

**LA LEXICOMÉTRIE HISTORIQUE : PERSPECTIVES POUR
L'HISTORIOGRAPHIE ALGÉRIENNE**

**HISTORICAL LEXICOMETRY : PERSPECTIVES FOR ALGERIAN
HISTORIOGRAPHY**

Chérif SOUTI

Université Larbi Ben M'hidi, Oum ElBouaghi

Laboratoire SLADD, Université Frères Mentouri Constantine 1

Résumé

Dans la présente publication, il est question de montrer comment la recherche historiographique algérienne pourrait s'approprier des méthodes d'analyse automatique de texte. Nous présentons, par ailleurs, des méthodes lexicométriques susceptibles d'être utilisées pour l'analyse du discours historique. Nous ouvrons également de nouvelles perspectives de recherche que les chercheurs algériens pourraient explorer dans cette optique. Pour étayer nos propos, nous avons effectué, à l'aide du logiciel Hyperbase, une analyse lexicométrique contrastive des discours de deux personnalités historiques algériennes : Ferhat Abbas et Hadj Messali. L'analyse lexicométrique, qui s'articule sur l'étude de la distance lexicale et l'étude des co-occurrences, donne à voir une réalité lexicale dichotomique correspondant au contraste politique qui oppose les deux hommes. En effet, F. Abbas a eu des débuts réformistes puis a évolué progressivement vers le nationalisme, tandis que H. Messali a toujours milité pour l'indépendance totale de l'Algérie. Notre travail montre ainsi que les frontières entre les deux champs de recherche ne sont pas totalement étanches et que des passerelles peuvent être créées dans une perspective historico-discursive. Soulignons enfin que l'approche lexicostatistique ne prétend pas se substituer à l'approche historique mais peut la compléter en ouvrant la voie à une panoplie d'analyses et en permettant de travailler sur des corpus plus étendus.

Mots-clés : lexicométrie, histoire, Algérie, applications, perspectives

Abstract

In this publication, we show how Algerian historiographical research could appropriate automatic text analysis methods. We also present lexicometric methods that can be used for historical discourse analysis. We also open new research perspectives that Algerian researchers could explore in this perspective. We have carried out a contrastive lexicometric analysis of the speeches of two Algerian historical figures: Ferhat Abbas and Hadj Messali, using Hyperbase software. The lexicometric analysis, which is based on the study of lexical distance and the study of co-occurrences, reveals a dichotomous lexical reality corresponding to the political contrast between the two personalities. Indeed, F. Abbas had reformist beginnings and then progressively evolved towards nationalism, whereas M. Hadj Messali always militated for the total independence of Algeria. Our work thus shows that the borders between the two fields of research are not totally watertight and that bridges can be created in a historical-discursive perspective. Finally, it should be emphasized that the lexicostatistical approach does not claim to replace the historical approach but can complement it by opening the way to a wide range of analyses and by allowing us to work on larger corpora.

Keywords: lexicometry, history, Algeria, applications, perspectives

La mise en ligne de corpus historiques grâce à la numérisation des archives a permis de lever une contrainte de taille - la saisie manuelle laborieuse des textes - et a rendu possible la prise en charge de ces corpus avec des outils lexicométriques de plus en plus performants.

Ces derniers ont fait leurs preuves dans plusieurs universités européennes où leur pratique est courante. Actuellement, les travaux les plus importants dans l'aire francophone sont ceux réalisés au sein du laboratoire Bases, Corpus, Langages (BCL) de l'université de Nice sous la direction de Damon Mayaffre. Dans le contexte algérien, elles demeurent peu connues.

Nous tenterons de répondre aux questions suivantes :

Les méthodes lexicométriques peuvent-elles être mises en œuvre dans les recherches historiques algériennes ? Comment ? Quelles pistes de recherche les chercheurs algériens peuvent-ils explorer dans cette perspective ?

L'objectif principal de cette publication est donc de montrer comment la lexicométrie, une branche de recherche qui relève du champ disciplinaire des sciences du langage en général et de l'analyse de discours en particulier, pourrait être mobilisée par les chercheurs algériens en histoire.

Le recours à la lexicométrie permet d'apporter un nouveau regard qui se veut le plus objectif possible afin de mieux appréhender le discours historique à travers son vocabulaire. En effet, grâce à ce que Damon Mayaffre appelle « La médiation entre le sujet analysant et l'objet analysé », au moyen de l'outil informatique, il est possible d'objectiver l'interprétation des textes. (Mayaffre, 2007 : p.171) Mayaffre explique cela en ajoutant que

Ni la lecture naturelle, ni l'interprétation, ni la subjectivité du chercheur ne sont abolies mais elles se trouvent encadrées par une phase de traitements systématiques et exhaustifs de la matière textuelle. (Ibid.)

Nous allons dans un premier temps faire une brève présentation de l'approche lexicométrique, cette dernière étant peu connue en Algérie et peu pratiquée. Nous présenterons ensuite le logiciel de lexicométrie le plus performant et le plus utilisé dans le monde francophone, à savoir Hyperbase*. Puis, nous explorerons les rapports entre lexicométrie et histoire.

