

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.

Résumé :

Cette contribution a pour objectif de présenter la lexicométrie, une discipline relativement récente qui demeure peu exploitée en Algérie. Nous nous proposons précisément, tout en brossant un tableau qui retrace succinctement son évolution à travers le temps, de présenter ses méthodes et ses fonctions ainsi que les outils qu'elle met en œuvre et les perspectives de son exploitation par les chercheurs algériens, en Sciences du langage et en littérature.

Chérif SOUTI
Laboratoire Slaad
Université Constantine 1

Abstract :

This contribution aims to present the lexicometry, a relatively new discipline that remains untapped in Algeria. We propose specifically, while painting a picture that succinctly traces its evolution through time, to present its methods and its functions as well as the tools it employs and the prospects of its exploitation by Algerian researchers in language sciences and literature.

Introduction :

La lexicométrie est une approche qui demeure peu exploitée au sein même de la discipline mère, c'est-à-dire les Sciences du langage, champ disciplinaire dont elle relève. Elle constitue une approche heuristique particulièrement utile : elle permet d'effectuer de nombreuses analyses sur des corpus étendus, et de dégager un large éventail d'axes de recherche qu'elle peut dégager. Ainsi pour Guiraud, « la

linguistique est la science statistique type ; les statisticiens le savent bien ; la plupart des linguistes l'ignorent encore » (in Muller, 1992 : 1). Le peu d'intérêt qu'on accorde à cette discipline demeure inexplicable. En fait, il faut noter qu'en France l'Analyse du discours a été, à ses débuts, essentiellement automatique. Elle a commencé à se développer grâce aux travaux du Centre de linguistique quantitative de Paris, créé en 1962. Il s'agit des travaux effectués en lexicométrie politique à l'ENS de Saint-Cloud et ceux réalisés par Michel Pêcheux de l'Université de Paris VII. On peut donc relever que les premières recherches s'inscrivent en partie dans une perspective quantitative et statistique du discours puisqu'elles font une large place à l'outil informatique (G.E. Sarfati, 2005 : 93-95).

Cette contribution répond à une motivation principale : faire connaître davantage cette discipline parmi la communauté universitaire de notre pays. En fait, l'Analyse du discours assistée par ordinateur est un domaine qui gagne à être développé en Algérie car il est encore peu connu et peu exploité. Il suffit pour le constater de remarquer le nombre de recherches et de publications quasi inexistantes dans ce domaine. En effet, si on excepte les travaux réalisés à l'E.N.S. d'Alger par Attika Kara et quelques autres travaux effectués à travers les universités algériennes, l'on peut facilement remarquer que la place réservée aux recherches lexicométriques demeure insuffisante. Par ailleurs, les travaux réalisés par A. Kara et ses collaborateurs sont principalement, à quelques exceptions près, des analyses lexicométriques d'œuvres littéraires d'écrivains algériens de langue française – Mouloud Mammeri, Mohammed Dib et Taher Djaout notamment-. Ces recherches s'inscrivent donc majoritairement dans le cadre de la lexicométrie littéraire. Quant à la lexicométrie politique, analyse du discours politique assistée par ordinateur, elle semble n'avoir jamais été exploitée. Toutefois, nous ne saurions être affirmatif sans un recensement exhaustif de tous les travaux réalisés en Sciences du langage et en littérature à travers les universités d'Algérie.

Qu'est-ce que la lexicométrie ?

