Corpus romanesque et lexicométrie

Ouafae Benzina*[1]

[1]Faculté des Lettres et des sciences Humaines -Université Moulay Ismail-Meknès – Maroc

Résumé

Nous nous proposons dans cette contribution d'étudier le vocabulaire du corpus romanesque de Guy de Maupassant: *Une vie*, *Bel-Ami*, *Mont-Oriol, Pierre et Jean, Notre cœur et Fort comme la mort*. Ce corpus, nous allons le décortiquer à travers des moyens scientifiques, à savoir la statistique et l'informatique. Comment l'ordinateur peut-il ouvrir de nouvelles perspectives à la littérature et aux études linguistiques ?
Parmi les outils informatiques qui traitent des corpus littéraires, nous avons opté pour le logiciel Hyperbase (version 7.1 pour windows). Un logiciel de statistique lexicale, élaboré par Etienne Brunet en 1989 et développé environ chaque année. Il en est actuellement à la version 10, disponible en 2018.

Le traitement statistique des données avec le logiciel Hyperbase permet d'étudier la structure du vocabulaire d'un corpus, étude qui s'intéresse en général à la distribution des fréquences, à l'étude des hapax, à la richesse lexicale, ainsi qu'à l'accroissement lexical. En plus de l'étude de la structure d'un corpus, Hyperbase permet, en effet, de déduire des conclusions également au niveau du contenu lexical en étudiant la distance lexicale, le vocabulaire spécifique ainsi que l'évolution d'un même auteur au cours de la période pendant laquelle il a produit son œuvre.

A cet égard, à l'aide de ce logiciel, nous avons constitué une base que nous avons nommée MAUPASSA.EXE à partir des six romans de l'auteur normand.

Les textes ont été hébergés par le Laboratoire d'Informatique de Besançon, par la suite, ils étaient placés, gracieusement, dans une bibliothèque numérique : Athéna, par Thierry Selva, sur son site " Maupassant par les textes " (http://maupassant.free.fr/).

Le travail se fera donc sur deux dimensions : l'une compte N, nombre de mots du texte (occurrences) ; l'autre dénombre V, vocabulaire du texte (vocables). Dans notre corpus, nous avons relevé une étendue N (corpus) de 549056 occurrences et 22932 vocables.

À cet effet, notre travail d'analyse statistique s'articulera en deux volets. Dans le premier, nous aurons à démonter ce qui particularise les romans de Guy de Maupassant par l'observation de son vocabulaire spécifique. Ce dernier est déterminé avec le logiciel Hyperbase en relevant le vocabulaire en excédent et le vocabulaire déficitaire dans le corpus. Le logiciel fait en même temps une comparaison avec le *Trésor de la langue française* en calculant l'écart réduit de chaque forme dans chacune des parties du corpus. Si le vocable en question est significativement sur-employé par notre auteur, nous déduisons que l'écrivain éprouve une certaine attirance pour ce vocable. A l'inverse, s'il est significativement sous-utilisé, par rapport à l'usage qu'en font les autres, l'auteur éprouve une certaine répulsion

*Intervenant

pour ce mot qu'il évite ou qu'il oublie.

Dans Hyperbase, les données du TLF sont insérées en tant que norme et servent de base de calcul en indiquant la différence entre notre corpus et celle de *Frantext.* Il s'agit donc de voir les mots préférés de Maupassant et la signification qu'il leur prête.

Dans notre étude du vocabulaire spécifique de Maupassant, nous nous intéresserons au vocabulaire positif et négatif. Par vocabulaire positif/négatif, nous entendons les vocables particulièrement suremployés et sous-employés qui vont nous donner une idée des thèmes traités ou négligés. Autrement dit, le vocabulaire positif représente le vocabulaire excédentaire qui, après une comparaison avec le TLF (XIXème et XXème siècle), a un écart réduit positif. Quand au vocabulaire négatif ou déficitaire, il désigne les formes qui n'apparaissent pas beaucoup dans le corpus et dont la présence par rapport au TLF est insignifiante. Ces formes ont un écart réduit négatif.

Dans ce sens, la lecture des listes des mots présentées, par le logiciel Hyperbase pour chaque texte, nous amène à constituer les champs associatifs qui nous permettrons d'interpréter ce vocabulaire.

Une autre façon d'approcher globalement le vocabulaire de notre corpus : c'est d'observer le vocabulaire le plus caractéristique dans chacun des textes qui le constituent, en prenant pour référence l'ensemble de ce corpus. Le vocabulaire spécifique est déterminé avec le logiciel Hyperbase par le calcul de l'écart réduit de chaque forme dans chacune des parties du corpus.

Dans le deuxième volet de notre étude, il sera question de l'étude de l'évolution du vocabulaire de notre corpus qui interroge l'ensemble du vocabulaire et mesure l'évolution de chaque terme, par le coefficient de corrélation de Bravais-Person. Avec ce calcul réalisé par le logiciel Hyperbase, nous saurons quels sont les mots que Guy de Maupassant emploie de plus en plus et ceux qu'il abandonne progressivement dans son œuvre romanesque.

