culture nationale, l'écrivain n'hésite pas à faire appel à l'emprunt pour nommer des concepts marocains

²⁶³ Edmond Amran EL MALEH, *Parcours immobile*, *op.cit.*, p.251

²⁶⁴ De nos jours, il y a un intérêt particulier pour les intraduisibles qui sont, en fait, liés à une volonté de défendre les singularités des peuples. Voir notamment : Barbara Cassin, *Vocabulaire européen des philosophies. Dictionnaire des intraduisibles*, Paris, Seuil / Le Robert, 2004.

²⁶⁵ Edmond Amran EL MALEH, *op. cit.*, 2004, p. 29

tels qu'ils sont dénommés dans leur langue d'origine, des dénominations porteuses de symboles culturels et de colorations locales authentiques.

Cependant, il y a des emprunts auxquels la langue française a procuré des équivalents sémantiques par le biais de la créativité de ses usagers ; mais l'auteur refuse tout de même de les utiliser. Par exemple, le mot « *msid* », que l'on traduit souvent par « école coranique », apparaît dans le texte d'El Maleh à maintes reprises et sans traduction. Cela nous pousse à nous interroger sur les motivations de l'écrivain à faire un tel choix. Peut-être que le néologisme « école coranique » ne connote pas le rôle que le *msid* a joué tout au long des siècles dans la transmission des valeurs de solidarité et de tolérance tant prônées par l'Islam. Aussi, il ne véhicule pas l'image de la flagellation associée à l'emploi de l'emprunt *msid*. Néanmoins, pour communiquer de telles connotations, l'auteur insère généralement ces emprunts non traduits, dans un contexte qui en éclaire suffisamment le sens, tel que le montre la phrase suivante citée dans *Parcours immobile* :

« un m'sid une petite baraque ouverte sur l'extérieur les gosses entassés de toute éternité assis à même le sol sur des nattes la tablette coranique tenue à deux mains se balançaient au rythme des versets sacrés sous la fêrule du Fqih. »²⁶⁶

De ce fait, l'inclusion de ces emprunts dans le texte répond à l'envie d'El Maleh de transmettre un imaginaire populaire nourri d'expériences personnelles et inspiré de souvenirs impressionnants de la vie traditionnelle des Marocains. En effet, des mots comme *hadaoua* et *aamou* ne sont pas porteurs d'une charge culturelle plurielle et ne viennent-ils pas incruster le texte littéraire français et marquer un effet de style à

²⁶⁶ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 227

résonance locale ? de même le fait de prendre une *dada* ou de confier ses petits enfants au *m'sid* ne relèvent-ils pas du vécu socioculturel marocain d'antan ?

À cet égard, la manifestation de la culture marocaine semble se réaliser à travers l'usage des emprunts dont l'emploi traduirait implicitement la volonté de l'écrivain de marquer son appartenance à la société marocaine et son attachement à la culture locale sous ses multiples facettes.

5. Champs sémantiques de l'emprunt

Les emprunts relevés dans l'écriture d'El Maleh paraissent couvrir plusieurs champs sémantiques. Toutefois, nous avons essayé de les cerner en vue d'identifier les plus productifs d'entre eux. La liste ci-après présentera les thèmes relatifs à chaque domaine identifié :

Spiritualité : *zaouïa, taïfa, soufi, hamdouchi, hmadcha, taréqa, aïssaoua, moussem, hadaoua, soufisme, soufistes, kabbaliste, jedba, soufis.*

Religion musulmane : *islam, muezzin, haj, msid, jamaa, shor, fqih, baten, rezq, mi'raj, hadith, coran, nabi, ramadan, dikr, minaret, din, kafer, msilmine, fatha, oulémas, alem, Iblis, daher, hram, ouerd, barzakh.*

Musique et folklore : *bendir, ghaïta, lira, ney, gnaoua, néfar, taaréja, guenbri, oud, zgharret/sgharet, halqa, hlaïqui, qaraqouz.*

Habits traditionnels : *farajia, burnous, djellaba, tchamir, selham, koumia, taguia, haïk, fouquia, mendil, zabador, choukara, dfina, caftan, saroual, bzioui, skali, metrouz, kettane, sbnia, babouches, kippas, taleth.*

Art culinaire : *couscous, tajine, mahraj, méchoui, skhina, hendia, chahma, safran, tehmira, cumin, ghriba, pstela, gousia, hargma, douaz, zafrane, kamoun,*

karouya, ibzar, debana, gouza, kharkoum, felfela, soudania, skingbir, madnouss, kasbour, lkma, teghmira, faduelos, mahia, melokhia, zaatar, taâm, chaqor, kanoun, mjamar, houta, bed, rabouz, maaza, jben, aajinn, mahrach, chréha, hlou, sanida, zerda, tehan, kermos, taïb, churros, leche, aknari.

Soins de beauté : *khôl , henné, hammam, ghassoul, harmal.*

Métiers : *fellah , tajer , adoul, kahouaji , khamès , kabla, toubib , moudiraam , mothasseb , berchman, hakawati , tbib, chaouch, hajjam, smaïtia , coutché , cadi, njara, medico, hafaf, mohil.*

Lieux et habitations : *noualla , kissaria, zriba , douar, khaïma, riad , souk , médina, casbah/Alkassabah/kasba , kharba, sérail , manzeh, douar, derb, stouan , douiria, mellah, makbara, guitoune.*

Vices : *rechoua, elghech, el falas, el bla, fadeha.*

Qualité et vertu : *el khair, el mhaba, tiqar, taïeb, el haq, hiba.*

Titres : *saïdati, lalla , chrif, hadja ,sidi, ba, ima, jda, rajoul, oum, rohé, aouitqa, saykh/cheikh, aami.*

Enseignement : *adab, ustad, arabia ,qari , alif /aleph, ba, noun, smaq, loha, katib, qasida/quasida/qsida/,ummî, el qalam/calame, maqamat, douaia, adib, qsida, logha.*

Croyances populaires : *ghoul, jnouns , mejnoun, meskoun,jawoui, jenn, aafrit, chouafa, achoubs , achoubs , sahar, aachoubi, aatar, majdoub.*

Insultes : *gheddar, khain , kelb , klab, dbâa , selgot, zoufri, zoufria, aarien, hmar, qhab, zanadek, kofar, hbil, kemara, mounafiquin, cheffar, quetaa, zbel, moulheddine.*

Politique : *frança, bled, siba, erraïs, harb, guirra, boulice, sultan, chiisme, chuoï, pacha, khalifa, mokhazni, macadam, caïd, chabakouni, merda, maghzen, askri, ouzir, cheikh, el maalem, malik, mamlaka, istiqlal, chiïtes, zaïm/ezzaïm, fidaïns/feddayin, feddaï, beïia, idrab, calife, vizir, goumier, marabout.*

Objets : *mica, carrossa, tivasa, merkeb, hraoua, la kfifa, bab, khamia, babour.*

Appartenances régionales ou ethniques : *nasrannia, nesrania, chleuhs, sahraoui, nasrani, zailachi, marrakchie, ndalousie, Lingliz, yahoud, souiri, ashkénaze.*

Religion et imaginaire juif : *seder, bar-mitsvah, pâque juive, hagada, mimosa, pessah, torah, talmud, guematria, tikkoun, tsimtsoum, zohar, shoah, guemara, hazan, kaddish, kippour, sofar, hiloula, yeshiva, galout, mimouna, kibboutz.*

Nature : *oued, sahara, nkhal, nouar, chergui, houa, le gharbi, berk, hamla, aar aar, ouarda, arganier, Bourghout, Zdioua, Jmel, deban, serraq-zit, elfaa-buska, dounia, khrief.*

Substance psychotrope : *maajoun, mokhadarat, sebsi, kif, tenfiha, haschich, joint, joinat.*

El Maleh, nous semble-t-il, s'est trouvé dans l'obligation de recourir à l'emprunt pour combler les lacunes de la langue française à transmettre certaines réalités. Ainsi, les champs sémantiques identifiés en matière d'emprunt rejoignent ceux que nous avons formés lors de l'examen du vocabulaire spécifique uniquement sur quelques domaines comme : la nature, la politique, l'appartenance régionale ou ethnique, etc. Néanmoins, la volonté de l'écrivain de transmettre un patrimoine national dans sa diversité culturelle, linguistique et ethnique l'a poussé à mobiliser des champs sémantiques comme : la spiritualité, la musique et le folklore, la religion musulmane,

la religion juive, les habits traditionnels, les croyances populaires, etc. Cependant, un domaine qui se démarque par rapport aux autres par le nombre d'emprunt et leur fréquence d'utilisation nécessite que l'on s'arrête là-dessus, il s'agit de l'art culinaire. L'émergence de ce champ sémantique dans l'œuvre d'El Maleh nous pousse à nous poser plusieurs questions sur la symbolique de ce choix thématique et l'importance que revêt la nourriture dans la vie de l'auteur.

En parcourant l'œuvre soumise à l'analyse, nous constatons que la nourriture est un vrai catalyseur de souvenirs. Elle revêt souvent une connotation culturelle et mémorielle. D'ailleurs, tout au long de ses récits, l'écrivain conjugue le mets au mot. Il nous présente des scènes du vécu quotidien parfumées par des saveurs renvoyant à un inconscient culturel et constituant un marqueur permettant de reconnaître un groupe social donné et de le distinguer par rapport à un autre. En effet, tout comme la langue et les coutumes, chaque société se définit par sa nourriture. D'ailleurs, Gilles Fumey dit à ce propos : « *nos nourritures sont identitaires, elles nous agrègent à une société, un pays, une tradition mieux que le meilleur d'une politique*²⁶⁷. »

Ainsi, il existe un rapport profond reliant les individus à la gastronomie au point qu'il y a des plats emblématiques qui peuvent différencier un groupe social d'un autre, les liens entre ces plats et la communauté sont parfois si étroits qu'il y a assimilation entre les deux, tel le couscous pour les maghrébins ou la pizza pour les italiens. Cependant la nourriture peut être vue sous plusieurs angles, comme l'affirme Saadia Lahlou :

« *Le repas lui-même devient polysémique : il peut être vu comme un comportement alimentaire avec une connotation socialisante (manger) ou comme un comportement social avec une connotation alimentaire (être ensemble). Cette*

²⁶⁷ Gilles FUMEY, « *la culture dans l'assiette* », Regards croisés, n° 20, 2006 : en ligne sur <http://escales.enfa.fr/wp-content/uploads/sites/7/2009/06/14-Fumey.pdf>, consulté le 23 octobre 2018.

polysémie se traduit par un même fonctionnement sur le mode de la participation dans laquelle l'objet mangé est un symbole de communauté, à la fois incorporé et partagé, c'est-à-dire incorporé par tous collectivement. »²⁶⁸

De ce fait, « *se nourrir* » est un acte biologique qui se développe au-delà de sa propre fin, celle de satisfaire un besoin naturel. En effet, il instaure tout un système de valeurs qui laissent transparaître plusieurs significations. De même, un plat partagé lors d'un repas familial ou pendant certaines cérémonies permet de sceller d'autres types de liens particulièrement sur le plan communautaire.

Edmond Amran El Maleh, qui se réclame comme un passionné de la cuisine, accorde une autre dimension à la nourriture. Dans ses entretiens avec Marie Redonnet, il affirme que : « *le lien avec la mère passe par la cuisine, et pas seulement pour la fille qui apprend. L'homme garde toute sa vie dans sa bouche le goût de la cuisine de sa mère* ». ²⁶⁹ Donc outre la dimension biologique et culturelle qu'on peut attribuer aux aliments, El Maleh ajoute une autre dimension d'ordre affective.

Dans les écrits d'El Maleh qui sont jalonnés par la recherche d'un passé perdu, l'évocation de réminiscences culinaires répond parfaitement au besoin de l'écrivain d'affirmer son identité judéo-marocaine. D'ailleurs, en analysant le champ sémantique de l'art culinaire, nous découvrons que la majorité des emprunts recueillis se rapportent à la cuisine marocaine en générale, sauf que les mots qui ont enregistré la fréquence la plus élevée sont ceux qui renvoient à la cuisine juive.

²⁶⁸ Saadia LAHLOU, *Penser manger*, Paris, PUF, 1998, p. 126

²⁶⁹ Marie REDONNET, *op.cit.*, 2005, p. 176

Parmi les scènes qu'El Maleh ne cesse d'évoquer dans ses romans, il y a celle qui se rapporte à la **Skhina**, la **Tafina** ou encore la **Dafina**, trois synonymes pour désigner un plat qu'on peut considérer comme un « archétype culinaire » de la cuisine juive.

La **Skhina** vient de la darija « **skhoun** » qui signifie chaud, elle se rapporte à un évènement bien particulier qui est le Shabbat, le samedi. Ce jour-là revêt un caractère religieux, il s'agit d'une journée réservée à la consécration de la création divine comme le traduit l'Exode : « *Pendant six jours on travaillera, mais le septième jour sera consacré à vos yeux un chômage total en l'honneur de l'Eternel - celui qui travaillera ce jour méritera la mort* » (35,2).

Il s'agit donc d'un plat très attaché à l'imaginaire juif. Généralement, il n'est pas consommé uniquement au samedi, mais figure dans bon nombre de menus des fêtes du calendrier hébraïque. Faut-il signaler que la **skhina** comporte plusieurs ingrédients : un pied de bœuf échaudé, des pois chiches, des pommes de terre, des œufs et du blé. Sa préparation obéit à un ancien procédé de cuisson bien spécifique au Shabbat. En effet, la tradition juive interdit l'allumage du feu le jour du samedi. De ce fait, les aliments consommés durant cette journée doivent être cuits à petit feu, la nuit du vendredi dans des ustensiles lourds et clos.

Dans *Parcours Immobile*, Edmond Amran El Maleh nous offre un passage qui illustre à merveille la valeur culturelle de la **skhina** pour les juifs marocains :

« La grâce, le repos du samedi, la maison suspendue à son silence, l'oreille collée à sa respiration. Encore et de loin des nappes affleurent sous la croûte du sol, crèvent à ciel ouvert : odeurs fades des pois chiches, des œufs durs de la **tafina**, la **skhina**, émotion gustative des papilles familiales, flots d'eau de Cologne, c'était samedi à n'en pas douter. »²⁷⁰

²⁷⁰ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 34

Telle la petite madeleine de Proust, ce plat occupe une place importante dans les textes d'El Maleh. Il incarne et symbolise son appartenance identitaire puisqu'il s'agit d'une préparation emblématique de la cuisine juive et dont la consommation n'est pas faite uniquement pour assouvir sa faim, mais surtout pour renouveler les accords et consolider l'esprit de collectivité et de partage.

Outre la *skhina*, EL Maleh se réjouissait d'évoquer la *mahia* dans ces récits. En effet, les juifs ont une longue tradition de distillateurs et de consommateurs d'alcool, surtout la *mahia*. Cette liqueur, interdite aux musulmans, est une boisson sacralisée chez les juifs, elle entre dans la composition principale des repas festifs et dans un ensemble d'usages rituels. C'est aussi le premier aliment dont la bénédiction ouvre le repas juif et plus particulièrement celui du Shabbat.

Dans les textes d'El Maleh, la *mahia* est souvent évoquée dans un contexte de gaieté et de partage, tout en mettant l'accent sur les bienfaits qu'elle procure à son consommateur :

« La conversation était fort animée, très gaie, on le devine, la mahia, eau de vie fabriquée à partir de figues y est pour quelque chose, ce n'est pas le seul effet de l'alchimie.²⁷¹ »

Le vin chez les juifs est introduit dans le rituel par la voie du sacré, tandis que c'est dans le même registre que l'islam l'interdit :

« Ils avaient avec les musulmans une horreur commune de la viande ou de la graisse de porc, animal réputé impur. Par contre ils buvaient volontiers du vin, et

²⁷¹ Edmond Amran EL MALEH, *op. cit.*, 1998, p. 72

surtout la **mahia**, liqueur faite avec des figues, des raisins secs ou des dattes, dont ils faisaient usage même pendant les repas.²⁷² »

De ce fait, la consommation d'alcool et du vin en particulier était le plus souvent conçue comme un signe de distinction entre les musulmans et les juifs, c'est à quoi El Maleh fait allusion par cette phrase dans *Mille ans, un jour* : « surtout quand aux cris et aux pleurs se mêlait la gaieté des hommes de la Hébra, buveurs de **mahia**²⁷³ ».

Le statut accordé au vin dans la tradition juive affecte même son procédé de fabrication qui, à l'image des aliments carnés, obéit à un ensemble de règles de pureté. En effet, les « hommes de la Hébra » ne peuvent boire du vin qui a été fabriqué ou touché par des personnes autres que les juifs. Edmond Amran El Maleh dans *Aïlen ou la nuit du récit* insiste sur ce point en disant d'un champ de figues que : « des mains juives avaient travaillé cette terre²⁷⁴ ». Ainsi, la **mahia** renvoie à cet esprit communautaire. C'est une substance de festivité, de gaieté et de rire, elle représente le lien de communion qui scelle les rapports sociaux et soutient une identité collective.

Il existe une autre caractéristique de l'alimentation chez les juifs, ce sont les lois de la *kacherut*. Cette dernière consiste à établir une distinction stricte entre les aliments purs et impurs. Ceux qui sont conformes à ces lois sont dit **casher**, c'est-à-dire convenables à la consommation, c'est l'équivalent en quelque sorte du label *halal* chez les musulmans. La plupart des lois de la *kacherut* sont tirées des livres de Lévitique. Leurs détails et leur application pratique sont toutefois énoncés dans la loi orale et développés dans la littérature rabbinique. A travers son personnage principal dans *Mille ans un jour*, El Maleh montre que l'alimentation représente un véritable

²⁷² Claude MARTIN, *Les israélites algériens de 1830 à 1902*, Paris, 1936, p. 35

²⁷³ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 106

²⁷⁴ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 85

code social et moral, une réglementation tirée des textes sacrés : « *Manger **non casher** au vu et au su de tout le monde était un défi insolent à la pieuse et traditionnelle communauté juive*²⁷⁵ ».

Plusieurs réminiscences culinaires, chargées de nostalgies, attestent de la volonté de l'écrivain de vouloir affirmer son origine juive. Mais pour rappeler sa double généalogie judéo-marocaine, El Maleh n'hésite pas à évoquer des aliments relatifs à la culture marocaine en général, tels : le couscous, le thé à la menthe, la figue de barbari, etc. Ainsi, tout au long de ses romans, il utilise la mémoire culinaire pour oblitérer un acte de reconnaissance et d'adoration envers sa terre natale. Le partage d'un aliment comme la *dafina* ou la *mahia* renvoie aussi bien à un inconscient qu'à une mémoire collective. Il s'ensuit donc que le discours gastronomique joue un rôle très important dans l'expression de l'identité culturelle.

Au cours de cette section, nous avons examiné les principales catégories d'emprunts utilisés à travers l'œuvre d'El Maleh. En plus, nous avons analysé les motivations possibles qui peuvent justifier l'utilisation de chaque type à part. Nous avons achevé ladite section en cernant les différents champs sémantiques des emprunts relevés dans le corpus soumis à l'étude. Dans la section suivante, il sera question du calque linguistique.

B. Calque linguistique

Les personnages d'El Maleh ne sont pas censés s'exprimer en français, ils sont en quelque sorte « traduits », ils usent d'une langue qui semble une transposition littérale de leur dialecte, d'où l'apparition d'un autre fait linguistique appelé calque.

²⁷⁵ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 48

Ce dernier consiste en une transgression des règles syntaxico-sémantiques d'un code linguistique. D'ailleurs, ce procédé de formation consiste à traduire littéralement dans une langue un mot simple ou composé emprunté à une autre, en s'inspirant davantage de sa lettre que de son esprit. Il est, aussi, comme le faisait souligner Bally « *un emprunt par traduction*²⁷⁶ ». Georges Mounin, à son tour, signale que le calque est : « *une forme d'emprunt d'une langue à une autre qui consiste à utiliser, non une unité lexicale de cette langue, mais un arrangement structural, les unités lexicales étant indigènes* »²⁷⁷. Cette définition, nous semble-t-il, adaptée à l'étude que nous envisageons entreprendre surtout que le mot « *indigène* »²⁷⁸ renvoie aux personnes, aux coutumes ou aux formes d'art propres à une population ou à une région donnée, c'est-à-dire, à ce qui relève de l'ordre culturel.

Dans la vie quotidienne le calque se rapporte souvent à une incompétence dans la langue cible, mais en littérature son utilisation est justifiée par la singularité d'un imaginaire qui s'exprime dans une langue hôte. Cela nous fait songer à une citation amusante de Paul Smaïl où il oppose un « nous » locuteur à un « vous » : « *Le foie est symboliquement pour nous, Arabes, ce que vous, roumis, appelez le cœur : Tu es le fils de mon foie* »²⁷⁹. Cette phrase souligne la distinction entre deux imaginaires par le biais du calque linguistique qui s'est construit sous une certaine forme de littéralité.

Théoriquement l'on peut distinguer deux types de calque : morphologique et sémantique. Or, s'agissant de la présente étude, il nous paraît intéressant de mettre en

²⁷⁶ Charles BALLY, *Linguistique générale et linguistique française*, Berne, 1950, p. 304, n° 499, cité par L. DEROY, op. cit., p. 216.

²⁷⁷ Georges MOUNIN, *Dictionnaire de la linguistique*, Paris, Presses universitaires de France, Quadrige, 1974, p. 58

²⁷⁸ Pour la définition du mot « *indigène* » nous faisons abstraction de sa connotation péjorative ancienne qui désigne une personne non civilisée.

²⁷⁹ Smaïl PAUL, *Ali le Magnifique*, Paris, Denoël, 2001, p. 116

exergue le rôle du calque phraséologique que nous pouvons étudier séparément du calque morphologique, vu la particularité de son emploi sous forme de proverbes et d'expressions idiomatiques dans le contexte étudié. Ainsi, nous comptons proposer une typologie en trois catégories : le calque phraséologique, le calque morphologique et le calque sémantique.

