

De la linguistique statistique à la logométrie : apports et limites de l'école française d'analyse du discours



Dr. Abdelhak Abderrahmane Bensebia

Université de Mostaganem Abdelhamid Ibn Badis, Algérie
abdelhakbensebia@yahoo.fr

Résumé : Dans le présent article, nous allons nous intéresser à l'analyse du discours, en tant que discipline qui connaît un essor notable, grâce à l'outil informatique, les différents développements que connaît le monde Internet aujourd'hui et les applications qui s'y associent. Le caractère statistique a rendu cette discipline de plus en plus appelée pour appréhender des textes et des corpus volumineux, afin d'exposer leurs différentes caractéristiques lexicologiques, et étudier ainsi les différentes stratégies discursives qui s'y imposent. Une telle démarche scientifique est devenue aussi accessible que nécessaire pour un chercheur qui tend à investir dans des résultats qualitatifs, en cherchant à associer linguistique à statistiques.

Mots-clés : logométrie, lexicométrie, lexicographie, discours, analyse assistée par ordinateur

المخلص : في هذا المقال، سنتطرق إلى دراسة طرق تحليل الخطاب والذي عرف تطورا كبيرا بفضل تقنيات الإحصائية من خلال الإعلام الآلي وتطور التطبيقات الخاصة بالانترنت. هاته التقنيات متعلقة خصوصا بالمدرسة الفرنسية في تحليل الخطاب لدراسة مختلف الظواهر اللغوية والاستراتيجيات الخطابية.

الكلمات المفتاحية: المدرسة الفرنسية في تحليل الخطاب، الظواهر اللغوية، الاستراتيجيات الخطابية، الخطاب، تطبيقات الإعلام الآلي.

Abstract : In this article, we will be interested in the analysis of the speech, as a discipline which makes notable great strides, thanks to the data-processing tool, the various developments which knows the world Internet today and the applications which join it. The statistical character returned this discipline increasingly called to apprehend long texts and corpora, in order to expose their various lexicological characteristics, and to thus study the various discursive strategies which are essential on it. Such a scientific step became as accessible as necessary for a researcher who tends to invest in qualitative results, while seeking to associate linguistic statistics.

Keywords: logometry, lexicostatistics, lexicography, speech, computer-assisted analysis

Introduction

L'analyse du discours, en tant que discipline qui s'intéresse à l'étude des phénomènes linguistiques et langagiers lors d'une situation de prise de parole, fait partie de ce que nous appelons aujourd'hui la linguistique du discours. Même si l'analyse du

discours peut être définie comme discipline, qui étudie et analyse les relations qui existent entre le contexte social et l'organisation syntaxique, la linguistique du discours est en train de se constituer comme bloc scientifique homogène qui regroupe toutes les sciences qui attachent au concept de discours une importance capitale, car il est aussi un phénomène social, qui exige une analyse fine, qui prend en charge ses visées perlocutoires.

Cependant, en parlant de la linguistique du discours, la notion d'analyse du discours apparaît comme ombre sans que les frontières séparant les deux concepts ne soient délimitées. Cette problématique émane du fait que le discours, en tant qu'entité linguistique, pourrait être pris scientifiquement par d'autres démarches, qui changent de buts, mais qui gardent ce discours objet de leurs analyses. Parmi ces sciences et disciplines, nous avons: l'analyse critique du discours, appelée aussi Critical Discourse Analysis (tradition anglo-saxonne), l'analyse de la conversation, les théories de l'argumentation... Toutes ces disciplines s'insèrent dans un cadre d'analyse unique. L'analyse du discours ne sera donc qu'une composante intégrante et indispensable de ce qui est la linguistique du discours. Par conséquent, l'analyse du discours, en tant que discipline à caractère scientifique, est en cours de se constituer comme une science homogène. Le seul handicap paraît dans la méthodologie, dans laquelle serait-elle inscrite. Il existe différentes voies et approches d'analyses, qui se mettent toutes d'accord sur son objet d'étude, qui demeure les énoncés (discours) à travers l'analyse du dispositif énonciatif.