Cette présentation en appellera une autre, celle des méthodes lexicométriques pouvant être mises à contribution dans une approche historique et les perspectives de leur exploitation dans le contexte algérien. Cela nous amènera à nous pencher sur la question des sources et des corpus dans le même contexte (cette question étant centrale en lexicométrie).

Enfin, pour illustrer nos propos, nous présenterons les résultats de l'étude lexicométrique sur les discours de Ferhat Abbas et Hadj Messali.

1.LA LEXICOMÉTRIE

La lexicométrie, étude quantitative du lexique par le biais des méthodes statistiques, est une discipline relativement récente qui reste peu connue même au sein de sa discipline-mère, l'analyse de discours. C'est certainement en raison de son caractère ésotérique qu'elle demeure aussi confidentielle encore que les études lexicométriques soient légion dans certaines universités notamment européennes. Née au milieu du XXe siècle, elle a dû surmonter des obstacles d'ordre épistémologique et matériel avant d'atteindre son état actuel (Sinclair, 2002). En effet, les premiers travaux étaient peu nombreux, l'équipement était primitif et dispendieux. En plus, il offrait des fonctionnalités restreintes (Souti, 2016 : p.80).

Il faut souligner, par ailleurs, l'originalité de cette approche utilisée essentiellement en sciences humaines et sociales mais qui

* Soulignons que cette présentation est inspirée de notre thèse de doctorat.

emprunte ses outils aux sciences exactes, principalement l'informatique et les mathématiques. C'est pourquoi on considère que son développement a toujours été tributaire de l'évolution de l'informatique qui lui fournit les outils nécessaires.

Chose peu connue, les premiers travaux d'analyse de discours en France ont été principalement automatiques. On peut citer ceux du philosophe Michel Pêcheux avec son projet *L'Analyse Automatique du Discours* (1969), entreprise innovante qui consacra l'apparition de l'Analyse de discours en France. On peut évoquer aussi les travaux initiés par le Centre de Linguistique Quantitative de Paris créé en 1962, et surtout ceux effectués au sein du laboratoire « Lexicométrie et textes politiques » de l'ENS des Sciences politiques de Saint-Cloud et dont les plus notables ont pour auteur Maurice Tournier, un des plus grands spécialistes et précurseurs de la lexicométrie (Souti, 2015 : p.04).

Le recours à l'outil informatique et aux méthodes statistiques s'explique par la préoccupation des premiers promoteurs de la discipline de garantir une certaine scientificité à leurs travaux et de défendre les Sciences humaines contre ce qu'on appelle « les sciences dures ».

En effet, la lexicométrie est née dans la remise en question de la lecture traditionnelle des textes. Par exemple, en histoire, la méthode traditionnelle consiste à rechercher et recueillir les sources et les textes, les lire puis les interpréter. Ce que les fondateurs de la lexicométrie reprochaient à cette méthode, c'était l'absence d'une prise en compte de la forme du texte et de son sens intrinsèque.

Les deux points essentiels qui font la force de la lexicométrie sont la possibilité de travailler sur des macro-corpus d'une part et de mener des études comparatives de lexique d'autre part. C'est pourquoi elle occupe une place constamment croissante dans les

sciences humaines et sociales. Ses domaines d'application sont de plus en plus nombreux et variés : la linguistique, la littérature, les sciences politiques, les sciences de la communication, l'histoire, la sociologie, etc.

2. LE LOGICIEL HYPERBASE

Hyperbase est un logiciel hypertexte d'exploration des textes et de traitement statistique des données textuelles. Il a été conçu et développé à l'Université de Nice depuis 1989 par Etienne Brunet et produit par l'UMR 6039 Bases, Corpus et Langage (CNRS-Université de Nice). Ce logiciel est actuellement le plus perfectionné des outils disponibles sur le marché eu égard à ses potentialités : c'est le seul programme lexicométrique qui permet un triple traitement du texte brut, du texte lemmatisé et du texte grammaticalement codé.

En effet, parmi la floraison des outils existants, Hyperbase est, grâce aux évolutions récentes, le seul logiciel d'exploration et d'analyse statistique textuelle qui permet de travailler non seulement sur le lexique mais également sur la composition grammaticale (catégories et codes grammaticaux), les structures syntaxiques, le style et la sémantique. Cela est possible grâce à l'intégration à la version récente d'un autre outil, un lemmatiseur connu d'ailleurs pour sa fiabilité. Il s'agit de l'outil Cordial de la société Synapse. En plus de la statistique lexicale traditionnelle sur formes graphiques brutes, Hyperbase propose une statistique grammaticale et syntaxique et ouvre la voie à l'interprétation sémantique. C'est notamment grâce aux développements récents d'Hyperbase que *la lexicométrie classique* a évolué vers la *logométrie* qui prend en compte d'autres éléments que les mots tels que les sèmes et les catégories grammaticales.