1. Calque phraséologique

Pour ce qui est du calque phraséologique, nous distinguons entre deux types : les proverbes et les expressions idiomatiques ou locutions figées. S'agissant de notre corpus, les structures calquées relevées proviennent en majorité de l'arabe dialectal et trouvent leur origine dans le référentiel marocain traditionnel.

1.1. Expressions calquées sur des proverbes

En parcourant l'œuvre d'El Maleh, nous avons pu relever quelques exemples d'expressions calquées sur des proverbes de l'arabe marocain. La caractéristique principale des structures relevées, c'est qu'elles ont pu conserver, malgré l'opération de traduction, le sens proverbial qui transmet des valeurs et des leçons de morale générales. En voici quelques exemples d'illustration :

- « *Aucune parenté avec le petit livre rouge, aucune vraiment si ce n'est celle qui unit la graisse au hachoir chahma elchaqor pour reprendre un de ces proverbes populaires dont il était si friand.*²⁸⁰ »

²⁸⁰ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, op.cit., p. 84

- « *Oua Allah ! oualou tart maaza, même si la chèvre devrait voler même s'il ne me reste qu'une goutte de sang dans les veines tu ne les prendras pas pour ce prix.*²⁸¹ »

Dans les exemples cités, l'auteur insère le proverbe original en arabe dialectale tout en fournissant au lecteur une traduction littérale et une autre intelligible pour rapprocher le lecteur du sens original.

D'autres expressions proverbiales, aussi imagées, sont insérées directement dans le texte en langue française sans la version arabe, comme : « *Grimpe au figuier pour en cueillir le fruit, mais qui donc t'as dit d'y aller*²⁸² ». Aussi bien : « *celui qui veut recueillir le miel des abeilles doit en souffrir les piqûres* »²⁸³ ; proverbe qui légitime la peine et la souffrance comme prix à payer pour atteindre un objectif. Nous avons noté, également, l'existence d'un dicton trivial que l'on peut rapprocher, en français, de l'expression « qui se ressemble s'assemble », il s'agit du proverbe suivant : « *La merde court toujours à la recherche de sa sœur jumelle jusqu'au fond des océans s'il le faut* »²⁸⁴.

Hérités de la culture marocaine orale, ces proverbes peuvent paraître bizarres pour un lecteur francophone natif non averti. Il nous semble que l'insertion de telles tournures qui relèvent de l'oralité de la culture marocaine dans un texte littéraire, cherche à rendre explicite, d'une manière originelle, les modes de pensée des marocains ainsi que la richesse et les spécificités de la société à laquelle ils

²⁸¹ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 117

²⁸² Edmond Amran EL MALEH, op. cit., 1998, p. 82

²⁸³ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 186

²⁸⁴ Ibid.

appartiennent

1.2. Expressions calquées sur des locutions figées

Une autre manière d'usage du calque mérite d'être soulignée, c'est le recours de l'auteur à la traduction de locutions figées ou d'expressions idiomatiques pour reproduire dans son texte l'imaginaire culturel de l'univers décrit. En explorant l'œuvre d'El Maleh, nous avons pu relever un ensemble considérable d'expressions calquées sur des locutions figées de l'arabe marocain. La lettre du grand père de Nessim, personnage principal de *Mille ans, un jour*, en est un modèle remarquable :

« À notre fils bien-aimé, *lumière de mon âme et de mes yeux* [...] les yeux baignés de larmes, *une lourde pierre sur le cœur*. [...], toi mon orgueil et ma joie, *goûter le miel de tes paroles* [...]. »²⁸⁵

L'emphase des images et le ton affectif de ses calques reproduisent un sentiment d'étrangeté même chez le lecteur francophone qui maîtrise le dialecte marocain, du moment que cette rhétorique a été abandonnée par la langue contemporaine. De ce fait, l'auteur esquisse à travers cette lettre une chaîne de transmission dans laquelle la langue occupe une place centrale.

Ces expressions calquées sur des locutions figées, nous fait songer à une autre caractéristique des dialectes arabes relatifs aux sociétés traditionnelles, c'est qu'ils ramènent souvent l'abstrait au concret. Ils exercent, en quelque sorte, un effet de matérialité sur le langage. Des représentations comme « *une lourde pierre sur le cœur* » et « *le miel de tes paroles* » tirent leur force de cette matérialité qui semble renvoyer le lecteur à l'expérience immédiate des choses.

²⁸⁵ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 11

Nous relevons également dans le corpus d'autres expressions qui vont dans le même sens que les exemples²⁸⁶ précédemment cités :

- « **Il faut ouvrir l'œil avec ces gens** »²⁸⁷ : être vigilant.
- « **Son cœur se coupait en morceaux, kelbha taï taktaa traf traf** », cette dernière expression apparaît dans un autre contexte avec une traduction différente : « elle était partie avec un légionnaire à Marrakech "Baz" ! quelle audace ! [...] **le cœur brisé de sa mère kalbaha taï tektaa le cœur coupé haché "Allah Ibqi Ster"** Dieu nous garde ! »²⁸⁸ : avoir le cœur brisé.
- « On n'agit pas autrement quand on s'attache aux pas de l'arpenteur, **tisserand des rues, raccommodeur des derbs**, artisan-araignée, espérant lire dans les traces laissées par son passage les repères des lieux de son inspiration. »²⁸⁹ : errer dans les rues, vagabonder.
- « Tu te dis, va toi aussi au **hammam**, un autre lieu parlant, va, assieds-toi sur un banc, **tes os sont fatigués, tu ne les entends plus, tu ne leur prêtes plus l'oreille** ni donc à toi-même, tu n'as pas oublié cette délicieuse expression : **Tsenat aala aadamk**, retrouve le repos de leur prêter attention. »²⁹⁰ : Se reposer, se relaxer.

Les autres expressions calquées sur l'arabe dialectal renvoient souvent à des champs sémantiques privilégiant les métaphores alimentaires. Pour l'expression

²⁸⁶ Sur le plan sémantique, de tels expressions constituent ce que Vladimir Kapor qualifie de « catachrèse interculturelle ». Voir notamment : KAPOR, Vladimir, « La catachrèse interculturelle », *Poétique* n° 148, 2006, p. 497-505.

²⁸⁷ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p.12

²⁸⁸ Edmond Amran EL MALEH, *Parcours immobile*, *op. cit.*, p. 31

²⁸⁹ Edmond Amran EL MALEH, *op. cit.*, 1998 p. 21

²⁹⁰ Edmond Amran EL MALEH, *op. cit.*, 1998, p. 138

suivante : « *Tatakoulli dmaghi ! tu me bouffes la cervelle [...]* »²⁹¹. La forme verbale initiale veut dire littéralement « *tu me manges* » ; or l’auteur a préféré la présenter dans un registre plutôt familier pour renforcer le caractère oral de l’énoncé, dans le but de rapprocher le lecteur de la signification originale de cette expression qui veut dire « *convaincre quelqu’un par des paroles qui endorment son cerveau* ».

En se basant essentiellement sur la traduction, tous ces calques sémantico-syntaxiques demeurent globalement transparents et marque une mémoire culturelle qui affirme sa différence. Cependant, si le principe du calque procède de la traduction, elle est quelques fois biaisée, du moment que le passage d’une langue à l’autre semble générer des interrogations sur la part idéologique du langage et sur le dynamisme créatif qui découle de cette translation approximative.

Par exemple, l’expression : « *on recueillait le miel de sa parole comme on disait* »²⁹² rappelle, pour un lecteur natif de la langue française, les « *paroles mielleuses* », avec la connotation péjorative que l’adjectif « *mielleux* » a couramment acquis dans l’usage contemporain. Or, cette expression calquée sur l’arabe signifie plutôt « *tirer profit de l’éloquence et de l’intelligence de quelqu’un* ».

Il n’en va pas de même pour l’expression « *tant de sel dans ses récits* »²⁹³ qui passe presque inaperçue en tant qu’arabisme aux yeux d’un lecteur francophone ; puisque cet emploi figuré²⁹⁴ est attesté aussi bien en français qu’en arabe.

Nous avons relevé d’autres calques où les idiotismes de la culture marocaine sont

²⁹¹ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 40

²⁹² Edmond Amran EL MALEH, op. cit., 2002, p. 68

²⁹³ Edmond Amran EL MALEH, op. cit., 2002, p. 147

²⁹⁴ Le sel est employé ici au sens figuré de « charme »

explicités par une retranscription de l'expression originale accompagnée d'une glose, plutôt que d'une traduction, à titre d'exemple : « *ce pain qu'il a partagé avec nous c'est le pain du flic, Taâm el boulice Yak* »²⁹⁵ ; cette expression métonymique induit l'idée de trahison, parce que le faite de « *manger le pain de la police* » c'est en quelque sorte « *être compromis avec la police* ». De même, cet idiotisme renvoie le lecteur explicitement à l'expression « *partager le pain* » avec ses connotations de cordialité, de fraternité et de confiance.

Parfois l'auteur demeure dans la neutralité de la traduction littérale pour rester fidèle à l'esprit marocain qu'il cherche à dupliquer, comme dans les expressions suivantes : « *bla hachma bla hia, sans honte et sans pudeur* »²⁹⁶, ou bien « *ana bellah ou be chrâa, au nom de Dieu et de la justice* »²⁹⁷. Nous avons repéré, au niveau du corpus, d'autres exemples du calque phraséologique où la traduction littérale tient une place privilégiée, tel l'exemple suivant : « *ach ka tqoul, que me dites-vous là, que me dis-tu là, le vous n'existe pas dans cette langue de tendresse* »²⁹⁸. En effet, l'auteur utilise tous d'abord une traduction intelligible est traduit la phrase suivante : « *ach ka tqoul* » en « *que me dites-vous là* » avant de se rattraper et faire une traduction littérale pour montrer que le vouvoiement n'existe pas dans sa langue mère qu'il désigne par « *langue de tendresse* ».

En parcourant l'œuvre d'El Maleh, nous avons remarqué également que les calques construits sur la base d'une traduction littérale infèrent au discours une dimension humoristique. En effet, l'auteur, qui est hanté par l'idée de transmettre les

²⁹⁵ Edmond Amran EL MALEH, *Parcours immobile*, op. cit., p. 168

²⁹⁶ Edmond Amran EL MALEH, op. cit., 1988, p. 47

²⁹⁷ Edmond Amran EL MALEH, op. cit., 2004, p. 11

²⁹⁸ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 213

valeurs et le mode de pensée d'un univers qui n'est pas francophone, s'amuse quelque fois à traduire littéralement des noms propres, comme : « *Ibn Muqla* »²⁹⁹ qu'il traduit par « *Fils du blanc de l'œil prune* », ou bien le surnom « *Sebaasoldi* »³⁰⁰ qu'il traduit par « *septsous* ». Nous avons relevé également des noms de lieux, comme : « *Jamaa el Fna* »³⁰¹ qu'il transpose en espagnol par « *Campo di Fiore* » et des noms communs tel « *Mokhadarat* »³⁰² qu'il traduit par « *les verdoyantes* ».

A partir des exemples cités, nous constatons que l'auteur recourt au calque en utilisant tantôt la traduction littérale, tantôt la traduction intelligible, en favorisant toutes sortes de manipulations syntaxico-sémantiques, de jeux de mots et de figures, comme si, dans le passage d'une langue à une autre, l'idée absolue de l'intraduisible, crée chez lui une envie démesurée de « traduire », autrement dit, d'explorer toutes les possibilités du code d'écriture. De plus, le calque de la phraséologie populaire et les catachrèses fossilisées se trouvent réinventées dans l'acte d'écriture littéraire et traductif. En effet, la traduction littérale permet à El Maleh d'introduire cette substance « exotique » dans la langue française et d'y inscrire une identité particulière.

²⁹⁹ Edmond Amran EL MALEH, *Parcours immobile*, op.cit., p. 174

³⁰⁰ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, op.cit., p.12

³⁰¹ Edmond Amran EL MALEH, op. cit., 1998, p. 153

³⁰² Edmond Amran EL MALEH, op. cit., 2002, p.126

2. Calque morphologique

Selon Boudjema AZIRI³⁰³, ce type de calque consiste en la formation et la conception de mots nouveaux par dérivation, par composition ou par affectation d'un signifié nouveau à un mot déjà existant dans la langue cible. De ce fait, ce procédé donne lieu à une nouvelle forme et un nouveau sens contribuant ainsi à l'enrichissement de la langue d'accueil. Sa formation s'opère sur le modèle phrastique de la langue d'origine et généralement sur un ordre syntagmatique ou phraséologique. C'est ainsi que nous avons pu avoir des expressions syntagmatiques comme : « *gratte-ciel* » à partir de « *sky-scraper* » et « *lune de miel* » calqué sur « *honey moon* », qui sont obtenues par traduction littérale à partir de l'anglais.

Pour ce qui est de l'emploi de ce type de calque dans le contexte littéraire étudié, nous avons relevé de nombreux néologismes qui sont formés suivant le procédé morphologique de l'anglais. Il s'agit de ***Tajine-party***³⁰⁴ qui est calqué sur des mots comme : *Lunch-party*, *Dinner-party* ou bien *Cocktail-party*, et de ***Buffet-tajine***³⁰⁵ calqué à son tour sur *Buffet-dinner*.

De telles constructions infèrent au texte malézien une sonorité étrangère. En effet, le rythme différent de deux langues réinvente la parole originelle, la réactive et lui procure une expressivité différente. D'où un appel implicite à l'ouverture sur l'autre pour favoriser des métissages culturels et linguistiques.

De même, nous relevons de nombreux calques morphologiques conçus suivant le

³⁰³ Boudjema AZIRI, *Néologismes et calques dans les médias amazighs : origines, formation et emploi*, Alger, Haut Commissariat à l'Amazighité, 2009, p.76

³⁰⁴ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, op.cit., p. 188

³⁰⁵ Ibid., p. 190

procédé de composition de l'arabe (standard et dialectal), c'est le cas de :

Tableau III. 50 : Calques morphologiques

Calques morphologiques	Équivalents fournis par
La rue des menuisiers	Rue des Njara
La science des lettres	Ilm al hourouf
L'œil de mica	Aïn el mica
Rue des anglais	Derb Lingliz
La grande rue des juifs	Le derb al yahoud
Les figues de nazareth, de la chrétienté	Kermos ennassara
Personnalité importante	Chakhssia oodma
Les gens du livre	Ahal El Kitab

Dans les exemples cités, l'auteur fournit l'équivalent de l'expression calquée en arabe standard ou dialectal, peut-être pour montrer au lecteur non averti qu'il s'agit d'une transposition de l'arabe en français de certaines appellations propres au contexte culturel maroco-musulman. Cependant, il y a des fois où l'auteur, ne donne aucune explication ni de traduction aux expressions calquées, comme :

Tableau III. 51 : Calques morphologiques non traduits

Calques morphologiques	Équivalents en arabe
Le calife des croyants	[xalyfat lmo'minīn]
La salle d'eau	[bīt lma]
La route du haschich	[trīq lhjī]

La particularité de ces exemples, qui infiltrent une nouvelle forme et un nouveau sens dans la langue d'accueil, demeure assurément dans le référent culturel transplanté dans la langue hôte ; un référent qui reste étrange à l'univers socioculturel

français.

Finalement, nous pouvons dire que la conscience de la singularité identitaire qui anime l'auteur, le pousse non seulement à transposer littéralement l'ordre des mots tel qu'il se manifeste dans la langue d'origine, mais à déconstruire la langue hôte (le cas des mots composés français/arabe) et de lui imposer un nouvel élément qui lui est étranger.

3. Calque sémantique

Pour Christian Nicolas³⁰⁶, le calque sémantique est « le processus de transfert de signification (d'une langue A à une langue B) et aussi, par métonymie, le résultat de ce transfert ». Il signale aussi la complexité de ce processus qui relève à la fois du linguistique et de l'extralinguistique.

Autrement dit, le calque sémantique se manifeste par l'ajout d'un nouveau sens (ou trait sémantique) au sens courant d'un terme et entraîne, par ricochet, un enrichissement sémantique de la langue d'accueil en adoptant une nouvelle acception empruntée au mot de la langue d'origine, avec lequel il partage une certaine ressemblance formelle.

Pour illustrer l'emploi du calque sémantique, nous citons un exemple³⁰⁷ en rapport avec le verbe « réaliser ». En effet, ledit verbe qui signifie traditionnellement « accomplir, concrétiser, etc. » a pris, depuis la fin du XIX^{ème} siècle, une autre acception, celle de "se rendre compte, comprendre" en calquant le sens du verbe « to

³⁰⁶ Christian NICOLAS, « *Le procédé du calque sémantique* », In : Cahiers de Lexicologie, n° 65, 1994, p. 75

³⁰⁷ Exemple donné par Dubois dans : J.DUBOIS et al. *Le Dictionnaire de Linguistique et des Sciences du Langage*, Paris, Larousse, 2012, p. 74.

realize » de l'anglais.

S'agissant de notre corpus nous avons retenu pour ce type de calque des emplois variés. Prenons tout d'abord le cas du mot **purification** pour lequel Le Petit Robert enregistre les acceptions suivantes :

1) Cérémonie, rite par lequel on se purifie « ablution » : *Purification par l'Eau. Les purifications prescrites par la loi islamique.*

2) Opération par laquelle on sépare une substance de ses impuretés ;

3) Purification ethnique.

Cependant, dans *Aïlen ou la nuit du récit*, El Maleh attribut au mot **purification** une signification toute particulière :

« Le ciel s'emplit soudain de l'éclat cuivré de la ghaïta, le cortège surgit à l'horizon : sur un cheval rouge miel, brides et selle brodées d'or, les pieds dans des étriers rectangulaires, ciselés d'argent, un homme, drapé d'un glorieux selham blanc, koumia en argent en bandoulière, tenait assis devant lui un enfant. Les drapeaux verts claquaient au vent comme des vagues brisées, le cortège suivait transpercé de youyous lancés vers le ciel, lentement se dirigeait vers le marabout. Circoncision ! Thara, **purification**³⁰⁸. »

Il apparaît, à travers cet exemple, que le sens du mot **purification** est ambigu, de plus, il ne réfère à aucune des significations retenues en français standard. Ainsi, nous proposons de recourir au contexte pour essayer de lever l'ambiguïté et découvrir le sens exact de son emploi.

Nous remarquons, tout d'abord, que l'utilisation du mot **purification** est liée à un contexte cérémonial et rituel pratiqué aussi bien chez les musulmans que chez les

³⁰⁸ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, op.cit., p.148

juifs, à savoir, la circoncision. De plus, le mot *purification* apparaît juste après *circoncision* et *Thara*³⁰⁹, cela suppose qu'il y a une relation sémantique entre ces trois mots. En cherchant la définition du mot *Thara* dans le dictionnaire Prémare nous aboutissons à ce qui suit :

1- Sainteté, pureté morale (en général)

2- Spéc. Etat de pureté rituelle exigée par la prière musulmane

3- Circoncision

De ce fait, nous remarquons que les définitions de *Thara* et de *purification* s'accordent, uniquement, sur une seule acception. Cependant, le contexte et l'ajout du mot *circoncision* au trio, laisse deviner que le mot *purification* n'est qu'un calque sémantique de *Thara* et que les trois mots renvoient au même référent qui est la *circoncision*.

Un autre exemple illustre le procédé du calque sémantique à travers l'emploi du mot *gazelle* :

« "Alors tu vas épouser la fille de Liahou, mbark messood, félicitations. Ghzala, ghzala, une *gazelle*" disait rabbi Chalom sur le ton familier et quelque peu complice d'un vrai connaisseur.³¹⁰ »

En fouillant dans Le Petit Robert, nous constatons que le mot *gazelle* provient de l'arabe gazâl, gazala ; il s'agit d'un mammifère (bovidé) à longues pattes fines, très répandue dans les déserts d'Afrique et d'Asie. Or, dans la citation ci-haut, il semble

³⁰⁹ « Thara » étant la traduction en arabe dialectal du mot circoncision

³¹⁰ Edmond Amran EL MALEH, *op.cit.*, 2002, p. 148

que **gazelle** ne réfère pas à ce qui a été mentionné dans Le Petit Robert. Cependant, en recourant au référentiel culturel marocain, nous pouvons constater que le Rabbi Chalom utilise ce mot pour louer la beauté de la fille de Liahou, en calquant la signification du mot **gazelle** sur le mot *ghzala*³¹¹ qui signifie en arabe dialectal « une jeune fille jolie et délicate³¹² ».

La même remarque reste valable pour le mot **poisson** évoqué dans la citation suivante :

*« Krimo emporté par la fièvre et l'ivresse imaginait, [...], des scènes inouïes accordées à la violence de son amour, quand elle arrivait, franchissant des siècles d'attente, le seuil de la petite chambre de Krimo sur la terrasse d'une maison enfouie dans un derb, alors la terre tremblait d'un long cri de désir aigu, voulu, d'une passion torrentielle de chair lancée sur ses flamboiements, Rabiaa déchirait l'anonymat de son haïk, le voile des conventions : "Houta, houta !" disait Krimo, **un poisson**, elle bondissait dans le tumulte des eaux vives, dans les vagues lancées à la conquête du ciel, chevaux de feu, elle le submergeait, l'abandonnait sur la rive heureux, anéanti, Krimo s'envolait, ses pieds ne touchaient plus l'arête dure du trottoir, [...].³¹³ »*

Dans cette citation l'auteur présente le mot **poisson** comme une traduction du mot *houta*. Alors qu'en est-il de la signification de ces deux mots, chacun dans sa langue de référence. Le Petit Larousse illustré définit le mot **poisson** de la manière suivante : « un vertébré aquatique, respirant toute sa vie au moyen de branchies et pourvu de nageoires locomotrices ». Tandis qu'en arabe dialectal le mot **houta**³¹⁴ est un nom unitaire de *hut* qui signifie *poisson* ; cependant, il a des emplois figurés qui peuvent

³¹¹ D'après *Lisanu Al-Arab*, le sens du mot *gazala* renvoie également au soleil pendant le moment de son lever. De ce fait, *gazala* serait l'incarnation d'une lumière éclatante et d'une beauté naturelle sublime, chose qui peut expliquer la relation entre le concept *gazala* et celui de la femme chez les arabes. (*Lisanu Al-Arab*, vol. 11, pp. 588)

³¹² PREMARE, A-L., et al, Dictionnaire ARABE-FRANÇAIS, Paris, l'Harmattan, 1993.

