

Immanence littéraire et thématique

Aujourd'hui où l'attention sur l'œuvre littéraire en elle-même demeure encore détournée par la recherche de ses déterminations extérieures, notamment psychologiques ¹, il n'est pas malvenu de poser de nouveau le problème du discours "critique qui ne s'intéresse qu'à l'*immanence de l'œuvre* – c'est-à-dire à l'œuvre elle-même, débarrassée de toutes considérations externes [...], des circonstances, historiques ou personnelles", comme l'écrivait Genette (2001, p. 131). Nous le citerons de façon détaillée, car son propos qui se centre sur une critique thématique aujourd'hui distincte de celle de J.-P. Richard, du fait des avancées de la sémantique linguistique, concerne aussi bien l'approche génétique du thème, que l'assistance lexicométrique requise par les études thématiques.

Se référant au Barthes de 1954 (*Michelet*) et de 1963 (*Essais critiques*), Genette énonce un clivage des plus nets et des plus reçus, en opposant "la démarche *transcendante* du positivisme explicatif et celle de l'*analyse immanente*, purement descriptive, en quête d'une cohérence ou d'une unité qui ne se soucie d'aucun *ailleurs* extérieur au texte" (pp. 136-7).

Soit d'une part, la critique positiviste, causaliste, "explicative par la recherche des causes extérieures" (p. 132), déterministe, puisqu'il s'agit "d'expliquer l'œuvre par un ailleurs extérieur à elle" (p. 135) ², comme le fait de façon emblématique la critique biographique, recourant à l'*alibi* de type idéologique, psychologique ou sociologique. Ou plus largement la critique universitaire, d'un Sainte-Beuve, Taine, Lanson, mais aussi freudo-marxiste d'un Mauron ou d'un Goldmann, qui "recourt à l'*ailleurs* de l'inconscient ou de l'appartenance sociale" (p. 137), jusqu'à la poétique qui "vise à travers les œuvres singulières des *essences* génériques délibérément trans-opérales" (p. 149) – telle l'énigmatique Littéarité.

Et d'autre part, la critique immanente, organiciste, internaliste, phénoménologique, "compréhensive, par la description, au plus l'interprétation, des seuls traits internes et de leurs rapports réciproques" (p. 132). Il s'agit de la "Nouvelle Critique" de Richard, Spitzer, Poulet, Starobinski, Barthes – et déjà Proust. Parce qu'elle est structurale, elle implique clôture et cherche ce qui est d'ordre invariant (p. 141) ³, à travers le repérage de constantes thématiques-

¹ Au nom de la maxime *le style c'est l'homme*, le texte est oublié au profit de l'auteur.

² Laquelle, bizarrement, n'est pas appelée critique transcendante, comme elle le devrait (p. 131).

³ Citant Proust, Genette rappelle pourquoi Poulet avait raison de le considérer comme le "fondateur de la critique thématique" (p. 142) : "la première tâche de la critique, qui se

psychologiques d'un "réseau organisé d'obsessions" (p. 133)⁴; "le thème est un réseau de relations, et donc une structure" (p. 146) – "sans doute sous l'influence de la pensée linguistique, la multiplicité des contextes organise dans l'œuvre une multiplicité de rapports" (p. 146).

Le propos n'est certes pas nouveau. Trente cinq ans auparavant, Genette opposait "le moyen de reconstituer l'unité d'une œuvre, son principe de cohérence, ce que Spitzer appelait son *etymon* spirituel" – les œuvres étant vues comme des *organismes* – au "danger d'effritement qui menace l'analyse thématique", venu du positivisme qui recherche "des déterminations extérieures" en ces termes : "la méthode structurale devrait intervenir pour donner à cette étude immanente une sorte de rationalité de compréhension qui remplacerait la rationalité d'explication abandonnée avec la recherche des causes. [...] L'analyse thématique tendrait donc spontanément à s'achever et à s'éprouver en une synthèse structurale où les différents thèmes se groupent en réseaux, pour tirer leur plein sens de leur place et de leur fonction dans le système de l'œuvre : c'est le dessein nettement formulé par Richard ou Rousset" (1966, pp. 156-7). Soit une manière de réconcilier "critique intersubjective" (à caractère thématique) et "objectivisme" (à caractère structural) – sans pour autant avoir, comme cela est le cas aujourd'hui, une théorie du sens qui articule ces deux principes.