1. Discours : Diversité des approches et méthodologies d'analyse

L'analyse du discours est définie comme une discipline qui s'inscrit dans le cadre de la linguistique et qui a pour but d'étudier l'ensemble des phénomènes langagiers et parfois extra langagiers qui se dégagent dans une situation de communication. Cette discipline aussi pluridisciplinaire soit-elle ne peut se limiter au seul fait d'interpréter ces phénomènes, elle est à la fois méthodologique qu'interprétative. Autrement dit, la qualité des interprétations dégagées reste conditionnée par la méthodologie empruntée.

La diversité de cette discipline peut aussi émaner des méthodologies et les matériaux réservés à l'analyse de corpus, et l'école dans laquelle vient s'inscrire le linguiste. Chaque méthode a ses propres avantages, ses caractéristiques, mais aussi ses limites. La méthodologie réservée pour un texte littéraire par exemple change en fonction de la problématique tracée au départ, comme le discours politique qui nécessite une attention particulière dans la mesure où certains phénomènes de différents types peuvent se manifester, et qui constituent un plus qui s'ajoutent aux ensembles des interprétations possibles.

La méthodologie en analyse du discours n'est pas un modèle d'analyse type, elle apparaît comme signe d'intelligence qui exige toute une littérature théorique en linguistique surtout, assez solide, qui s'ajoute à l'ensemble des autres littératures théoriques et pratiques.

À la diversité des approches et ces littératures, il existe désormais des analyses de discours, dont le but demeure le même : appréhender l'ensemble de ces phénomènes qui se dégagent et qui attribuent au corpus ses propres caractéristiques qui l'associent ou qui le distinguent des autres discours.

Dans le présent document, l'accent sera mis sur ce que nous appelons aujourd'hui l'école française en analyse de discours. Je préfère l'appeler ainsi et non pas l'analyse de discours à la française (Dufour Françoise et Rosier Laurence, 2012 : 5-13) car cette école a beaucoup donné en matière de concepts et de méthodologies, elle reste aussi unique dans sa philosophie d'analyse de corpus qui puise dans les textes institutionnels d'où l'intérêt de cette école qui a permis à toute une culture institutionnelle de s'installer à côté d'une autre linguistique. L'autre force de cette école se trouve aussi ancré dans son caractère linguistique, qui s'éloigne de certaines cultures qui font l'éloigner de ses considérations analytiques et interprétatives. Or, l'originalité de cette école reste sa capacité à s'associer à ces cultures, aux nouveaux outils d'analyses sans que l'aspect linguistique ne soit relégué au second plan.

Ce qui caractérise cette école serait aussi son aptitude à créer de nouveaux concepts, recourir à des méthodes plus originales dans la constitution des corpus, dégager une nouvelle réflexion sur les corpus qui peuvent être étudiés mais ouverte sur de nouveaux phénomènes linguistiques et langagiers. L'ingéniosité de linguiste dans l'appréhension d'un corpus est par conséquent engagée.

2. Une analyse de discours à la française

L'analyse de discours s'est constituée et développée comme un bloc scientifique en France dans les années 1960-1970 (Francine Mazière, 2010 : 3). L'analyse de discours est par définition multidisciplinaire qui fait appel aux différentes sciences notamment humaines et sociales pour appréhender un fait, même si sa vocation primordiale reste linguistique. Selon (Francine Mazière, 2010 : 3), si l'analyse de discours résiste, c'est parce que

« toute AD tient compte de la langue en tant qu'objet construit du linguiste, et des langues particulières en tant que dotées de formes particulières ; -toute ad a une relation double aux héritages descriptifs des langues. Elle prend en compte la grammaire des langues particulières contre une syntaxe logique universelle. Elle ne

prend en compte que des productions attestées ; -elle configure les énoncés à analyser en corpus construits, souvent hétérogènes, selon un savoir revendiqué, linguistique, historique, politique et philosophique ; - elle construit ses interprétations, ses « lectures », par une critique matérialiste et des méthodes assumées, en tenant compte des données de langue(s) et d'histoire, en prenant en compte les capacités linguistiques réflexives des sujets parlants, mais aussi en refusant de poser à la source de l'énoncé un sujet énonciateur individuel qui serait « maître en sa demeure. »

