

La lexicométrie au service de l'analyse de discours: exploration automatique d'un corpus sur les élections présidentielles de 2014

Dr. Khelfi Chahrazed, Université Oran1 Ahmed Ben Bella, , Algerie.

date de soumission: 13/02/2018 date d'acceptation: 03/06/2018 date de publication: 12/06/2018

Résumé

L'objectif de notre article est de faire état d'une approche automatique, relativement méconnue en Algérie pour l'analyse du discours médiatique. Il s'agit de la lexicométrie ou la recherche sur corpus. En appliquant une analyse assistée par ordinateur, nous avons construit un large corpus de discours médiatique en Algérie durant la période du printemps arabe (2011-2015) afin d'analyser le comportement des journalistes vis à vis de la mémoire de la guerre d'Algérie. L'exploration lexicométrique de ce corpus s'est révélée intéressante dans le repérage des spécificités linguistiques, typographiques et idéologiques du discours de la presse francophone pendant les présidentielles de 2014. Au moyen de programmes statistiques notamment le calcul des spécificités développé par le laboratoire de Saint Cloud, nous avons repéré les régularités significatives (discursives) susceptibles de nous informer sur les positions de la presse et les grandes thématiques caractérisant son discours durant les élections.

Mots clés: lexicométrie, calcul des spécificités, occurrence, discours, corpus.

ملخص

الغرض موضوع بحثنا هذا، عن منهجية جديدة منتشرة منذ زمن ليس بالقليل في الدول الغربية، ولكن استخدامها في الأوساط البحثية العربية لا يزال محتشما، يتعلق الأمر بتحليل الإحصائي المعجمي للنصوص Lexocométrie. وإيماننا منا أن استغلال إمكانيات الحواسيب والبرمجيات في عصرنا يساهم في تطوير البحث العربي في مجال تحليل الخطاب بكل أنواعه الأدبية والسياسية والصحفية، ارتأينا أن نسلط الضوء على هذه المنهجية وذلك من خلال دراسة خطاب الصحف الجزائرية المكتوبة باللغة الفرنسية لإبراز نجاعة هذه الطريقة الحاسوبية، ودور برنامج Lexico3 في تحليل وسير فور المتون الرقمية كبيرة الحجم، وتحديد الخصائص المعجمية والنحوية والأسلوبية للصحف خلال الانتخابات الرئاسية لعام 2014، ومن ثم تفسير اتجاهاتها الأيديولوجية ومواقفها السياسية.

التمارين الصعبة المتمثلة في الانتقال بالأنكار والكلمات من لغة إلى لغة أخرى.

الكلمات المفتاحية: المتون الرقمية، الخطاب، البرمجيات، الصحافة، الإحصاء، المعجم.

1. Un bref rappel historique:

Le monde arabe est secoué par des mouvements de contestations déclenchés au Maghreb notamment en Tunisie en Décembre 2010 et qui se sont propagés pour toucher le reste des pays du Moyen Orient. Ces révoltes populaires qui brandissaient le slogan « dégage », exigeaient le départ des dictateurs, l'instauration d'une démocratie et un partage des richesses qui assure aux populations des emplois et de meilleures conditions de vie. Après une répression sanglante, les manifestations ont conduit au départ de Benali et Moubarak en Tunisie et en Egypte. Mais elles ont tourné à la guerre civile en Libye et en Syrie. Les autres régimes ont réussi à désamorcer la contestation en faisant des concessions pour maintenir la paix comme le Maroc, l'Arabie Saoudite, Oman et le Koweït.

Dans cette situation politique en pleine mutation, les algériens aspiraient à percevoir une dynamique en faveur du renouveau politique, et pourquoi pas l'instauration d'une vraie démocratie. A l'approche de l'échéance électorale, il était presque acquis que le président âgé de 76 ans, au pouvoir depuis 1999, victime d'un accident vasculaire cérébral et qui n'a pu reprendre ses fonctions qu'à l'issue d'une longue hospitalisation en France, ne se représenterait pas pour un quatrième mandat. Mais la réalité était tout autre, le 22 Février 2014 le premier ministre Abdelmalek Sellal officialise la candidature du président. Décision qui a suscité l'indignation de la classe politique et poussé certaines formations et personnalités à boycotter le scrutin qu'ils jugeaient joué d'avance en faveur du président sortant. A commencer par le principal parti islamiste, le Mouvement de la société pour la paix(MSP) et le rassemblement pour la culture et la démocratie(RCD), le parti d'opposition kabyle. La société civile n'était pas en reste puisqu'elle s'est mobilisée à travers les réseaux sociaux en constituant le mouvement « Barakat » qui a organisé des rassemblements à travers le territoire pour exprimer son ras le bol.