Discussion

La conclusion, paradoxale, de l'article de Genette est que pour opérer les nécessaires "vérifications structurales" des sens contextuels d'un mot vedette dont parlait Richard (*étoile* chez Mallarmé, p. 145), "ce mouvement de navette, même lorsqu'il travaille dans une immanence, y établit une transcendance interne" (p. 146). Or ces "rapprochements toujours hasardeux", "en sautant ou en glissant d'un prédicat thématique à un autre" – car "il n'est de thème que prédiqué" (p. 148) –, témoignent d'une "désinvolture à l'égard du cheminement mot à mot d'un texte", laquelle "distingue clairement la critique thématique de l'explication de texte traditionnelle", Lanson étant alors "le plus sévère gardien d'une critique presque vraiment immanente". Presque, car la fin de l'article avoue "qu'aucune critique, et sans doute aucune lecture, et en tout cas aucune

révèle peut-être la seule, consiste pour l'essentiel à dégager les traits permanents d'une œuvre, en notant leurs récurrences et leurs échos fraternels qui retentissent d'une occurrence à l'autre" (p. 143). Il cite aussi le Richard de 1967 écrivant : "L'œuvre ne saurait trouver hors d'elle-même ni les prémisses de son sens, ni le principe de son ordre. La critique sera donc une activité pleinement immanente à l'œuvre" (p. 136).

⁴ Sans être une "pré-critique", elle précède une "critique véritable à visée explicative [...] cette visée sera en quelque sorte déléguée à des disciplines extra-littéraires", comme celles des historiens ou des psychanalystes (pp. 133-4).

explication de texte ne peut être vraiment immanente" (p. 146). L'argument est que celle-ci, de par la relation *intratextuelle* qui lui est consubstantielle, révèle l'effort mnémonique du lecteur qui établit une cohérence interne dans le texte expliqué.

Or, en relativisant ainsi l'opposition immanence *vs* transcendance, et en voulant apaiser la controverse ("il n'y a plus là motif à bien vives querelles : la guerre est finie, peut-être", derniers mots de l'article, p. 149), Genette ne parvient pas à faire oublier combien le *semiotic turn* des années soixante était justifié, quand Barthes notamment récusait "une critique fondée tout entière sur le rapport d'extériorité", déterministe, positiviste, que cet ailleurs soit la vie de l'auteur, les forces socio-historiques s'exerçant sur sa production littéraire, ou les *essences* génériques de la poétique. Même quand le thuriféraire du sens "pluriel", corollaire de l'inachèvement, optera pour un déconstructionnisme prenant le contre-pied des certitudes structurales et scientifiques, reléguées au rang de mythe durant les années 70, cela ne l'empêchera pas de plaider pour l'immanence, via le concept de *connotation*. Ainsi l'étincelante analyse de contenu d'une nouvelle balzacienne que fut *S/Z* concevait la nomination d'un thème comme un acte de dispersion : "lire, c'est trouver des sens, et trouver des sens, c'est les nommer ; mais ces sens nommés sont emportés vers d'autres noms [...] je nomme, je dénomme, je renomme : ainsi passe le texte" (1970, p. 17). Soit un plaidoyer pour le fuyant lexical, qui aboutissait à la conception d'une "thématique infinie, proie d'une nomination sans fin", laquelle se trouve, de fait, perfectivée par "la fatalité d'un coup de dés qui arrête et fixe le glissement des noms : c'est la thématique" (*ibid.* p. 100). Mais au-delà de la paire traditionnelle objectivité *vs* subjectivité, lesquelles "sont toutes deux des imaginaires" (*ibid.* p. 17), c'est la "systématique" du sens textuel qui doit s'imposer à la lecture, par cet outil nécessaire qu'est la connotation ; cela, à condition de "ne pas la confondre avec l'association d'idées : celle-ci renvoie au système d'un sujet ; celle-là est une corrélation immanente au(x) texte(s) ; ou encore, si l'on veut, c'est une association opérée par le texte-sujet à l'intérieur de son propre système" (*ibid.* pp. 14-15).