Hyperbase[†] offre une palette de fonctions documentaires disposées horizontalement en haut de l'écran : les index, les concordances, les contextes, les cooccurrences ; et des fonctions statistiques disposées verticalement à gauche de l'écran : les spécificités, les analyses factorielles de correspondance, la richesse lexicale, l'évolution du vocabulaire, la distance, la connexion textuelle, etc. (Brunet, 2001)

3. LEXICOMÉTRIE ET HISTOIRE

L'intérêt de l'histoire pour les méthodes lexicométriques de manière particulière et pour la linguistique et l'analyse de discours de manière plus générale n'est pas nouveau[‡]. Il remonte aux années 1960 au moment même de la naissance de cette nouvelle approche qui provoqua un grand enthousiasme chez beaucoup d'historiens. En effet, habitués à une méthode hypothético-déductive devenue insuffisante, ces derniers ont vu dans la nouvelle discipline une approche plus globale des textes permettant de rendre compte de leur structure et par là même un moyen de faire de leur domaine une véritable science par l'introduction des méthodes statistiques et quantitatives. Régine Robin écrit à ce propos :

Si les linguistes fournissent à l'historien des techniques, grâce à ce protocole méthodologique, il va pouvoir lire de façon neuve et interpréter les textes. (...) On pense pouvoir pallier ces manques, ces absences, par le recours à des techniques dûment éprouvées, dont le caractère formalisé offre toutes les garanties. (Robin, 1973 : p.16)

Concrètement, l'objectif consistait essentiellement à donner un « tour lexical » aux connaissances historiographiques par le

[†]D'autres logiciels tels que TXM et Lexico 5 offrent à peu près les mêmes fonctionnalités.

[‡]Sur ce sujet, voir entre autres les travaux de Jacques Guilhaumou, Denise Maldidier, Régine Robin ainsi que la revue *Histoire et mesure*.

déploiement dans le temps et l'espace des éléments de contenu du corpus. (Guilhaumou, 2007 : p. 185)

Mais certains historiens, souhaitant conserver la maîtrise du lieu d'où ils observaient les faits historiques, arguant de la transparence du mot face au référent (Ibid. : p. 177), sont demeurés défiants parfois même hostiles face à l'introduction de ces nouvelles méthodes dans leur domaine. D'autres, moins défiants, maintenaient les méthodes lexicométriques en tant qu'outils de dépouillement des archives sans accorder une grande importance à la linguistique au sens large. (Ibid. : p. 177)

Durant les années 70-80, les travaux se concentraient surtout au sein du laboratoire « Lexicométrie et textes politiques » de l'ENS de Saint-Cloud et à l'Université de Nanterre (Paris X). Dans ces deux espaces, linguistes et historiens collaboraient ensemble. Les travaux portaient majoritairement sur les discours des hommes politiques français des XIX^e et XX^e siècles.

L'intérêt pour la lexicométrie historique redoubla au cours des années suivantes avant de commencer à s'essouffler en raison de plusieurs contraintes. Il est utile de souligner que les études lexicométriques se bornaient au début à l'étude des mots en tant que formes graphiques, ainsi qu'à leur répartition et leur fonctionnement dans un corpus. C'est ce qu'on reprochait à la nouvelle discipline.

Plus tard, le renouvellement épistémologique et le développement des outils d'analyse a permis la prise en charge de la question du sens, jusque-là négligée. Cela est devenu possible grâce à la contextualisation (étude de l'environnement linguistique des mots). L'on a même dépassé le niveau du mot pour atteindre celui de la phrase, ce qui permet l'étude de la syntaxe et de la grammaire.

Il faut savoir également que la lexicométrie historique ne prétend pas se substituer à la démarche historique telle que pratiquée

depuis toujours. Elle se conçoit plutôt comme une étude complémentaire permettant non seulement de dégager de nouvelles perspectives de recherche mais aussi de conférer aux recherches plus d'objectivité en raison du recul que le chercheur peut prendre. En effet, celui-ci ne fait qu'interpréter des sorties-machines fournies par le programme lexicométrique. Certes, l'interprétation des résultats issus du logiciel pourrait être empreinte de subjectivité, mais il n'en demeure pas moins que l'intervention du chercheur se situe en aval et permet ainsi de limiter considérablement son degré d'implication.

4. QUELQUES MÉTHODES LEXICOMÉTRIQUES POUR LES HISTORIENS

Les méthodes lexicométriques visent toutes à caractériser un ensemble lexical en le confrontant à d'autres ensembles qui présentent des caractéristiques communes. Toutes ces méthodes peuvent être appliquées dans les études de textes historiques.

Nous nous limitons ici à en présenter les plus importantes : l'étude de la distance lexicale, l'étude des spécificités lexicales et l'étude des co-occurrences.