2. Présentation du corpus et méthodologie:

Notre corpus est composé de l'ensemble des éditoriaux parus pendant la campagne et élections présidentielles de 2014, sur une période qui s'étend du 23 Mars au 24 Avril 2014. Nous avons privilégié les éditoriaux qui sont les mieux indiqués pour refléter la ligne éditoriale du journal. La collecte s'est faite sur 4 quotidiens francophones Le choix s'est fait sur les journaux au plus grand tirage Elwatan (200 000) exemplaires, le Quotidien d'Oran (200 000) et Liberté à (125 000).

Pour constituer notre corpus nous nous sommes inscrit à la newsletter et ouvert un compte dans les différents journaux afin d'accéder aux archives.

Les éditoriaux des 4 journaux sont présentés de différentes manières et mal organisés sur la page du journal, ce qui a nécessité de longues heures de navigation afin de retrouver le texte de l'édito parmi la multitude des rubriques.

Les textes sont publiés sous format HTML, nous avons choisi de les enregistrer d'abord sous format WORD pour pouvoir les mettre par la suite en code Txt compatible avec le logiciel dont nous avons choisi de travailler.

Nous avons retenu 100 éditos consacrés à la campagne et élections présidentielles, 28 textes pour El watan, 26 textes dans le Quotidien d'Oran, 26 et 20 pour Liberté et El Moudjahid respectivement.

L'observation de notre corpus a révélé une certaine disparité en nombre d'articles ; le journal El Moudjahid présente le plus grand écart. De surcroit la taille des éditos est plus longue dans El watan et le Quotidien d'Oran que dans les 2 autres journaux.

2.1 La lexicométrie:

La lexicométrie est une méthode qui s'est développée grâce aux progrès de l'informatique et la disponibilité des textes sur Internet. Les chercheurs dont l'objet d'étude est le texte (le discours), s'y intéressent de plus en plus, en constituant des corpus numériques pour les étudier avec des outils informatiques. C'est une discipline qui allie la lexicologie, la statistique et l'in-

formatique pour l’analyse de corpus étendus. Ou, comme l’exprime (Rastier, 2005), “la textométrie fait partie des techniques capables de plonger dans des dimensions profondes du matériau textuel”. Son intérêt réside dans les traitements quantitatifs qu’elles opèrent sur des corpus étendus, et la mobilisation de calculs statistiques pour mesurer le degré de significativité du lexique d’un texte. Ce qui favorise la lecture, la compréhension et l’interprétation en analyse de discours.

La méthode e a été développée au Laboratoire de lexicologie de l’école normale supérieure de Saint-Cloud à partir de 1970 sous la direction de Maurice Tournier. De nombreux travaux ont été réalisés dans ce laboratoire sur les discours politiques et syndicaux.

3. Analyse du corpus:

Pour le traitement de notre corpus nous avons choisi le logiciel Lexico 3, c’est un programme mis au point par André Salem professeur et chercheur de l’équipe SYLED-CLA2T Paris3.

Le logiciel procède à la segmentation automatique du texte en occurrences de formes graphiques, qui permet de découper le texte en unités minimales sur la base d’un ensemble de caractères délimiteurs choisis parmi l’ensemble des caractères du texte (Lebart et Salem, 1994). Les occurrences sont les entrées nécessaires pour le traitement lexicométrique.

Le dépouillement de notre corpus en formes graphiques donne les résultats suivants:

Nombre d'occurrences	51130
Nombre de formes	8627
Nombre d'hapax	5094
Fréquence maximale	2560

Après la segmentation, le logiciel génère toute une série de documents qui permettent de mieux appréhender le vocabulaire du corpus. Dans l’index hiérarchique par exemple les formes sont classées par fréquence décroissante

qui aide à examiner les formes les plus utilisées. L'index nous informe sur les thématiques du corpus. On trouve en tête de liste : campagne, Bouteflika, électoral, candidat, pouvoir, système, régime, président, mandat, Algérie...

En consultant les concordances, programme lexicométrique documentaire qui facilite le retour au contexte pour chaque mot afin de comprendre le sens des unités choisies, On observe une grande variabilité de la fréquence d'emploi des formes, d'un journal à l'autre. Ce qui ne suffit pas de mesurer et ordonner les contrastes entre différentes parties (journaux). Peut-on trouver un instrument apte à expliciter les spécificités des différents quotidiens? Le programme d'analyse des spécificités tente de répondre à cette question.

4. Le programme d'analyse des spécificités:

L'analyse des spécificités permet de porter un jugement sur la fréquence de chacune des unités textuelles dans chacune des parties du corpus.