On désigne sous le vocable « lexicométrie » la discipline qui prend en charge l'analyse informatisée du discours et du lexique. Cette jeune discipline est appelée également « analyse du discours assistée par ordinateur », ou encore « traitement automatisé du discours ». Néanmoins, les appellations proposées pour désigner « l'étude scientifique du discours faite avec l'outil informatique » sont aussi nombreuses que diverses et témoignent de l'état de fluctuation dans lequel se trouve la nouvelle discipline qui cherche encore ses contours. En voici quelques-unes : « Textométrie, Statistique textuelle » (André Salem), « Logométrie » (Damon Mayaffre), « Statistique lexicale (ou lexicostatistique) », « Statistique linguistique », etc. G. Mounin (1974 : 203), la définit comme le « Domaine de la lexicologie dans lequel les procédures de la statistique sont utilisées pour l'étude quantitative du lexique. ». Selon Leimdorfer et Salem (1995 : 133), on regroupe sous le terme de lexicométrie « toute une série de méthodes qui permettent d'opérer, à partir d'une segmentation, des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire ». Enfin, pour Maingueneau, (2009 : 81), il s'agit d'une « Discipline auxiliaire de l'analyse du discours qui vise à caractériser un ensemble discursif (souvent un positionnement) par rapport à d'autres appartenant au même espace grâce à l'élaboration informatique de réseaux quantifiés de relations significatives entre ses unités. Il s'agit par conséquent d'une démarche essentiellement comparative. »

En revanche, si diverses et si nombreuses que soient les dénominations et les définitions de la jeune discipline, il faut retenir qu'elle s'intéresse au discours à travers l'analyse de son lexique.

Pour ce faire, elle fait appel à des méthodes et des outils mathématiques et informatiques : des logiciels de statistique lexicale. Ceux-ci sont si nombreux, les plus connus dans l'univers francophone sont Hyperbase d'Etienne Brunet et Lexico d'André Salem. Citons aussi : Cordial, Sphinx, Alceste, Tropes, Sato, Weblex, Prospero, Astartex, etc. Mais il ne faut pas perdre de vue que cette discipline ouvre la voie à d'autres analyses, qualitatives notamment. L'originalité de cette approche réside dans le fait qu'elle permet d'effectuer des analyses dont l'entreprise était jusque-là impossible. En effet, la conjugaison à l'analyse traditionnelle d'une analyse automatique assistée par ordinateur a introduit une rupture totale dans le champ de l'analyse des données. Grâce à la mathématisation de la recherche linguistique, on peut travailler sur des corpus plus étendus et assurer une certaine objectivité des traitements. Le linguiste prend plus de recul par rapport à l'objet de sa recherche. Son intervention n'est que partielle et sa tâche principale consiste en l'interprétation des sorties-machines. Avant l'apparition de la statistique lexicale, se posait encore le problème de la représentativité des échantillons étudiés (Il était alors impossible de travailler sur des macro-corpus). Grâce à la lexicométrie, on peut par exemple appréhender l'œuvre complète d'un auteur sans se soucier du problème de la représentativité du corpus puisque l'analyse prend en charge l'intégralité de l'œuvre. En effet, comment garantir, lorsqu'il s'agit par exemple d'aborder l'œuvre d'un auteur, que les échantillons soumis à l'analyse soient représentatifs de toute sa production littéraire puisqu'on sait que même si elles appartiennent au même auteur, des œuvres peuvent présenter des dissemblances discursives? Avec l'avènement de l'approche lexicométrique, ces obstacles sont levés. L'analyse automatique est donc une autre manière, inhabituelle et originale d'appréhender les discours.

La lexicométrie s'est scindée, depuis son avènement, en deux grandes branches. La première s'est constituée autour de Charles Muller de l'Université de Strasbourg et s'intéresse au traitement statistique des textes littéraires, d'où son nom de lexicométrie littéraire. La revue *Lexicometrica* est actuellement le principal organe de diffusion de ses travaux. La lexicométrie littéraire s'intéresse notamment au style d'un écrivain, à la richesse et l'évolution de son vocabulaire. La seconde, la lexicologie politique, pour sa part, s'est développée vers la fin des années 1960 grâce aux travaux de Maurice Tournier au laboratoire Lexicométrie et textes politiques de l'ENS de Fontenay-Saint-Cloud. La tâche principale d'une telle démarche est d'examiner les rapports entre politique et phénomènes langagiers. Lorsqu'il s'agit d'appréhender un texte sociopolitique, la lexicométrie est chargée « d'examiner, à partir de corpus de textes soumis à comparaison, comment les termes échangés dans l'espace public autour des enjeux de pouvoir rendent compte des luttes d'appropriation ou de dépossession symboliques qui se jouent dans le lieu même de l'échange. », (Bonnaïfous et Tournier, 1995 : 69). La revue *Les Langages du politique* a constitué pendant longtemps le principal organe de diffusion des travaux effectués dans cette branche.