2.1. La distance lexicale

Elle permet de mesurer la proximité ou l'éloignement des sous-parties du corpus sur le plan lexical les uns par rapport aux autres. Elle peut être mesurée selon le vocabulaire commun aux textes étudiés et selon le vocabulaire spécifique à chaque texte. Un lexique commun à deux textes les rapproche tandis qu'un lexique spécifique à l'un d'eux, donc non-spécifique à l'autre, les différencie. (Souti, 2016 : p. 113)

2.2. L'étude des spécificités lexicales

Elle permet de déterminer les lexies qui caractérisent le plus le lexique et celles qui le caractérisent le moins. Une spécificité positive signifie que la lexie en question est sur-employée par

rapport à l'ensemble du lexique. Une spécificité négative ne signifie pas que la lexie n'est pas employée, mais qu'elle est sous-employée en comparaison avec les autres items lexicaux du corpus étudié. (Ibid. p.114)

2.3. L'étude des co-occurrences

Elle sert à repérer et à présenter les mots qui apparaissent ensemble dans les mêmes phrases ou les mêmes paragraphes. Ces mots sont appelés « co-occurents ». Elle permet de déterminer les thèmes abordés dans le texte et par là même d'en délimiter l'univers thématique. Le mot dont on cherche le contexte est appelé « mot-pôle ». (Ibid., p.111)

5. PERSPECTIVES POUR L'HISTORIOGRAPHIE ALGÉRIENNE

Dans le cadre des études historiques, ces méthodes pourront s'avérer utiles si elles sont mises à contribution dans une approche contrastive de discours de formations politiques différentes ou opposées ou encore dans une perspective contrastive chronologique lorsqu'il s'agit de la même formation politique.

Dans cette optique, des études contrastives du discours des différents courants politiques algériens de l'époque coloniale pourraient apporter plus d'éclairages sur les enjeux sociopolitiques de cette période historique et de mieux la comprendre. En effet, le paysage politique algérien - appelé à l'époque *musulman* - se caractérisait par la présence de plusieurs formations politiques défendant des projets différents, parfois même opposés.

Ces courants sont le nationalisme-indépendantisme représenté par le PPA-MTLD (Parti du Peuple Algérien) de Hadj Messali ; le réformisme modéré proche du nationalisme incarné par les AML (Amis du Manifeste et de la Liberté) puis par l'UDMA (Union Démocratique du Manifeste Algérien) de F. Abbas ; le

réformisme religieux de l'Association des Oulémas Musulmans ainsi que le communisme du Parti Communiste Algérien (PCA).

Les corpus peuvent être constitués de discours des leaders de ces courants et d'autres textes en tous genres notamment la presse partisane particulièrement prolifique et dont une partie est disponible même sur le net.

Le même travail pourrait porter sur la période post-indépendance : par exemple sur l'évolution chronologique du discours du Parti unique (FLN) depuis l'indépendance jusqu'à l'institution du pluralisme politique ; le discours présidentiel algérien dans une perspective comparative confrontant les discours des présidents de l'Algérie indépendante ; le discours syndical ; le discours journalistique (presse écrite) ; etc.

Les méthodes lexicométriques étant essentiellement comparatives, il est même possible d'envisager des études dans une perspective épistémologique portant sur l'écriture de l'histoire de l'Algérie, celle-ci continuant d'attiser les passions de part et d'autre de la Méditerranée. Les corpus sont abondants : des centaines de livres ont été publiés depuis l'indépendance sur l'Algérie dite *française*.

6. LES HISTORIENS SONT-ILS SUFFISAMMENT OUTILLÉS POUR MAITRISER L'OUTIL LEXICOMÉTRIQUE ?

La lexicométrie étant une approche fondamentalement automatique et quantitative, la manipulation de l'outil lexicométrique requiert l'acquisition d'un ensemble de savoirs et de techniques qui relèvent à la fois de la linguistique et de l'informatique.

Il est vrai que les outils récents sont dotés d'une ergonomie telle que leur utilisation devient aisée pour n'importe quel chercheur quel que soit son niveau d'initiation à l'informatique pourvu qu'il sache manier un ordinateur. Néanmoins, une initiation aux

méthodes lexicométriques et leur exploitation dans le cadre des études historiques semble indispensable.

Du reste, ce n'est pas le maniement du programme lexicométrique qui pourrait poser problème. C'est plutôt, la non-maitrise des méthodes lexicométriques et des notions théoriques fondamentales préalables à toute utilisation du logiciel en vue de son exploitation dans une perspective historique. C'est pourquoi une formation pratique à ces méthodes est requise.

Soulignons enfin que les outils disponibles, conçus dans des laboratoires de linguistique, sont destinés initialement aux chercheurs en Sciences du langage. Les historiens devraient peut-être créer leurs propres logiciels et les adapter aux spécificités de leur démarche.

7. LA PROBLÉMATIQUE DES SOURCES ET DES CORPUS

Comme souligné précédemment, la démarche lexicométrique a pour objectif de dégager les traits spécifiques d'un ensemble discursif (un texte, une œuvre, un auteur, une époque, etc.) par rapport à d'autres ensembles discursifs.

Les études lexicométriques requièrent des corpus assez importants afin de conférer aux résultats obtenus suffisamment de fiabilité et de signification.

Le traitement lexicométrique se fait sur des macro-corpus constitués de textes numérisés enregistrés en format TXT. C'est pourquoi, la question des sources et des corpus a toujours été essentielle en lexicométrie. Ce n'est pas tant l'existence de ces sources et corpus qui préoccupait les chercheurs mais leur disponibilité en format numérique en vue d'une prise en charge par le logiciel.