Le calcul des spécificités se fait à partir du nombre d'occurrences d'une forme donnée dans une partie donnée en prenant en compte la fréquence totale de la forme, sa fréquence dans les sous-parties, la taille de la partie et celle du corpus. Selon les mots de Guyard Marie-Renée:

Cette méthode permet de classer les formes d'un texte en trois lots différents : — toute forme dont la sous-fréquence dans un échantillon et toute sous-fréquence supérieure à elle sont, ensemble, assorties d'une probabilité inférieure à 5%, sera considérée comme spécifique positive de cet échantillon ; — toute forme dont la sous-fréquence et toute sous-fréquence inférieure à elle n'atteignent ensemble pas 5 % de chances de se produire, sera considérée comme spécifique négative de l'échantillon ; — toute forme dont la sous-fréquence n'entre pas dans un des deux cas précédents est considérée comme banale.(Guyard 1981 : 95-122)

Revenons à notre corpus pour mesurer la valeur des spécificités des formes et commençons par analyser:

4.1 Quatrième mandat:

Se trouve dans notre corpus sous deux formes, en adjectif numéral

(quatrième) et en chiffre (4e). Afin de regrouper ces formes en une seule entité, nous nous sommes servis du programme groupe de formes de Lexico3 qui permet la manipulation des groupes comme des entités uniques.

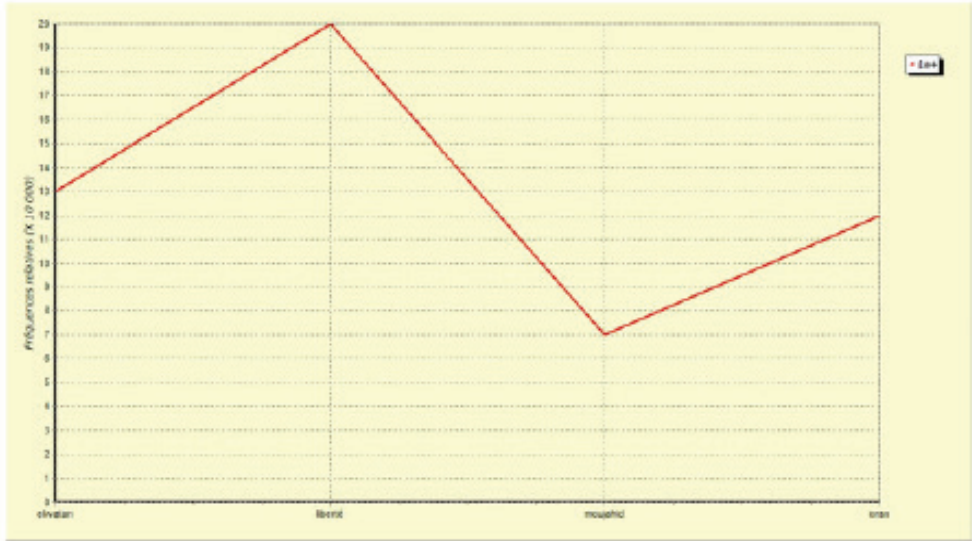


Figure1 : graphique de ventilation du groupe de forme « quatrième mandat »

Figure1: graphique de ventilation du groupe de forme «quatrième mandat»

En termes de fréquence relatives on constate que la récurrence de quatrième mandat dans Liberté est la plus élevée par rapport aux autres journaux. La fréquence la moins importante est constatée dans Moudjahid et se rapproche de très près dans Elwatan et le quotidien d’Oran. Sauf que le calcul de spécificité nous donne d’autres données qui soulignent une apparente contradiction concernant le journal Elmoujahid.

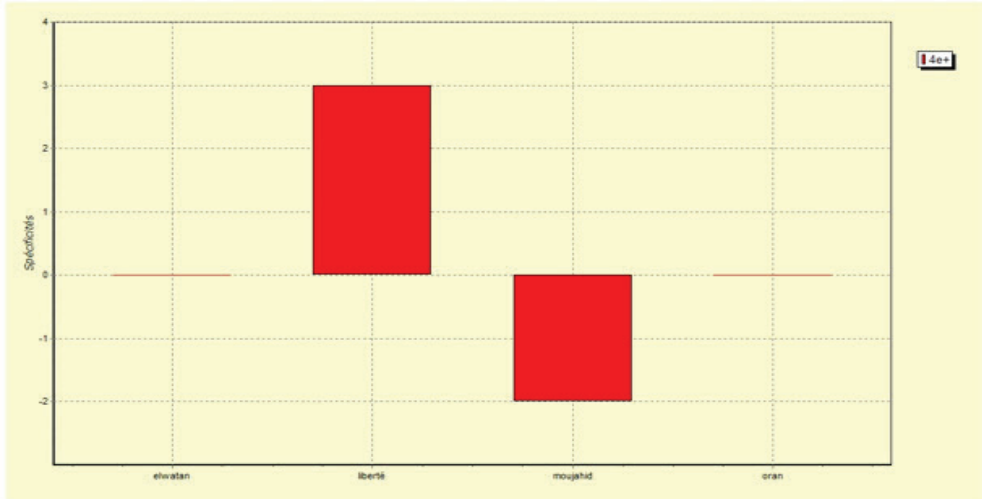


Figure2 : graphique des spécificités du groupe de forme « quatrième mandat

Figure2: graphique des spécificités du groupe de forme «quatrième mandat.