Cette analyse du discours s'est développée en creusant dans l'héritage de la linguistique structurale qui investit dans la langue comme un ensemble de pratiques, autrement dit, sur l'usage de la langue, sur des énoncés porteurs de sens, en investissant aussi dans toutes les linguistiques dites externes, notamment la sociolinguistique, l'ethnologie et l'ethnographie de la communication, la sociologie de la communication, dans les outils informatiques pour une meilleure quantification des corpus.

Or ce qui mérite d'être rappelé se trouve dans la dimension historique de cette pratique. L'analyse de discours, c'est une écriture de l'Histoire, qui s'appuie sur l'histoire pour dégager des lectures, sans que l'histoire ne soit son premier souci. Elle étudie l'énoncé en travaillant sur sa structure lexicale, les conditions de sa production, l'auditoire, le sens et la connotation...tous les éléments qui peuvent affecter cet énoncé seront repérés, et par conséquent analysés.

Comme toute discipline à caractère scientifique, l'analyse de discours a pour objet d'étude le discours, pris dans ses différentes dimensions sociales et individuelles. À la différence des autres lectures, le discours n'est défini que comme étant une manifestation d'un ensemble de référents sociaux collectifs par le biais de la parole individuelle. La dimension individuelle du/dans le discours constitue un corpus intéressant sur laquelle s'appuie cette pratique pour dégager des lectures qui s'inspirent de la psychologie, de la sociologie ou encore de la philosophie.

L'autre objet d'étude de l'analyse de discours demeure le dispositif énonciatif spécifique à une situation de communication, ce dispositif n'est qu'une association d'une organisation textuelle à laquelle vient s'ajouter la situation de communication. Ce dispositif permet de rendre compte de la situation de communication à travers l'analyse de tous les éléments qui l'affectent (lieu, locuteur, auditoire...) et les agencements ou l'organisation syntaxique. Selon GRAWITZ (Grawitz, 1990 : 345) les énoncés « *ne se présentent pas comme des phrases ou des suites de phrases mais comme des textes. Or un texte est un mode d'organisation spécifique qu'il faut étudier comme tel en le rapportant aux conditions dans lesquelles il est produit. Considérer la structure d'un texte en le rapportant à ses conditions de production, c'est l'envisager comme discours.* »

3. Comment analyser un discours ?

Comment analyser un discours ? Pour certains linguistes, l'analyse du discours n'est qu'une partie intégrante de ce qui est la linguistique générale, du fait que le discours est l'usage de la langue dans un contexte particulier, indépendamment de l'instance où il est produit. Cette conception, de ce qui est «discours», est le fruit de la pensée saussurienne qui oppose ce concept à la langue, mais aussi qui insiste sur la présence d'une force sociale, dans le discours, qui permet à la langue d'être actualisée ou de s'inscrire dans un contexte social particulier. Autrement dit, le discours n'est que la somme de deux constituants; les unités linguistiques et la composante sociale.

De ce fait, faire de l'analyse du discours n'est pas une tâche facile. Il s'agit d'un travail fastidieux et coûteux. Cet avis est partagé par MAINGUENEAU (Maingueneau, 1996 : 8) qui affirme que

«...les difficultés que l'on rencontre pour délimiter le champ de l'analyse de discours viennent pour une part d'une confusion fréquente entre analyse du discours et ses diverses disciplines du discours (analyse de la conversation, analyse du discours, théories de l'argumentation, théories de la communication, sociolinguistique, ethnolinguistique...- la liste n'est pas exhaustive).Chacune étudie ce discours à travers un point de vue qui lui est propre.»