³¹³ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 180

³¹⁴ PREMARE, A-L., *op.cit.*, 1993

variées selon qu'il est attribué au sexe féminin ou masculin. S'agissant de la femme, l'insertion du mot **houta** dans les expressions *Houta shabla* ou *houta bouria* signifie « une belle femme au teint éclatant », néanmoins dans l'usage courant nous pouvons dire uniquement **houta** pour décrire la beauté d'une femme. Pour ce qui est de l'extrait cité *supra*, Krime, qui est charmé de sa bien-aimée, il l'assimile à un **poisson** pour décrire sa beauté physique. De ce fait, il attribue au mot *poisson* une nouvelle acception, un nouveau trait sémantique celui de beauté, en calquant le sens du mot *houta*.

Un autre exemple qui peut illustrer davantage l'usage du calque sémantique est le mot **frère** utilisé, dans la citation suivante, avec une acception tout à fait différente de celle du français standard :

*« Mounir, tu vas sans doute être étonné, j'aurais pu commencer cette lettre par cher très cher ami, **frère**, à quoi bon? J'aurais pu aussi commencer par la formule consacrée "Al hamdou li Allah", mais en français, cette langue que nous utilisons, ça paraît absurde ! »*

En arabe dialectal le mot **frère** ne désigne pas uniquement une personne née d'un même père ou d'une même mère, mais il s'étend à toute personne de sexe masculin avec qui l'on se sent des affinités en rapport avec la tribu, la race, la nationalité, la religion, les liens politiques ou affectifs, etc. Dans l'extrait cité ci-dessus, il s'agit de deux amis qui partagent un intérêt profond pour le militantisme national. C'est la raison pour laquelle l'émetteur de la lettre a songé utiliser le mot **frère** pour interpeler son ami, cependant il s'est rendu compte qu'« *en français, [...], ça paraît absurde !* ».

Nous avons repéré un autre concept qui enrichit, de plus en plus, le procédé du calque sémantique, dans l'œuvre d'El Maleh, et ce par l'utilisation du mot **louve** auquel correspond la définition suivante :

1) Femelle du loup ;

2) Fig : 1- Outil pour le levage des pierres de taille (levier) 2- Filet de pêche, verveux à deux entrées opposées.

Dans le corpus, nous avons relevé l'emploi du mot **louve** à travers l'extrait cité *infra* :

« Saïdati, Lalla, [...], si le froid pénètre vos os, si la diba, **la louve**, la crampe mord vos mollets, [...], nous apportons apaisement à la douleur, consolation, guérison, [...].³¹⁵ »

Ici, le mot **louve** est employé avec une charge sémantique d'origine dialectale, en effet, l'auteur lui procure le sens du mot **diba**³¹⁶, auquel le dictionnaire Prémare réserve la définition suivante : « Fourmillements ou ankylose dans un membre ; crampe dans les doigts ou les orteils ». Ainsi, l'auteur a conféré au mot **louve** un sens plus étendu en lui rajoutant une acception qui lui est étrange, celle de crampe.

Un dernier exemple tiré de *Parcours Immobile* est celui de **marieuse** que Mercier³¹⁷ traduit par « zuwwaja »³¹⁸. Il s'agit là aussi d'un calque sémantique qui correspond à une réalité étrangère à la France contemporaine, alors qu'au Maroc et au Maghreb en général, c'est un rôle social très bien codifié.

En somme, nous constatons que les calques sémantiques jalonnent le texte d'El Maleh, traduisant ainsi des notions spécifiques à la culture marocaine. Dans tous les

³¹⁵ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 18

³¹⁶ PREMARE, A-L, *op. cit.*, 1993

³¹⁷ MERCIER H, *Dictionnaire ARABE-FRANÇAIS*, RABAT, éditions la porte, 1951

³¹⁸ C'est une femme qui, grâce à son expérience, joue les entremetteuses afin de rapprocher deux personnes en vue de conclure un mariage.

exemples précédemment cités, l’auteur tantôt, il explique le mot calqué, comme dans le cas de *louve* où il montre qu’il s’agit d’une crampe, tantôt il laisse au lecteur de deviner le sens en recourant au contexte extralinguistique. Cependant, quelques fois la signification du mot pourra être inaccessible, par exemple un lecteur francophone natif n’ayant pas l’habitude de concevoir une quelconque relation entre une *femme* et une *gazelle* ou bien même une *femme* et un *poisson*, ne comprendra certainement pas le sens de cet usage. Mais ceci, ne réduit en rien l’apport du calque sémantique dans la transposition, en langue française, de la culture populaire marocaine et dans l’enrichissement de certains mots qui ont vu leur champ sémantique s’étendre pour englober des acceptions différentes à celle décrites dans le français standard.

C. Alternance codique

Nous avons constaté d’autres particularités langagières oscillant entre plusieurs catégories linguistiques, à savoir l’emprunt, le calque et d’alternance codique³¹⁹. D’où la difficulté de leur tracer des contours définitoires. En effet, déterminer si de telles occurrences appartiennent à telle ou telle catégorie nécessite une étude approfondie qui ne relève pas de notre objet de recherche. Toutefois, nous nous contenterons de dresser un bilan des différentes langues utilisées dans l’AC pour mettre en évidence l’hétérogénéité linguistique de l’écriture d’El Maleh qui est, aussi, révélatrice de l’hétérogénéité culturelle du Maroc.

Ainsi, l’analyse du corpus révèle que l’arabe est pourvoyeur de 70% des alternances codiques recueillies dont 63% sont en arabe marocain. Nous retenons également que l’usage de l’arabe standard en alternance avec le français est lié surtout à la religion musulmane, la spiritualité et le savoir. Nous avons choisi les exemples

³¹⁹ Désormais AC

suivants à titre d'illustration :

- **Religion musulmane**, ex :

« *Sa propre main l'avait écrit dans ce cahier rouge, **dikr**, répétition du nom de Dieu, sacrilège !* ³²⁰ »

- **Spiritualité**, ex :

« *Khalili chante, psalmodie, accompagné ponctué par un battement léger rythmé sur le bendir, récitatif **quasida** célèbre dans la confrérie des Hmadcha* ³²¹ »

- **Savoir**, ex :

« *Il écrit, penché sur une feuille de parchemin, trempant le roseau, le **qalam**, dans une fiole d'encre préparée par ses soins* ³²² »

Quant à l'alternance codique français/arabe marocain, elle est surtout liée aux croyances populaires (sorcellerie, médecine traditionnelle, etc.) et à certains aspects de la vie quotidienne (Folklore, cérémonie, commerce, etc.) tel que le montrent les extraits suivants :

³²⁰ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 63

³²¹ Ibid., p. 19

³²² Edmond Amran EL MALEH, *op. cit.*, 1998, p. 158

- **Médecine traditionnelle**, ex :

« **Saïdati, Lalla**, il s'adresse aux femmes, voix métallique, technique de pointe, si vous avez mal aux dents, ici on n'arrache pas on calme, si le froid pénètre vos os, si la **diba**, la louve, la crampe mord vos mollets, si vos cheveux cassent sous le peigne, tombent, si votre enfant fait pipi au lit, si les nerfs **aasseb** vous tourmentent de jour et de nuit, si vous avez des vertiges, **dokha**, si votre matrice nécessite un désinfectant, je vous le jure, je vous le dis, nous ne sommes pas venus ici pour de l'argent, avec l'aide de Dieu et de son prophète **Mohammed Rassoul Allah**, nous apportons apaisement à la douleur, consolation, guérison, la voix intarissable, flot métallique dominant la rumeur du souk, les femmes voilées se pressent, de main en main passent les petites fioles magiques³²³ »

- **Sorcellerie**, ex :

« **Aachoubi, aatar**, épicier, herboriste, **sahar**, intermédiaire à l'occasion, Moha, [...] entretient plusieurs conversations en même temps: une cuiller de **tehmira** rouge pour la petite fille effrontée, du cumin pour l'homme digne, silencieux, du **ghassoul**, le shampoing invincible, pour la jeune femme voilée, [...] la jeune femme, celle-là à côté, elle parle à voix basse[...] " il faut que ce soit du vrai, pas de la tricherie, **yak**, n'est-ce pas", non discret, rassurant, murmure " **Kouni hania, gheda Inchallah**, sois tranquille, demain si Dieu veut..." [...]. De la poudre de vagin de porc-épic femelle, **derbana**, Moha est convaincu, convaincant : le lion se soumettra comme un enfant timide et sans force. Elle en fera ce qu'elle voudra. Moha laisse filtrer sa certitude à

³²³ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, op.cit., p. 18

travers ses yeux mi-clos: **Dbâa** ! idiot, tête vidée, hagard, hébété, envoûté, pantin jeté à la merci des caprices de la femme [...]!³²⁴ »

- **Cérémonie**, ex :

« il voyait une immense salle de réception, un orchestre de musique andalouse sur une estrade, des tables rondes, des **méchouis**, des agneaux entiers superbes, des montagnes de **couscous**, des serveurs muets en **tarbouche** rouge, gilet brodé, ample **saroual**, silencieux, solennels, efficaces, le festin, des nus, des morts, des mendiants hâves en haillons erraient dans les couloirs, ramassaient les restes dans des bidons, portaient à leur bouche impatiente des poignées de **couscous**, leur main tremblait, la graine se répandait par terre, ces gens-là des bêtes, il était dans la salle d'audience, est-ce qu'il rêvait, c'était lui ou un autre, en **djellaba** blanche, **burnous** blanc, **bzioui**, laine tissée blanche de finesse et de légèreté sur un cœur lourd gorgé de noirceur, **beïia**, serment d'allégeance, soumission, des siècles de tradition, la main qu'il venait d'embrasser lui intimait l'ordre de se relever.³²⁵ »

- **Commerce**, ex :

« ses clients de vrais amis, "**Aioua Yeshuaa ! aamal maana taman**, fais-nous un prix: **khlass, khlass**, allez va, je jure par Dieu, **khlass!**" les rires, les plaisanteries, la mise en scène, le marchandage de l'amitié et de la confiance, Yeshuaa qui ne le connaissait sur le souk, la confiance, la limpidité de la parole donnée, combien de

³²⁴ Ibid., p. 150

³²⁵ Ibid., p. 47

fois il avait prêté de l'argent, vendu à crédit "**gheïr aabih**" ne t'inquiète pas, tu me paieras quand Dieu voudra.³²⁶ »

- **Folklore**, ex :

« Juan Goytisolo, en son lieu d'élection *Jamaa el Fna*, vit dans la proximité de ces **hlaiqui**, ces poètes conteurs, narrateurs qui font cercle autour d'eux, la **halqa**, ce cercle aimanté par la magie de la parole, attirant les foules charmées, suspendues à cette parole éphémère, renaissante à chaque fois dans sa vierge vivacité, célébrant de grandes épopées.³²⁷ »

Notons que l'auteur alterne souvent entre le français et l'arabe dialectale en utilisant des expressions idiomatiques formées à partir du nom d'Allah et qui, sur le plan pragmatique, n'ont plus grand-chose de religieux. Lesdites expressions sont fréquemment insérées dans les textes d'El Maleh tantôt sans traduction, tantôt avec traduction, par exemple³²⁸ :

- *Inchallah, si Dieu veut*
- *Dieu nous garde, Allah ihfed*
- *Allah yaoudi*
- *Allah ou Akbar, Dieu est le plus grand*

³²⁶ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 154

³²⁷ Edmond Amran EL MALEH, *op. cit.*, 1998, p. 149

³²⁸ Pour les exemples qui suivent nous avons mentionné uniquement les expressions qui sont source d'AC français/arabe pour éviter l'encombrement de la thèse par les extraits.

- *Allah irhmo , que Dieu ait son âme*
- *Allah, Allah !*
- *OuAllah*
- *Allah ihedihoum, que Dieu les remette dans le droit chemin*
- *Hamdou li Allah : louange à Dieu*
- *Allah ijaal salama, Que Dieu nous assure le salut*
- *Tbark Allah aaleh, que la bénédiction de Dieu descende sur lui.*
- *Allah Ibqi ster, Dieu nous garde !*
- *Allah ister, que Dieu nous protège*
- *Allah yarhmo, que Dieu l'ait en sa miséricorde*
- *Allah irhmo que Dieu ait son âme*
- *Les jours ont fui, Soubhana Allah [...] El bla*
- *Allah IHdihoum*
- *Allah, Allah ya rbbi, mon Dieu, pourquoi ? Alash, Alash ?*

Des locutions exclamatives sont aussi répertoriées dans les alternances codiques français/arabe (standard ou marocain), comme :

- *Ha al aar ! Ya latif, ya latif !*
- *aajib ! là ? étonnant, non ?*

- *ihstra ! daba tchouf, tu vas voir !*
- *ouili, ouili !*
- *Ya latif, ya latif ! Malheur, malheur !*
- *Ihsra, ihsra ! mchat el yam ! tqada dak che ! où est ce temps ? tous cela est fini.*
- *Hiba, hiba ! quelle prestance !*
- *khlass, khlass !*
- *yaouili ! yaouili !*
- *khlass a mama ! cela suffit*
- *Elle était partie avec un légionnaire à Marrakech "Baz" ! quelle audace !*

Les alternances codiques français/arabe marocain est une occasion pour El Maleh de nous offrir un champ fertile d'insultes typiquement marocaines, comme :

- *Oulad el hram ! fils du péché !*
- *Oulad el khab ! fils de pute !*
- *L'homme de la rue, oueld souk, sans famille*
- *Hmar, hmar, âne, âne, ignorant*
- *Sabagnolé merqaa esseroual, cet espagnol au pantalon rapiécé, zbel, ordure*

- *Koufar billah [...]tfou !*
- *Allah yaatikoum el bla, soyez maudits, que la malédiction de Dieu s'abatte sur vous.*
- *Ces athées fourvoyés, kofar billah ! zanadek !*
- *Yallah zid limak, allez viens, la putain de ta mère*
- *Din el klab, oulad el qhab, cette race de chiens, ces fils de pute*
- *Kafer bi Allah*
- *Naalat Allah Aala ek-kadibin, que Dieu maudisse les menteurs !*

Quoique le lecteur ne soit pas natif du dialecte marocain, il lui sera aisé de deviner la dimension injurieuse de certaines insultes. Par exemple, la transposition de certaines expressions en français comme : « fils du péché », « fils de pute » ne les dépayse pas trop, du moment que dans la langue française l'insulte populaire se construit volontiers sur le modèle « fils de N ». De même, cette alternance codique qui fait une large place aux exclamations, interjections et insultes, est une manière de figurer la résonnance authentique du parler populaire.

Quant aux alternances codiques français/judéo-arabe, ils renvoient surtout au domaine liturgique :

*« C'était donc, à suivre ce récit, en dépit de ses imperfections, de ses maladresses et de son incapacité à faire revivre la fête de ce dîner la veille de la **pâque juive**, le seder, la nuit de la **Hagada**, réunissant l'hôte, son illustre invité et nombre de convives, fête religieuse, marquante dans l'histoire des Hébreux, symbole d'une*

*convivialité sans distinction de confession*³²⁹ . »

Mais ce sont les AC qui relèvent de la vie quotidienne qui sont les plus attrayants vu leur sonorisation particulière³³⁰ :

*Ex 1 : « Ouhad **Zdioua** lijabli baba bi trezo ime, le petit chevreau que mon père m'a apporté pour le prix de trois deniers*³³¹ »

*Ex 2 : « "**Maiitzbach**" qu'il ne sorte pas! cela ne figurait pas dans le dictionnaire il ne savait pas d'ailleurs à quoi servait ce livre.*³³²»

*Ex 3 : « Ghadda bzahd Allah, demain si Dieu veut »*³³³

Pour ce qui est de l'AC français/espagnol, elle apparaît notamment en concomitance avec le parler *jebli* utilisé au nord du Maroc. Elle révèle l'une des traces du brassage linguistique et culturel des deux communautés marocaine et espagnole au Maroc. D'ailleurs, Edmond Amran El Maleh signale à maintes reprises l'existence de citoyens espagnols qui ont vécu sur le territoire marocain. De *Aïlen ou la nuit du*

³²⁹ Edmond Amran EL MALEH, *op. cit.*, 1998, p. 158

³³⁰ Nous reviendrons sur le sujet de la prononciation du judéo-marocain dans le point réservé aux interférences phonétiques

³³¹ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 158

³³² Edmond Amran EL MALEH, *Parcours immobile, op.cit.*, p. 44

³³³ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit, op.cit.*, p. 111

récit, nous choisissons l'extrait ci-dessous où l'espagnol est source d'AC avec le français :

*« prosaïquement il (don Fernando) est un des derniers Espagnols restés sur place après le départ des **conquistadors**, l'un des derniers survivants avec Antonio, sorte d'oiseau déplumé, silencieux, qui veille sur la caisse, une vraie **caja** en métal qui se ferme et s'ouvre juste le temps de rendre la monnaie, **señora**, la **dueña**, de noir vêtue, les cheveux gris tirés en chignon, menue, le visage travaillé par la beauté des rides, dans sa minuscule cuisine, s'active invariablement, **boquerones**, **pescado frito**, **paella**, colonisation espagnole de petites gens entre la **zarzuela** et le drame, [...]. »³³⁴*

Quant à l'AC français/berbère, elle apparaît d'une manière occasionnelle. Elle se manifeste dans des énoncés où les locuteurs sont des sujets berbères :

*« Hier il suffisait de louer le couteau pour s'en mettre plein le ventre, aujourd'hui il faut payer à la pièce, **hendia**, **chumbo**, **aknari**, **kermos** ennassara, fruit du destin, fruit métaphysique »³³⁵*

L'AC français/anglais est rarement utilisé du moment que l'anglais est réservé uniquement à une certaine élite lettrée. Les séquences qui contiennent des mots anglais se rapportent généralement aux discussions politiques, par exemple :

³³⁴ Ibid., p. 15

³³⁵ Ibid., p. 53

« Le planton entre un plateau à la main, raideur militaire stricte, ghriba, petits gâteaux d'Iblis, une bouteille de Chivas, **Muslims we are** ! rire énorme, **you're lucky, my fellow**, notre **Parliament** est tellement démocratique qu'il ne nous a pas donné l'autorisation de tirer, **terrific** ! rire à faire trembler les murs ! »³³⁶

La juxtaposition des codes revêt dès lors l'aspect d'une entreprise musicale où il s'agit de faire sonner les mots, de créer une échoïté particulière. En plus, l'intérêt de cette alternance codique ne réside pas tant dans la reproduction des langues qui jalonnent l'univers linguistique marocain que dans son inscription en tant que marque d'oralité au sein de l'écrit ; une oralité qui trouble le rythme et la syntaxe certes, mais qui marque la résonance authentique du parler marocain.

D. Interférences phonétiques

L'empreinte de la culture marocaine dans l'œuvre d'El Maleh n'est pas suggérée uniquement par les mots étrangers qui jalonnent le dire des personnages. Elle est mise en avant dans le texte de manière plus ou moins mimétique par le biais de ce que les linguistes appellent les interférences phonétiques.³³⁷

Généralement, dans un texte littéraire, les interférences phonétiques sont aisément perceptibles, elles résultent d'un gauchissement de la prononciation interprété orthographiquement dans le texte.

S'agissant du corpus étudié, l'apparition des interférences est souvent liée à la

³³⁶ Ibid., p. 37

³³⁷ J. DUBOIS et al, **Dictionnaire de linguistique et des sciences du langage**, Paris, Larousse, 1994, p. 252 : *On dit qu'il y a interférence quand un sujet bilingue utilise dans une langue-cible A, un trait phonétique, morphologique, lexical ou syntaxique caractéristique de la langue B.*

volonté de l'écrivain de mettre en exergue quelques timbres vocaliques relevant de certaines catégories sociales. De *Aïlen ou la nuit du récit* nous citons le passage suivant :

*« A l'Alliance israélite il avait appris le français, lui Baba, mais il zozotait un peu : **Ze** vous fais une friction ? Quelle émotion, il en bafouillait presque, la France entière qu'il drapait dans l'ample peignoir dont il serrait les bouts autour du cou du colonel, **rien di to** ! Baba avait un peu l'accent, peut-être pas aussi prononcé que cela quand même, **quamime** : rien du tout, est-ce qu'il pouvait faire payer le colonel ! **rien di to**, même qu'il avait donné une bouteille d'eau de Cologne pour madame, " malins **ces petits Juifs** et rapides avec ça ! il vous a taillé une brosse comme pas un" disait le colonel. »³³⁸*

Cet extrait rappelle ce que nous avons évoqué à propos du judéo-marocain qui se caractérise par une prononciation particulière. En effet, Baba quoi qu'il ait appris le français à l'alliance israélite, il n'arrive pas à se débarrasser du [z], qui est caractéristique du parler juif au Maroc, ni de son accent dialectal. Ainsi, au lieu de dire « [ʒə] vous fais une friction ? » il remplace la chuintante par la sifflante en prononçant la phrase ainsi : « [zə] vous fais une friction ? ».

Soulignons, aussi, que l'auteur lui-même montre, en fin de citation, que Baba est un juif, ceci n'a d'intérêt que de mettre en relation l'accent du protagoniste avec son origine ethnique.

De même, le protagoniste d'El Maleh, qui est aussi un marocain, trouve une difficulté à articuler correctement certains phonèmes français qui sont difficilement prononçable par un arabophone. Il s'agit de la voyelle nasale [ã] et la de voyelle orale [ɛ] dans le cas de « quand même / quamime » qu'il remplace

³³⁸ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, op.cit., p. 106

par les voyelles orales [a] et [i]. De même, pour dire « rien du tout / rien di to » il substitue les voyelles orales [y] et [u] par [i] et [o].

