

Christian MILAT

L'écriture leclézienne auscultée par ordinateur

L'originalité de l'ouvrage de Margareta Kastberg Sjöblom tient à ce qu'il se propose d'étudier à la fois l'ensemble de l'œuvre de Le Clézio et l'évolution de celle-ci non pas en fonction de tel ou tel thème plus ou moins préétabli, mais à partir de son vocabulaire et de sa langue. Pour ce faire, l'auteure utilise les méthodes lexicométriques de la statistique lexicale traditionnelle, telles que Muller et Guiraud les ont vulgarisées. Elle y a cependant ajouté le recours à des logiciels comme *Hyperbase* et *Cordial* qui, en permettant « non seulement d'analyser la surface matérielle ou graphique du texte, mais encore de traiter le texte lemmatisé, ses codes grammaticaux, ses combinaisons et enchaînements syntaxiques, voire ses isotopies sémantiques » (p. 36), font qu'à la lexicométrie de la première génération succède désormais la logométrie, laquelle rend possible un traitement complet du discours.

Deux corpus ont été pris en compte. Le premier comprend la quasi-totalité des textes parus de 1963 à 1999 — quatre ouvrages ayant été écartés en raison des particularités typographiques ou langagières qu'ils présentent —, soit 31 textes qui, appartenant à des genres divers (romans, nouvelles, récits poétiques, essais littéraires, biographies, ouvrages ethnologiques, récits de voyage, biographies, livres pour les enfants) et présentant une taille très variable, comportent 51 009 vocables apparaissant dans 2 281 659 occurrences. Le second corpus ne retient que les œuvres de fiction : regroupant 17 livres de taille sensiblement égale, il compte 41 317 vocables et 1 717 096 occurrences.

L'étude aborde tout d'abord la structure du vocabulaire leclézien. On y apprend que les hapax se trouvent en plus grand nombre dans les textes où la syntaxe est moins recherchée, à savoir les ouvrages ethnologiques, les biographies ainsi que les essais littéraires et, s'agissant des œuvres de fiction, les textes expérimentaux, proches du

Nouveau Roman, publiés de 1963 jusqu'au milieu des années soixante-dix. L'étendue et la diversité du vocabulaire suivent une tendance analogue, avec néanmoins une propension à une richesse accrue et à une moindre répétition des mêmes mots dans les textes publiés après 1986. La même tripartition de l'œuvre peut en outre être repérée sur le plan de l'accroissement lexical, lequel va de pair avec le renouvellement thématique : présent dans la première période, celui-ci cesse pendant la deuxième, mais réapparaît au cours de la troisième.

La phrase leclézienne fait l'objet de la section suivante. En ce qui concerne la répartition entre les mots courts (de une à quatre lettres) et les mots longs (de plus de huit lettres), l'analyse lexicométrique fait observer une même ventilation et une même évolution : les mots longs se trouvent surtout dans les ouvrages non fictionnels et les premiers livres du corpus romanesque, mais ils s'effacent ensuite pour voir de nouveau leur nombre s'accroître dans les derniers textes de fiction. Quant aux mots de longueur moyenne (de cinq à huit lettres), alors qu'ils sont peu nombreux dans les textes de la première période, ils enregistrent par la suite une progression constante. L'étude examine ensuite les signes de ponctuation forte. Si la fréquence du point connaît des variations successives qui empêchent d'en tirer des conclusions vraiment intéressantes, celle des points de suspension, en revanche, est tout à fait significative : leur disparition, après le bref usage qui en est fait dans les seuls trois premiers livres de *Le Clézio*, manifeste la volonté de l'écrivain de ne laisser aucune place au non-dit. De même, le fait que le point d'interrogation est plus utilisé que le point d'exclamation — les deux étant par ailleurs moins nombreux dans les livres non fictionnels que dans les textes de fiction — est riche de sens : « l'auteur, à travers ses différents personnages, semble surtout [se] poser des questions et chercher des réponses, plutôt qu'exprimer des affects » (p. 100). Le recensement de ces quatre signes de ponctuation permet de déterminer la longueur moyenne de la phrase leclézienne, qui est de 21,09 mots, ce qui, selon des études entreprises sur d'autres corpus, inscrit celle-ci dans la moyenne du mouvement littéraire français de la seconde moitié du XX^e siècle. Cependant, cette valeur moyenne générale recouvre des moyennes qui, d'un livre à l'autre, varient de 15,81 à 34,59 mots, les phrases les plus