Origines et évolution de l'approche lexicométrique

Si l'on fait remonter les premiers balbutiements de l'analyse automatique du discours au milieu du XX^e siècle, il n'en demeure pas moins que beaucoup de savants, à travers l'Histoire, ont effectué des mesures sur des textes et en ont calculé les mots, surtout dans le but d'étudier le style de leurs auteurs. Mais la discipline telle qu'on la connaît aujourd'hui a traversé plusieurs phases et a dû surmonter des obstacles d'ordre épistémologique et matériel avant d'atteindre son état actuel. Par ailleurs, on ne peut nier le fait que le développement de cette branche de recherche ait toujours été tributaire de l'évolution de l'informatique, laquelle lui

fournit les outils statistiques nécessaires. Au XX^e siècle, les premiers projets d'analyse automatique du discours sont dus notamment à l'italien Roberto Busa qui fut le premier à utiliser l'ordinateur dans l'analyse linguistique et littéraire (Busa, 1998). Mais il faut évoquer également le nom de Georges Kingsley Zipf, linguiste et philosophe américain qui a inventé la loi qui porte son nom – la Loi de Zipf –. Celle-ci est largement utilisée dans les recherches de statistique linguistique. Jusqu'aux années 1960, les travaux étaient peu nombreux et menés par un nombre très restreint de chercheurs épars. En 1962, est créé le Centre de linguistique quantitative de Paris. Il donnera naissance peu après à la revue *Langages* qui fut, des décennies durant, un lieu de rencontre important pour les linguistes et les analystes du discours. Dans les années 1970 et 1980, une amélioration nette est à signaler. Les recherches ont fait un bond en avant grâce à l'apparition du micro-ordinateur au début de la décennie 1970, car il faut savoir qu'au début, le matériel était très dispendieux, donc difficile à acquérir. En France, les précurseurs sont notamment Pierre Guiraud et Charles Muller qui ont développé la lexicométrie littéraire. Guiraud développa une approche scientifique qui prend en charge l'examen du style d'un auteur en mettant en œuvre une méthode statistique. Les travaux de Charles Muller portèrent essentiellement sur le théâtre classique. Nous pouvons citer à titre d'exemple *Etude de statistique lexicale*. Le vocabulaire du théâtre de Pierre Corneille, Larousse, Paris, 1967.

Michel Pêcheux est lui aussi un des promoteurs de la discipline avec sa fameuse *Analyse Automatique du Discours*. Il faut rappeler aussi les recherches sur le discours politique initiées par Maurice Tournier et son équipe de recherche au Laboratoire de l'ENS de Saint-Cloud. Ce laboratoire se dota en 1981 d'une revue : « *Mots/Ordinateurs/Textes* ». Le choix de Pêcheux et Tournier pour l'automatisation de l'analyse du discours était motivé par leur préoccupation de défendre les Sciences humaines contre ce qu'on appelle les « Sciences dures ». Il faut retenir aussi et surtout le nom d'Etienne Brunet de l'Université de Nice qui a contribué fortement au développement de la démarche. Ses recherches portèrent sur le lexique des grands auteurs classiques de la littérature française. On lui doit surtout le très remarquable logiciel *Hyperbase*. Les recherches ont connu un déclin pendant les dernières années de la décennie 1980. Les contraintes épistémologiques et matérielles étaient parfois insurmontables. La quête d'outils plus performants commença à se faire sentir parmi les chercheurs. C'est dire que l'équipement utilisé était peu développé et n'avait rien à voir avec les outils très sophistiqués de ces dernières années. De nos jours, les obstacles auxquels la jeune discipline a fait face pendant les premières décennies de sa vie ont été en partie levés grâce à la disponibilité de bases de données contenant des textes numérisés en grandes quantités, disponibles sur le web. Il faut savoir que la numérisation des corpus prenait une grande partie du temps du chercheur puisqu'elle se faisait manuellement.