Cela explique, ne serait-ce que partiellement, qu'en dépit de l'engouement suscité par son apparition, l'évolution de la discipline a été freinée dans son élan durant les années 1970 et

1980. Certes, son originalité intéressait beaucoup de chercheurs, mais la problématique des corpus a contribué fortement à l'affaiblissement de cet intérêt[§]. En effet, il n'était pas possible d'imaginer, en ces temps-là, la disponibilité de corpus en format numérique en grand nombre, de quelque type qu'ils fussent (littéraires, politiques, journalistiques, etc.). Les chercheurs devaient, par conséquent, procéder eux-mêmes à la saisie laborieuse de leurs corpus ou la confiaient à d'autres personnes. Note à ce propos : « Dans le domaine des travaux sur le discours qui sont encore menés en linguistique, les préoccupations historiques ont reculé au profit de perspectives formalisantes ou sociolinguistiques. » (Courtine, 1994 : p.07)

C'est l'apparition d'internet et sa généralisation au fil des années qui a rendu possible le redémarrage de la lexicométrie. Plusieurs contraintes ont été ainsi levées. Effectivement, depuis le début du nouveau millénaire, les corpus numériques sont de plus en plus accessibles en ligne. De nos jours, des bases de données impressionnantes sont en accès libre sur le net et les textes sont directement téléchargeables.

La plus importante de ces bases est la base *Goofre* de l'entreprise américaine Google. Cette dernière a lancé à partir de 2003 un projet gigantesque de numérisation de tous les livres publiés depuis l'avènement de l'imprimerie jusqu'à nos jours. Il s'agit de plusieurs millions de titres. Par ailleurs, Google Labs, filiale de Google, développe et met en ligne un outil de traitement statistique des textes appelé *Google NgramViewer*. La nouveauté tient dans le fait que le nouvel outil permet d'effectuer des analyses lexicométriques sur la base *Goofre*. Il permet surtout de « retracer sur une très longue durée l'évolution des fréquences d'emploi de termes ou d'expressions » (Chateauraynaud et Debaz, 2010).

[§] Il faut y ajouter les limites des programmes lexicométriques de l'époque.

Il faut savoir également que *Goofre* est intégrée dans les dernières versions du logiciel Hyperbase. En plus de *Goofre*, celui-ci met à la disposition des lexicométriciens des dizaines de bases constituées d'œuvres littéraires complètes d'auteurs français et algériens. Les auteurs algériens sont Mohammed Dib et Mouloud Mammeri. Cela rend l'exploitation de ces œuvres simple et possible sans que les chercheurs ne soient contraints de les numériser puis de créer leurs propres bases sur le logiciel.

Il n'en demeure pas moins que certains corpus ne sont pas encore disponibles en version numérique et qu'ils nécessitent une saisie manuelle particulièrement chronophage. C'est le cas des corpus anciens car ils ont été publiés à une époque où seule la version papier était disponible. La version numérique requiert un travail de saisie manuelle ou avec une numérisation au scanner, puis une conversion en format TXT. Après avoir acquis les textes, le chercheur peut les soumettre à des logiciels de reconnaissance de caractères de plus en plus performants rendant la numérisation moins pénible.

En ce qui concerne le sujet de cette publication, une grande partie des corpus intéressant les historiens spécialistes de l'Algérie a été numérisée et mise en ligne en accès libre ou payant sur la base de données Gallica** de la BNF (Bibliothèque Nationale de France). Lancée en 1997, Gallica regroupe des millions de documents dont ceux se rapportant à l'époque coloniale (1830-1962). Les collections sont variées et multiples : livres, journaux, revues, cartes administratives et géographiques, émissions radiophoniques, vidéos, etc.)

En plus de Gallica, les historiens ont à leur disposition le site des Archives Nationales d'outre-mer (ANOM)†† sur lequel ils peuvent trouver des documents sur l'époque coloniale.

** www.gallica.bnf.fr

†† www.archivesnationales.culture.gouv.fr

Certes, l'existence de ces bases de données est une aubaine pour les chercheurs leur évitant de faire des déplacements coûteux et leur épargnant la saisie au clavier. Il n'en demeure pas moins qu'elles ne contiennent pas tout ce qui a été écrit durant la période coloniale. C'est pourquoi la numérisation manuelle est nécessaire pour certains corpus notamment ceux disponibles uniquement en Algérie.

8. ÉTUDE CONTRASTIVE DES DISCOURS DE DEUX FIGURES DU NATIONALISME ALGÉRIEN

Pour illustrer nos propos, nous présentons deux analyses lexicométriques portant sur le discours nationaliste algérien à travers ses deux principales figures, Ferhat Abbas et Messali Hadj. Cela nous amènera à opérer des comparaisons entre les deux discours en sachant que les deux hommes politiques en question avaient des lignes politiques différentes, parfois même opposées.

En effet, Hadj Messali est considéré comme le père du nationalisme algérien. Il est le fondateur du premier mouvement nationaliste algérien revendiquant l'indépendance pure et simple. Néanmoins, ses différends avec le FLN l'ont amené à s'opposer à la guerre de libération et à fonder le MNA.