Revenons à Genette. La transcendance qu'il revendique ressortit au vocabulaire philosophique. Il objecte ainsi que "prédiquer (nommer) un objet, c'est toujours l'assigner à une classe [...] la relation au singulier comme tel est toujours indicible, elle ne se dit qu'en se généralisant" ("toute critique, littéraire ou artistique, et plus généralement encore toute relation esthétique en tant qu'elle *identifie* ses objets, c'est-à-dire qu'elle les perçoit, les distingue, les nomme et les définit comme tels, inévitablement les conceptualise et par là même les transcende au profit de catégories plus générales – je dirais volontiers plus

génériques"), et que donc "toute relation consciente à cet objet transcende son immanence" (p. 148), par abstraction, en dégagant le prototype de cet objet. En concluant ainsi à la "transcendance du signifié" (p. 139), Genette ignore délibérément l'acte de sémantique textuelle qui identifie des signifiés particuliers, au terme de parcours interprétatifs. Or selon cette théorie linguistique, le sens peut être dit, avec F. Rastier, transcendant au texte, mais immanent à une situation d'énonciation (1989, p. 16⁵), et à "une pratique d'interprétation" (2001, p. 277).

Quand Genette concluait *Figures I* sur l'idée de Valéry selon laquelle la littérature n'est que l'exploitation de certaines propriétés du langage (1966, p. 265), après avoir déclaré dans son sillage : "La littérature étant d'abord œuvre de langage, et le structuralisme, de son côté, étant par excellence une méthode linguistique, la rencontre la plus probable devait évidemment se faire sur le matériel linguistique [...]" (*ibid.*, p. 149), il est pour le moins paradoxal que Genette ne reconnaisse pas cette autonomie des signifiés linguistiques, dans le texte, lui qui ne pouvait pourtant ignorer Greimas (1966) et sa théorie de l'isotopie.

En effet, bien qu'il définisse l'adjectif "*thématique* comme équivalent de *qui porte sur le contenu*" (p. 141), Genette ne dissocie pas cette étude des "signifiés, ou contenus" (p. 140) du point de vue phénoménologique et psychologique, fût-il structural. La métaphore musicale qu'il reprend du couple thème/variante ne suffit pas elle non plus pour appréhender "la variance des occurrences dispersées" (p. 141). A lui comme à Richard pour Mallarmé, il manque une théorie contextuelle du sens linguistique.

I. Thématique et génétique

Si l'on prend le cas particulier du texte *in statu nascendi*, comme disent les généticiens littéraires, faut-il voir une antinomie entre l'immanence du texte définitif et la transcendance du Sujet écrivain (auteur, scripteur) qui fait recourir aux brouillons antérieurs ? A. Compagnon (1992, pp. 59-60) n'hésitait pas à opposer la critique génétique à "une critique de type philosophique, ou encore à une critique narratologique ou structuraliste, pour laquelle le texte est le texte, à expliquer dans son immanence, indépendamment du contexte et de l'histoire. [...] Toutes ces variantes, ces transformations mettent en cause l'idée reçue de la

⁵ Cf. *ibid.* p. 15 son rejet de l'unicité vs l'infinité, décrétées *a priori* : "Ainsi, postuler un sens fondamental immanent ne limite pas plus la diversité des lectures *structurales*, que postuler un sens pluriel, en fait transcendant, n'évite les ressassements d'une psychanalyse vieillissante."

consistance théorique et structurale, de la cohérence philosophique et narratologique de la *Recherche*, construite comme un tout, un système dont les exposés les plus fameux auraient été donnés par G. Deleuze dans *Proust et les Signes* et par G. Genette dans *Discours du récit*." De façon convergente, une généticienne spécialiste de Proust, E. Marantz (1989 p. 53) reprochait cet esprit de système philosophique : "A. de Lattre prétend que la singularité de celui-ci s'explique par la prise de position métaphysique de Proust, par son ontologie de l'absence et de la séparation. Selon A. de Lattre, la coupure métaphysique et la disjonction psychologique qui en résulte détermine, chez Proust, non seulement toute perception d'un personnage, mais sa fabrication. Or, on n'a qu'à parcourir les avant-textes de la *Recherche* pour découvrir que la vérité est tout autre et que le personnage proustien se constitue par le savant montage de traits épars, disséminés dans tout l'ensemble des brouillons, traits, qui, fort souvent, avaient été attribués au préalable à de tout autres personnages."