De la même manière, la disponibilité de logiciels assez perfectionnés a rendu possible le redémarrage de la discipline et renforcé sa fiabilité scientifique. La décennie 1990 est marquée par deux moments saillants: la publication en 1994 par Ludovic Lebart et André Salem de leur ouvrage intitulé *Statistique textuelle* et la création en 1997 par André Salem de la revue *Lexicometrica*. Au début du XXI^e siècle, on retient surtout le nom de Damon Mayaffre. Celui-ci propose une refonte de la discipline par une évolution de la lexicométrie traditionnelle vers une « logométrie » qui croise différents niveaux de traitements documentaires et statistiques du texte. La logométrie étend son analyse à d'autres unités du discours en plus des formes graphiques. Elle prend en compte alors d'autres unités linguistiques du corpus analysé : les lemmes (formes canoniques des mots dans le dictionnaire : l'infinitif pour les verbes, le masculin

singulier pour les noms et les adjectifs, etc.) ; les structures grammaticales; sémantiques et rhétoriques, etc.

Débat autour de la démarche lexicométrique

On a souvent reproché à la lexicométrie son caractère purement matériel. Ainsi, selon ses détracteurs, l'approche lexicométrique serait sans pertinence scientifique étant donné que le traitement statistique se borne à une description de la matérialité graphique des textes sans vouloir rendre compte du sens de ce matériel. Il serait donc, sans utilité scientifique d'appréhender des textes dans un sens graphique, visuel, matériel et informatique.

Pour ses défenseurs, la lexicométrie constitue, au contraire, un outil heuristique d'une grande utilité ouvrant la voie à des analyses lexico-syntaxiques. En effet, grâce au développement de certains outils qui ne s'arrêtent pas à la matérialité du lexique d'un texte mais s'enfoncent davantage dans sa structure, la question de la pertinence des analyses et de la fiabilité des résultats ne se pose presque plus

L'analyse lexicométrique

Après avoir soigneusement constitué et enregistré en ordinateur son corpus d'analyse, le chercheur procède à une division de ce corpus en plusieurs sous-parties comparables. Cette distribution tient compte des hypothèses d'analyse. Celles-ci sont formulées en fonction d'une variable de l'énonciation déterminée préalablement par le lexicométricien (auteur, époque, lieu, genre littéraire, etc.) La division du corpus vise à examiner les ressemblances ou les oppositions lexicales entre les sous-corpus. Ainsi, le chercheur peut vérifier la pertinence de son découpage et partant valider ou invalider ses hypothèses. Par exemple, si on divise un corpus constitué de l'œuvre complète d'un auteur en fonction de la variable « temps », c'est-à-dire en fonction des époques pendant lesquelles les textes de l'écrivain ont été produits, on peut distinguer plusieurs époques, donc plusieurs sous-parties. L'analyse lexicométrique fait ressortir les ressemblances ou les clivages lexicaux entre les différentes sous-parties (donc les différentes époques). L'hypothèse de recherche serait donc axée sur l'effet du temps sur l'évolution du lexique d'un auteur.

Après la division en sous-corpus du corpus global, le linguiste met en œuvre des programmes lexicométriques qui effectuent des calculs statistiques de lexique. Mais il ne faut pas pour autant perdre de vue le fait qu'il appartient au linguiste d'interpréter les faits quantitatifs décelés par la machine. Les programmes de statistique lexicale sont nombreux et divers. Ils offrent ainsi des fonctions différentes. Nous nous contentons ici de présenter les outils mis au point et utilisés par les chercheurs du laboratoire de Saint-Cloud et intégrés au logiciel Hyperbase. On distingue parmi ceux-ci deux types principaux (cf. Tournier, 1996).