Pour sa part, F. Abbas a eu des débuts politiques plutôt réformistes car il voulait combattre le système colonial de l'intérieur. En revanche, dès les années 20 jusqu'aux années 50, il n'a pas cessé d'évoluer politiquement (de l'assimilationnisme au nationalisme en passant par le fédéralisme) jusqu'à devenir en 1958 le premier président du Gouvernement Provisoire de la République Algérienne (GPRA).

Les deux hommes ont entamé leurs carrières politiques à la même époque (années 1920) mais le nationalisme de Messali remonte à ses débuts politiques (années 1920) tandis que celui de F. Abbas a un peu tardé à naître. On peut le faire remonter aux années 1940.

Le corpus est constitué de textes produits entre 1922 et 1962. Pour garantir une certaine homogénéité à notre corpus, nous avons retenu 103 textes pour chacun des deux. Les textes de Messali sont extraits de deux ouvrages : *Messali Hadj par les Textes* de Jacques Simon et *Les Archives de la révolution algérienne* de Mohammed Harbi. Les textes de F. Abbas sont extraits de son livre *Le Jeune Algérien* sorti en 1931 ainsi que de la presse des années 30, 40 et 50^{‡‡}.

Nous avons distribué le corpus global en 07 sous-parties. Ce découpage nous a permis de saisir les contrastes lexicométriques entre les deux hommes politiques et entre les différentes périodes.

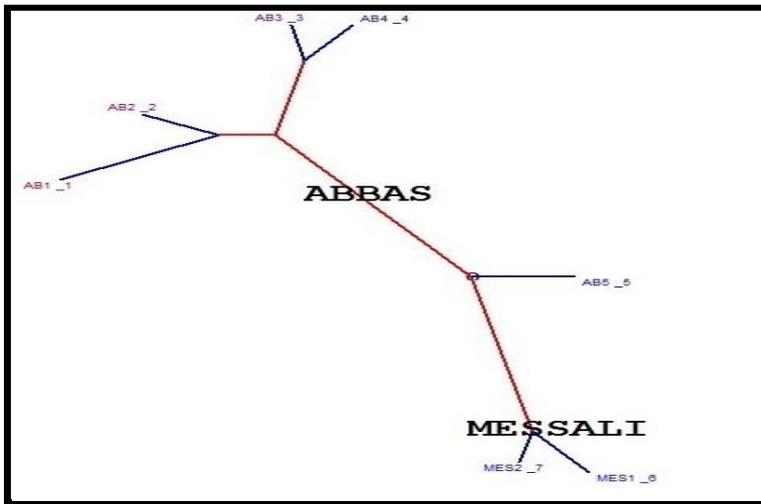
Nous avons mis en œuvre deux méthodes lexicométriques: la distance lexicale intertextuelle et l'étude des co-occurrences.

8.1. La distance lexicale intertextuelle

L'étude fait apparaître les résultats représentés ci-dessous sous forme d'arbre obtenus avec l'analyse arborée effectuée sur Hyperbase :

^{‡‡} Les journaux *L'Entente Franco-Musulmane* (1935-1941), *Egalité* (1944-1948), *La République Algérienne* (1948-1955) et *El Moudjahid* (1958-1951).

Graphique 1 : la distance lexicale intertextuelle



Politiquement et heuristiquement parlant, les résultats obtenus sont intéressants. En effet, le logiciel ne connaît pas les configurations sociopolitiques du corpus mais le distribue en fonction de deux paramètres : 1. l'auteur ; 2. le facteur chronologique.

Les sous-corpus Abbas 1 et Abbas 2 regroupant des textes de F. Abbas des années 1920 et 1930 à une époque où il militait pour l'assimilation (revendications réformistes dans le cadre de la souveraineté française) occupent la même branche sur l'arbre. Ils sont donc proches lexicalement. Lexicométriquement, cela s'explique par le fait qu'ils partagent un lexique commun qui les rapproche. Il en est de même des textes Abbas 3 et Abbas 4 représentant les années 1940 et 1950 durant lesquelles F. Abbas évolue vers le nationalisme sans revendiquer l'indépendance de l'Algérie de manière explicite. Le mot qui caractérise ces années-là, c'est le fédéralisme. F. Abbas lutte pour la création d'un Etat algérien autonome dans le cadre d'une fédération française.

Ce n'était pas le cas de Hadj Messali qui militait de manière constante pour l'indépendance pure et simple. C'est pourquoi les

résultats de l'analyse abordée donnent à voir une dichotomie nette et précise entre les deux ensembles car les sous-corpus Messali 1 et Messali 2 se trouvent très éloignés des sous-corpus Abbas1, Abbas2, Abbas3 et Abbas 4. Lexicométriquement, cette dichotomie s'explique par l'existence d'un lexique différent qui différencie les deux ensembles discursifs et les éloigne.