Il existe une proximité entre Elwatan et le quotidien d’Oran, la forme ne représente pas une spécificité du discours mais elle est considérée en terme lexicométrique une forme banale puisque il est question de quatrième mandat si le président sortant venait à être réélu et que sa récurrence dans la presse est tout à fait normal.. Mais ce qui est frappant c’est son caractère spécifique négatif dans Moujahid et spécifique aussi, mais en sens inverse, positif dans Liberté. Ce qui est en apparence contradictoire ne l’est pas. Durant toute la période de la campagne et même avant, la forme a véhiculé une connotation péjorative, d’ailleurs l’examen de son champ sémantique dans les journaux d’opposition comme ElWatan (sinistre, péché originel, suicidaire, imposé) et Le quotidien d’Oran (supercherie, coquille, opposition) le prouve. Ce qui explique son sous-emploi par le journal Moujahid.

4.2 Analyse du champ sémantique de 4mandat:

Le tableau des spécificités ci-dessus nous informe sur la fréquence de la forme mais n’est pas suffisant pour révéler la position politique de chaque journal, c’est pour cette raison une étude plus attentive de la forme dans les

concordances est indispensable à travers les réseaux tissés par les mots qui l'entourent.

Pour Elwatan la forme est qualifiée de sinistre, péché originel, suicidaire, imposé, coup de force. Elle est l'action de Bouteflika, le pouvoir, les directeurs de campagne (Benyounes, Cherif Abbès), les partisans, les télévisions privées.

Liberté aussi utilise le même ton dans sa dénomination du 4 mandat. Il est la cause de rejet massif des présidentielles, un leurre de plus, incertain, fraude, danger. Les agents de ce mandat sont personnifiés : Bouteflika, Sellal, Amar Ghoul, soutiens de Bouteflika, partisan de chef de l'état.

Quant au Quotidien d'Oran il est dans une position mitigée concernant les agents de ce mandat, il n'est pas l'œuvre de Bouteflika mais du : camp du 4, les pros du 4, la machine du 4, les partisans du 4mandat. Cependant il est qualifié de supercherie, coquille vide, déconfiture, n'a pas l'efficacité, controversé.

Dans Elmoujahid bien évidemment la forme est associée à l'édification, réélection, constance, continuité, consécration.

L'examen exhaustif des qualifications et associations de la forme « quatrième mandat » atteste l'existence d'une opposition politique claire et directe contre le candidat président par elwatan et liberté et légèrement conciliante dans le quotidien d'Oran et progouvernemental dans Elmoujahid.

4.3 Les acteurs politiques:

Attachons nous maintenant à la catégorie des noms propres. Ce n'est pas un commentaire sur l'emploi de tel ou tel nom qui nous intéresse ici mais une appréciation sur l'ensemble de la catégorie qui s'impose. Et ce pour vérifier l'intérêt que porte chaque quotidien à la campagne et les élections.