longues se trouvant dans les ouvrages ethnologiques — le discours descriptif y dominant —, tandis que les phrases les plus courtes figurent dans les dialogues des textes de fiction. L'analyse de la segmentation interne de la phrase s'effectue, elle, au moyen du repérage des signes de ponctuation faible. De ceux-ci, c'est la virgule qui est la plus fréquemment utilisée, son usage devenant néanmoins plus rare après le milieu des années soixante-dix : alors que les premières œuvres romanesques multiplient les énumérations, les juxtapositions, les accumulations de termes, les textes qui suivent marquent la volonté de Le Clézio d'abandonner « la phrase compliquée comprenant plusieurs parties ou thèmes, dans sa recherche d'un langage simple et pur » (p. 111). S'agissant du point-virgule, l'évolution est encore plus accentuée : surtout employé dans les tout premiers livres, il tend bientôt à disparaître. Le deux-points, lui, est surtout fréquent, hormis dans les romans de la première période, dans les essais littéraires et les ouvrages ethnologiques, riches en énumérations et en explications, ces deux types de livres contenant en outre la quasi-totalité des parenthèses. Compte tenu de ces différents signes de ponctuation faible, la longueur moyenne des séquences intérieures de la phrase leclézienne est de 7,40 mots. Quant à ses facteurs de complexité, ils sont évalués au travers de l'inventaire des marqueurs de subordination et de coordination. Particulièrement nombreux dans les ouvrages riches en dialogues et plus rares dans les livres à discours narratif dominant, les subordinants évoluent selon trois périodes : fréquents dans *Le Procès-verbal*, premier roman, ils sont pratiquement abandonnés par la suite avant d'être réactivés au cours de la deuxième période, tandis qu'ils deviennent épisodiques durant la troisième période. Le nombre encore plus important des coordonnants manifeste que la longueur des phrases leclésiennes est « une longueur par “empilement” et non par “imbrication” » (p. 124). Cela dit, une fois encore, la distribution des coordonnants connaît une évolution chronologique : nombreux dans les premiers romans, ils se raréfient à partir du milieu des années quatre-vingt.

Consacrée aux parties du discours, une autre section s'intéresse à la structure syntaxique de l'écriture leclézienne. Catégorie grammaticale la plus nombreuse (20,5 % du corpus), le substantif est plus fréquent

dans les ouvrages non fictionnels, surtout à partir de 1980, au moment où Le Clézio traite du monde et de la culture des Amérindiens. Dans les romans, la fréquence des substantifs, faible dans les premiers textes, s'accroît au fur et à mesure que l'œuvre progresse — la description ayant tendance à l'emporter sur l'action —, sauf dans les récits où les dialogues sont nombreux. Parmi ces substantifs, Le Clézio, à l'opposé de nombreux autres romanciers, emploie plus de mots masculins (56,5 %) que de mots féminins (43,5 %). C'est que, selon l'auteur, l'écrivain manifeste là sa volonté « de rester dans le concret et de garder un langage simple et "terre à terre" » (p. 144-145). S'agissant du nombre, si, dans les ouvrages ethnologiques, le pluriel prime sur le singulier, la situation est en revanche l'inverse dans les romans, lesquels expriment souvent la solitude d'un individu face à la nature. Quant aux adjectifs, très fréquents dans les livres non fictionnels, ils le sont tout autant — description oblige — dans les textes proches du Nouveau Roman, mais à compter du recueil de nouvelles *Mondo* (1978), leur nombre se réduit considérablement. Par ailleurs, l'analyse de la distribution des pronoms personnels permet de dégager, globalement, une domination du récit à la troisième personne. L'indice pronominal (quotient entre les première et deuxième personnes des pronoms personnels et les mêmes personnes des possessifs) permet, lui, de relever la présence d'un style soutenu ou élevé dans les essais littéraires, les ouvrages ethnologiques et biographiques, à la facture poétique très marquée, et celle d'un style plus familier dans les œuvres romanesques des première et troisième périodes. Enfin, les occurrences verbales étant principalement reliées aux séquences caractérisées par l'action, les verbes sont moins nombreux dans les essais et les ouvrages d'ethnologie que dans les textes de fiction — romans et nouvelles —, et l'évolution de leur nombre suit inversement celle du nombre des substantifs. Dans les premiers romans, ceux proches du Nouveau Roman, les verbes sont fréquemment conjugués au subjonctif, au conditionnel ou à l'impératif alors que, dans les textes suivants, plus conventionnels, l'indicatif, le mode de la réalité, domine. Importants sont également les effectifs de l'infinitif : « Il semble », fait remarquer Sjöblom, « que ce mode impersonnel, "distant" et qui permet la création d'un espace intemporel, convienne parfaitement à l'écriture leclézienne. » (p. 179)