L'emplacement des textes de F. Abbas écrits après 1958, c'est-à-dire à partir de son élection à la tête du Gouvernement Provisoire de la République Algérienne (GPRA) et donc du mouvement révolutionnaire est curieux. Ces textes regroupés dans le sous-corpus Abbas5 se placent dans une position médiane entre les textes de Messali et ceux de F. Abbas des années 20, 30 et 40. Politiquement, à partir de 1956, année où il rejoint officiellement le FLN, F. Abbas abandonne définitivement les idées réformistes pour embrasser le nationalisme et l'indépendantisme. Plus curieux encore, en observant l'arbre, on peut constater que les textes de F. Abbas le nationaliste sont plus proches des textes de Messali que de ceux de F. Abbas l'assimilationniste ou le fédéraliste.

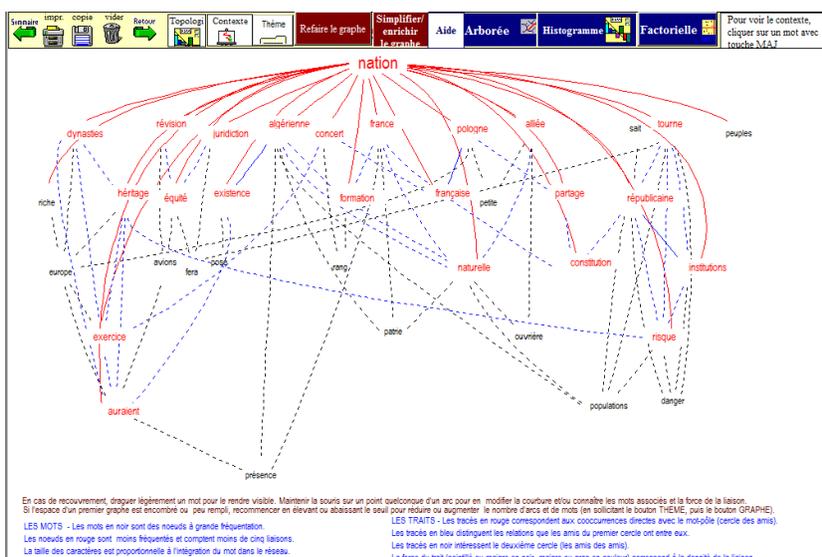
8.2. L'étude des co-occurrences

Elle consiste à étudier le nationalisme à travers sa lexie représentative ou mot-pôle « nation » dans les deux corpus. Il s'agit de cerner la distribution du mot-pôle et l'évolution de son univers lexical immédiat – cotexte – à travers les lexies qu'il attire le plus.

Dans le corpus de F. Abbas, l'étude des co-occurrences donne à voir une évolution lexicale chronologique en ce qui concerne les lexies attirées dans l'univers immédiat du mot-pôle *Nation*. Celui-ci possède un univers lexical qui change d'une période à une autre. Ainsi, durant les années 20-30, les mots les plus caractéristiques de cet univers sont relatifs au loyalisme et à l'assimilation. Si on exclut les verbes, on peut citer « Occident »,

cédé leur place à d'autres lexies caractéristiques de la période du fédéralisme. Pour la première fois, un mot renvoyant à l'Algérie en tant que nation apparaît dans l'univers immédiat de la lexie « nation ». Il s'agit de la lexie « algérienne » qui entretient un rapport direct avec le mot-pôle. Par ailleurs, la lecture de ce graphe fait apercevoir un rapport significatif triangulaire entre « nation », « algérienne » et « existence ».

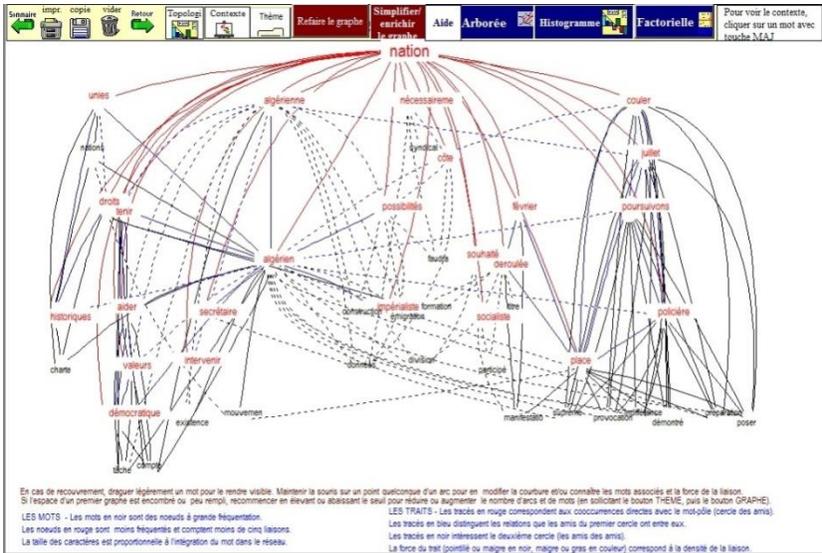
Graphique 3 : les co-occurents de « Nation » dans le corpus Abbas2