Il semble important de rappeler la présence de quatre méthodes pour exercer et pratiquer l'analyse du discours (désormais AD). La première conception part du principe que l'analyse du discours est par définition empirique, qui a besoin de plusieurs compétences permettant, tant pour le destinataire que pour l'émetteur, de déchiffrer le discours, comme texte qui n'est produit que pour émettre ce que l'énonciateur a besoin d'annoncer. D'autre part, le discours n'est qu'un produit social exigeant la présence de méthodes et de procédés pour l'analyser. L'AD dans cette perspective n'est que la somme des interprétations produites en vue de construire le sens.

La troisième conception consiste à montrer que l'AD fait partie des sciences sociales, et que seules les disciplines attachées à ces sciences humaines, ont la faculté à gérer le sens du discours. La société ne peut pas fonctionner sans communication, le sens de chaque discours relève de la compétence sociale et le degré de la sociabilité des individus. Cette pratique fera aussi appel à la société pour comprendre les activités discursives, spécifiques à chaque communauté, et d'autre part, cette démarche scientifique n'est par définition empirique que parce qu'elle vise la compréhension de l'Homme à travers le procédé de parler et de se comporter.

Une dernière conception. L'AD relève de la sociolinguistique, et si nous disons sociolinguistique, nous ferons aussi allusion mais implicitement au domaine social. De ce

point de vue, elle ne se réduit pas à une discipline empirique mais à une démarche qui cherche à comprendre la société à travers le fait linguistique et langagier.

L'École française de l'analyse de discours demeure originale par sa capacité à s'orienter vers tout ce qui est social, sa capacité à mettre en place différents procédés afin d'analyser un corpus, par forcément à caractère institutionnel, en trouvant aussi utile de s'inspirer de différentes théories et lectures philosophiques. L'originalité de cette école se trouve indiscutablement dans son recours aux méthodes statistiques et à l'outil informatique comme moyen efficace pour dégager différentes lectures et interprétations.

Dans une présentation d'un numéro de la revue *Langage et Société*, Françoise Dufour et Laurence Rosier (2012) ont dressé un panorama assez complet mais non exhaustif de tous les articles et les textes consacrés à l'analyse de discours française qui semble être une science en constant mouvement ; une analyse de discours avec des outils opératoires et des méthodes de plus en plus versées dans la logique, les mathématiques, l'informatique et les statistiques (Maingueneau, 1995 : 5-11) .

Cette école cherchait à comprendre les énoncés ; ce que les mots veulent dire, loin de toute autre considération liée même aux conditions de production, pour se pencher dans un second temps sur une analyse axée sur ce qui façonne ces énoncés, une analyse à la fois dynamique et qui s'éloigne de plus en plus des considérations purement linguistiques.

L'héritage de cette école est considérable, et témoigne d'une école en pleine mutation, d'une dynamique logique qui s'adapte à toute typologie de discours, d'où l'intérêt de mettre en place de nouveaux concepts et méthodes capables de prendre en charge de nouvelles questions et répondre de façon correcte à toute interrogation, en vue d'analyser un texte.

4. De la linguistique à la logométrie

L'apparition de l'outil informatique et des chercheurs qui s'intéressent à la statistique textuelle a eu un impact positif sur l'évolution de l'analyse de discours en tant que pratique récente qui s'appuie davantage sur les progrès de l'informatique.

De cette linguistique qui s'intéresse au recensement des mots qui composent et qui structurent des énoncés est apparue la lexicométrie pendant les années 70, appelée aujourd'hui la logométrie, un prolongement de la première pratique, qui vise l'analyse des unités linguistiques jugées plus pertinentes (Mayaffre, 2009) dans un texte, et qui devient de plus en plus une méthode d'analyse très fréquentée pour étudier les grands

corpus, ériger des comparaisons entre corpus, investir dans des méthodes s'inspirant des mathématiques et des statistiques en vue de proposer un modèle d'analyse qui permet d'appréhender au mieux un texte. De ce point de vue, nous estimons que l'analyse de discours connaît un grand bouleversement avec l'apparition des logiciels qui offrent toute une panoplie de méthodes et des graphes qui prennent en charge les énoncés dans leurs dimensions linguistique et autres, et en cherchant à investir dans le lexique et les différentes formes qui les composent. Ces méthodes aussi diversifiées constituent des pistes de recherche et des axes non négligeables qui pourraient éventuellement un jour constituer une menace pour les jeunes linguistes qui s'intéressent peu aux sciences mathématiques, et rendent l'analyse de discours française inaccessible, sinon très compliquée.