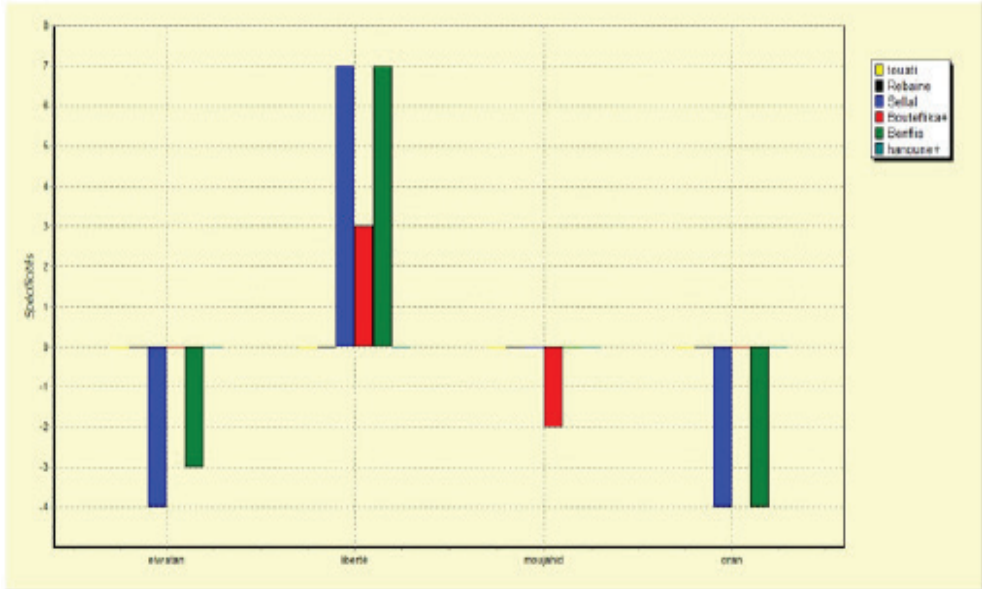


Figure 3 : graphique des spécificités des acteurs politiques

Figure 3: graphique des spécificités des acteurs politiques

Après examen de ce graphique, force est de constater que la lecture de ces données n’est pas évidente. On remarque que les acteurs de la campagne sont spécifiques positifs dans Liberté uniquement, c’est-à-dire sur employés par les journalistes et ce par rapport au reste du corpus. La spécificité positive est d’interprétation aisée : elle nous indique que le journal Liberté est le plus impliqué et intéressé par la campagne électorale. Et le fait que la fréquence de Benflis et Sellal soit abondante, nous pousse et avec beaucoup de précaution à penser que Liberté, en plus de s’intéresser à la campagne des deux candidats, perçus comme les grands favoris, soutient Benflis. Quant à la spécificité négative, elle est plus difficile à interpréter puisqu’elle nous dit que cette forme est sous-utilisée dans une partie par rapport au corpus. Autant la spécificité positive est le résultat d’une action du locuteur, autant la spécificité négative est le résultat de l’examen des fréquences en comparaison avec ce qu’ont dit les autres locuteurs. On pourrait penser que les autres journaux

sont moins emballés par la campagne puisque quelques parties ont boycotté le scrutin comme le RCD dont le journal Elwatan est porte parole.

On examinant les spécificités dans Moudjahid, on se serait attendu à ce que les noms Bouteflika ou Sellal soient spécifiques positifs puisque le quotidien est d'obédience FLN et soutient le candidat sortant, mais on constate que ce dernier a dans le discours du journal une valeur spécifique négative. Cela pourrait être le résultat, à notre avis, d'une stratégie délibérée du journal pour donner plus de légitimité à ces présidentielles et davantage d'objectivité au journal. Toutefois, le discours mérite une analyse plus approfondie pour déceler cette stratégie.

Une forme qui n'est spécifique d'aucune partie est une forme banale. Elle renseigne sur le champ commun à tous les quotidiens. On trouve dans cette catégorie les autres acteurs Touati, Rebaine, Hanoune.

4.4 Les formes fonctionnelles:

Pour exprimer un message, le locuteur fait appel au vocabulaire qui est de deux types : les mots pleins et les mots outils. On trouve dans la première catégorie les substantifs (par exemple : réalité, paix, maison...), les adjectifs ou adverbes (par exemple : grand, favorable, récemment...) et les verbes(par exemple : partir, servir...). Afin de donner sens et cohésion aux structures textuelles, on a besoin de mots fonctionnels qui lient entre les différentes formes lexicales pleines. Dans cette catégorie on peut énumérer : les formes des verbes être et avoir, les pronoms personnels, relatifs, déterminant, conjonctifs...

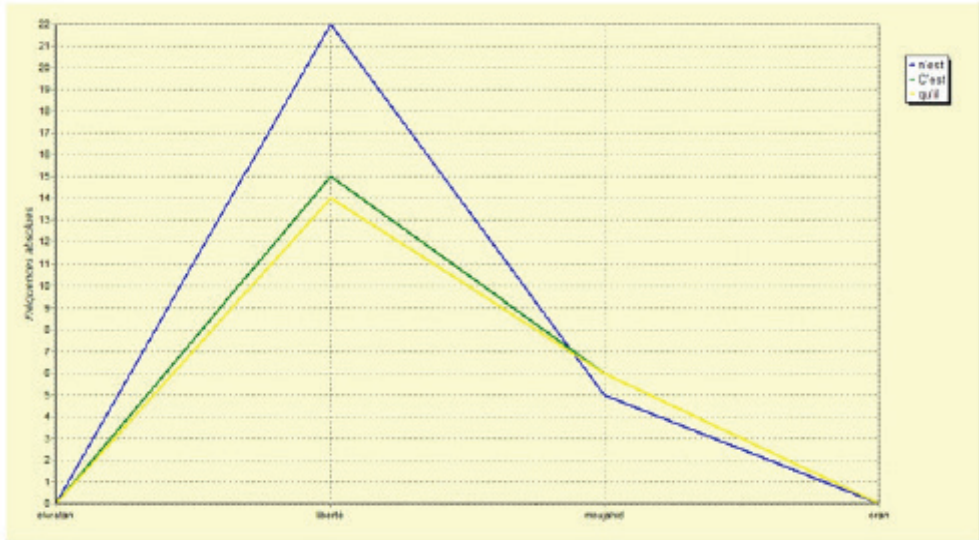


Figure 4 : graphique de fréquences des formes fonctionnelles

Figure 4: graphique de fréquences des formes fonctionnelles

L'examen exhaustif des formes fonctionnelles de notre corpus révèle une forte parenté en termes de constructions syntaxiques entre Liberté et E IMoujahid. Le verbe être et avoir à la forme négative « n'est » « n'a » est spécifique des deux quotidiens et est à zéro fréquence dans le reste des journaux. Idem pour « c'est » le verbe être précédé du pronom démonstratif « ce » éliidé.