Si ce reproche vaut pour un système conçu comme intentionnel par l'auteur, nous avons démontré ⁶ qu'une approche structurale d'un corpus d'auteur n'est nullement incompatible avec la critique génétique. Sur un état du texte final, désormais "modifié" par la confrontation avec le texte initial dans un avant-texte ayant le "statut scientifique", qui contredit cet inachevé apparent, comme l'affirme De Biasi (1987) : "un *avant-texte* n'existe nulle part hors du geste théorique qui le constitue : c'est le dossier de rédaction rendu intégralement lisible et reconstruit comme un système cohérent de transformations successives", donc dans un état de clôture temporaire, un réseau thématique peut tout à fait être cerné, et décrit selon une analyse sémique, sans que cela ne remette en cause la "consistance" du propos dont parle Compagnon.

Quant à l'immanence, pourquoi devrait-elle être indépendante "du contexte et de l'histoire" ? La sémantique intègre des disciplines voisines, comme le pose Rastier : "pour décrire la richesse des relations contextuelles, la linguistique ne peut en rester à l'espace douillet mais confiné de la phrase; elle s'ouvre aux textes, et par là aux cultures et à l'histoire, en réaffirmant son statut de science sociale" (1989, p. 7). L'intégration a cependant des limites, et refuse celle du Sujet : ce n'est pas parce qu'on prête attention aux premiers-jets scripturaux qu'on serait plus proche des intentions, du vécu que l'auteur aurait voulu transmettre dans le texte. Le degré de proximité entre intention, vécu et premiers-jets est d'ailleurs très difficilement mesurable (cf. Rastier, 1997, p. 196 : "Un dossier génétique se prête à deux usages principaux : soit l'on utilise les brouillons pour étayer par ces documents des conjectures sur la version finale;

⁶ Dans notre thèse sur Proust (Paris IV Sorbonne, 1993, refondue dans *Le thème de la blondeur*, L'Harmattan, 2003).

soit, avec plus d'ambition, l'on tente d'en surprendre la genèse – bien que les brouillons ne soient sans doute pas plus proche que l'état final d'un hypothétique moment pré-linguistique de l'énonciation." Rastier y étudie "les opérations de réécriture" successives "sans postuler une intentionalité de l'auteur" p. 195 ⁷).

L. Jenny le dit bien (2002, p. 56) : "Le brouillon n'est pas la vérité profonde du texte achevé [...] Car on ne peut logiquement fonder le sens d'un texte sur les refus à l'aide desquels il s'est constitué." Une telle assertion va d'ailleurs à l'encontre des théories conjointement cognitive et "de la cyberculture" de J.-L. Lebrave, pour qui il y aurait *isomorphisme* entre la structure du brouillon et celle de l'hypertexte" qui le retranscrit; et "fidélité du brouillon et de l'hypertexte aux processus mentaux" des scripteurs; si bien que "plus *naturel*, le brouillon est ainsi implicitement présenté comme plus *vrai* que le texte" (*ibid.* pp. 59, 61). Deux théories conjointes qui n'ont que l'apparence de la scientificité, conclut Jenny : "Plutôt qu'une science, au prestige incertain malgré son appareil technologique, il me semble que cette branche de la génétique [...] ne vise pas à la constitution d'un sens mais à la reviviscence d'un trajet mental, simulé sur écran par des parcours hypertextuels." Or pareille posture "étonnamment antiherméneutique" – puisque "née de l'évolution de disciplines du sens comme la philologie, la critique et la poétique" (*ibid.* pp. 64-5) – laisse toute sa place à une génétique fondée sur la sémantique du texte, dont l'effort d'objectivation du sens demeure sinon scientifique du moins rationnel.

Et comme le posait R. Debray-Genette, non seulement il y aurait compatibilité entre la clôture de la poétique narrative et l'arbitraire, le hasard, les variations qui hantent la génétique textuelle, mais le texte définitif comporte de façon hétérogène des passages conservés depuis le premier jet et d'autres retravaillés (1977, p. 21). Bref, cela incline à conclure qu'il est délicat et peu intéressant de savoir quel écrit (extrait ou unité textuelle) est le plus proche de la source énonciative. Seule la concomitance des théories de l'énonciation et de la pragmatique avec l'essor des travaux de génétique depuis les années 80 semble justifier la nouvelle alliance de ce secteur de la linguistique avec cette approche de la genèse littéraire. Nous prétendons que l'alliance peut *aussi* se fonder sur une sémantique qui diffère de tels modèles déterministes.