Les programmes de segmentation et d'inventaire des formes utilisées :

Le logiciel procède au départ à une segmentation du corpus en formes graphiques ou unités linguistiques. Il effectue ensuite une indexation des formes relevées. L'indexation est alphabétique lorsque les occurrences sont classées par ordre alphabétique, ou hiérarchique lorsqu'il est tenu compte de l'ordre croissant ou décroissant des fréquences des formes répertoriées. Après l'indexation, le chercheur effectue un découpage en segments répétés. Ceux-ci sont des chaînes ou des séries de formes contigües non interrompues par des signes de ponctuation. Ces séries se répètent, semblables, plusieurs fois dans le même texte. Les outils font, dans un deuxième temps, la mise en « concordance » des unités linguistiques (En délimitant

le contexte immédiat de chaque forme : la phrase dans laquelle une forme est employée) et la mise en contexte (plusieurs phrases avant et après la forme ou le segment répété)

Les programmes de calculs statistiques

Ils servent à comparer les fréquences comptabilisées. On distingue, sans prétendre être exhaustif, deux opérations principales mises au point au Laboratoire de lexicologie politique de Fonteney-Saint-Cloud: l'AFC (Analyse Factorielle par Correspondance) et l'analyse des Spécificités.

L' AFC : analyse lexicale contrastive

Elle sert à faire une analyse lexicale contrastive entre les sous-corpus. Le but de l'AFC est de « déterminer, dans les correspondances quantitatives entre un ensemble d'émetteurs (les différents états de la variable qui partitionne le corpus en « textes ») et un ensemble de termes (le vocabulaire de ces textes), quels sont les clivages les plus pertinents. » (Tournier, 1996 : 185). L'AFC, certes, fait ressortir des clivages, des ressemblances, des regroupements et des oppositions lexicaux mais elle est insuffisante. Pour la renforcer et l'affiner, on fait appel à l'analyse des spécificités.

L'analyse des spécificités

Elle consiste à remplacer l'examen des fréquences par l'examen des probabilités dont ces fréquences sont assorties. En effet, il se trouve qu'une comparaison des fréquences est parfois sans valeur : lorsque les sous-corpus sont de longueurs inégales, « Il s'agit de substituer à chaque fréquence locale un jugement sur elle qui tienne compte des rapports quantitatifs dans lesquels elle a été constatée. » (Tournier, 1996 : 186). Grâce à la probabilisation, on peut minimiser les écarts de longueurs entre les textes étudiés. C'est là le rôle de la mathématique. En effet, pour qualifier des faits littéraires ou linguistiques, les logiciels de lexicométrie s'articulent dans leurs calculs sur des théories hypergéométriques (donc mathématiques).

L'analyse des spécificités permet de caractériser les occurrences présentes dans un corpus. Ainsi, une forme à spécificité positive est une forme « sur-employée », une forme à spécificité négative est une forme « sous-employée ».

Que peut apporter la lexicométrie à l'analyse du discours en Algérie ? Perspectives.

En Algérie, l'analyse statistique du vocabulaire peut être d'un grand intérêt pour l'analyse du discours et la lexicologie. Les corpus peuvent être puisés dans deux domaines qui se caractérisent par une production prolifique : la littérature et le domaine sociopolitique.