Le « En » utilisé comme préposition dans des locutions adverbiales (par exemple : en tout état de cause, en attendant) avec les constructions relatives avec « que » sont spécifiques aussi des deux journaux et inexistantes dans le reste du corpus.

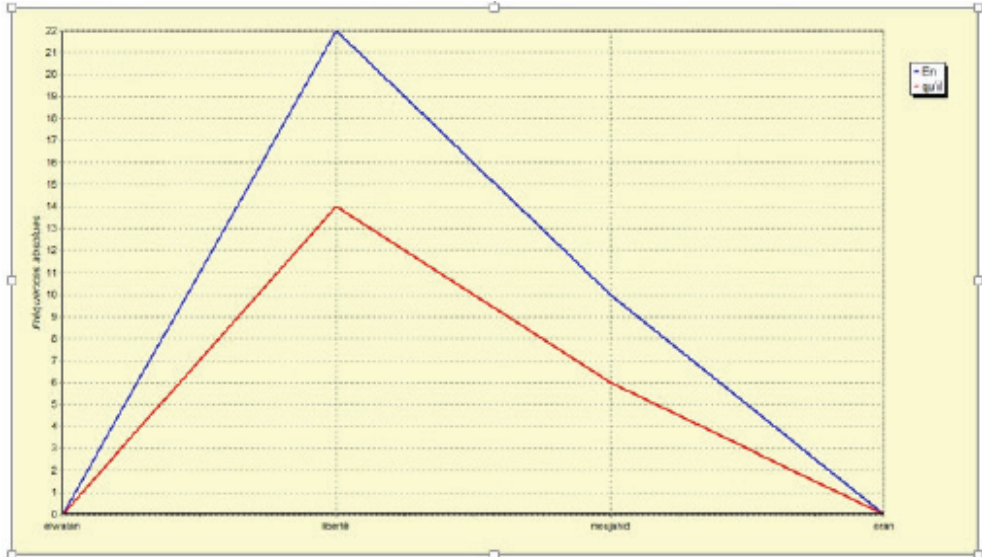


Figure 5 : graphique de fréquences des formes fonctionnelles

Figure 5 : graphique de fréquences des formes fonctionnelles

Ce constat nous conduit à suggérer que les journalistes de Liberté et Elwatan pourraient appartenir à la même école en matière de style de rédaction.

4.5 Conventions d'écritures:

Ce qui nous frappe dans les spécificités linguistiques de notre corpus, c'est la forte similitude entre Elwatan et le Quotidien d'Oran à propos des conventions d'écriture. Ils désignent le peuple en employant l'adjectif avec minuscule « les algériens » alors que Liberté et Elwatan le désignent au moyen de sa nationalité avec majuscule « les Algériens ».

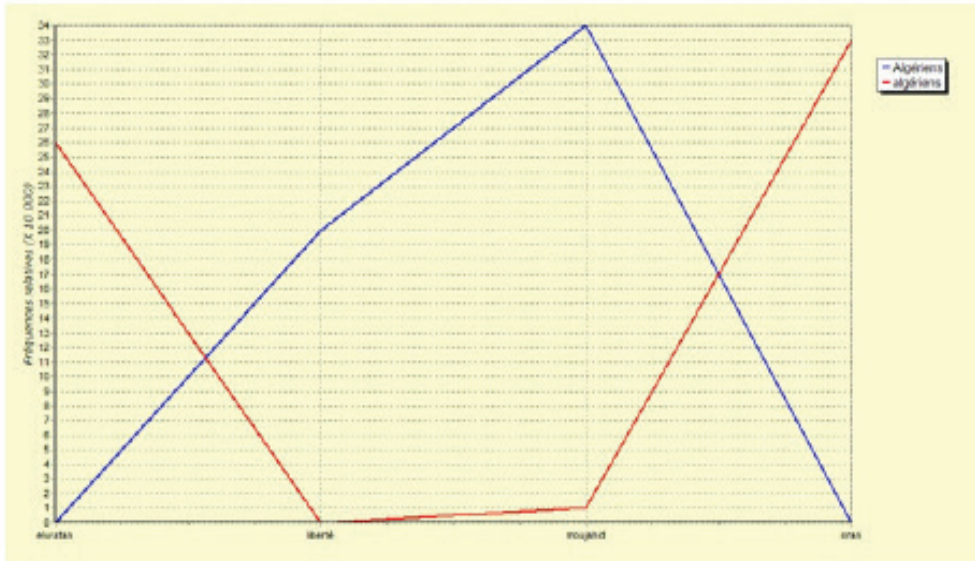


Figure 4 : graphique de fréquences de « Algériens et algériens »

Figure 4: graphique de fréquences de « Algériens et algériens »

4.6 Positions politiques et thèmes:

Les journalistes des différents quotidiens abordent plusieurs sujets, nous avons jugé utile d'examiner leurs fréquence dans notre corpus, ils pourraient nous informer sur la position de chaque journal et sa ligne éditoriale.