Dans le même ordre d'idées, l'auteur utilise le mot « sampagna³³⁹ », dans un autre passage de *Mille ans, un jour*, au lieu de « champagne ». Ici, le remplacement de la chuintante [ʃ] par la sifflante [s] est aussi un cas de réalisation phonétique communautairement marquée et traditionnellement réputée propre au judéo-marocain.

Nous citons, également un autre passage de *Aïlen ou la nuit du récit* qui illustre la contamination de la prononciation de Titi, par « du berbère, du chleuh » :

« Les portraits de la famille royale, impérissable empire, **empaïre** comme dit Titi, elle avait fait un peu d'anglais, des leçons chez Mrs Simpson, n'est-ce pas, isn't it ? it was ! prononcez comme Titi en nasillant un peu : **isentititwas**, nasillement contamination **du berbère, du chleuh**, [...] maintenant Baba, l'œil encore vif, mange sa *skhina*, une table en acajou massif, **mahagni**, aventure bâtarde de l'anglais mahogany, cette même table sur laquelle feu Sir Harold prenait son breakfast, **ilbrikfast**, porrezde, comme dit Titi ! »³⁴⁰

Dans cet extrait, El Maleh fait également la liaison entre l'origine ethnique de Titi et sa prononciation altérée par le nasillement qui caractérise sa langue-mère. De même, cette citation nous permet de relever, également, un autre marquage de l'influence de l'arabe dialectal et de sa gutturalité présumée, c'est-à-dire de sa tendance à privilégier les articulations postérieures. Il s'agit de l'extraction d'un pseudo-article « il » de « breakfast » en le prononçant « *ilbrikfast* ».

³³⁹ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 39

³⁴⁰ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 105

L'auteur dans les deux extraits a mentionné que ces protagonistes ont reçu, tous les deux, un enseignement dans une langue étrangère (le français pour le cas de Baba et l'anglais pour le cas de Titi). Malgré ce constat, les deux personnages gardent toujours l'accent qui caractérise leurs parlers d'origine (le judéo-arabe pour Baba et le berbère pour Titi).

De ce fait les interférences phonétiques peuvent être révélatrices de l'accent³⁴¹ qui est, à son tour, une forme de marquage individuel³⁴² et (ou) sociocommunautaire.

Dans les textes d'El Maleh, ces interférences phonétiques marquent également le niveau d'instruction des personnages. En effet, il y a des cas où l'auteur fait parler ses protagonistes des langues variées en toute aisance. Tel Rachid, qui a été formé à Harvard dans un business school high top et qui « *se plaisait à mêler des mots d'anglais à la conversation* »³⁴³. Ou bien Azzedine, professeur d'anglais agrégé, qui a « *une prononciation impeccable* », tel que le montre la citation suivante :

« *Qu'est-ce que je vous sers, menthe, tea, dear Harrison, prononciation impeccable, Azzedine est professeur d'anglais agrégé sans cette misère des traitements.* »³⁴⁴

El Maleh se montre particulièrement sensible à ces fluctuations du langage et commente sans cesse la manière de prononcer des uns et des autres. Dans un passage

³⁴¹ Nous considérerons le terme « accent » au sens extensif d'« ensemble des traits de prononciation qui s'écartent de la prononciation considérée comme normale et révèlent l'appartenance d'une personne à un pays, une province, un milieu déterminés » *TLFI*, article « accent ».

³⁴² Individuel parce que généralement ces réalisations « modulées » aux phonétismes des protagonistes se rapportent également à leur niveau d'instruction.

³⁴³ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 33

³⁴⁴ *Ibid.*, p. 37

de *Aïlen ou la nuit du récit*, l’auteur, encore une fois, relie l’accent à l’appartenance géographique, mais cette fois-ci, c’est de l’arabe dialectal qu’il s’agit : « *Abdelkrim rit, les mots explosent drus, saveur, mélange d’espagnol, d’argot marin, déjà l’accent du Nord, Abdelkrim s’adresse à lui, comment, tu es de là-bas ? hadik dounia, ce monde-là* »³⁴⁵. Ainsi, l’auteur transcrit **dounia** avec allongement du [u] qui est particulier à l’arabe dialectal parlé au nord du Maroc.³⁴⁶

Parfois, les interférences phonétiques sont liées, dans l’œuvre d’El Maleh, à l’évocation nostalgique et humoristique du parler populaire. Prenons à titre d’illustration les exemples suivants :

- « **chabo**³⁴⁷ » : au lieu de chapeau. Cette prononciation mime le parler populaire et les ajustements de timbres vocaliques auxquels procède le dialecte marocain qui est dépourvu de l’occlusive bilabiale sourde [p].

– « **grama**³⁴⁸ » : c’est une déformation-troncation de « grand-maman », dépossédé de ses voyelles nasales imprononçables pour un arabophone.

- « **Tivi**³⁴⁹ », « **pidiji**³⁵⁰ », « **sitiem**³⁵¹ » : au lieu de TV, PDJ, CTM. Ce sont des sigles qui ont été altérés par l’accent de l’arabe dialectal.

³⁴⁵ Ibid., p. 12

³⁴⁶ Quoique cet exemple n’entre pas dans le cadre des interférences phonétiques français/arabe, mais nous avons jugé judicieux de le relever, justement, pour mettre l’accent sur l’intérêt que l’auteur accorde à la sonorité des énoncés de ses protagonistes.

³⁴⁷ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 55

³⁴⁸ Edmond Amran EL MALEH, *Parcours immobile*, *op.cit.*, p. 27

³⁴⁹ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit*, *op.cit.*, p. 28

³⁵⁰ Ibid.

³⁵¹ Edmond Amran EL MALEH, *op. cit.*, 1988, p. 46

– « **guirra**³⁵² », au lieu de « guerre ». Ici, l’adaptation de l’emprunt a affecté également la morphologie de la langue hôte, par le redoublement orthographique de la consonne « r » pour suggérer le [r] roulé et la substitution du « e » muet final par un « a » qui marque le féminin en arabe. Dans un passage de *Mille ans, un jour*, El Maleh s’attarde sur cette réalisation phonétique en vue de souligner le rapport sensuel que l’auteur/narrateur entreprend avec les mots :

« *Kiji dansait autour de ce mot, roulement des r dans sa bouche : guirra ! cri rude jailli de la gorge : harb ! Kiji passait au registre de la langue classique*³⁵³ ».

Pour Jean-Louis Maume³⁵⁴, le **r** que « roulent » avec affectation les hommes maghrébins est une sorte d’affirmation de la virilité par l’accent, une manière d’imprimer son empreinte sur le corps de la langue française. Ainsi, Embarek Kiji, conteur qui semble maîtriser l’arabe classique, recourt néanmoins à ces emprunts au français qu’il adapte au dialecte marocain par le roulement des « **r** » et fait résonner les mots pour le plaisir échoïque et la connotation de leurs sonorités.

- « **Tivasa**³⁵⁵ » : pour télévision. Il s’agit d’une adaptation phonétique qui rime heureusement avec le mot féminin arabe [talfza]. Comme pour le mot « guirra » on a rajouter un « a » qui marque le féminin en arabe. Notons également que le dialecte marocain privilégie les mots qui finissent par des voyelles.

³⁵² Edmond Amran EL MALEH, *op. cit.*, 2002, p. 40

³⁵³ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 40.

³⁵⁴ Jean-Louis MAUME, « L’apprentissage du français chez les arabophones maghrébins », *Langue française*, n° 19, 1973, p. 101

³⁵⁵ Edmond Amran EL MALEH, *Aïlen ou la nuit du récit, op.cit.*, p. 28

– « *sounima*³⁵⁶ », pour cinéma. C’est aussi un emprunt qui a subi une adaptation phonétique du parler populaire.

–« *bissitat*³⁵⁷ » : au lieu de « visites » : la fricative labiodentale sonore [v] qui n’existe pas dans le système phonologique de l’arabe est remplacée par l’occlusive bilabiale sonore [b]. De même, après la consonne finale, est rajouté le morphème arabe de féminin pluriel [āt].

Ces variations phonostylistiques résultent notamment de l’interférence entre deux systèmes phonétiques, celui de la langue d’origine et celui de la langue d’accueil. Elles constituent un marquage de l’appropriation de l’autre langue suite à un évènement donné. Dans le cas de l’arabe dialectal l’incident qui a suscité l’apparition de tels phénomènes est lié à l’établissement du Protectorat. D’ailleurs, El Maleh parle souvent de cette période de l’histoire marocaine comme le commencement de la confrontation avec une modernité qui allait faire trembler les bases d’une culture.

En somme, ces variations de la phonétique française reproduisent le dialecte marocain et son appropriation de certains emprunts au français³⁵⁸. Le fait de les réimprimer dans un texte littéraire fait chanter le français sur un rythme marocain. Ainsi, lorsque l’auteur reproduit l’accent de ses protagonistes, c’est l’hétérogénéité des parlers et des ethnies qui se reflètent autant dans la prononciation que dans les choix lexicaux. En effet, l’auteur quand il fait parler un juif, ce dernier n’articule pas les mots de la même façon qu’un berbère ou bien qu’une personne issue du nord du

³⁵⁶ Edmond Amran EL MALEH, *op. cit.*, 2002, p. 51

³⁵⁷ Edmond Amran EL MALEH, *Parcours immobile*, *op.cit.*, p. 35.

³⁵⁸ Il faut souligner que l’appropriation de l’arabe dialectal de certains emprunts français présente l’interférence comme un véritable mécanisme de naturalisation.

Maroc, etc. De ce fait, le parler des personnages d'El Maleh accuse leur origine géographique et leur statut socioculturel par la figuration de l'accent, qui elle-même traduit, d'une certaine manière, l'identité marocaine.

A la fin de ce chapitre, nous pouvons conclure que les manifestations multilingues dans le texte d'El Maleh se proclament à plusieurs niveaux : emprunts, calques linguistiques, alternances codiques, interférences phonétiques, etc. Bien sûr nous n'avons pas cerné tous les phénomènes langagiers émergeant dans l'œuvre étudiée, mais nous nous sommes penchée sur ceux qui nous semblent utiles pour répondre à notre problématique de recherche.

Ainsi, nous avons articulé ce chapitre en quatre sections. Dans la première, nous avons pu retenir que les emprunts répertoriés, à travers le corpus d'El Maleh, proviennent de diverses langues. Ainsi, nous avons classifiés lesdits emprunts en trois catégories : les emprunts de luxe, les emprunts naturalisés et les emprunts de nécessité.

Nous avons indiqué au niveau de la première catégorie que l'emprunt de luxe, quoiqu'il vienne doubler son équivalent dans la langue française, il ajoute une légère nuance sémantique à la réalité exprimée. De même, nous avons souligné au niveau de la deuxième catégorie que le recourt à des emprunts naturalisés ne pose aucun problème d'ordre sémantique pour le lecteur francophone natif, du moment qu'ils sont déjà attestés dans les dictionnaires usuels du français. Tandis que dans la troisième catégorie, nous avons montré que le maniement d'emprunts de nécessité se justifient par leur utilité, vu l'incapacité de la langue française à transférer un certain contenu sémantique. Le dénominateur commun de tous ces types d'emprunt c'est qu'ils permettent de traduire un univers culturel et linguistique propre à l'auteur.

Au niveau de la deuxième section, il était question du calque linguistique. Ce procédé qui est basé essentiellement sur la traduction recouvre plusieurs types : des calques phraséologiques, des calques morphologiques et des calques sémantiques.

Lors du classement de l'inventaire des calques, nous avons constaté que l'arabe

dialectal est la première langue pourvoyeuse de ce phénomène. L'utilisation du dialecte marocain englobe ainsi toutes les caractéristiques de la langue populaire : grivoiseries, insultes cocasses, surnoms imagés, ce qui confère au texte d'El Maleh truculence et vivacité.

Cet emploi de la langue, bien qu'il peut paraître bizarre pour un lecteur francophone natif mais il permet de traduire les caractéristiques d'une culture indigène sous ses multiples facettes : croyances populaires et religieuses, traditions locales, valeurs morales, pratiques quotidiennes, etc.

Au niveau de la troisième section nous avons montré qu'il existe un autre fait langagier figurant dans notre corpus d'analyse, caractérisé par le passage dynamique d'une langue à une autre ; il s'agit de l'alternance codique. Ainsi, nous avons constaté que l'alternance du français avec des langues en présence dans le paysage linguistique marocain est surtout lié à la religion musulmane, aux croyances populaires et à certains aspects de la vie quotidiennes, etc. De ce fait, dans le texte d'El Maleh, les langues s'entrelacent plus intimement, comme pour rendre compte d'une réalité autre, mais qui reste propre à l'écrivain.

Au-delà de la simple juxtaposition codique, nous avons vu, au niveau de la quatrième section, que l'auteur use de l'interférence phonétique pour rendre un accent, dialectal en l'occurrence, au sein du français auquel il imprime une couleur particulière. En effet, El Maleh actualise des emprunts qui sont adaptés au système phonétique et (ou) morphologique de l'arabe dialectal, avec ses multiples variantes, comme : la sonorisation du [p] en [b] et du [ʒ] en [z]. Ainsi la tendance de l'auteur est de rester au plus près de la matière sonore d'origine. D'ailleurs, les personnages à qui sont attribuées ces réalisations « déformées » du français, sont bien évidemment des représentants d'une culture nationale traditionnelle, en plus, cette exhibition de

l'accent mime le timbre vocalique de certains parlers marocains, elle peut même servir à suggérer, avec humour ou ironie, l'appartenance géographique, l'origine ethnique, le niveau d'instruction ou le statut socioculturel d'un personnage quelconque.

Témoins d'une richesse linguistique et culturelle, ces pratiques langagières concourent de différentes manières à la mise en évidence de la culture marocaine. Elles contribuent au portrait d'un patrimoine national sous ses multiples facettes. Cependant, ces phénomènes langagiers quand ils émergent dans le contexte littéraire, requièrent une double compétence de la part du lecteur : une linguistique et une autre culturelle pour mieux cerner les débouchés de chaque usage.

Dans cette dernière partie nous nous sommes intéressée à l'identité du vocabulaire et à son sémantisme en empruntant trois voies : l'étude des rapprochements et des éloignements entre les différents textes, l'examen du vocabulaire spécifique d'El Maleh et finalement l'analyse du plurilinguisme dans l'œuvre en question.

Ainsi, nous avons commencé par l'examen de la connexion lexicale. Tout d'abord en utilisant la méthode Jaccard qui ne considère que la présence ou l'absence des mots dans les textes considérés. Ensuite, nous avons eu recours à une autre méthode qui tient compte de la fréquence des mots. La première méthode nous a permis de dresser une liste de couples qui s'attirent soit parce qu'ils partagent le même contenu lexical, le même genre ou bien la même période de production. Il s'agit bien évidemment des couples (GENET, ZRIREK) ; (AILEN, MILLE) et (ELHAKI, PARCOURS).

Pour compléter l'analyse et exprimer la connexion lexicale de manière plus détaillée, nous nous sommes intéressée à la distance intertextuelle. Également dans cette étape, la parenté des textes GENET et ZRIREK était prouvée, du moment qu'ils se partagent le même genre mais aussi des thèmes comme : la littérature, les écritures mystiques et l'allégorie. Cependant, les textes qui restent les plus éloignés sont ABOUNOUR et FEMME. Ce qui signifie qu'ils sont caractérisés par un vocabulaire privatif spécifique à chacun d'eux.

Ensuite, nous avons calculé l'indice de connexion lexicale qui a également montré une forte parenté entre GENET et ZRIREK. Par ailleurs, tous les binômes qui ont intégré le texte FEMME ont enregistré un indice de connexion trop faible. Ceci reste normal du moment que l'écrivain, vers la fin de sa vie, avait le souci de sombrer dans une sempiternelle répétition. C'est la raison pour laquelle il a entamé l'écriture du texte FEMME qui reste plus ou moins différent des autres, particulièrement sur le plan thématique.

Pour remédier aux insuffisances de la méthode Jaccard qui ne tient pas compte de la fréquence des mots, nous avons testé une autre approche empruntée à Ch. MULLER, à savoir la loi binomiale. Cependant, cette dernière n'a pas satisfait notre curiosité du moment que les résultats obtenus n'étaient fiables que pour des couples de textes dont les étendues étaient voisines.

L'examen de la connexion lexicale a été complété par l'étude du vocabulaire spécifique qui a mis en valeur les penchants thématiques de l'œuvre étudiée. Soulignons que nous avons analysé les spécificités lexicales d'El Maleh suivant deux logiques. La première, de nature exogène, a mis l'accent sur les spécificités externes du corpus global, en considérant le *T.L.F* comme norme et base de calcul. La deuxième, de nature endogène, a pris en charge les spécificités de chaque texte par rapport au corpus entier.

Ainsi, l'étude exogène a révélé une forte inclination de l'auteur pour les thèmes de la littérature, de l'écriture, de l'allégorie, de la mémoire, de l'histoire, de la quête, de l'identité, de l'art culinaire, etc. Tous ses thèmes, et bien d'autres, participent à la mise en évidence de la mémoire et de l'histoire marocaine, du moment qu'ils renvoient le lecteur à un espace précis et à une époque révolue. D'ailleurs, notre auteur qui écrit à partir d'un choc émotionnel ne veut pas laisser le passé se dérober sous silence, bien au contraire, il le déploie dans tous les sens et sur tous les registres, parce qu'un passé remémoré est synonyme d'avenir. C'est la raison pour laquelle l'auteur donne la possibilité à ses lecteurs de capter les frémissements de la vie et les choses de tous les jours, il leur laisse transparaître des scènes du vécu marocain quotidien pour assurer la transmission de la tradition et des valeurs ancestrales aussi bien culturelles que religieuses aux générations futures et, par là même, communiquer de façon subtile et efficace, le bonheur de vivre sur une terre hospitalière et généreuse. Ainsi l'écriture chez El Maleh se conçoit comme l'opposé de la perte de soi qui

entraîne l'être dans la quête d'un ailleurs meilleur et qui aboutit, à son tour, à une réelle découverte de soi et donne naissance à une identité nouvelle.

S'agissant de l'étude endogène nous avons regroupé les textes qui hébergent plus ou moins les mêmes thèmes. Ainsi, nous avons obtenu trois blocs. Le premier qui est constitué par les premières productions d'El Maleh est surtout dominé par une question centrale, à savoir, la disparition de la communauté juive marocaine et les retombées de cette tragédie sur le devenir judéo-arabe. Lesdits textes ont soulevé une prépondérance des thèmes de la littérature, du militantisme, de la politique, de la religion juive, de la misère sociale, de la pluralité ethnique, du savoir, etc. Dans le deuxième bloc nous retrouvons la tragédie palestinienne avec les champs sémantiques de la mort et de la souffrance. Cependant la réflexion sur l'écriture et la littérature reste la plus développée. Le troisième bloc enfin s'écarte légèrement des autres textes en abordant des thèmes comme l'enseignement et l'exil d'El Maleh.

Pour ce qui est du troisième chapitre, il était réservé à l'étude du plurilinguisme dans l'œuvre d'El Maleh. En effet, le corpus soumis à l'étude se caractérise par une profusion de mots provenant de diverses langues en présence dans le paysage linguistique marocain, sachant bien que, comme le soutient Nancy Huston : « *Les langues ne sont pas seulement des langues ; ce sont aussi des world views, c'est-à-dire des façons de voir et de comprendre le monde*³⁵⁹ ». De ce fait, à travers les langues mises en valeur dans le texte d'El Maleh, nous avons essayé de découvrir le monde que l'auteur cherche à faire voir.

Ainsi, nous avons soumis l'ensemble des occurrences présentes dans le corpus à l'étude pour relever les principales manifestations langagières multilingues et essayer

³⁵⁹ Nancy HUSTON, *Nord perdu, suivi de Douze France*, Paris, Actes Sud, Babel, 1999, p. 51.

de dégager une typologie. Nous avons conclu que le plurilinguisme d'El Maleh se manifeste, à plusieurs niveaux, à travers l'utilisation des emprunts, des calques linguistiques, des alternances codiques, des interférences phonétiques, etc. Le dénominateur commun de tous ces phénomènes langagiers, c'est qu'ils concourent de différentes manières dans la mise en évidence de la culture marocaine. Ils contribuent au portrait d'un patrimoine national sous ses multiples facettes. Donc, les mots d'El Maleh sont là à titre de mémoire, pour rappeler l'origine partagée par les marocains. Ils présentent le texte maléhien comme un lieu de recueillement où se manifeste une volonté de revenir aux origines.

Ces phénomènes linguistiques peuvent se manifester naturellement dans une société plurilingue, cependant dans un texte littéraire, ils sont récupérés ou détournés à des fins propres au scripteur. En effet, ces pratiques langagières, quand elles émergent dans le texte maléhien, libèrent le français de sa dimension culturelle étrangère et le plongent dans un univers spécifique à la langue maternelle de l'auteur/narrateur. D'ailleurs, El Maleh déclare, en abordant la question de la langue dans ses entretiens avec Marie Redonnet, qu'il s'agit d'

« une expérience humaine ancrée dans la réalité marocaine. Lisant n'importe quel livre que j'ai écrit, ça raconte autre chose que ce qu'on attendait... Profondément, tout en étant tributaire de la langue française, la façon dont je vis cette langue, dont je m'en sers n'est pas la même que celle d'un écrivain français. »³⁶⁰.

³⁶⁰ Marie REDONNET, *Entretiens avec Edmond Amran El Maleh*, Grenoble/Rabat, La pensée sauvage/Fondation Edmond Amran El Maleh, 2005, p. 56.

Donc le texte malézien semble mimer cette parole marocaine où l'individualité s'efface devant la polyphonie de la masse et c'est au lecteur de tenter de retrouver cette parole pure cachée dans la langue d'écriture malézienne.