Ces méthodes restent désormais incontournables pour dégager le sens, comme l'analyse des hapax qui sont des mots qui n'apparaissent qu'une seule fois dans un corpus, par opposition aux sous-fréquences qui permettent de désigner les mots qui n'ont qu'une seule occurrence. Les mots de fréquences¹ ont pour fonction de participer à la richesse lexicale d'un texte (Bensebia, 2011 : 162), l'analyse statistique des caractéristiques lexicométriques selon les balises introduites sur le corpus, analyse de la structure de vocabulaire en vue d'appréhender l'accroissement de vocabulaire, les hautes et basses fréquences et leurs rapports avec les autres éléments textuels, les digrammes de ventilation des mots par partition et par balises.

Le diagramme de Pareto offre la possibilité de classer les différents phénomènes linguistiques par ordre d'importance (Lebart.et Salem., 1994 : 47) Il

«...fournit une représentation très synthétique des renseignements contenus dans la gamme des fréquences. Ce diagramme est constitué par un ensemble de points tracés dans un repère cartésien. Sur l'axe vertical, gradué selon une échelle logarithmique, on porte la fréquence de répétition F , qui varie donc de F_{max} , la fréquence maximale du corpus. Sur l'axe horizontal, gradué selon la même échelle logarithmique, on porte, pour chacune des valeurs de la fréquence F comprise entre 1 et F_{max} , le nombre $N(F)$ des formes répétées au moins F fois dans le corpus. La courbe obtenue est donc une courbe cumulée.»

L'inventaire des segments répétés permet de caractériser tous les éléments qui se répètent et qui se présentent comme les plus utilisées, indépendamment de leur longueur ou caractère. Ils sont aussi définis comme des «*suites de formes dont la fréquence est supérieure ou égale à 2 dans le corpus.*» (Lebart.et Salem., 1994 : 60). Il s'agit d'une étape cruciale dans les méthodes lexicométriques, permettant ainsi de déterminer la nature des lexies employées, et le fonctionnement internes des discours et « c'est au sein des segments répétés qu'on a le plus de chance de trouver des

expressions propres au domaine étudié.» (Habert. et all : 1998).

L'analyse factorielle des correspondances des tableaux croisés est un outil statistique qui permet de repérer les spécificités et la fréquence des unités textuelles dans chacune des parties choisies. Il est avantageux de les étudier, pour se rendre compte de la répartition des nouvelles formes. Cet outil met aussi en évidence les proximités qui peuvent apparaître ou exister entre discours ou parties de discours, réparties selon des balises.

Les outils statistiques énumérés ci-dessous sont fournis par le logiciel Lexico3, qui reste original et prisé pour différentes typologies de discours, et tout spécifiquement, les discours de presse ou à caractère institutionnel. Ces outils statistiques offrent des lectures systématiques qui permettent des analyses contrastives, et différentes fonctionnalités capables de rendre le corpus exploitable *lexicométriquement*. Ce logiciel est développé par l'équipe SYLED-CLA2T de l'Université de la Sorbonne nouvelle.

À côté de ce logiciel, il existe d'autres qui cherchent à investir dans d'autres éléments comme des lemmatiseurs, la structure de vocabulaire, l'analyse arborée, analyse des co-occurents, et analyse de thèmes récurrents... Certains dégagent des lectures de style et de contextes.