Ce coefficient examine donc une éventuelle relation entre une distribution et la chronologie. Il évolue entre deux limites : +1 (pour une progression) et -1 (pour une régression), comme l'indique Brunet : " la valeur du coefficient est négative ou positive selon que le mot est en régression ou en progression. "[**1**]

Le calcul de corrélation nous permet donc de suivre, globalement, la trajectoire des mots dans les différents textes du corpus. Le tri des résultats permet de constituer la liste (décroissante) des mots en progression et la liste (croissante) des mots en régression. Cet examen du contenu lexical nous permettra donc d'enrichir notre connaissance de la spécificité stylistique des textes de notre corpus et de préciser l'évolution de l'écriture de Guy de Maupassant.

A cet égard, nous avons remarqué qu'il y a eu un renouvellement lexical chez Maupassant. L'auteur passe du lexique des parties du corps, des verbes d'action et de l'argent (termes concrets) au vocabulaire de la beauté, de l'art, des salons des sentiments et des préoccupations morales (termes abstraits).

En effet, le thème de l'argent, très dominant dans les quatre premiers romans de Maupassant, disparaît complètement dans ses deux derniers romans, *Fort comme la mort* et *Notre cœur.* À partir de *Pierre et Jean*, le romancier observateur devient romancier psychologue.

En supprimant tout problème matériel à ses personnages et en les faisant évoluer dans des milieux favorisés, Maupassant estime mieux s'attarder sur leur personnalité. Dans *Fort comme la mort*, à une seule exception près, on ne trouvera aucune somme d'argent numériquement définie ; il en va de même dans *Notre cœur.*

Enfin, nous verrons dans quelle mesure les résultats obtenus sont en adéquation avec les analyses stylistiques classiques sur l'œuvre romanesque de l'auteur normand.

Références bibliographiques

1- Œuvre de Guy de Maupassant

- Maupassant, Guy, *Bel-Ami*, Paris, Pocket, 2006.

- Maupassant, Guy*, Fort comme la mort*, Paris, Albin Michel, 1983.

- Maupassant, Guy, *Mont-Oriol*, Paris, Gallimard, 2002.

- Maupassant, Guy, *Notre cœur*, Paris, Gallimard, 1993.

- Maupassant, Guy, *Pierre et Jean*, Paris, Albin Michel, 1999.

- Maupassant, Guy, *Une vie*, Paris, Presses-Pocket, 1977.

2- Ouvrages et articles sur Maupassant et son œuvre

- Amis de Flaubert et Maupassant, *Maupassant 2000*, *Bulletin Flaubert-Maupassant* N∘9, Dieppe, 2001.

- BENAMRHAR, Abdelkrim, " Le miroir dans les romans de Guy de Maupassant ", in *Langues et littératures,* vol. XI, Rabat, 1993, p. 125-138.

- BESNARD-COURSODON, Micheline, *Étude thématique et structurale de l'œuvre de Maupassant : le piège,* Paris, Nizet, 1973.

- BONNEFIS, Phillipe, *Comme Maupassant*, Lille, Presses Universitaires de Lille, 1981.

– BURY, Marianne, *La Poétique de Maupassant*, Paris, SEDES, 1994.

- BURY, Marianne, *Maupassant*, Paris, Nathan, 1991.

- BURY, Marianne, Une Vie *de Guy de Maupassant*, Paris, Gallimard, 1995.

- CLERET, Anne-Marie et Réauté, Brigitte, Bel-Ami *de Maupassant*, Paris, Hachette, 1999.

- DELAISEMENT, Gérard, *Maupassant Journaliste et Chroniqueur*, Paris, Albin Michel, 1956.

– DIZOL, Jean – Marie, *Guy de Maupassant*, Toulouse, Milan, 1997.

- FONYI, Antonia, *Maupassant* 1993, Paris, Kimé, 1993.

- FREBOURG, Olivier, *Maupassant, le clandestin*, Paris, Mercure de France, 2000.

- GIACCHETTI, Claudine, " Les hauts et les bas : la conquête de l'espace dans *Bel-Ami* de Maupassant ", *in : Revue romane*, XXVI, 2, 1991, p. 219-229.

- GIACCHETTI, Claudine, *Maupassant, espaces du roman*, Paris, Droz, 1993.

- MALRIEU, Joël, Bel-Ami *de Guy de Maupassant*, Paris, Gallimard, 2002.

- MORAND, Paul, *Vie de Guy de Maupassant*, Paris, Flammarion, 1942.

- ROCHEFORT-GUILLOUET, Sophie, *Étude sur Maupassant et le roman*, Paris, Ellipses, 1999.

- SALEM, Jean, " Le bestiaire imaginaire de Guy de Maupassant ", *in Maupassant et l'écriture*, Actes du colloque de Fécamp 21-22-23 mai 1993, Paris, Nathan, 1993, pp.129-138.

- SALEM, Jean, *La philosophie de Maupassant*, Paris, ellipse, 2000.

– SANTELLI, Claude, *Mon Ami Maupassant*, Paris, Éditions 1, 1998.