Si, comme le prétend Compagnon, l'inachèvement textuel ⁸, voire la déconstruction, "accueille la contingence et l'indéterminisme" (1992, p. 61), tout

⁷ L'étude fut précédée par une première publication sur Flaubert (1992).

⁸ Faux postulat de la critique génétique, puisque, comme le précise Rastier : "la seule originalité d'une série génétique – du premier brouillon au texte final – est d'être close, ou du moins finie, et d'émaner du même scripteur" (1997, p. 194).

déterminisme extra-textuel doit être combattu. Non seulement le déterminisme psychologique de l'auteur, mais aussi celui de la référence au monde. Ainsi, à propos de la genèse d'*Hérodias*, Rastier conclut : "La croyance en un réalisme de Flaubert, invétérée malgré ses protestations, s'appuie certes sur le fait qu'il constituait des dossiers de notes admirés de Zola. Mais, outre que le processus génétique consiste à les négliger ou à les détruire, ces notes restent d'autres textes, qui ne sont pas plus près d'une réalité non-sémiotique que le texte final" (1997, p. 214). Cela revient à réaffirmer, selon un principe structuraliste que la critique transcendante et positiviste voudrait invalider et rendre dépassé, qu'il convient de récuser la fusion du sens textuel avec les deux réalités supposées en amont du texte produit : le monde réel, concret ou mythique, et la psychologie du producteur du texte.

L'analyse sémique à la refondation de la génétique et de la thématique

Bien avant le "code sémique" de *S/Z*, appliqué à Balzac, Barthes déjà, à propos des noms propres chez Proust (1967), "accumule en eux des éléments sémiques nombreux, dénotatifs mais surtout connotatifs", tel notamment ce sème /aristocratie/ afférent aux noms propres terminés "en longues finales muettes", dont *Guermantes* est l'emblème, comme le rappelle J. Milly, qui commence ainsi son article, avant de le poursuivre sur "le rapport cratyléen d'imitation" qui unit de tels signifiés avec des signifiants sonores, phonèmes et syllabes (1974, p. 65). La visée n'y est pas explicitement génétique, même s'il est question de Proust se pastichant lui-même.

C'est au cours d'une table ronde, à laquelle Barthes participe, que Genette pose le principe génétique du "caractère évolutif de l'œuvre de Proust, [...] de cette succession ininterrompue d'états divers d'un même texte, depuis *Les Plaisirs et les Jours* jusqu'au *Temps retrouvé*", afin de "passer à une herméneutique nouvelle, qui serait syntagmatique" : "je veux dire qu'il ne suffit plus avec Proust de noter des récurrences de motifs et d'établir à partir de ces répétitions, par empilement et homologation, des objets thématiques dont on dressera ensuite le réseau idéal, selon une méthode dont Ch. Mauron a donné la version la plus explicite, mais qui est au fond de toute critique thématique. Il faut aussi tenir compte des effets de distance ou de proximité, bref de *place* dans le texte, entre les divers éléments de contenu" (1975, pp. 91-2). Si, lors de cette table ronde, J.-P. Richard adopte cette "thématique contextuelle" (incluant les avant-textes), il lui adjoint aussitôt le principe interprétatif de la migration sémique, de type déconstructionniste dans le sillage de *S/Z* : "ce qui m'apparaît comme proprement thématique, dans le donjon de Roussainville, c'est la possibilité qu'il nous offre de l'ouvrir, presque de l'éclater, d'opérer en tout cas une libération disséminante de ses différents traits constitutifs (qualités ou fonctions), de le

dissocier en somme, pour le relier à d'autres objets présents et actifs dans l'étendue de la fiction proustienne." Telle "la rousseur (suggérée par le signifiant *Roussainville*) qui renvoie le donjon à la libido de toutes les petites filles rousses. [...] La modulation du thème peut même apparaître ici comme très authentiquement freudienne" (1975, p. 95). En effet, dispersion sémique et transcendance psychanalytique se conjuguent dans son ouvrage sur Proust (1974), en ces années 70 où le post-structuralisme tend à dépasser la linguistique (le niveau d'analyse sémique) par l'alibi d'une interprétation avant-gardiste venue d'une autre science humaine.