La littérature

La production littéraire algérienne francophone constitue un potentiel important vu le nombre des œuvres produites depuis les débuts de la littérature algérienne francophone (Dib, Kateb, Mammeri, Feraoun, etc.) jusqu'à nos jours (Mimouni, Djebbar, Boudjedra, etc.) En utilisant l'approche lexicométrique, on peut appréhender ces œuvres d'une autre manière et faire des comparaisons jusque-là impossibles puisque l'analyse statistique du vocabulaire prend en compte des lexiques assez larges. Ainsi, il serait intéressant de confronter les vocabulaires des différentes époques afin de déceler les proximités ou les distances, les ressemblances ou les divergences et d'en déterminer les raisons. Ces études permettraient également de rendre compte des rapports entre l'écriture des différents écrivains et le contexte sociohistorique caractérisant chaque époque. Elles offriraient aussi la possibilité de cerner les dimensions qui marquent chaque génération. Ces études peuvent porter sur la structure du vocabulaire, le style, la rhétorique, les thèmes, etc. Quelques pistes de recherche peuvent être explorées par des analyses

lexicométriques : évolution de la littérature algérienne francophone à travers le temps (analyses lexicales); analyses lexicales ou stylistiques contrastives entre les différentes générations de cette littérature (première, deuxième et troisième générations), etc.

Le discours sociopolitique

En ce qui concerne les discours sociopolitiques, l'approche statistique peut se révéler, comme ce fut le cas en France et ailleurs, d'un grand intérêt dans les traitements. Le pays n'a pas cessé de vivre, depuis l'indépendance et même avant, une situation sociopolitique houleuse qui se caractérise par des fluctuations, des agitations, des crises et des changements parfois violents. La mise en œuvre des méthodes modernes de la statistique lexicale garantirait une objectivation des analyses. Elle permettrait surtout d'effectuer des comparaisons lexicales et par là même de déceler les stratégies discursives qui sont employées et qui trahissent des aspirations de pouvoir chez les hommes politiques algériens. Dans ce domaine, plusieurs pistes de recherche sont à explorer lexicométriquement. On pourrait ainsi s'intéresser à l'analyse du discours politique algérien de la colonisation, Indigène¹ et colonialiste, pour mieux comprendre l'histoire de cette période. La presse de cette époque constitue une source inépuisable de ce discours. Des analyses comparatives entre le lexique utilisé dans les discours des différentes tendances politiques et idéologiques : réformistes (Oulémas) ; UDMA (Union Démocratique du Manifeste Algérie, parti de Ferhat Abbas), etc. ; communiste, le PCA (Parti Communiste Algérien) et nationaliste : PPA (Parti du Peuple Algérien)² contribueraient à une meilleure compréhension de la situation sociopolitique de l'époque.

De la même manière, il serait utile de mettre à contribution cette approche pour mieux rendre compte de l'évolution du discours politique algérien depuis l'indépendance jusqu'à nos jours, en tenant compte de l'évolution de la situation sociopolitique (acteurs, événements, périodes, etc.). Depuis Ben Bella (premières années de l'indépendance), Boumediène (le socialisme, la révolution agraire, les nationalisations, etc.), Bendjedid, (ouverture démocratique de 1989, multipartisme, etc.), décennie noire (affaiblissement de l'Etat, etc.), jusqu'à l'arrivée de Bouteflika au pouvoir et les changements qu'il a introduits : concorde civile, réconciliation nationale, crise politique des dernières années. Tous ces événements ont accompagné l'évolution sociopolitique du pays et n'auraient pas manqué d'affecter profondément les discours sociopolitiques.

Le discours syndical : il est possible de remarquer, pendant les dernières années, l'apparition de syndicats autonomes de défense des intérêts des travailleurs de plusieurs secteurs (Education, santé, etc.) Beaucoup de ces syndicats se situent en opposition de la Centrale Syndicale l'UGTA (Union Générale des Travailleurs Algériens) qu'ils accusent de connivence avec le pouvoir. Il serait important d'effectuer des analyses contrastives entre les discours de ces différents syndicats et celui de l'UGTA afin de mieux comprendre les stratégies énonciatives mises en œuvre et les préoccupations sociopolitiques qu'elles traduisent.

¹ Nous utilisons ce mot pour distinguer les Algériens musulmans (colonisés) des colons européens.