CONCLUSION GÉNÉRALE

Après ce voyage fascinant dans le monde lexical d'Edmond Amran El Maleh, une rétrospective des tenants et aboutissants de cette étude s'impose. Rappelons à ce stade que nous avons essayé d'approcher l'écriture d'El Maleh via une démarche différente de l'analyse littéraire traditionnelle ; il s'agit de la statistique lexicale qui consiste à aborder un matériau littéraire subjectif à travers une méthode et un outil objectif. Notre but était de chercher « ***comment Edmond Amran El Maleh a pu faire de son œuvre, à travers la langue d'écriture et les unités lexicales employées, un outil pour la redécouverte de la mémoire, de l'histoire et de la culture marocaine.*** »

Pour répondre à cette problématique, notre thèse a été structurée en trois parties. La première était réservée à la présentation du cadre théorique dans lequel s'inscrit notre recherche. Il était question alors de présenter le corpus de travail ainsi que la méthode d'analyse. Tandis que les deux dernières parties ont abordé successivement la structure et le contenu du lexique.

En effet, l'étude du contenu lexical nous a permis de répondre à notre problématique d'un point de vue sémantique, stylistique et thématique. La structure du vocabulaire, fondée sur l'examen de plusieurs paramètres structurels, nous a donné l'occasion d'explorer la structure formelle de l'œuvre en dégagant des propriétés linguistiques intéressantes, indépendamment du contenu lexical. Notons que cette analyse a été largement favorisée par l'outil informatique qui nous a permis de prendre en considération l'intégralité de la production littéraire maléhiennne. Travail qui n'a jamais été abordé jusqu'à présent pour l'œuvre en question.

Ainsi, au niveau de la première partie, nous avons commencé par présenter la vie et l'œuvre de l'auteur. Celle-ci nous a permis de postuler que son identité ancrée dans les racines judéo-arabes et berbères occupe une place centrale dans sa vie. En abordant son œuvre littéraire nous avons montré qu'il essaye d'apporter des

éclaircissements sur la question juive, afin qu'il n'y ait pas d'amalgame entre sionisme et judaïsme tout en condamnant l'asservissement politique des consciences et des passions. La première partie était aussi une occasion pour présenter la méthode de travail, à savoir la statistique lexicale. Ainsi, nous avons fait un tour d'horizon sur ses origines et ses fondements théoriques tout en abordant quelques notions de base. Ensuite nous nous sommes intéressée à la présentation de la base de données « HYPEREDM » qui est formée par l'ensemble de la production littéraire d'El Maleh. Cette partie a débouché sur une démonstration des performances du logiciel *Hyperbase* quant au traitement automatique des données textuelles : exploration, exploitation statistique, calcul hypergéométrique, richesse lexicale, connexion lexicale, etc.

Lors de nos investigations sur la structure, nous avons commencé le premier chapitre par l'étude de la longueur des mots. Nous avons alors établi que la période de production n'a pas d'effet sur la distribution des classes de longueur. De même, nous avons constaté des régularités entre les romans et l'utilisation des mots courts et moyens qui sont en faveur d'un registre familier, et entre les essais et la tendance de l'auteur à utiliser des mots longs, ce qui relève souvent du style soutenu. Néanmoins, nous avons soulevé quelques exceptions, par exemple ELHAKI, qui est un roman, privilégie à la fois les mots longs (lg9, lg10, lg11) et les mots courts et moyens (lg3, lg5, lg6 et lg7). Ceci pourrait se justifier par l'utilisation de l'auteur d'un discours soutenu abordant les écritures mystiques de certaines figures intellectuelles prestigieuses.

L'examen de l'étendue N (occurrences) a permis de constater l'existence d'une dynamique tripartite. En effet, l'étendue des quatre premiers textes évolue selon une tendance haussière, ce qui pourrait être expliqué par l'acharnement d'El Maleh pour l'écriture. Cependant, nous avons enregistré par la suite une décadence de l'étendue

qui a pris fin avec un second pic dans ZRIREK. Ce pic peut être justifié par des facteurs d'ordre thématique. En effet, quoique la majorité des textes d'El Maleh abordent plus ou moins les mêmes thèmes, dans *Le Café bleu*, *Zrirek* nous assistons à une concentration d'idées sur la littérature et l'écriture allégorique. Finalement nous repérons une décroissance relative aux textes FEMME et LETTRES, ceux dont l'étendue est la plus faible de toutes les partitions de notre corpus.

A travers l'étude de l'étendue V (vocables), nous avons déduit que le vocabulaire d'El Maleh est déficitaire dans l'ensemble des textes. Néanmoins, nous avons mis en évidence les segments les moins déficitaires, à savoir : ELHAKI, ABOUNOUR, FEMME, PARCOURS. Cela dit que l'auteur aurait usé de presque tout son lexique dans ces textes.

En appliquant le coefficient de corrélation des rangs à notre corpus, pour comparer l'ordre chronologique de parution des textes avec l'étendue des occurrences N et celle des vocables V, nous nous sommes rendu compte que le lexique personnel d'El Maleh ne dépend pas de la date de parution des textes, ce qui pourrait laisser entendre que l'auteur diversifie ses choix thématiques.

Le second chapitre s'est attardé sur la richesse lexicale et l'accroissement du vocabulaire. Abordant le premier volet, nous avons classé les textes en partant de celui qui a le vocabulaire le plus riche à celui qui a le vocabulaire le plus faible :

**ELHAKI > AÎLEN > PARCOURS > MILLE > ZRIREK > ABOUNOUR
> GENET > FEMME > LETTRES**

En analysant l'accroissement du vocabulaire, nous avons décelé un renouvellement lexical au niveau des textes ELHAKI, ZRIREK et LETTRES. Pour les deux premiers textes nous avons montré que ce constat est surtout lié à leur

contenu sémantique et thématique. Cependant pour *Lettres à moi-même* le renouvellement lexical est dû à la fois au thème et au genre, car ce texte est le premier à aborder l'exil d'El Maleh, mais aussi le premier et le dernier récit épistolaire de toute la production littéraire de l'auteur.

Signalons que nous avons établi un autre classement basé sur le calcul de l'écart réduit. L'analyse de ce dernier rejoint ce que nous avons dit à propos de l'étendue des occurrences N. En effet, les premiers romans se caractérisent par un vocabulaire riche avec des apports lexicaux importants, ce qui reflète surtout l'intention manifeste de l'écrivain de vouloir tout exprimer sur l'exode des juifs marocains et de faire jaillir la spontanéité créatrice de son imagination. Cependant nous avons repéré une chute au niveau des quatre derniers textes, ce qui correspond à l'abandon du genre romanesque mais aussi de certains thèmes. Donc, il semble que les récits où nous suivons le destin des juifs, qui sont partis à la recherche de la Terre promise, apportent plus de vocabulaire que les parties qui sont consacrées à d'autres thèmes.

Ensuite, pour mieux apprécier le phénomène de l'accroissement du vocabulaire, nous avons mesuré l'apport lexical à partir du dernier texte jusqu'au premier. Le résultat était frappant parce que les textes FEMME et LETTRES qui ont l'étendue la plus faible et qui sont aussi les derniers à être publiés étaient classés dans la première et la deuxième position. Cela signifie qu'ils actualisent un nombre de vocables qui n'existent pas dans les textes précédents. Nous avons justifié ce constat par la volonté de l'écrivain de créer un changement radical dans son œuvre que ce soit au niveau de la forme qu'au niveau du contenu. D'ailleurs, l'avant dernier texte LETTRES est le seul récit épistolaire qui existe dans tout le corpus, de même, c'est le seul qui parle de l'exil d'El Maleh, de son arrivée à Paris. Tandis que le dernier texte FEMME est le seul qui aborde sa découverte de la ville d'Asilah et sa rencontre

avec la mère de son fils spirituel Khalil El Ghrib. De ce fait, le classement auquel nous avons abouti confirme parfaitement bien cette diversification dans l'écriture maléhiennne.

Nous avons achevé ce chapitre par l'application d'une autre démarche à notre corpus, à savoir l'accroissement inverse. Ce dernier a montré que la tendance régulière à la baisse est inversée dans ELHAKI et ZRIREK. Donc ces deux textes signalent la fin d'un mouvement et l'épuisement de l'inspiration de l'auteur. Le mouvement dont il s'agit ici est surtout d'ordre générique et thématique : le texte ELHAKI marque une rupture avec le genre romanesque. D'ailleurs aucun des cinq textes publiés après n'admet l'étiquette de roman, du moment que l'auteur a délaissé ce genre en faveur des essais et des nouvelles. ZRIREK est aussi le dernier essai qui aborde l'écriture comme thème principale.

Toutefois, le classement établi sur la base de l'écart réduit a dévoilé un éclatement lexical au niveau du texte ABOUNOUR, chose que l'ordre chronologique n'a pas pu relever. Ceci pourrait être dû aux effets du genre ; d'ailleurs ce texte est le seul recueil de nouvelles qui existe dans tout le corpus d'El Maleh.

L'analyse du profil grammatical de l'œuvre d'El Maleh a permis de relever une opposition entre deux pôles, à savoir la catégorie verbale et la catégorie nominale. Cette opposition³⁶¹ substantif /verbe nous renvoie à la binarité espace/temps. L'univers espace est rendu en langue française par le substantif, tandis que l'univers temps est restitué par le verbe. Ce constat rejoint ce que nous avons avancé au niveau de la première partie. En effet, l'objectif même de l'écriture maléhiennne est de restituer la mémoire séculaire de sa terre natale (espace) à une époque révolue

³⁶¹ Véronique MAGRI-MOURGUE, *op.cit.*, 2009, pp. 55-85.

(temps). Par l'acte scriptural, notre auteur fraye un passage délicat dans les labyrinthes du passé et les dédales de la mémoire. Le sillage formé par son écriture est, en fait, un parcours sinueux où la dimension temporelle n'est plus une suite linéaire, mais plutôt une mosaïque de moments retrouvés et revécus par un esprit en quête continuelle d'une explication de ce qui résiste à toute analyse rationnelle. Ce qui est surtout particulier, c'est que le temps chez El Maleh se transforme en un certain nombre de lieux fragmentaires et fragmentés au point de constituer un vrai puzzle d'espaces remémorés et différemment connotés.

Ajoutons que l'utilisation des verbes dans le texte maléhien est aussi motivée par le mouvement et le déplacement. D'ailleurs, voyage et errance structurent l'ensemble des récits d'El Maleh par le biais de ses protagonistes qui nous offrent une sorte de voyage à la fois physique et mental. Ils parcourent l'espace et le temps, nous communiquant ainsi une représentation des événements vus à travers un homme éclairé.

Dans la troisième et dernière partie qui était réservée à l'analyse du contenu lexical, nous avons opté pour trois chapitres où nous avons étudié successivement le phénomène de la connexion lexicale, les spécificités de l'écriture maléhienne et les manifestations du plurilinguisme dans l'œuvre en question.

Ainsi, dans le premier chapitre l'étude des rapprochements entre les textes de notre corpus a été effectuée, tout d'abord, en utilisant la méthode Jaccard, ensuite en ayant recours à la loi binomiale. La première méthode nous a permis de dresser une liste de couples qui s'attirent soit parce qu'ils partagent le même contenu lexical, le même genre ou bien la même période de production. En effet, dans les binômes (AILEN, MILLE) et (ELHAKI, PARCOURS) qui sont tous des romans, nous trouvons les thèmes de l'exode juif, de la guerre, du colonialisme, du militantisme

marocain, etc. Aussi, si nous prenons en considération leur chronologie de parution, nous remarquerons que les quatre romans se succèdent. En plus, le texte ELHAKI qui est le dernier à être publié, constitue une récapitulation des sujets traités dans les trois autres romans. Nous avons aussi décelé un rapprochement entre GENET et ZRIREK. Ces deux textes, qui sont des essais, se partagent aussi les thèmes de la littérature, des écritures mystiques et de l'allégorie.

Encore une fois nous avons repéré l'éloignement du texte FEMME de toutes les autres partitions du corpus. Ceci reste normal du moment que l'écrivain, vers la fin de sa vie, avait le souci de sombrer dans une sempiternelle répétition ; c'est la raison pour laquelle il a entamé l'écriture du texte FEMME qui reste plus ou moins différent des autres textes surtout sur le plan thématique. Donc l'éloignement du texte FEMME des autres est dû au besoin de l'écrivain de créer un changement dans son œuvre.

Pour remédier aux insuffisances de la méthode Jaccard qui ne considère que la présence ou l'absence des mots dans les textes considérés, abstraction faite de leur fréquence, nous avons testé une autre approche empruntée à Ch. MULLER, à savoir la loi binomiale. Cependant, cette dernière n'a pas satisfait notre curiosité du moment que les résultats obtenus n'étaient fiables que pour des couples de textes dont les étendues sont voisines.

Le deuxième chapitre s'est intéressé aux spécificités lexicales d'Edmond Amrane El Maleh en s'attardant sur l'étude du vocabulaire excédentaire et celui déficitaire. En d'autres termes, nous avons mis en lumière les mots-clés qui se regroupent sous les mêmes champs sémantiques pour relever les thèmes privilégiés de l'auteur. Ainsi, nous avons pu dévoiler l'attachement d'El Maleh aux thèmes de la littérature, de l'écriture, de l'histoire, de la mémoire, de la quête d'identité, des appartenances linguistiques et ethniques, etc. Tous ces thèmes et bien d'autres participent à la mise

en évidence de la mémoire et l'histoire marocaine. D'ailleurs, en lisant ses œuvres, nous voyons El Maleh en train de « coudre des débris de temps »³⁶² immémorial et mémoriel, mais aussi des fragments d'espaces oubliés, occultés et surtout séparés les uns des autres. Il procède à un minutieux travail de creusement dans la mémoire, il revient sur les traces de ses ascendants et évoque, et invoque, l'image d'un peuple vivant en pleine sérénité malgré les divergences et la multiplicité qui caractérisent les groupes ethniques et religieux qui le compose.

Soulignons que la permanence des thèmes traités, qui retient intuitivement l'attention du lecteur malézien, est toujours très présente dans l'œuvre, mais l'exploitation de l'outil informatique et statistique a confirmé de façon quantitative ces intuitions, notamment dans l'examen des spécificités internes où nous avons décelé une redondance des thèmes abordés. Néanmoins, cette étude a permis de détecter la même dynamique tripartite qui a émergé lors de l'examen de la structure du lexique. En effet, quoique les écrits d'El Maleh traitent plus ou moins les mêmes thèmes, nous avons mis en évidence une certaine spécialisation dans le contenu sémantique de chaque genre. Notons que cette spécialisation est aussi liée à la période de production. Ainsi, nous avons obtenu trois blocs. Le premier qui est constitué par les premières productions d'El Maleh (PARCOURS, AÏLEN, MILLE, ELHAKI) est surtout dominé par une question centrale, à savoir : la disparition de la communauté juive marocaine et les retombés de cette tragédie sur le devenir judéo-arabe. Lesdits textes, qui appartiennent d'ailleurs au genre romanesque, ont soulevé une prépondérance des thèmes de : la littérature, le militantisme, la politique, la religion juive, la misère sociale, la pluralité ethnique, le savoir, etc.

³⁶² Nous empruntons cette expression au titre d'un ouvrage collectif dédié à Khalil El Ghrib, *Coudre le temps*, Edition de La Maison d'à côté, 2007.

Dans le deuxième bloc, composé de deux essais et un recueil de nouvelles (ZRIREK, GENET, ABOUNOUR), nous retrouvons la tragédie palestinienne avec les champs sémantiques de la mort et de la souffrance. Cependant, les réflexions sur l'écriture et la littérature restent les plus développées. Le troisième bloc formé par les textes FEMME et LETTRES s'écarte sensiblement du reste de l'œuvre maléhienne en abordant des thèmes comme l'enseignement et l'exil d'El Maleh. De ce fait, la rupture que nous avons pu constater pour ces deux textes dans l'étude de la structure est aussi forte dans le contenu qui marque une rupture thématique.

Nous avons signalé au niveau de la première partie que l'auteur laisse planer le doute quant à la nature générique de ses œuvres. Cependant, notre étude a permis de mettre en exergue l'opposition générique qui s'observe à tous les niveaux du corpus, que ce soit dans la structure ou dans le contenu. De ce fait, l'appartenance à un genre précis de chacun de ses livres est bien réelle. Chaque genre littéraire a en fait son anatomie, sa physiologie et son fonctionnement et cela transparaît très clairement dans les différents textes qui forment l'œuvre maléhienne.

Pour ce qui est du troisième chapitre, il était réservé à l'étude du plurilinguisme dans l'œuvre d'El Maleh. En effet, cette dernière devient quelques fois une mosaïque, où plusieurs langues se mêlent. Ainsi, la mission que nous nous sommes fixée est de décrire et comprendre le mode d'apparition de ces langues pour déceler leur contribution dans la mise en évidence de la mémoire, l'histoire et la culture marocaine.

À travers l'examen de l'œuvre en question nous avons conclu que les manifestations multilingues dans le texte maléhien se proclament à plusieurs niveaux : emprunts, calques linguistiques, alternances codiques, interférences phonétiques, etc. Le dénominateur commun de tous ces phénomènes langagiers est

qu'ils permettent de traduire les caractéristiques d'une culture indigène sous ses multiples facettes : croyances populaires et religieuses, traditions locales, valeurs morales, pratiques quotidiennes, habitudes culinaires et vestimentaires, etc.

Soulignons d'emblée que l'emploi de ces pratiques langagières présuppose un background interférentiel soumis à la fois à des influences d'ordre linguistique et à des considérations culturelles. Ces dernières sont indispensables à la fois pour l'auteur et pour le lecteur parce qu'elles permettent de décoder, d'analyser et aussi d'interpréter la charge sémantique contenue dans le texte et, par conséquent, de déceler les caractéristiques authentiques de chaque culture.

Généralement, l'écriture littéraire émane d'un acte réfléchi de la part de l'écrivain. D'ailleurs, aucun mot n'est innocent et encore moins en littérature où le scripteur est censé choisir consciemment ses stratégies discursives. Ceci nous pousse à nous demander si l'emploi de ces pratiques langagières émane toujours d'un acte conscient de la part d'El Maleh. Peut-être serait-il nécessaire d'évoquer ce propos de l'auteur cité dans *Le Magazine littéraire* pour comprendre son attitude face à la langue d'écriture : « *écrivain en français, je savais que je n'écrivais pas en français, il y avait cette singulière greffe d'une langue sur l'autre, ma langue maternelle l'arabe, ce feu intérieur*³⁶³ ». Ainsi, ce n'est pas gratuit qu'El Maleh fait en sorte que les traces d'autres langues figurent dans ses textes ; en fait, cette technique reste ajustée au besoin de l'auteur de vouloir mettre en évidence la diversité culturelle marocaine. Sa représentation des variations linguistiques parlées au sein du Maroc se colore d'un contexte géographique et ethnologique et s'enrichit de la multiplicité des registres linguistiques et des états diachroniques de la langue. De ce fait, c'est une volonté de remonter aux origines qui fait l'essence de ce choix d'écriture qui entraîne le lecteur

³⁶³ Voir : https://fr.wikipedia.org/wiki/Edmond_Amran_El_Maleh, consulté le 28/12/2019 à 23h

sur les traces d'un passé collectif. En effet, l'auteur, dans son œuvre, retrace presque toute l'histoire du Maroc et y glisse un trait culturel marocain important, celui de l'attachement fort de l'homme à sa terre-mère.

Vers la fin de ce périple, nous pouvons dire qu'El Maleh a mobilisé plusieurs démarches pour *rendre compte de la mémoire, de l'histoire et de la culture marocaine* et ce à travers la diversité générique de ses textes, la nature des thèmes abordés, mais surtout, à travers la langue d'écriture qui a fait émerger la richesse linguistique, ethnologique et culturelle de la société marocaine. D'ailleurs, comme disait Emile Benveniste « *Langue et société ne se conçoivent pas l'une sans l'autre* »³⁶⁴. De ce fait, chaque langue, en tant que structuration d'un ensemble d'éléments et de règles, est spécifique à l'image de chaque société dans la mesure où elle se façonne pour refléter les mutations et les mouvances de cette dernière et permet par ailleurs d'assimiler et de véhiculer la culture qui lui est inhérente.

Ainsi, en se laissant prendre par les délices de l'activité mémorielle et nostalgique, El Maleh partage avec nous lecteurs, à travers la langue d'écriture, les souvenirs de la communauté marocaine. A défaut de pouvoir changer le cours de l'histoire, il essaye au moins de réveiller le passé en attribuant à ses récits, le rythme, la musique, les odeurs, les couleurs, les voix et les résonnances d'une époque donnée. Il s'inscrit dans une généalogie bien spécifique et nous transmet de la manière la plus subtile possible ses us et coutumes, ses habitudes culinaires, ses rituels... choses qui renvoient aussi bien à un inconscient culturel qu'à une mémoire collective. De ce fait, le but derrière le multilinguisme suggéré par El Maleh est de restituer le sens et

³⁶⁴ Émile BENVENISTE, *Problèmes de linguistique générale*, 1, Paris, Gallimard, 1966, p. 29.

l'essence de la société marocaine dans son histoire glorieuse mais aussi dans sa diversité linguistique, ethnologique et culturelle.

Quant à nous, nous avons su apprécier, tout au long de cette étude, l'interaction entre la littérature, les sciences du langage et les outils informatiques, qui apporte à l'étude classique de l'écriture littéraire un nouvel enrichissement et un éclairage pertinent. Notons que nos différentes analyses nous ont amenée à produire des conclusions non seulement sur l'écriture maléhiennne, mais aussi sur l'outil et la méthode utilisée. Cela étant dit, cette expérience nous a apporté non seulement de l'assurance quant à la fiabilité de nos conclusions, mais nous a également convaincue de la robustesse de la technique lexicométrique.