C'est d'ailleurs à ce même ouvrage que se réfère R. Debray-Genette pour étudier la genèse d'un motif, depuis les "embryons romanesques dans *Les Plaisirs et les Jours*" jusqu'à la fragmentation contextuelle de ce motif dans *Combray*. Elle y observe, après les travaux de Ph. Hamon, que leur description procède par "un épuisement de sèmes" (1976, p. 137). Elle reproche toutefois à la méthode de Richard de "dégager les thèmes fondamentaux" à partir du corpus entier d'un auteur, sans moduler sur telle ou telle "œuvre particulière", ni tenir compte des ébauches scripturales, précisément en raison du postulat de l'unité du Sujet et de sa production textuelle (1977, p. 32) : "Si la thématique d'un écrivain est consubstantielle à son être d'écrivain et d'homme, les éléments qui la spécifient devraient garder une sorte de constance depuis la première feuille de brouillon jusqu'au texte final. Or c'est ce que contredisent souvent les brouillons." Soit une contradiction entre les variations particulières, immanentes à l'écriture, et la transcendance d'un principe général, ontologique, qu'elle recèle.

C'est bien ce dernier qui se trouve nié par une spécialiste proustienne à qui Debray-Genette rend hommage : Cl. Quémard. Cette généticienne procède, comme la thématique richardienne, par comparaison des contextes, fût-ce pour les unifier⁹. Elle collationne les différentes versions d'un motif, par exemple les "rêveries onomastiques proustiennes" (1979), ayant pour objet les toponymes normands, Bayeux, Coutances et autre Lamballe. Toutefois, dans un respect du sens contextuel, l'onirisme du narrateur n'y est plus rapporté à la loi libidinale. Celle-ci était formulée on ne peut plus clairement dès la première page de *Proust et le monde sensible* : "On décrirait chaque *moindre désir* afin de dégager à travers lui, en lui, les quelques grandes figures, sensibles ou libidinales, qui en organisent de manière spécifique l'émergence. On en viendrait ainsi à dessiner les directions significatives d'une présence au monde" (1974, p. 7). En dépit du fait que Richard s'appuie sur une proposition de lecture du narrateur de la *Recherche* lui-même, faut-il pour autant réduire voire identifier les signifiés du

⁹ Par cette *isonomie* qui nie leurs "inégalités qualitatives", et qui est un corollaire de l'organicisme structural, dont parle aujourd'hui Rastier (2001, p. 44).

roman à ces deux types de "figures" ? Poser la question, c'est y répondre, par la négative. On ne voit pas pourquoi les sèmes matérialistes qui organisent son exposé ("le velouté, le marbré, le naissant, l'écumeux, etc.") devraient se ramener au Désir dont ils seraient *a priori* l'expression.

Même M. Collot, quand il refuse la "relation d'extériorité" du thème au profit de son étude internaliste, se coupe d'une origine désirante : "Le thème selon la thématique, c'est plutôt l'ensemble des significations qu'une œuvre prête à ces référents ou à ses références : il s'agit moins d'un objet extérieur à l'œuvre que d'une catégorie sémantique qui lui est propre. Par exemple, le thème de l'*aéré* étudié chez Proust par J.-P. Richard ne se confond ni avec l'air que nous respirons ni avec ce que nous en lisons par ailleurs : il est fait des connotations spécifiques dont se charge dans la *Recherche* cette qualité élémentaire. Ce signifié est donc idiomatique, individuel, ce qui le distingue d'une conception répandue du thème comme *topos*, lieu commun légué par la tradition littéraire." L'*aéré* n'étant qu'une des "modulations" du signifié ainsi abordé dans ses "rapports réciproques" avec d'autres traits matérialistes (1988, pp. 81-2).