Historiquement, le nationalisme de F. Abbas naît à cette époque et commence à se radicaliser progressivement. Il reconnaît désormais l'existence de la nation algérienne : « La personnalité algérienne, la patrie algérienne que je n'ai pas découverte en 1936 dans les masses musulmanes, je la découvre aujourd'hui. La nier, c'est tourner le dos à la vérité. » (*Egalité*, 06 septembre 1946)

Quelques mois plus tard, il réitère son affirmation « On ne tourne pas impunément et infiniment le dos à la Vérité. La NATION algérienne naîtra avec la France ou sans la France. Avec elle et pour elle ou sans elle et contre elle. » (*Egalité*, 29 novembre 1946)

changé durant sa longue carrière politique. Les co-occurents les plus caractéristiques de *nation* sont « algérienne », « nécessairement », « unies », « algérien », « droits », « poursuivons », etc.



En 1936, lors du Congrès Musulman regroupant les formations réformistes musulmanes (Elus, Oulémas et Communistes) et revendiquant le rattachement de l'Algérie à la France, Messali Hadj intervient et fait une déclaration qui sera reprise dans tous les livres d'Histoire car considérée comme un geste fondateur : « Cette terre bénie est la nôtre, elle n'est ni à vendre, ni à acheter, ni à hypothéquer, ses héritiers sont là et l'Etoile nord-africaine y veillera. »

On peut remarquer que l'étude des co-occurences confirme les résultats de l'étude de la distance lexicale. Les deux études lexicométriques donnent à voir une réalité dichotomique entre les textes des deux hommes politiques et évolutive chronologique dans les textes de F. Abbas. Il s'agit pour ce dernier d'une évolution chronologique lexicale correspondant à une évolution politique reconnue par les historiens.

L'étude de la distance lexicale montre que le choix du lexique est tributaire de l'époque et des éléments conjoncturels notamment chez F. Abbas dont les idées et les projets politiques en dépendent également. Il s'agit en effet d'un vocabulaire évolutif qui ne résiste pas aux contraintes imposées par la chronologie et la conjoncture changeantes.

L'étude des co-occurrences, quant à elle, nous a permis de constater que la physionomie de l'univers lexical de la lexie étudiée change d'une époque à l'autre puisque de nouvelles lexies apparaissent tandis que d'autres disparaissent dans son contexte. Les lexies du nationalisme attirent des mots différents en fonction du projet et des idées politiques de l'auteur notamment en ce qui concerne F. Abbas.

L'étude effectuée nous permet ainsi de penser que l'analyse automatique est une manière originale d'appréhender les discours que les historiens peuvent mettre en œuvre dans leurs travaux. Avant l'apparition de la statistique lexicale, se posait le problème de la représentativité des échantillons étudiés. Il était alors impossible de travailler sur des macro-corpus. Grâce aux progrès récents, ce problème a été résolu et il est désormais possible de travailler sur des corpus plus étendus.

Par ailleurs, la mathématisation de la recherche permet la consolidation de l'analyse traditionnelle par une analyse automatique quantitative. L'historien peut ainsi prendre plus de recul par rapport à l'objet de recherche.

BIBLIOGRAPHIE

BRUNET, Etienne, 2001, « Le Logiciel Hyperbase », L'Astrolabe, disponible sur <http://www.uottawa.ca/academic/arts/astrolabe/auteurs.htm>)

Consulté le 20 décembre 2013.

CHATEAURAYNAUD, Francis et JOSQUIN, Debaz, 2010, « Prodiges et vertiges de la lexicométrie. », disponible sur <http://socioargu.hypotheses.org>) Consulté le 20 février 2011.

COURTINE, Jean-Jacques, « Le tissu de la mémoire : quelques perspectives de travail historique dans les sciences du langage ». Langages, n° 114, 1994, pp. 5-12.

GUILHAUMOU, Jacques, « L'analyse de discours du côté de l'histoire. Une démarche interprétative », Langage et société, n° 121-122, 2007, pp. 177-187.

MAYAFFRE, Damon, « Analyses lexicométriques et logométrie », in OLIVESI, S. (dir.), *Introduction à la recherche en SIC*, Presses de l'Université de Grenoble, Grenoble, 2007, pp.153-180.

ROBIN, Régine, *Histoire et linguistique*, Paris, Armand Colin, 1973, 306 p.

SINCLAIR, Stéphane, 2002, « Quelques obstacles historiques et épistémologiques dans le développement de l'analyse de texte informatisée », in L'Astrolabe. Recherche littéraire et Informatique (revue électronique) disponible sur <http://www.uottawa.ca/academic/arts/astrolabe/articles>)

Consulté le 15 mars 2012.

SOUTI, Chérif, *Texte et contexte : évolution du lexique dans l'œuvre politique, littéraire et journalistique de Ferhat Abbas. Approche lexicométrique*, thèse de doctorat, Université Frères Mentouri Constantine, 2016

SOUTI, Chérif « La lexicométrie ou l'analyse du discours assistée par ordinateur : ce que l'informatique et les mathématiques peuvent apporter à la littérature et la linguistique », *Expressions*, n°01,2015, pp. 122-130.