Certes nous avons abordé plusieurs aspects de l'écriture maléhiennne mais nous ne pouvons pas prétendre être exhaustive. Faute de temps, nous n'avons pas exploré d'autres recherches telles que l'étude de l'onomastique ou des numéraux. Aussi, nous n'avons pas étudié d'autres niveaux de la syntaxe tel que la longueur des phrases qui est très significative dans l'écriture maléhiennne ; d'ailleurs cet écrivain asthmatique comblait les dyspnées dont il souffrait par une ponctuation essoufflée ou encore une absence de ponctuation. Nous n'avons pas non plus exploité d'autres outils quantitatifs comme Alceste ou Lexico, qui serait sûrement un travail de méthodologie très utile permettant d'affiner encore l'analyse textuelle. Ce sont autant de pistes que nous ouvrons pour d'éventuelles recherches que ce soit pour nous-même ou pour d'autres chercheurs en sciences du langage.

Bibliographie et Sitographie

1. Œuvre d'Edmond Amran El Maleh

➤ Romans :

Le retour d'Abou El Haki, Grenoble, La Pensée Sauvage, 1998.

Aïlen ou la nuit du récit, Paris, André Dimanche, 2000.

Parcours immobile, Paris, André Dimanche, 2000.

Mille ans, un jour, Paris, André Dimanche, 2002.

➤ Essais :

Jean Genet, Le Captif amoureux et autres essais, Grenoble/Casablanca, La Pensée sauvage/Toubkal, 1990.

Café bleu, Zrerek, Paris, La Pensée sauvage/Le Fennec, 1998 1999.

Une femme, une mère, Grenoble, La pensée sauvage, 2004 2002.

La malle de Sidi Maâchou, Casablanca, Al Manar, 1998.

➤ **Recueil de nouvelles :**

EL MALEH, Edmond Amran, *Abner, Abounour*, Thionville/Casablanca, La Pensée sauvage/Le Fennec, 1995 1996.

➤ **Récit épistolaire :**

EL MALEH, Edmond Amran, *Lettres à moi-même*, Casablanca, Le Fennec, 2010

2. Linguistique et statistique lexicale

AZIRI Boudjema, *Néologismes et calques dans les médias amazighs : origines, formation et emploi*, Alger, Haut-Commissariat à l'Amazighité, 2009.

ALBERT Henry. BALLY Charles, Linguistique générale et linguistique française. In : Revue belge de philologie et d'histoire, tome 25, fasc. 1-2, 1946

BARTHELEMY J.P. & LUONG X, *Sur la topologie d'un arbre phylogénétique : aspects théoriques, algorithmiques et applications à l'analyse de données textuelles*, Mathématiques et Sciences Humaines, n° 100, 1987.

BARTHES Roland, *Linguistique et littérature*, In : *Langage*, Volume 3, N° 12, 1968.

BRUNET Etienne, *Le Vocabulaire de Jean Giraudoux, Structure et Evolution*, Genève, Slatkine, 1978.

BRUNET Etienne, *Le vocabulaire de Proust I, Étude quantitative*, Genève, SLATKINE, 1983.

BRUNET Etienne, *Le vocabulaire français de 1789 à nos jours, d'après les données du Trésor de la langue française*, Genève-Paris, Slaktine-Champion, 1981, Volume I.

- BRUNET Etienne, les Rougons- Macquart : Aspects quantitatifs, Extrait de la revue: *Informatique et Statistique dans les sciences Humaines XXI*, de 1 à 4, université de liège, C.I.P.L.
- BULLY Philippe Zipf, *Créateur de la linguistique statistique*, In : Communication et langages. N°2, 1969, pp. 23-28
- CASSIN Barbara, *Vocabulaire européen des philosophies. Dictionnaire des intraduisibles*, Paris, Seuil / Le Robert, 2004.
- POUDAT Céline, LANDRAGIN Frédéric, *Explorer un corpus textuel : Méthodes – pratiques – outils*, Louvain-la-Neuve, De Boeck Supérieur, 2017.
- NICOLAS Christian, *Le procédé du calque sémantique*, dans Cahiers de Lexicologie, n° 65, 1994.
- BUTLER Christopher S, *Statistics In: Linguistics*, Oxford, Blackwell Publishers, 1985.
- COSSETTE A, *La richesse lexicale et sa mesure*, Paris, Champion, 1994.
- DEROY Louis, *L'emprunt linguistique*, Paris, Les Belles Lettres, 1956.
- DUBOIS.J et al. *Le Dictionnaire de Linguistique et des Sciences du Langage*, Paris, Larousse, 2012
- GUIRAUD P, *Les Caractères Statistiques du Vocabulaire*, Paris, Presse Universitaire de France, 1954.
- GUIRAUD. P, *Problèmes et Méthodes de la Statistique Linguistique*, Dordrecht, D.Reidel Publishing Company, 1959.

- HABERT B, NAZARENKO A et SALEM A., *Les linguistiques de corpus*, Paris, Armand Colin, 1997.
- HAMERS Josiane F, « Emprunt », in MOREAU Marie Louis, *Sociolinguistique : Concepts de base*, Bruxelles, Mardaga, 1997.
- HATCH Evelyn and LAZARATON Ann., *The research manual: Design and Statistics for Applied Linguistics*, Los Angeles, Heinle Publishers, 1991.
- HEFIED Ali, *Statistique linguistique : Aspects stylo-statistique du vocabulaire dans quinze « Voyages Extraordinaires » de Jules Verne*, Thèse de doctorat d'Etat, Université Sidi Mohamed Ben Abdallah, Faculté des lettres et des sciences humaines, Dhar El Meraz, Fès, 1999, Vol 1.
- IRAQUI SINACEUR-Zakia , *Histoire et emprunt linguistique* in : *Trames de langues : usages et métissages linguistiques dans l'histoire du Maghreb* / Institut de recherche sur le Maghreb contemporain sous la direction de Jocelyne Dakhli, Paris, Maisonneuve et Larose, 2004.
- KANNAS Claude, *Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse, 1994.
- MAGRI-MOURGUE, Véronique, *Le Voyage à Pas Comptés, Pour une Poétique du Récit de Voyage au XIX^e Siècle*, Paris, champion, 2009.
- MAGRI-MOURGUE, Véronique, *Le discours sur l'Autre à travers quatre récits de voyage en orient*, Paris, Honoré champion, 1995.
- MAGRI-MOURGUES, Véronique, « *Distance intertextuelle et connexion lexicale : outils de catégorisation générique ou stylistique ?* », In : International

Conférence on *Statistical Analysis of Textual Data*, n°10, Paris, 2010, pp. 334-340.

MOUELHI Zoubeïr, *Essai de lexicométrie d'une œuvre arabe : le vocabulaire de Tawhîdî dans al- 'Imtâ' wa-l-Mu'ânasa*, Thèse de Doctorat, Université Lumière-Lyon 2, 2008, Vol 1.

MOUELHI Zoubeïr, *La Richesse Lexicale dans une Perspective de Lexicométrie Arabe : Étude contrastive de 5 méthodes de mesure : Application au « Imtâ wa-l-muânasa de Tawhîdî »*, in : revue électronique Texte et corpus, n°3, Août 2008, Actes des journées de la linguistique de Corpus, 2007, pp. 271-284.

MOUNIN Georges, *Dictionnaire de la linguistique*, Paris, Presses universitaires de France, Quadriga, 1974.

MULLER Charles, *Essai de Statistique Lexicale, l'illusion Comique de pierre corneille*, Paris, Librairie Klincksieck, 1964.

MULLER Charles, *Étude de Statistique Lexicale, le Vocabulaire du Théâtre de Pierre de Corneille*, Paris, Larrousse, 1967.

MULLER Charles, *Initiation à la statistique linguistique*, Paris, Larrousse, coll. "langue et langage", 1968, Revue Romane, Bind 5, 1970, pp. 147-151.

MULLER Charles, *Initiation aux méthodes de la statistique linguistique*, Paris, Honoré Champion, 1992.

MULLER Charles, *Initiation aux méthodes de la statistique linguistique*, Paris, Hachette, 1973.

MULLER Charles, « *La statistique lexicale* », In : Langue française. N°2, 1969.

- MULLER Charles, *Langue française. Débats et bilans. Recueil d'articles*, Paris, Champion, 1993.
- MULLER Charles, *Principes et méthodes de la statistique lexicale*, Paris, Hachette, 1977.
- MULLER Charles, *Principes et Méthodes de la Statistique Lexicale*, Paris, Classiques Hachette, 1992.
- MULLER Charles, *Statistique lexicale et théorie du lexique*, in : *Langue Française et linguistique quantitative*, Genève, Slatkine, 1979.
- OAKES P. Michael, “*Statistics for corpus linguistics edinburgh textbooks*”, In : *empirical linguistics*, Cambridge, Edinburgh university press, 1998
- PARADA RAMIREZ José Gregorio, *Lecture documentée et analyse textométrique de l'œuvre de Jules Verne. Les influences de la Franc-maçonnerie dans son œuvre*, Thèse rédigée pour l'obtention du doctorat en langue et littérature française, Nice, 2013.
- RENS Bod & all, *Probabilistic Linguistics*, London, The MIT Press, 2003.
- SALEM André, *Pratique des segments répétés. Essai de statistique textuelle*, Paris, INalf, coll. "SAINT-CLOUD", 1987.
- VERBOROM RATIO, P. SALAT, “*Exemples d'Études Statistiques portant sur le Vocabulaire Latin*”, Clermont-Ferrand, Faculté des lettres et sciences humaines de l'université de Blaise-Pascale, 1991, pp. 33-53.

3. Statistique

ALALOUF Serge, *Méthodes statistiques*, Québec, Edition Loze-Dion, 2014.

ARMELLE Mathé, *L'essentiel de la statistique descriptive*, Issy-les-Moulineaux, Gualino, 2016.

ARMELLE Mathé, *L'essentiel des probabilités*, Issy-les-Moulineaux, Gualino, 2016.

BENZECRI J.P, *L'analyse des données*, Tome I et II, Paris, Dunod, 1973.

CANTONI Eva, HUBER Philippe, RONCHETTI Elvezio, *Maîtriser l'aléatoire*, Paris, Springer, Collection statistique et probabilités appliquées, 2006.

DANIEL Caumont, SILVESTER Ivanaj, *Analyse des données – Méthodes et outils statistiques, applications à la prise de décision, cas d'entreprises détaillés*, Paris, Dunod, 2017.

GIRARDIN Valérie & LIMNIOS Nikolaos, *Probabilités en vue des applications : introduction aux processus et à la statistique*, Paris, Vuibert, 2008.

HASSENE Siby, *Introduction à la statistique et aux probabilités*, Québec, Edition Loze-Dion, 2017.

HURLIN Christophe et al, *Statistique et probabilités en économie-gestion*, Paris, Dunod, 2015.

LECOUTRE, Jean-Pierre, *Statistique et probabilités*, Paris, Dunod, 6^{ème} édition, 2016.

LEJEUNE Michel, *Statistique : La théorie et ses applications*, Paris, Springer, 2^{ème} édition, Collection statistique et probabilités appliquées, 2010.

LETHIELLEUX Maurice, *Statistique descriptive*, Paris, Dunod, 8^{ème} édition, 2016.

OLIVIER Elisabeth, *L'essentiel de la statistique descriptive*, Issy-les-Moulineaux, Gualino, 2008.

REISCHER Corina, *Théories des probabilités : problèmes et solutions*, Sainte-foy, Presses de l'Université du Québec, 2002.

REVUZ Daniel, *Probabilités*, Paris, Hermann, Collection Méthodes, 1997.

YADOLAH Dodge, GIUSEPPE Melfi, *Premiers pas en simulation*, Paris, Springer, Collection Statistique et probabilités appliquées, 2008.

4. Dictionnaires

MERCIER H, *Dictionnaire ARABE-FRANÇAIS*, Rabat : éditions la porte, 1951

PREMARE Alfred-Louis, et al, *Dictionnaire ARABE-FRANÇAIS*, Paris, l'Harmattan, 1993.

5. Dictionnaires en ligne

la Base de données lexicographiques panfrancophone,

http://www.bdlp.org/recherche.asp?base=*&AC=oui&AL=oui&AN=oui&BE=oui&BU=oui&CA=oui&CE=oui&CB=oui&CI=oui&FR=oui&LO=oui&MD=oui&MA=oui&IM=oui&NC=oui&QU=oui&RE=oui&RW=oui&SU=oui&TC=oui

Le Centre nationale des ressources textuelles lexicales, <https://www.cnrtl.fr/portail/>

Lisanu Al-Arab : <http://arabeclassique.forumactif.com/t2524-lisan-al-arab-pdf-ibn-manr-muammad-ibn-mukarram>

TLFi : Trésor de la langue française informatise, <http://atilf.atilf.fr>

6. Littérature et Arts

ADAMS C and all., *The Performance Prism in practice, Measuring Business Excellence*, volume 5, N° 2, 2005.

BARTHES Roland, *La Chambre claire, Notes sur la photographie*, Cahiers du cinéma, Paris, Gallimard/Seuil, 1980

BAUDELAIRE Charles, *l'Art Poétique*, Paris, Conard, 1925.

BEN MSILA Anouar, *L'écrivain, le peintre et l'allégoricien*, in : **Ben Msila** Anouar. Edmond Amran El Maleh, *Art, culture et écriture* : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des langues et Sciences humaines de Meknès, 2013.

BEN-AMI Issachar, *Culte des saints et pèlerinages judéo-musulmans au Maroc*, Paris, Maisonneuve et Larose, 1990.

BROWN Kenneth, *People of Sale: Tradition and Change in a Moroccan City, 1830-1930 (Harvard Studies in Urban History) Hardcover* , Harvard Univ Press; 1st edition, 1976

CAMPS G, *Réflexion sur l'origine des juifs des régions nord-sahariennes*, in Michel Abitbol (éd), *Communautés juives des marges sahariennes*, Jérusalem, 1982.

- CHOURAQUI André, *La condition juridique de l'Israélite marocain*, Paris, éd. Presses du Livre français, 1950.
- COMMELIN Bertrand., *Le gouvernement d'entreprise, La Bourse et les entreprises*, Cahiers Français N° 277, 2001.
- DUFOUR Marie-Cécile, *L'écriture allégorique*, Grenoble, La Pensée sauvage, 1999.
- DUGAS Guy, *La littérature judéo-maghrébine ! d'expression française. Entre Djéba et Cagayous*, Paris, L'Harmattan, 1990.
- EL HADDAD Yasmina, POEL Ieme van der, *Variations sur la figure du double : la mémoire judéo-marocaine chez E.A. El Maleh et M. Bennaroch*, In : **Ben Msila** Anouar. Edmond Amran El Maleh, *Art, culture et écriture* : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des langues et Sciences humaines de Meknès, 2013.
- EL MALEH, Edmond Amran, *Essaouira, l'oubliée*, in revue Autrement, Paris, n° 11 (Hors-série), 1985.
- EL MALEH, Edmond Amran, *La figure singulière du juif marocain qui écrit en langue français*», *les cahiers d'Études Maghrébines*, n°3, juin 1991.
- EL MALEH, Edmond Amran, *La peinture marocaine au rendez-vous de l'histoire*, Wafabank, Casablanca, 1988.
- EL MALEH, Edmond Amran, « *Bouchta El Hayani* », In : *Edmond Amran El Maleh, Parcours mobile*.
- EL MALEH, Edmond Amran, « *Flammes et cendres* », In *Edmond Amran El Maleh, parcours mobile*, Casablanca, Ed. Galerie 38, Europrint, 2010.

- FLAMAND Pierre, *Diaspora en terre d'Islam*. Les communautés israélites du sud marocain, éd. Imprimeries réunies, casablanca, 1958.
- PAUL A. Fortier , *Etat présent de l'utilisation des ordinateurs pour l'étude de la littérature française*, Computers and the Humanities, Vol.5, N° 3, January 1971.
- FRAUTSCHI R. L, *A Project for Author Discrimination in the Encyclopédia*, SAMLA Bulletin 32, Novembre 1967, 14- 17, communication par Frantschi, « Diderot in Computerland », 23 rd Kentucky Conférence.
- FUMEY Gilles, « *la culture dans l'assiette* », Regards croisés, n° 20, 2006 : en ligne sur <http://escales.enfa.fr/wp-content/uploads/sites/7/2009/06/14-Fumey.pdf>, consulté le 23 octobre 2018 à 23h.
- GERBER Jane, *Jewish society in Fez : 1465-1700*, Leiden. E. J. Brill, 1980.
- GRYNBERG Anne, *Vers la terre d'Israël*, Paris, Gallimard, 2008.
- HADOUCH Ghyslaine, *Entre Mémoire et Oubli, L'écriture de l'Histoire*, in : BEN MSILA Anouar. *Edmond Amran El Maleh, Art, culture et écriture* : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013.
- IKKEN Aïssa, *Quelques réflexions d'E. A. El Maleh sur l'art et les peintres marocains*, in : Ben Msila Anouar. *Edmond Amran El Maleh, Art, culture et écriture* : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des langues et Sciences humaines de Meknès, 2013.
- KAPOR, Vladimir, *La catachrèse interculturelle*, Poétique n° 148, 2006.

- LAHLOU Saadia, *Penser manger*, Paris, PUF, 1998.
- LEON l'Africain, Description de l'Afrique, traduit de l'italien par Epaulard A, Paris. Librairie d'Amérique et d'Orient, 1956, tome 1, p. 234. Cité d'après KENBIB Mohammed, Juifs et Musulmans au Maroc, 1859-1948. Rabat. Publication de la Faculté des Lettres et des Sciences humains, 1994.
- LETOURNEAU Roger, *La vie quotidienne à Fès en 1900*, Paris, éd. Hachette, 1965.
- MARTIN Claude, *Les israélites algériens de 1830 à 1902*, Paris, 1936.
- MAUME Jean-Louis, « **L'apprentissage du français chez les arabophones maghrébins** », *Langue française*, n° 19, 1973.
- NISSELON Rachel, *Locating the Productive* in the Inthinkable in : *Expressions maghrébines*, vol. 09, n° 2, hiver 2010.
- PAUL Smaïl, *Ali le Magnifique*, Paris, Denoël, 2001.
- REDONNET Marie, *Entretiens avec Edmond Amran El Maleh*, Grenoble/Rabat, La pensée sauvage/Fondation Edmond Amran El Maleh, 2005.
- RICOEUR Paul, *La mémoire, L'Histoire, L'Oubli*, Paris, Editions du seuil, collection, l'ordre philosophique, 2000.
- ROSEMARY Marangoly George, *The Politics of Home: postcolonial Relocations and Twentieth-Century Fiction*, Berkeley (cal), University of California Press, 1996.
- SCHROETER Daniel, *Merchants of Essaouira; urban society and imperialism in south-western Morocco. 1844-1866*. Cambridge, Cambridge University Press, 1988.

- SEMACH Y. D, *Une chronique juive de Fès*, Hespéris, Institut des Hautes Etudes Marocaines, p. 91-92, in : KENBIB Mohammed, Juifs et Musulmans au Maroc 1859-1948. *Contribution à l'histoire des relations inter-communautaires en Terre d'Islam*, Rabat, publication de la faculté des Lettres et des Sciences humaines, 1994.
- SERHANE Abdelhak, **notre voix**, in: *Horizons Maghrébins* - Le droit à la mémoire, N°27, 1994. Présences d'Edmond Amran El Maleh.
- SPIRE Marie-Brunette, *Des écrivains juifs du Maghreb à la recherche de leur identité*, dans LASRY Jean-Claude et TAPIA Claude, *Les juifs du Maghreb : diasporas contemporaines*, Paris, L'Harmattan, 1989.
- ZAFRANI Haïm, Lieux de rencontre et de dialogue, in : *Horizons Maghrébins* - **Le droit à la mémoire**, N°27, 1994. Présences d'Edmond Amran El Maleh.
- ZIGHIGHI Zouhir, *Une poétique de la matière*, in : Anouar Ben Msila. Edmond Amran El Maleh, *Art, culture et écriture* : Acte de colloque 38/ 2013, 1-2 décembre 2011, Meknès. Faculté des Lettres et Sciences humaines de Meknès, 2013.