À noter l'usage des participes qui se révèle, au fil des textes, de plus en plus fréquent, en particulier celui des participes passés qui, dans la seconde moitié de l'œuvre, sont très souvent associés au recours répété au passé composé. Moins littéraire, celui-ci remplace de plus en plus le passé simple, principalement utilisé dans les romans considérés par l'auteure, comme par bien d'autres critiques, comme proches du Nouveau Roman, ce qui ne représente pas un mince paradoxe quand on sait que les Nouveaux Romanciers, eux, ont fortement critiqué l'emploi du passé simple... À ce temps se substitue également un présent intemporel qui contribue notamment à donner à l'œuvre une dimension à la fois onirique et mythique. La fréquence de l'imparfait est tout aussi significative : suggérant une durée indéterminée, celui-ci devient chez Le Clézio le temps de la « nostalgie » (p. 189).

La dernière section de l'étude a trait au contenu du discours qui implique la signification des mots et des différentes catégories lexicales. En d'autres termes, il s'agit ici de dégager les principales thématiques ou isotopies récurrentes dans l'œuvre leclézienne et, pour ce faire, d'y extraire les mots particulièrement suremployés ou sous-employés par rapport aux 86 millions d'occurrences du corpus *Frantext* de 1900-1990. Parmi les mots les plus suremployés figurent des substantifs qui ont trait à la nature (*mer, lumière, vent, plage, ciel, soleil, terre, montagne, eau, nuages*), des verbes qui se rapportent au regard (*briller, brûler, regarder, voir*) et au mouvement (*bouger, marcher, glisser, traverser, aller, monter, retourner*) et des adjectifs de couleur (*noir, bleu, blanc, gris, rouge*). Parmi les mots les plus sous-employés, on relève les pronoms personnels des première et deuxième personnes du singulier et du pluriel (les dialogues sont peu nombreux), les pronoms possessifs (Le Clézio rejette la possession matérielle) et le verbe *faire* (la pensée et la sagesse l'emportent sur l'action désordonnée). Diachroniquement, il est de nouveau possible de distinguer des évolutions : certaines thématiques (la ville, le corps, les insectes, le minuscule), très présentes au début de l'œuvre, perdent progressivement de l'importance.

Globalement, l'auteure produit là une étude tout à fait pertinente, manipulant les concepts et les méthodologies avec la précision ainsi

que la prudence qui s'imposent et apportant à ses analyses les nuances indispensables. Néanmoins, ces qualités ne vont pas sans certaines faiblesses. Ainsi, Sjöblom affirme que Le Clézio « passe les vingt premières années de sa vie à Nice » (p. 17) alors que l'écrivain, âgé de sept ans, a séjourné pendant une année au Nigéria, cette première rencontre, décisive, avec son père et le continent noir ayant généré à la fois un roman, *Onitscha* (1991), et un texte (auto)biographique, *L'Africain* (2004). Pareillement, *Voyage à Rodrigues* semble abusivement qualifié de « roman » (p. 56 et *passim*) puisque ce texte constitue en fait la « version (auto)biographique » du roman *Le Chercheur d'or*. Enfin, il est à maintes reprises question de « l'école du "nouveau roman" » (p. 18 et *passim*). Or, le Nouveau Roman n'est rien moins qu'une école, ainsi qu'il est timidement formulé à la seule page 263. À ces erreurs de détail s'ajoutent quelques coquilles (par exemple, *vocable* est employé, aux pages 41 et 202, comme un mot féminin) ou défauts typographiques.

Néanmoins, ces réserves ne doivent pas faire oublier l'intérêt d'un ouvrage où le recours à l'informatique génère une description rigoureuse et objective d'un très large corpus, dont il permet de saisir l'écriture à la fois dans ses variations génériques et ses évolutions temporelles. Confirmant de façon formelle et impartiale une grande part des conclusions présentées par des études antérieures, d'inspiration parfois plus stéréotypée, voire intuitive, il apporte également des éclaircissements inédits à une œuvre sur laquelle, année après année, Le Clézio continue de greffer de nouveaux textes.

Référence : Margareta Kastberg Sjöblom, *L'Écriture de J.M.G. Le Clézio. Des mots aux thèmes*, Paris, Honoré Champion, coll. « Lettres numériques », 297 p.