Il est question dans les éditos de fraude, électeurs, opposition, pouvoir, violence, transition, violence, consensus, électeurs, crise, stabilité, réconciliation.

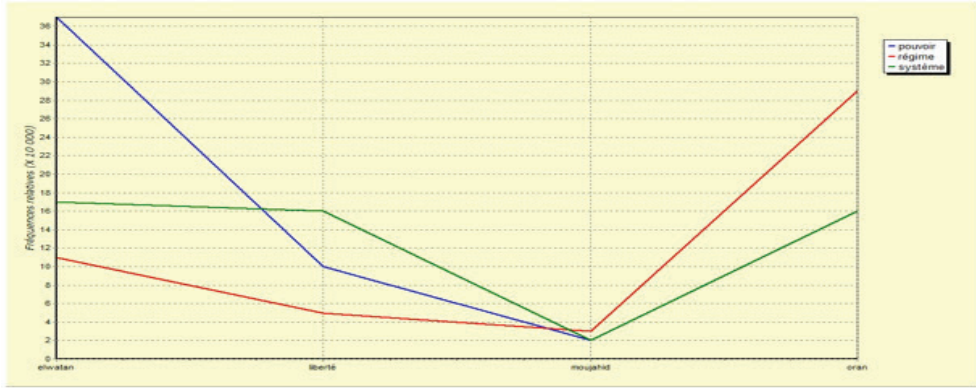


Figure 5 : graphique de fréquences « thèmes »

Le système en place est présent sous plusieurs appellations : pouvoir, régime et système, il occupe une place importante dans les éditos. La fréquence de ces formes est élevée dans tous les journaux sauf Elmoujahid. La récurrence de ce thème s’explique par l’obligation qui lui(le régime) est faite de rendre compte de son mandat. Il est selon les journaux d’opposition, à l’origine du blocage politique que connaît le pays.

Nous avons constaté l’absence totale, zéro fréquence de la forme « Etat » dans Elwatan, Liberté et le Quotidien d’Oran, alors qu’elle est présente dans Elmoujahid. L’Etat algérien n’existe pas pour la presse d’opposition.

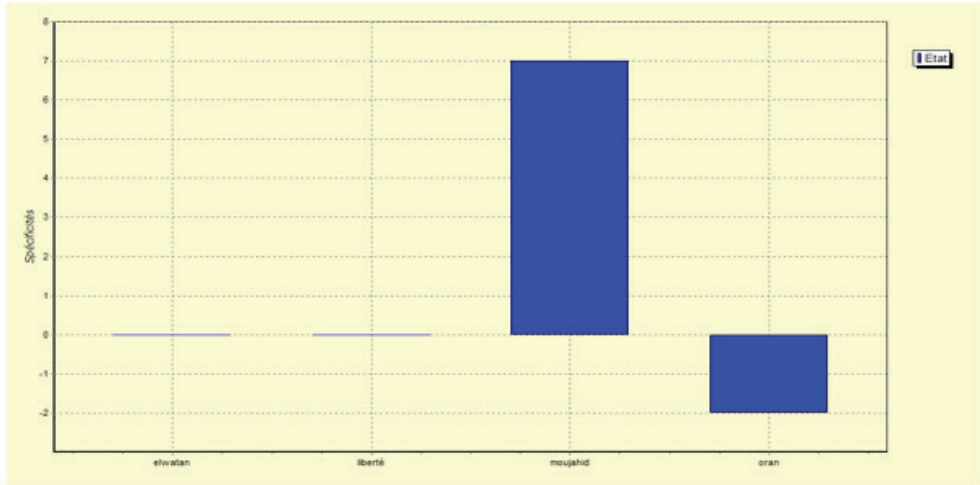


Figure 6 : graphique de fréquences « Etat »

Figure 6 : graphique de fréquences « Etat »

Parmi les formes extrêmement sur employées, on repère fraude et transition. Fraude est spécifique positive dans Liberté et d’emploi banal dans Elwatan et Le quotidien d’Oran mais elle est manifestement sous employé par Elmoujahid.

On peut remarquer que l’écart accentue fortement l’opposition entre le quotidien d’Oran et les deux autres journaux d’opposition en ce qui concerne la fréquence de « consensus » et « transition » qui sont sur employés par le journal. Cela nous conduit à suggérer que le Quotidien est dans une opposition mitigée et se montre plus conciliant avec le système en appelant au changement progressif et à la conciliation entre le régime et le peuple..