Table des figures

Figure I. 1 : Fonction de répartition de la loi normale générale	107
Figure I. 2 : Création de base de données.....	117
Figure I. 3 : Sélection de la langue des textes	117
Figure I. 4 : Sélection du lemmatiseur	118
Figure I. 5 : Introduction des titres longs des textes.....	119
Figure I. 6 : Saisie des titres longs, courts et abrégé.....	120
Figure I. 7 : Aperçu sur les textes	121
Figure I. 8 : Boite de dialogue rappelant le nombre des textes constitutifs du corpus	122
Figure I. 9 : Lemmatisation à la volée.....	123
Figure I. 10 : Importation des textes	124
Figure I. 11 : Importation du dernier texte	125
Figure I. 12 : Fin de la lemmatisation et démarrage de l'indexation	126
Figure I. 13 : Phase d'extraction des passages spécifiques	127
Figure I. 14 : Sélection du corpus de référence	128
Figure I. 15 : Validation des titres des textes	129
Figure I. 16 : Étape de sauvegarde des modifications	130
Figure I. 17 : Réinitialisation de la base.....	131
Figure I. 18 : Menu principal.....	132
Figure I. 19 : Mot « aller » dans le dictionnaire.....	135
Figure I. 20 : Dictionnaire hiérarchique.....	136
Figure I. 21 : Contexte du mot « maison » dans les textes du corpus	138
Figure I. 22 : Texte d'origine de présence du mot « maison » dans l'ouvrage « Parcours ».....	139
Figure I. 23 : Concordance du mot « maison »	140
Figure I. 24 : Étendue relative des textes du corpus.....	141
Figure I. 25 : Répartition des textes du corpus selon l'étendue relative	142

Figure I. 26 : Sélection de l'objet de la représentation	143
Figure I. 27 : Distribution du mot « nuit »	144
Figure I. 28 : Distribution conjointe des deux substantifs « nuit » et « patriote ».....	145
Figure I. 29 : Représentation de l'écart entre les deux occurrences « nuit » et « maison »	146
Figure I. 30 : Évolution du lexique par ordre alphabétique	147
Figure I. 31 : Évolution hiérarchique du lexique.....	148
Figure I. 32 : Graphe de l'analyse factorielle.....	150
Figure I. 33 : Vocabulaire et Hapax.....	151
Figure I. 34 : Accroissement du lexique dans l'ordre normal	152
Figure I. 35 : Distributions des fréquences.....	152
Figure I. 36 : Accroissement inverse du lexique	153
Figure I. 37 : Analyse factorielle de la distance lexicale établie sur N	155
Figure I. 38 : Connexion lexicale sur l'ensemble	155
Figure I. 39 : Comparaison des textes.....	156
Figure I. 40 : Comparaison interne.....	157
 Figure II. 1 : Analyse factorielle des classes de longueurs	 193
Figure II. 2 : Répartition des textes selon le nombre d'occurrences	200
Figure II. 3 : Progression chronologique de N'et V' réel	220
Figure II. 4 : Richesse lexicale selon la formule de P. GUIRAUD	237
Figure II. 5 : Richesse lexicale (Vocabulaire déficitaire)	249
Figure II. 6 : Richesse lexicale selon l'indice w de BRUNET.....	255
Figure II. 7 : Accroissement lexicale	266
Figure II. 8 : Accroissement du vocabulaire selon l'ordre inverse.....	276
Figure II. 9 : Distribution des hapax.....	278
Figure II. 10 : Répartition des catégories grammaticales dans les textes	291
Figure II. 11 : Répartition des catégories grammaticales dans le corpus.....	292
Figure II. 12 : Répartition des catégories grammaticales dans le corpus.....	293
Figure II. 13 : Répartition des textes selon l'écart réduit des verbes	302

Figure II. 14 : Répartition des textes selon l'écart réduit des adverbes.....	303
Figure II. 15 : Répartition des textes selon l'écart réduit des mots de relation	304
Figure II. 16 : Répartition des textes selon l'écart réduit des substantifs.....	305
Figure II. 17 : Confrontation des textes selon les catégories verbales et nominale....	307
Figure II. 18 : AFC de la répartition des catégories grammaticales.....	318
Figure III. 1 : Effectifs des formes communes dans le corpus.....	340
Figure III. 2 : Analyse factorielle de la distance lexicale (lemmes)	341
Figure III. 3 : Analyse arborée de la distance lexicale (lemmes).....	343
Figure III. 4 : Origines linguistiques des emprunts.....	454
Figure III. 5 : Répartition des emprunts à l'arabe	455

Table des tableaux

Tableau I. 1 : Extrait de la table de la loi normale centrée réduite.....	109
Tableau I. 2 : Étendue des textes et du corpus.....	112
Tableau II. 1 : Répartition des lettres dans le corpus.....	168
Tableau II. 2 : Répartition du corpus selon le nombre d’occurrences	170
Tableau II. 3 : Longueur moyenne des mots dans les textes du corpus.....	171
Tableau II. 4 : Calcul de la variance dans PARCOURS.....	175
Tableau II. 5 : Calcul de la variance dans AÎLEN	176
Tableau II. 6 : Table de la loi normale centrée et réduite.....	179
Tableau II. 7 : Écart réduit des textes pris deux à deux.....	181
Tableau II. 8 : Application du test de Khi-deux.....	186
Tableau II. 9 : Calcul des écarts réduits	188
Tableau II. 10 : Distribution des classes de n lettres (vocables)	190
Tableau II. 11 : Classement des textes selon le nombre d’occurrence.....	198
Tableau II. 12 : Calcul de la corrélation des rangs des textes	204
Tableau II. 13 : Application de la loi binomiale à PARCOURS.....	209
Tableau II. 14 : Écart entre la valeur réelle et la valeur théorique.....	212
Tableau II. 15 : Nombres de mots qui manqueront dans le corpus	217
Tableau II. 16 : Test de Spearman	221
Tableau II. 17 : Calcul du rapport V/N	231
Tableau II. 18 : Calcul du rapport $V1/N$	232
Tableau II. 19 : Calcul du rapport $V1V$.....	234
Tableau II. 20 : Richesse lexicale selon la formule de GUIRAUD	236
Tableau II. 21 : Vérification des conditions de la fréquences moyennes	242
Tableau II. 22 : Corrélation entre N et V	245
Tableau II. 23 : Richesse lexicale selon la loi binomiale.....	248
Tableau II. 24 : Richesse lexicale selon l’indice w de BRUNET.....	254
Tableau II. 25 : Comparaison des méthodes de mesure de la richesse lexicale	258

Tableau II. 26 : Accroissement du vocabulaire dans l'ordre chronologique	263
Tableau II. 27 : Accroissement lexical	267
Tableau II. 28 : Accroissement du vocabulaire selon $w = 9,09$	271
Tableau II. 29 : Accroissement du vocabulaire selon l'ordre inverse	275
Tableau II. 30 : Répartition des textes selon les catégories grammaticales.....	288
Tableau II. 31 : Répartition des textes selon les catégories grammaticales.....	290
Tableau II. 32 : Effectifs des catégories par textes & leurs probabilités	295
Tableau II. 33 : Répartition des textes selon les effectifs théoriques des catégories grammaticales	297
Tableau II. 34 : Répartition des textes selon les effectifs théoriques des substantifs	298
Tableau II. 35 : Répartition des textes selon les effectifs théoriques des verbes	298
Tableau II. 36 : Répartition des textes selon les effectifs théoriques des adjectifs	299
Tableau II. 37 : Répartition des textes selon les effectifs théoriques des adverbes....	299
Tableau II. 38 : Répartition des textes selon les effectifs théoriques des mots de relation	300
Tableau II. 39 : Traduction binaire de la significativité du test	301
Tableau II. 40 : Répartition des textes et catégories grammaticales selon l'écart réduit	306
Tableau II. 41 : Test de Khi2 appliqué à la classe des substantifs dans PARCOURS	310
Tableau II. 42 : Test de Khi2 appliqué à toutes les catégories grammaticales dans PARCOURS	311
Tableau II. 43 : Test de Khi2 appliqué à toutes les catégories grammaticales dans tous les textes	312
Tableau III. 1 : Effectifs des formes communes dans le corpus	339
Tableau III. 2 : Confrontation des textes deux à deux selon la distance globale	345
Tableau III. 3 : Exemple de la connexion lexicale du texte Parcours avec les autres textes du corpus.....	349
Tableau III. 4 : Connexion lexicale des binômes.....	350
Tableau III. 5 : Connexion lexicale (indice brut)	353
Tableau III. 6 : Effectifs réels des textes EL HAKI et LETTRES	356

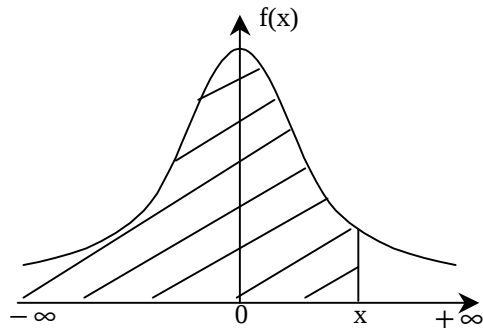
Tableau III. 7 : Effectifs théoriques des textes LETTRES et EL HAKI	358
Tableau III. 8 : Ecart entre effectifs théoriques et réels des textes EL HAKI et LETTRES	359
Tableau III. 9 : Écart au carré des textes EL HAKI et LETTRES	360
Tableau III. 10 : Khi-deux des textes EL HAKI et LETTRES.....	361
Tableau III. 11 : Effectifs réels des textes MILLE et ÀILEN.....	363
Tableau III. 12 : Effectifs théoriques des textes MILLE et ÀILEN.....	364
Tableau III. 13 : Écart entre effectifs théoriques et réels des textes MILLE et ÀILEN	365
Tableau III. 14 : Écart au carré des textes MILLE et ÀILEN.....	366
Tableau III. 15 : Khi-deux des textes MILLE et ÀILEN	367
Tableau III. 16 : Appartenances linguistiques et ethniques (Substantifs).....	374
Tableau III. 17 : Appartenances linguistiques et ethniques (Adjectifs).....	377
Tableau III. 18 : Littérature	378
Tableau III. 19 : Vie.....	379
Tableau III. 20 : Joie.....	380
Tableau III. 21 : Souffrance.....	381
Tableau III. 22 : Langage.....	382
Tableau III. 23 : Nature	383
Tableau III. 24 : Parties du corps.....	384
Tableau III. 25 : Lieux et habitations	385
Tableau III. 26 : Peinture.....	386
Tableau III. 27 : Passion.....	387
Tableau III. 28 : Art culinaire	388
Tableau III. 29 : Politique	389
Tableau III. 30 : Fiction	391
Tableau III. 31 : Verbes de déplacement.....	392
Tableau III. 32 : Substantifs relatifs au voyage	393
Tableau III. 33 : Verbes de réflexion	397
Tableau III. 34 : Vérité.....	398

Tableau III. 35 : Oubli.....	399
Tableau III. 36 : Communauté.....	400
Tableau III. 37 : Famille.....	401
Tableau III. 38 : Temps.....	402
Tableau III. 39 : Vocabulaire spécifique PARCOURS	405
Tableau III. 40 : Vocabulaire spécifique AILEN.....	411
Tableau III. 41 : Vocabulaire spécifique Mille.....	416
Tableau III. 42 : Vocabulaire spécifique d’El HAKI	421
Tableau III. 43 : Vocabulaire spécifique ZRIREK.....	425
Tableau III. 44 : Vocabulaire spécifique GENET	432
Tableau III. 45 : Vocabulaire spécifique ABOUNOUR	436
Tableau III. 46 : Vocabulaire spécifique de FEMME	439
Tableau III. 47 : Vocabulaire spécifique LETTRES	443
Tableau III. 48 : Emprunts de "luxe".....	460
Tableau III. 49 : Emprunts naturalisés	467
Tableau III. 50 : Calques morphologiques	492
Tableau III. 51 : Calques morphologiques non traduits	492

Annexes

Loi Normale centrée réduite

Probabilité de trouver une valeur inférieure à x.



$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2p}} e^{-\frac{u^2}{2}} du$$

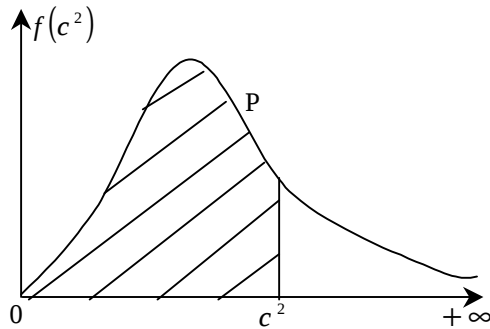
X	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Table pour les grandes valeurs de x :

x	3	3,2	3,4	3,6	3,8	4	4,2	4,4	4,6	4,8
F(x)	0,99865003	0,99931280	0,99966302	0,99984085	0,99992763	0,99996831	0,99998665	0,99999458	0,99999789	0,99999921

Loi du c^2

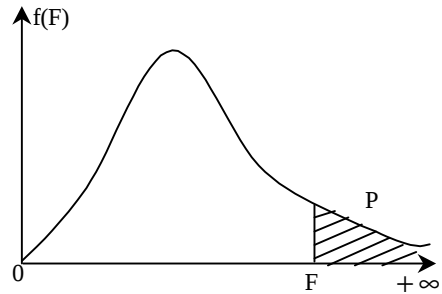
Valeur de C^2 ayant la probabilité P d'être dépassée.



ddl/P	0,5%	1,0%	2,5%	5,0%	10,0%	50,0%	90,0%	95,0%	97,5%	99,0%	99,5%
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,041	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290
27	11,808	12,878	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,994
29	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,335
30	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672
31	14,458	15,655	17,539	19,281	21,434	30,336	41,422	44,985	48,232	52,191	55,002
32	15,134	16,362	18,291	20,072	22,271	31,336	42,585	46,194	49,480	53,486	56,328
33	15,815	17,073	19,047	20,867	23,110	32,336	43,745	47,400	50,725	54,775	57,648
34	16,501	17,789	19,806	21,664	23,952	33,336	44,903	48,602	51,966	56,061	58,964
35	17,192	18,509	20,569	22,465	24,797	34,336	46,059	49,802	53,203	57,342	60,275

Lorsque $n > 30$ on peut admettre que la quantité $\sqrt{2c^2} - \sqrt{2n-1}$ suit une loi normale centrée réduite.

Loi de Fisher-Snedecor



Valeurs de F ayant 5% de chances d'être dépassées.

$n_2 \backslash n_1$	1	2	3	4	5	6	8	10	12	18	24	30	50	60	120
1	161,446	199,499	215,707	224,583	230,160	233,988	238,884	241,882	243,905	247,324	249,052	250,096	251,774	252,196	253,254
2	18,513	19,000	19,164	19,247	19,296	19,329	19,371	19,396	19,412	19,440	19,454	19,463	19,476	19,479	19,487
3	10,128	9,552	9,277	9,117	9,013	8,941	8,845	8,785	8,745	8,675	8,638	8,617	8,581	8,572	8,549
4	7,709	6,944	6,591	6,388	6,256	6,163	6,041	5,964	5,912	5,821	5,774	5,746	5,699	5,688	5,658
5	6,608	5,786	5,409	5,192	5,050	4,950	4,818	4,735	4,678	4,579	4,527	4,496	4,444	4,431	4,398
6	5,987	5,143	4,757	4,534	4,387	4,284	4,147	4,060	4,000	3,896	3,841	3,808	3,754	3,740	3,705
7	5,591	4,737	4,347	4,120	3,972	3,866	3,726	3,637	3,575	3,467	3,410	3,376	3,319	3,304	3,267
8	5,318	4,459	4,066	3,838	3,688	3,581	3,438	3,347	3,284	3,173	3,115	3,079	3,020	3,005	2,967
9	5,117	4,256	3,863	3,633	3,482	3,374	3,230	3,137	3,073	2,960	2,900	2,864	2,803	2,787	2,748
10	4,965	4,103	3,708	3,478	3,326	3,217	3,072	2,978	2,913	2,798	2,737	2,700	2,637	2,621	2,580
11	4,844	3,982	3,587	3,357	3,204	3,095	2,948	2,854	2,788	2,671	2,609	2,570	2,507	2,490	2,448
12	4,747	3,885	3,490	3,259	3,106	2,996	2,849	2,753	2,687	2,568	2,505	2,466	2,401	2,384	2,341
13	4,667	3,806	3,411	3,179	3,025	2,915	2,767	2,671	2,604	2,484	2,420	2,380	2,314	2,297	2,252
14	4,600	3,739	3,344	3,112	2,958	2,848	2,699	2,602	2,534	2,413	2,349	2,308	2,241	2,223	2,178
15	4,543	3,682	3,287	3,056	2,901	2,790	2,641	2,544	2,475	2,353	2,288	2,247	2,178	2,160	2,114
16	4,494	3,634	3,239	3,007	2,852	2,741	2,591	2,494	2,425	2,302	2,235	2,194	2,124	2,106	2,059
17	4,451	3,592	3,197	2,965	2,810	2,699	2,548	2,450	2,381	2,257	2,190	2,148	2,077	2,058	2,011
18	4,414	3,555	3,160	2,928	2,773	2,661	2,510	2,412	2,342	2,217	2,150	2,107	2,035	2,017	1,968
19	4,381	3,522	3,127	2,895	2,740	2,628	2,477	2,378	2,308	2,182	2,114	2,071	1,999	1,980	1,930
20	4,351	3,493	3,098	2,866	2,711	2,599	2,447	2,348	2,278	2,151	2,082	2,039	1,966	1,946	1,896
21	4,325	3,467	3,072	2,840	2,685	2,573	2,420	2,321	2,250	2,123	2,054	2,010	1,936	1,916	1,866
22	4,301	3,443	3,049	2,817	2,661	2,549	2,397	2,297	2,226	2,098	2,028	1,984	1,909	1,889	1,838
23	4,279	3,422	3,028	2,796	2,640	2,528	2,375	2,275	2,204	2,075	2,005	1,961	1,885	1,865	1,813
24	4,260	3,403	3,009	2,776	2,621	2,508	2,355	2,255	2,183	2,054	1,984	1,939	1,863	1,842	1,790
25	4,242	3,385	2,991	2,759	2,603	2,490	2,337	2,236	2,165	2,035	1,964	1,919	1,842	1,822	1,768
26	4,225	3,369	2,975	2,743	2,587	2,474	2,321	2,220	2,148	2,018	1,946	1,901	1,823	1,803	1,749
27	4,210	3,354	2,960	2,728	2,572	2,459	2,305	2,204	2,132	2,002	1,930	1,884	1,806	1,785	1,731
28	4,196	3,340	2,947	2,714	2,558	2,445	2,291	2,190	2,118	1,987	1,915	1,869	1,790	1,769	1,714
29	4,183	3,328	2,934	2,701	2,545	2,432	2,278	2,177	2,104	1,973	1,901	1,854	1,775	1,754	1,698
30	4,171	3,316	2,922	2,690	2,534	2,421	2,266	2,165	2,092	1,960	1,887	1,841	1,761	1,740	1,683
31	4,160	3,305	2,911	2,679	2,523	2,409	2,255	2,153	2,080	1,948	1,875	1,828	1,748	1,726	1,670
32	4,149	3,295	2,901	2,668	2,512	2,399	2,244	2,142	2,070	1,937	1,864	1,817	1,736	1,714	1,657
33	4,139	3,285	2,892	2,659	2,503	2,389	2,235	2,133	2,060	1,926	1,853	1,806	1,724	1,702	1,645
34	4,130	3,276	2,883	2,650	2,494	2,380	2,225	2,123	2,050	1,917	1,843	1,795	1,713	1,691	1,633
35	4,121	3,267	2,874	2,641	2,485	2,372	2,217	2,114	2,041	1,907	1,833	1,786	1,703	1,681	1,623
40	4,085	3,232	2,839	2,606	2,449	2,336	2,180	2,077	2,003	1,868	1,793	1,744	1,660	1,637	1,577
50	4,034	3,183	2,790	2,557	2,400	2,286	2,130	2,026	1,952	1,814	1,737	1,687	1,599	1,576	1,511
80	3,960	3,111	2,719	2,486	2,329	2,214	2,056	1,951	1,875	1,734	1,654	1,602	1,508	1,482	1,411
100	3,936	3,087	2,696	2,463	2,305	2,191	2,032	1,927	1,850	1,708	1,627	1,573	1,477	1,450	1,376
120	3,920	3,072	2,680	2,447	2,290	2,175	2,016	1,910	1,834	1,690	1,608	1,554	1,457	1,429	1,352

Valeurs de F ayant 2,5% de chances d'être dépassées.

$n_2 \backslash n_1$	1	2	3	4	5	6	8	10	12	18	24	30	50	60	120
1	647,793	799,482	864,151	899,599	921,835	937,114	956,643	968,634	976,725	990,345	997,272	1001,405	1008,098	1009,787	1014,036
2	38,506	39,000	39,166	39,248	39,298	39,331	39,373	39,398	39,415	39,442	39,457	39,465	39,478	39,481	39,489
3	17,443	16,044	15,439	15,101	14,885	14,735	14,540	14,419	14,337	14,196	14,124	14,081	14,010	13,992	13,947
4	12,218	10,649	9,979	9,604	9,364	9,197	8,980	8,844	8,751	8,592	8,511	8,461	8,381	8,360	8,309
5	10,007	8,434	7,764	7,388	7,146	6,978	6,757	6,619	6,525	6,362	6,278	6,227	6,144	6,123	6,069
6	8,813	7,260	6,599	6,227	5,988	5,820	5,600	5,461	5,366	5,202	5,117	5,065	4,980	4,959	4,904
7	8,073	6,542	5,890	5,523	5,285	5,119	4,899	4,761	4,666	4,501	4,415	4,362	4,276	4,254	4,199
8	7,571	6,059	5,416	5,053	4,817	4,652	4,433	4,295	4,200	4,034	3,947	3,894	3,807	3,784	3,728
9	7,209	5,715	5,078	4,718	4,484	4,320	4,102	3,964	3,868	3,701	3,614	3,560	3,472	3,449	3,392
10	6,937	5,456	4,826	4,468	4,236	4,072	3,855	3,717	3,621	3,453	3,365	3,311	3,221	3,198	3,140
11	6,724	5,256	4,630	4,275	4,044	3,881	3,664	3,526	3,430	3,261	3,173	3,118	3,027	3,004	2,944
12	6,554	5,096	4,474	4,121	3,891	3,728	3,512	3,374	3,277	3,108	3,019	2,963	2,871	2,848	2,787
13	6,414	4,965	4,347	3,996	3,767	3,604	3,388	3,250	3,153	2,983	2,893	2,837	2,744	2,720	2,659
14	6,298	4,857	4,242	3,892	3,663	3,501	3,285	3,147	3,050	2,879	2,789	2,732	2,638	2,614	2,552
15	6,200	4,765	4,153	3,804	3,576	3,415	3,199	3,060	2,963	2,792	2,701	2,644	2,549	2,524	2,461
16	6,115	4,687	4,077	3,729	3,502	3,341	3,125	2,986	2,889	2,717	2,625	2,568	2,472	2,447	2,383
17	6,042	4,619	4,011	3,665	3,438	3,277	3,061	2,922	2,825	2,652	2,560	2,502	2,405	2,380	2,315
18	5,978	4,560	3,954	3,608	3,382	3,221	3,005	2,866	2,769	2,596	2,503	2,445	2,347	2,321	2,256
19	5,922	4,508	3,903	3,559	3,333	3,172	2,956	2,817	2,720	2,546	2,452	2,394	2,295	2,270	2,203
20	5,871	4,461	3,859	3,515	3,289	3,128	2,913	2,774	2,676	2,501	2,408	2,349	2,249	2,223	2,156
21	5,827	4,420	3,819	3,475	3,250	3,090	2,874	2,735	2,637	2,462	2,368	2,308	2,208	2,182	2,114
22	5,786	4,383	3,783	3,440	3,215	3,055	2,839	2,700	2,602	2,426	2,332	2,272	2,171	2,145	2,076
23	5,750	4,349	3,750	3,408	3,183	3,023	2,808	2,668	2,570	2,394	2,299	2,239	2,137	2,111	2,041
24	5,717	4,319	3,721	3,379	3,155	2,995	2,779	2,640	2,541	2,365	2,269	2,209	2,107	2,080	2,010
25	5,686	4,291	3,694	3,353	3,129	2,969	2,753	2,613	2,515	2,338	2,242	2,182	2,079	2,052	1,981
26	5,659	4,265	3,670	3,329	3,105	2,945	2,729	2,590	2,491	2,314	2,217	2,157	2,053	2,026	1,954
27	5,633	4,242	3,647	3,307	3,083	2,923	2,707	2,568	2,469	2,291	2,195	2,133	2,029	2,002	1,930
28	5,610	4,221	3,626	3,286	3,063	2,903	2,687	2,547	2,448	2,270	2,174	2,112	2,007	1,980	1,907
29	5,588	4,201	3,607	3,267	3,044	2,884	2,669	2,529	2,430	2,251	2,154	2,092	1,987	1,959	1,886
30	5,568	4,182	3,589	3,250	3,026	2,867	2,651	2,511	2,412	2,233	2,136	2,074	1,968	1,940	1,866
31	5,549	4,165	3,573	3,234	3,010	2,851	2,635	2,495	2,396	2,217	2,119	2,057	1,950	1,922	1,848
32	5,531	4,149	3,557	3,218	2,995	2,836	2,620	2,480	2,381	2,201	2,103	2,041	1,934	1,905	1,831
33	5,515	4,134	3,543	3,204	2,981	2,822	2,606	2,466	2,366	2,187	2,088	2,026	1,918	1,890	1,815
34	5,499	4,120	3,529	3,191	2,968	2,808	2,593	2,453	2,353	2,173	2,075	2,012	1,904	1,875	1,799
35	5,485	4,106	3,517	3,179	2,956	2,796	2,581	2,440	2,341	2,160	2,062	1,999	1,890	1,861	1,785
40	5,424	4,051	3,463	3,126	2,904	2,744	2,529	2,388	2,288	2,107	2,007	1,943	1,832	1,803	1,724
50	5,340	3,975	3,390	3,054	2,833	2,674	2,458	2,317	2,216	2,033	1,931	1,866	1,752	1,721	1,639
80	5,218	3,864	3,284	2,950	2,730	2,571	2,355	2,213	2,111	1,925	1,820	1,752	1,632	1,599	1,508
100	5,179	3,828	3,250	2,917	2,696	2,537	2,321	2,179	2,077	1,890	1,784	1,715	1,592	1,558	1,463
120	5,152	3,805	3,227	2,894	2,674	2,515	2,299	2,157	2,055	1,866	1,760	1,690	1,565	1,530	1,433

Valeurs de F ayant 1% de chances d'être dépassées.