² Le Parti du Peuple Algérien, fondé par Messali Hadji. Il donna naissance en 1954 au Front de Libération Nationale (FLN) qui déclara et mena la guerre d'indépendance contre la France.

Le discours journalistique : l'ouverture démocratique de 1989 a rendu possible la parution d'une presse écrite privée. Les journaux francophones sont nombreux et se comptent par dizaines. Des études comparatives peuvent être menées pour mettre le point sur les orientations de chaque catégorie de la presse existante: presse partisane, presse d'opposition, presse privée, presse publique, etc. De la même façon, la parution ces deux dernières années de chaînes privées de télévision dont certains programmes sont diffusés en français constitue un terrain de recherche intéressant.

Enfin, les chercheurs en Sciences du langage peuvent utiliser les logiciels lexicométriques pour repérer, dans leurs dépouillements, les emprunts et les néologismes. Hyperbase est le programme le plus approprié car il confronte le lexique du corpus soumis à l'analyse au corpus du Trésor de la langue française (TLF) qu'il contient et fait ressurgir les anomalies lexicales.

Conclusion

L'outil lexicométrique s'avère donc d'un intérêt particulier puisqu'il ouvre des perspectives nouvelles à la lexicologie et à l'analyse du discours. Il permet notamment de « déconstruire le plus objectivement possible un texte pour accéder à son sens. » (Mayaffre, 2005).

Il faut avouer pourtant que l'analyse statistique du vocabulaire, qui conjugue analyse linguistique et méthodes mathématico-informatiques, demeure encore une discipline ésotérique. Ce qui explique en partie le peu d'intérêt qu'elle suscite auprès des chercheurs qui sont versés dans les Sciences du langage ou dans les Sciences des textes littéraires. En Algérie, ce serait se priver des atouts multiples d'une démarche scientifique qui a déjà fait ses preuves que de continuer à se désintéresser de la lexicométrie. Nous tenterons, à travers nos prochaines publications, d'en illustrer les applications par des exemples concrets. Il s'agira d'analyses lexicométriques qui porteront sur le discours politique algérien, la littérature et la presse algérienne francophones.

Bibliographie :

- BONNAFOUS S. et TOURNIER M., 1995 : « Analyse du discours, lexicométrie, communication et politique », in *Langages*, n°117. pp. 67-81.
- BUSA R., 1998 : «Dernières réflexions sur la statistique textuelle », communication faite aux JADT 1998³, <http://lexicometrica.univ-paris3.fr/jadt/jadt1998/busa.htm>
- LEIMDORFER F. et SALEM A., 1995 : « Usage de la lexicométrie en Analyse de discours », in *Cahiers des Sciences humaines*, n°31 (1). pp. 131-143.
- MAINGUENEAU D., 2009 : *Les Termes clés de l'analyse de discours*, Seuil, Paris.
- MAYAFFRE D., 2005 : « De la lexicométrie à la logométrie » in, *L'Astrolabe*, <http://www.uottawa.ca/academic/arts/astrolabe/articles/art0048/logometrie.htm>.
- MOUNIN G. (dir.), 1974 : *Dictionnaire de la Linguistique*, PUF, Paris.
- MULLER Ch., 1992 : *Initiation aux méthodes de la statistique linguistique*, Editions Champion, Paris.
- PÊCHEUX M., 1981 : « Analyse de discours et informatique », in *Actes du Congrès international informatique et sciences humaines - L.A.S.L.A. - Université de Liège* -
- SARFATI G.-E., 2005 : *Eléments d'analyse du discours*, Armand Colin, Paris.

³ Journées Internationales d'Analyse Statistique des Données Textuelles, Paris, France.

TOURNIER M., 1996 : « Les discours sociopolitiques et l'analyse lexicométrique », in BOYER H. (dir.), Sociolinguistique, Territoire et objets, Delachaux et Nestlé, Lausanne