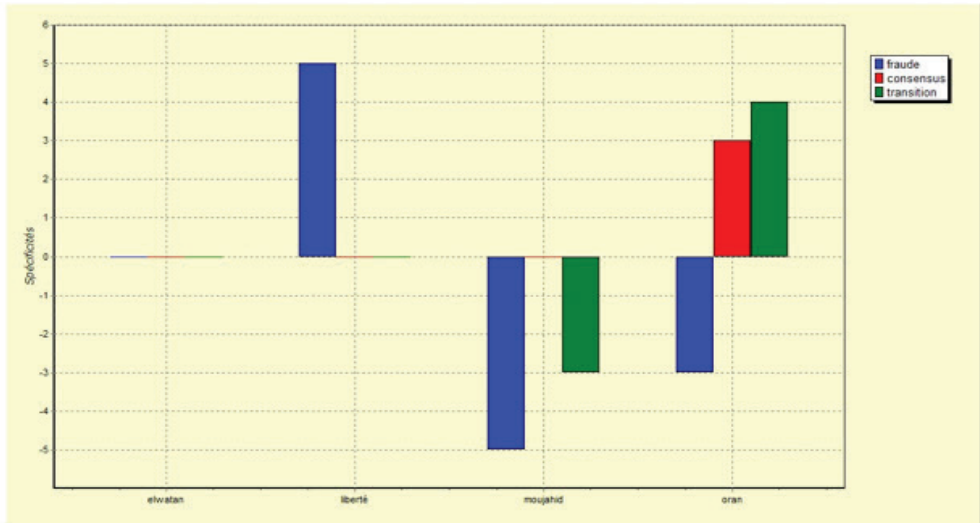


Figure 7 : graphique de fréquences « thèmes »

Figure 7: graphique de fréquences « thèmes »

5. Conclusion:

L’analyse des spécificités de notre corpus a mis en évidence la forte fréquence de « Quatrième mandat » notamment dans la presse d’opposition. L’examen du champ sémantique de la forme dans les différents quotidiens dénote une opposition claire contre le candidat président par El watan et Liberté et position légèrement conciliante dans le Quotidien d’Oran. Et le même constat est fait concernant les formes « transition » et « consensus » qui sont spécifiques à ce dernier et d’emploi banal dans le reste du corpus.

Quant aux acteurs de la campagne, le calcul a relevé un suremploi de Benflis, Bouteflika et Sellal, ce qui nous a poussé à conclure que le journal soutient le candidat Benflis et s’intéresse au déroulement de la campagne plus que les autres journaux.

En matière de constructions syntaxiques et style, on a constaté une forte parenté entre Liberté et El Moudjahid au point de nous laisser croire qu’il s’agirait des mêmes journalistes. Comme il existe une similitude entre El wa-

tan et le Quotidien d'Oran en matière de conventions d'écritures.

A partir de ce qui a été avancé ci-dessus, on peut dire que La lexicométrie permet des traitements quantitatifs qui favorisent la lecture, la compréhension et l'interprétation de grands corpus. Très efficace pour entrer dans les spécificités lexicales des textes, mais elle n'est pas pour autant adaptée à toutes les recherches en analyse de discours.

Bibliographie:

- Arabic and the Media: Linguistic analyses and applications, edited by Reem Basiouney, Leiden. Boston, Brill, 2010.
- Arab Media and political renewal community: legitimacy and public life, edited by Naomi Sakr, London, New York, I.B. Tauris & Co. Ltd, 2007.
- Bowker, Lynne et Pearson, Jennifer. Working with specialized language: a practical guide to using corpora. Routledge, 2002.
- Dijk Van, Discourse analysis, its development and application to the structure of news, Journal of communication, vol 33, N°2, 1983.
- Dijk Van, Opinions and Ideologies in Editorials, Papers for the 14th international symposium of critical discourse analysis, language, social life and critical thought, Athens, 14-16 December, 1995.
- Fairclough Norman, Language and power, New York, Longman Inc, 1989.
- Glasgow University Media group, Getting the message: News, truth and Power, London, Routledge, 1993.
- Geffroy Annie. Trois successeurs de Marat pendant l'été 1793 : analyse lexicométrique des spécificités. In: Mots, n°1, 1980, pp. 167-187.
- Guilhaumou Jacques, L'historien du discours et la lexicométrie [Étude d'une série chronologique : le « Père Duchesne » d'Hébert (Juillet 1793 - mars 1794)]. In: Histoire & Mesure, 1986 volume 1 - n°3-4. Varia. pp. 27-46.
- Guyard Marie-Renée. Spécificités d'auteurs dans Le Surréalisme au service de la Révolution. In: Mots, n°2, mars 1981. Qu'est-ce que le vocabulaire spécifique d'un texte politique? pp. 95-122.
- Lebart, L., Salem, A., Statistique textuelle, Paris, Dunod, 1994.
- Les corpus en linguistique et en traductologie, édité par Michel Ballard et Carmen Pineira-Tremontant, Artois Presses Université, 2007.

- Mellor Noha, *Modern Arab journalism: problems and prospects*, Edinburg University Press Ltd, 2007.
- Rafferty Kean, *That's what they said about the press*, New York, Vantage press, 1975.
- Raštier, François. *Enjeux épistémologiques de la linguistique de corpus*, Presses Universitaires de Grenoble, 2005.