– SATIAT, Nadine, *Maupassant*, Paris, Flammarion, 2003.

- SAVINIO Alberto, *Maupassant et l' "Autre"*, Paris, Gallimard, 1977.

– TASSART, François, *Nouveaux souvenirs intimes sur Guy de Maupassant* (inédit), Paris, Nizet, 1962.

– TOLSTOÏ, Léon, *Guy de Maupassant*, Montpellier, L'ANABASE, 1995.

- TREVOR A. Le V. Harris, *Maupassant et* Fort comme la mort : *Le roman contrefait*, Paris, Nizet, 1991.

- VIAL, André, *Guy de Maupassant et l'art du roman*, Paris, Nizet, 1954.

3- Ouvrages de statistique lexicale

- BENZECRI, Jean-Paul, *L'analyse des données : I La Taxinomie*, Paris, Dunod, 1976.

- BENZECRI, Jean-Paul, *L'analyse des données :2 L'analyse des correspondances,* Paris, DUNOD, 1976.

- BERNET, Charles, *Le Vocabulaire des tragédies de Jean Racine. Analyse statistique*, Genève-Paris, Slatkine-Champion,1983.

- BERNET, Charles, " Faits lexicaux. Richesse du vocabulaire. Résultats ", in *Études sur la richesse et la structure lexicales,* Paris-Genève, Champion-Slatkine, 1988, pp. 1-11.

- BERNARD, Michel, *Introduction aux études littéraires assistées par ordinateur,* Paris, PUF, 1999.

- BERNARD, Michel, " Rêvons un peu... Essai de prospective sur les études littéraires assistées par ordinateur " ? *in* : *Mesures et démesure dans les lettres françaises au XXe siècle*, Hommage à Henri Béhar, Paris, Honoré Champion, 2007, pp. 359-369.

- BRUNET, Étienne, *Le Vocabulaire français de 1789 à nos jours,* Genève-Paris, Slatkine-Champion, vol. I, 1981.

- BRUNET, Étienne, *Le Vocabulaire de Proust I*, *étude quantitative*,

Genève, Slatkine, 1983.

- BRUNET, Étienne, " La structure lexicale dans l'œuvre de Hugo ", *in Études sur la richesse et la structure lexicales,* Paris-Genève, Champion-Slatkine, 1988, pp. 23-42.

- BRUNET, Étienne, *Le Vocabulaire de Victor Hugo*, Genève - Paris, Slatkine - Champion, vol. I, 1988.

- BRUNET, Étienne, *Compte d'auteurs*, Paris, Honoré Champion, 2009.

- ÉVRARD Étienne et Mellet Sylvie " Les Méthodes quantitatives en langues anciennes ", *in* : *Lalies* 18, Actes des sessions linguistique et littérature, Paris, Presses de l'École Nor-

male Supérieure, 1998, pp.111-155.

- GICQUEL, Bernard, *Stylistique littéraire et Informatique*, (publ. par le) Centre d'études et de recherches sur les textes électroniques littéraires, Arras : Artois presses université, 1999.

- HEFIED, Ali, *Statistique linguistique : Aspects stylostatistiques du vocabulaire dans quinze voyages extraordinaires de Jules Verne*, Thèse de doctorat d'État, Fès, 1999.

- JUILLARD, Michel, " Du bon choix d'un corpus et de son bon usage, ", *in : Mots chiffrés et déchiffrés* (Mélanges offerts à Étienne Brunet, Textes rassemblés par Sylvie Mellet et Marcel Vuillaume), Paris, Honoré Champion, 1998, pp. 139-116.

- KASTBERG SJÖBLOM, Margareta, *L'écriture de J.M.G. Le Clézio : Des mots aux thèmes*, Paris, H. Champion. 2006.

- LEBART, L. et Salem, A., *Statistique textuelle,* Paris, Dunod, 1994.

- LENOBLE, Michel, " Statistique lexicale et critique littéraire le mariage impossible ?" *in* : *Méthodes quantitatives et informatique dans l'étude des textes*, Genève-Paris, Slatkine-Champion, 1986, pp. 565-574.

- LUONG, Xuan Nhuam et NOVI, Michel, " Représentations arborées de données textuelles ", in *Méthodes quantitatives et informatique dans l'étude des textes*, Genève-Paris, Slatkine-Champion, 1986, pp. 575-586.

- MAGRI, Véronique, " Stylistique générique et statistique pour une poétique du récit de voyage ", *in* : *JADT' 06*, Volume II, Actes des 8es journées internationales d'analyse statistique des données textuelles, Besançon, Presses Universitaires de Franche-Comté, 2006, pp. 655-666.

- MAGRI, Véronique, *Le Voyage à pas comptés*, Paris, Honoré Champion, 2009.

- MAYAFFRE, Damon, *Le Poids des mots*, Paris, Honoré Champion, 2000.

É. Brunet, *Le Vocabulaire de Victor Hugo*, Paris - Genève, Champion – Slatkine, 1988, vol. I, p. 391.

Mots-Clés: corpus, romans, Maupassant, hypebase, lexicométrie