$n_2 \backslash n_1$	1	2	3	4	5	6	8	10	12	18	24	30	50	60	120
1	4052,185	4999,340	5403,534	5624,257	5763,955	5858,950	5980,954	6055,925	6106,682	6191,432	6234,273	6260,350	6302,260	6312,970	6339,513
2	98,502	99,000	99,164	99,251	99,302	99,331	99,375	99,397	99,419	99,444	99,455	99,466	99,477	99,484	99,491
3	34,116	30,816	29,457	28,710	28,237	27,911	27,489	27,228	27,052	26,751	26,597	26,504	26,354	26,316	26,221
4	21,198	18,000	16,694	15,977	15,522	15,207	14,799	14,546	14,374	14,079	13,929	13,838	13,690	13,652	13,558
5	16,258	13,274	12,060	11,392	10,967	10,672	10,289	10,051	9,888	9,609	9,466	9,379	9,238	9,202	9,112
6	13,745	10,925	9,780	9,148	8,746	8,466	8,102	7,874	7,718	7,451	7,313	7,229	7,091	7,057	6,969
7	12,246	9,547	8,451	7,847	7,460	7,191	6,840	6,620	6,469	6,209	6,074	5,992	5,858	5,824	5,737
8	11,259	8,649	7,591	7,006	6,632	6,371	6,029	5,814	5,667	5,412	5,279	5,198	5,065	5,032	4,946
9	10,562	8,022	6,992	6,422	6,057	5,802	5,467	5,257	5,111	4,860	4,729	4,649	4,517	4,483	4,398
10	10,044	7,559	6,552	5,994	5,636	5,386	5,057	4,849	4,706	4,457	4,327	4,247	4,115	4,082	3,996
11	9,646	7,206	6,217	5,668	5,316	5,069	4,744	4,539	4,397	4,150	4,021	3,941	3,810	3,776	3,690
12	9,330	6,927	5,953	5,412	5,064	4,821	4,499	4,296	4,155	3,910	3,780	3,701	3,569	3,535	3,449
13	9,074	6,701	5,739	5,205	4,862	4,620	4,302	4,100	3,960	3,716	3,587	3,507	3,375	3,341	3,255
14	8,862	6,515	5,564	5,035	4,695	4,456	4,140	3,939	3,800	3,556	3,427	3,348	3,215	3,181	3,094
15	8,683	6,359	5,417	4,893	4,556	4,318	4,004	3,805	3,666	3,423	3,294	3,214	3,081	3,047	2,959
16	8,531	6,226	5,292	4,773	4,437	4,202	3,890	3,691	3,553	3,310	3,181	3,101	2,967	2,933	2,845
17	8,400	6,112	5,185	4,669	4,336	4,101	3,791	3,593	3,455	3,212	3,083	3,003	2,869	2,835	2,746
18	8,285	6,013	5,092	4,579	4,248	4,015	3,705	3,508	3,371	3,128	2,999	2,919	2,784	2,749	2,660
19	8,185	5,926	5,010	4,500	4,171	3,939	3,631	3,434	3,297	3,054	2,925	2,844	2,709	2,674	2,584
20	8,096	5,849	4,938	4,431	4,103	3,871	3,564	3,368	3,231	2,989	2,859	2,778	2,643	2,608	2,517
21	8,017	5,780	4,874	4,369	4,042	3,812	3,506	3,310	3,173	2,931	2,801	2,720	2,584	2,548	2,457
22	7,945	5,719	4,817	4,313	3,988	3,758	3,453	3,258	3,121	2,879	2,749	2,667	2,531	2,495	2,403
23	7,881	5,664	4,765	4,264	3,939	3,710	3,406	3,211	3,074	2,832	2,702	2,620	2,483	2,447	2,354
24	7,823	5,614	4,718	4,218	3,895	3,667	3,363	3,168	3,032	2,789	2,659	2,577	2,440	2,403	2,310
25	7,770	5,568	4,675	4,177	3,855	3,627	3,324	3,129	2,993	2,751	2,620	2,538	2,400	2,364	2,270
26	7,721	5,526	4,637	4,140	3,818	3,591	3,288	3,094	2,958	2,715	2,585	2,503	2,364	2,327	2,233
27	7,677	5,488	4,601	4,106	3,785	3,558	3,256	3,062	2,926	2,683	2,552	2,470	2,330	2,294	2,198
28	7,636	5,453	4,568	4,074	3,754	3,528	3,226	3,032	2,896	2,653	2,522	2,440	2,300	2,263	2,167
29	7,598	5,420	4,538	4,045	3,725	3,499	3,198	3,005	2,868	2,626	2,495	2,412	2,271	2,234	2,138
30	7,562	5,390	4,510	4,018	3,699	3,473	3,173	2,979	2,843	2,600	2,469	2,386	2,245	2,208	2,111
31	7,530	5,362	4,484	3,993	3,675	3,449	3,149	2,955	2,820	2,577	2,445	2,362	2,221	2,183	2,086
32	7,499	5,336	4,459	3,969	3,652	3,427	3,127	2,934	2,798	2,555	2,423	2,340	2,198	2,160	2,062
33	7,471	5,312	4,437	3,948	3,630	3,406	3,106	2,913	2,777	2,534	2,402	2,319	2,176	2,139	2,040
34	7,444	5,289	4,416	3,927	3,611	3,386	3,087	2,894	2,758	2,515	2,383	2,299	2,156	2,118	2,019
35	7,419	5,268	4,396	3,908	3,592	3,368	3,069	2,876	2,740	2,497	2,364	2,281	2,137	2,099	2,000
40	7,314	5,178	4,313	3,828	3,514	3,291	2,993	2,801	2,665	2,421	2,288	2,203	2,058	2,019	1,917
50	7,171	5,057	4,199	3,720	3,408	3,186	2,890	2,698	2,563	2,318	2,183	2,098	1,949	1,909	1,803
80	6,963	4,881	4,036	3,563	3,255	3,036	2,742	2,551	2,415	2,169	2,032	1,944	1,788	1,746	1,630
100	6,895	4,824	3,984	3,513	3,206	2,988	2,694	2,503	2,368	2,120	1,983	1,893	1,735	1,692	1,572
120	6,851	4,787	3,949	3,480	3,174	2,956	2,663	2,472	2,336	2,089	1,950	1,860	1,700	1,656	1,533

Valeurs de F ayant 0,5% de chances d'être dépassées.

$n_2 \backslash n_1$	1	2	3	4	5	6	8	10	12	18	24	30	50	60	120
1	16212,463	19997,358	21614,134	22500,753	23055,822	23439,527	23923,814	24221,838	24426,728	24765,730	24937,093	25041,401	25212,765	25253,743	25358,051
2	198,503	199,012	199,158	199,245	199,303	199,332	199,376	199,390	199,419	199,449	199,449	199,478	199,478	199,478	199,492
3	55,552	49,800	47,468	46,195	45,391	44,838	44,125	43,685	43,387	42,881	42,623	42,466	42,211	42,150	41,990
4	31,332	26,284	24,260	23,154	22,456	21,975	21,352	20,967	20,705	20,258	20,030	19,892	19,667	19,611	19,469
5	22,785	18,314	16,530	15,556	14,939	14,513	13,961	13,618	13,385	12,985	12,780	12,656	12,454	12,402	12,274
6	18,635	14,544	12,917	12,028	11,464	11,073	10,566	10,250	10,034	9,664	9,474	9,358	9,170	9,122	9,001
7	16,235	12,404	10,883	10,050	9,522	9,155	8,678	8,380	8,176	7,826	7,645	7,534	7,354	7,309	7,193
8	14,688	11,043	9,597	8,805	8,302	7,952	7,496	7,211	7,015	6,678	6,503	6,396	6,222	6,177	6,065
9	13,614	10,107	8,717	7,956	7,471	7,134	6,693	6,417	6,227	5,899	5,729	5,625	5,454	5,410	5,300
10	12,827	9,427	8,081	7,343	6,872	6,545	6,116	5,847	5,661	5,340	5,173	5,071	4,902	4,859	4,750
11	12,226	8,912	7,600	6,881	6,422	6,102	5,682	5,418	5,236	4,921	4,756	4,654	4,488	4,445	4,337
12	11,754	8,510	7,226	6,521	6,071	5,757	5,345	5,085	4,906	4,595	4,431	4,331	4,165	4,123	4,015
13	11,374	8,186	6,926	6,233	5,791	5,482	5,076	4,820	4,643	4,334	4,173	4,073	3,908	3,866	3,758
14	11,060	7,922	6,680	5,998	5,562	5,257	4,857	4,603	4,428	4,122	3,961	3,862	3,697	3,655	3,547
15	10,798	7,701	6,476	5,803	5,372	5,071	4,674	4,424	4,250	3,946	3,786	3,687	3,523	3,480	3,372
16	10,576	7,514	6,303	5,638	5,212	4,913	4,521	4,272	4,099	3,797	3,638	3,539	3,375	3,332	3,224
17	10,384	7,354	6,156	5,497	5,075	4,779	4,389	4,142	3,971	3,670	3,511	3,412	3,248	3,206	3,097
18	10,218	7,215	6,028	5,375	4,956	4,663	4,276	4,030	3,860	3,560	3,402	3,303	3,139	3,096	2,987
19	10,073	7,093	5,916	5,268	4,853	4,561	4,177	3,933	3,763	3,464	3,306	3,208	3,043	3,000	2,891
20	9,944	6,987	5,818	5,174	4,762	4,472	4,090	3,847	3,678	3,380	3,222	3,123	2,959	2,916	2,806
21	9,829	6,891	5,730	5,091	4,681	4,393	4,013	3,771	3,602	3,305	3,147	3,049	2,884	2,841	2,730
22	9,727	6,806	5,652	5,017	4,609	4,322	3,944	3,703	3,535	3,239	3,081	2,982	2,817	2,774	2,663
23	9,635	6,730	5,582	4,950	4,544	4,259	3,882	3,642	3,474	3,179	3,021	2,922	2,756	2,713	2,602
24	9,551	6,661	5,519	4,890	4,486	4,202	3,826	3,587	3,420	3,125	2,967	2,868	2,702	2,658	2,546
25	9,475	6,598	5,462	4,835	4,433	4,150	3,776	3,537	3,370	3,075	2,918	2,819	2,652	2,609	2,496
26	9,406	6,541	5,409	4,785	4,384	4,103	3,730	3,492	3,325	3,031	2,873	2,774	2,607	2,563	2,450
27	9,342	6,489	5,361	4,740	4,340	4,059	3,687	3,450	3,284	2,990	2,832	2,733	2,565	2,522	2,408
28	9,284	6,440	5,317	4,698	4,300	4,020	3,649	3,412	3,246	2,952	2,794	2,695	2,527	2,483	2,369
29	9,230	6,396	5,276	4,659	4,262	3,983	3,613	3,376	3,211	2,917	2,759	2,660	2,492	2,448	2,333
30	9,180	6,355	5,239	4,623	4,228	3,949	3,580	3,344	3,179	2,885	2,727	2,628	2,459	2,415	2,300
31	9,133	6,316	5,204	4,590	4,195	3,918	3,549	3,314	3,149	2,855	2,697	2,598	2,429	2,385	2,269
32	9,090	6,281	5,172	4,559	4,166	3,889	3,521	3,286	3,121	2,828	2,670	2,570	2,401	2,356	2,240
33	9,049	6,248	5,141	4,531	4,138	3,861	3,495	3,260	3,095	2,802	2,644	2,544	2,374	2,330	2,213
34	9,012	6,217	5,113	4,504	4,112	3,836	3,470	3,235	3,071	2,778	2,620	2,520	2,350	2,305	2,188
35	8,976	6,188	5,086	4,479	4,088	3,812	3,447	3,212	3,048	2,755	2,597	2,497	2,327	2,282	2,164
40	8,828	6,066	4,976	4,374	3,986	3,713	3,350	3,117	2,953	2,661	2,502	2,401	2,230	2,184	2,064
50	8,626	5,902	4,826	4,232	3,849	3,579	3,219	2,988	2,825	2,533	2,373	2,272	2,097	2,050	1,925
80	8,335	5,665	4,611	4,028	3,652	3,387	3,032	2,803	2,641	2,349	2,188	2,084	1,903	1,854	1,720
100	8,241	5,589	4,542	3,963	3,589	3,325	2,972	2,744	2,583	2,290	2,128	2,024	1,840	1,790	1,652
120	8,179	5,539	4,497	3,921	3,548	3,285	2,933	2,705	2,544	2,251	2,089	1,984	1,798	1,747	1,606

Pour les grands échantillons, $\frac{s_1 - s_2}{s \sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}} \rightarrow N(0,1)$ avec $s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$.

Table des matières

INTRODUCTION	1
PREMIERE PARTIE. Cadre théorique et méthodologique	1
CHAPITRE PREMIER. L'homme et l'œuvre	18
A. Notice biographique sur Edmond Amran El Maleh	19
B. Edmond Amran El Maleh et l'Art.....	21
1. Enfance : esquisse d'un esthéticien.....	22
2. Esthétique moderne	23
3. Attitude esthétique d'Edmond Amran El Maleh	25
4. Écrits d'Edmond Amran El Maleh sur la peinture.....	27
5. Amitiés d'Edmond Amran El Maleh.....	31
6. Peinture et littérature : deux facettes d'une seule entité	32
C. Toile de fond des œuvres d'Edmond Amran El Maleh	35
1. Œuvre maléhiennne	36
2. Exode et la littérature de l'exil	41
3. Imaginaire judaïque	45
4. Histoire des juifs au Maroc.....	47
CHAPITRE 2. De la statistique à la statistique lexicale.....	57
A. Analyse statistique en linguistique.....	58
1. Statistique et Linguistique	58
1.1. Statistique	58
a. Statistique descriptive	60
b. Statistique théorique ou mathématique	65
1.2. Statistique linguistique	70
2. Linguistique et informatique	72
B. Statistique lexicale	74
1. Dépouillement et quantification du corpus	78
1.1. Norme lexicologique	80
1.2. Lemmatisation du corpus.....	88
2. Présentation de la méthode d'analyse	92

2.1. Probabilités.....	92
a. Notes historiques.....	93
b. Calcul des probabilités	95
2.2. Lois de probabilités discrètes usuelles.....	100
a. Loi de Bernoulli	100
b. Loi binomiale.....	102
2.3. Lois continues.....	104
a. Loi uniforme	104
b. Loi normale (ou loi de LAPLACE GAUSS).....	105
CHAPITRE 3. Corpus et méthodes de découpage et d'analyse.....	111
A. Délimitation du corpus et préparation de la base de données	111
1. Présentation du corpus	112
2. Préparation des données textuelles : Base de données « HYPEREDM.EXE ».....	113
B. Hyperbase	114
1. Création de la base	115
2. Menu principal	131
3. Exploration	133
4. Exploitation statistique	140
5. Calcul hypergéométrique et écart réduit.....	142
6. Corrélation.....	146
7. Analyse factorielle.....	148
8. Richesse lexicale, hapax, accroissement	150
9. Connexion lexicale (ou distance intertextuelle)	153
10. Spécificités.....	155
DEUXIÈME PARTIE. Structure du lexique	164
CHAPITRE PREMIER. Étendue du corpus	165
A. Longueur du mot	165
1. Mesure de la longueur moyenne.....	167
1.1. Calcul du nombre de lettres dans le corpus	167
1.2. Calcul de la longueur moyenne des mots	171
2. Comparaison des moyennes	172
2.1. Calcul de l'écart-type	173
2.2. Calcul de l'écart réduit.....	177

3. Effectifs théoriques	185
4. Groupes de longueurs	188
5. Analyse factorielle.....	192
B. Étendue du corpus.....	197
1. Étendue N (occurrences).....	198
1.1. Classement des textes selon l'étendue N	198
1.2. Test sur le coefficient de corrélation des rangs (Analyse bivariée).....	201
2. Étendue de V (vocabulaire)	206
2.1. Modèle binomial d'une distribution théorique	207
2.2. Test de l'écart réduit ou z score	213
CHAPITRE 2. Richesse lexicale et accroissement du vocabulaire.....	228
A. Richesse lexicale	229
1. Appréciation qualitative de V d'après N.....	231
1.1. Rapport V/N	231
1.2. Rapport V1/N	232
1.3. Rapport V1/V	233
1.4. Formule de GUIRAUD	235
1.5. Méthode binomiale	239
1.6. Indice w d'Etienne Brunet	250
B. Accroissement du vocabulaire	259
1. Accroissement chronologique	261
1.1. Accroissement chronologique selon la formule $y = axb$	261
1.2. Indice w d'Étienne Brunet	268
2. Accroissement inverse	273
CHAPITRE 3. Catégories grammaticales.....	286
A. Distribution des catégories grammaticales	287
B. Comparaison externe	294
C. Effectifs théoriques des catégories grammaticales	294
D. Test de Pearson	308
E. Analyse factorielle des correspondances	317
TROISIEME PARTIE. Contenu lexical.....	333
CHAPITRE PREMIER . Connexion lexicale	336

A. Méthode Jaccard.....	338
1. Vocabulaire commun à deux textes	338
2. Distance intertextuelle	344
3. Indice de connexion lexicale	347
B. Essai de la loi binomiale.....	354
1. Application de la loi binomiale aux textes EL HAKI et LETTRES.....	355
2. Application de la loi binomiale aux textes MILLE et AÏLEN.....	362
CHAPITRE 2. Vocabulaire spécifique du corpus	372
A.Vocabulaire spécifique excédentaire et déficitaire : Étude exogène	372
1. Appartenances linguistiques et ethniques	374
2. Littérature	377
2.1. Écriture	379
2.2. Allégorie	381
3. Art culinaire.....	388
4. Faits historiques et politiques dans l'œuvre d'El Maleh : entre réel et fictif	388
5. Quête de l'identité.....	391
6. La mémoire dans l'œuvre d'El Maleh	394
7. Communauté	399
8. La notion de temps	402
B. Analyse du vocabulaire spécifique de chaque texte : étude endogène	403
1. Parcours Immobile / Aïlen ou la nuit du récit / Mille ans, un jour / Le retour d'Abou El Haki..	404
2. Le café bleu, Zrirk / Jean Genet, Le Captif amoureux et autres essais / Abner Abounour.....	425
3. Lettres à moi-même / Une femme, une mère.....	438
CHAPITRE 3. Plurilinguisme dans l'œuvre d'EDM	451
A. Emprunts	452
1. Origines linguistiques et catégories grammaticales des emprunts	454
1.1. Origines linguistiques de l'emprunt.....	456
1.2. Catégories grammaticales de l'emprunt :	457
2. Emprunts de luxe.....	458
3. Emprunts naturalisés	467
4. Emprunts de nécessité	471

5. Champs sémantiques de l'emprunt	474
B. Calque linguistique	482
1. Calque phraséologique	484
1.1. Expressions calquées sur des proverbes.....	484
1.2. Expressions calquées sur des locutions figées	486
2. Calque morphologique	491
3. Calque sémantique	493
C. Alternance codique	499
D. Interférences phonétiques	509
CONCLUSION GÉNÉRALE	525
Bibliographie et Sitographie	538
Table des figures	551
Table des tableaux	554
Annexes	558
Table des matières	565