5. Logométrie : attentes et limites

La logométrie a rendu l'analyse de discours accessible par le biais de différents outils et logiciels qui puisent davantage dans les progrès actuels dans le domaine des sciences informatiques. Cette approche d'analyse ne peut désormais être critiquée dans la mesure où cette pratique se démocratise de plus en plus, et beaucoup d'études s'y intéressent, et cherchent à fournir d'autres applications qui répondent mieux aux exigences multiples, qui se diversifient en fonction des problématiques à répondre, et la diversité des langues. Malgré son apport, cette lexicométrie a laissé place à une dimension machine, qui gouverne plus le chercheur, en l'éloignant de toute approche dite empirique, et/ou de recherche qui le laisse piocher dans les autres lectures philosophiques qui sont l'essence même de l'analyse de discours française, ou encore linguistiques qui permettent d'appréhender la structure du texte ou de l'énoncé, ou tous les autres domaines qui accordent au discours sa dimension linguistique, psychologique et sociale.

Cette logométrie a mis en place aussi une génération de chercheurs qui ne s'intéressent qu'à l'aspect extérieur de la langue, en ignorant que cette méthode est beaucoup plus un outil de description statistique qu'un outil d'analyse, qui cède sa place à un bilan parfois mitigé.

Aucune méthode, originale soit-elle, ne peut négliger l'apport des écoles et les courants

linguistiques, le génie du chercheur dans l'analyse des faits linguistiques ; objets de la recherche. Les compétences d'un linguiste se dégagent et se manifestent au moment où ce dernier s'intéresse à ce qui fait un texte un texte, ce qui fait un énoncé un ensemble des idées qui apparaissent en un monde structuré.

Le texte ou le discours ne pourront aussi se définir au profit d'un traitement automatique qui épargne la dimension humaine de l'acte de parler ou de communiquer. Ce traitement n'est pas une lecture naturelle, même s'il est mis en place pour se dégager de toute appréhension subjective. Or, comme nous l'avions toujours dit, la subjectivité dans l'analyse ne peut être effacée que partiellement, au profit d'une objectivité souvent contestée.

Cependant, une lecture traditionnelle reste néanmoins difficile à réaliser quand le chercheur est confronté à un corpus composé de centaines de milliers de mots, de plusieurs locuteurs...ou pour mettre en place des analyses contrastives ou autres. Aucune lecture n'est originale si elle n'investit pas dans cet héritage de la linguistique et les autres sciences connexes.

Si l'outil informatique a permis de réaliser des lectures originales, et de décrypter des corpus aussi importants sur le plan quantitatif, la logométrie est avant tout une méthode qui fait entrer le chercheur dans la sphère de la modernité, et propose des lectures qui doivent à la fois faire appel à cet héritage ancien de la linguistique et à une lecture scientifique conventionnelle.

La logométrie s'est limitée à une description statistique d'un texte, et n'intervient pas pour expliciter ce que qui fait le choix de lexies par le locuteur, et sa valeur paraît par conséquent interrogative.

Bibliographie

Bensebia, A. A. 2011. *Les stratégies discursives dans le discours présidentiel algérien de 1996 à 2006- Typologie textométrique des textes en langue française*. Éditions EUE, Allemagne. N° ISBN 978-3-8417-8318-9.

Damon, M. 2009. *L'analyse du discours assistée par ordinateur*, Polycopés - Formation - Alexandrie - 01-11 décembre 2009 (CNRS - UMR 6039 « Bases, corpus et Langage »).

Dufour, F., Rosier, L. 2012. « Introduction. Héritages et reconfigurations conceptuelles de l'analyse du discours « à la française : perte ou profit ? ». *Langage et société*, 2012/2 n° 140, pp. 5-13.

Grawitz, M. 1990. *Méthode des sciences sociales*, Édition Dalloz.

Habert, B., Fabre, C., Issac Fabrice. 1998. *De l'écrit au numérique : constituer, normaliser et exploiter les corpus économiques*. Paris : Éd. Masson.

Lebart, L., Salem A. 1994. *Statistique textuelle*, version électronique. Paris : Dunod.

Mangueneau, D. 1996. *Les termes-clés de l'analyse du discours*, Paris : Seuil.

Mangueneau, D. 1995. Présentation. *Langages*, 29e année, n° 117, pp. 5-11.

Mazière, F. 2012. *L'analyse du discours*. Paris : P.U.F. « Que sais-je ? ».