Par ailleurs, pour donner une légitimité à ces derniers, Collot qui cite Greimas – sans doute parce que tous deux partagent la même base d'une phénoménologie de la perception –, lui reprend sa terminologie avec ses deux types de composants : d'une part les *sèmes nucléaires* qui traduisent des "qualités sensibles" qui seraient autant de "propriétés objectives des choses" ; d'autre part les *sèmes contextuels* : "Le travail proprement critique de la thématique commence en effet à mes yeux à partir du moment où l'on définit quelles sont, parmi les virtualités sémantiques du thème, celles qui sont effectivement actualisées dans une œuvre", ce qui implique "la *mise en texte* du thème", d'après Collot (1988, pp. 84-5). Cela renvoie en effet à cette page de *Sémantique structurale* qui définissait le contenu du signe minimal (ou sémème) "comme la combinaison du noyau sémique et des sèmes contextuels" (Larousse, 1966, p. 50). Or sans entrer ici dans la discussion, il s'avère que Rastier a montré (1987, pp. 74-5, 117-9, 172-3) que ce dualisme millénaire des *sèmes figuratifs* (ou extéroceptifs, correspondant aux qualités sensibles du monde) *vs sèmes abstraits* (ou intéroceptifs, ces *classèmes* qui servent à catégoriser le monde et ne se réfèrent à aucune extériorité) ne permettait pas une théorie sémantique unifiée. La principale objection étant que les sèmes contextuels ou classèmes ne sont précisément pas les seuls à "être récurrents dans le discours et à en garantir l'isotopie", comme le prétend pourtant Greimas (1979). De plus, sur le plan philosophique, ces catégories du *sensible vs intelligible* constituent une transcendance par rapport à la nature linguistique des composants, laquelle se trouve alors dépourvue de réalité immanente.

Du point de vue génétique, pour approfondir le "cratyisme instinctif" promu par Barthes (*supra*) et surtout Genette (1976), par lequel la synopsis de ces Noms amène à l'attribution imitative de signifiés à des signifiants par le narrateur, Cl. Quémar se concentre sur le "rapport des différents sèmes entre eux" (1979, p. 93). Ce faisant, non seulement elle met le plan référentiel – celui de l'expérience onirique – en sourdine, mais, devant de longues versions retravaillées par Proust, elle rend attentive au travail de "polissage stylistique" (p. 73). En sorte que sa génétique thématique y gagne en autonomie, par rejet du déterminisme psychologique, fût-il freudien, mais aussi du déterminisme d'une réalité concrète extra-textuelle – celui-là même qui imposerait d'aller voir *in situ* les lieux normands et bretons pour vérifier la perception auditive-visuelle, telle que le narrateur en fait état dans les brouillons.

Enfin, la sémantique interprétative a permis à F. Rastier d'approfondir les parcours interprétatifs de l'avant-texte d'*Hérodias* (1992, 1997), grâce à sa théorie de l'isotopie, conçue comme "effet de la récurrence d'un même sème" (1987). De là la redéfinition d'une notion littéraire très psychologisée (chez Richard, Mauron, etc.) : "on nomme *thème* une structure stable de traits sémantiques (ou *sèmes*), récurrente dans un corpus, et susceptible de lexicalisations diverses" (2001, p. 197). En sorte que le déterminisme change de camp, par les contraintes qu'exercent les diverses réécritures successives sur une "impression référentielle", du fait qu'elles engendrent chaque fois de nouvelles "images mentales".¹⁰

Ainsi, pour reprendre l'exemple de l'audition colorée chez Proust, le texte final propose *Coutances, cathédrale normande, que sa diphtongue finale, grasse et jaunissante couronne par une tour de beurre* ; il se distingue du premier jet (Cahier 29, in Quémar, p. 78) par l'abandon du sème /taillant/: *Coutances, dont le grand nom mélancolique coupait le ciel comme un couteau, avec ce reflet doré de sa dernière syllabe, un peu soleil couchant, un peu beurre, dans cette Normandie d'art et de prose, de gothique et de fermes* ; et d'un brouillon intermédiaire (Cahier 32, in Quémar, p. 79) par l'abandon du sème /roture/ : *Coutances, dont le clocher reçoit de cette grasse finale le luisant du beurre normand*; et par l'insertion de son antonyme /noblesse/ ("couronne", "tour", "cathédrale" substituée à "clocher" ; oubliés les "fermes" et le prosaïsme). Cela dit, sans prendre ici en compte les migrations contextuelles aux autres

¹⁰ Sans doute est-ce là une reformulation théorique du principe de M. Riffaterre, selon lequel la *sémiosis* détermine la *mimésis* : "Toute œuvre littéraire forme un système. Dans la sémantique de ce système, les rapports entre les mots du texte l'emportent sur les rapports que ces mots entretiennent avec les choses, ou même, s'y substituent entièrement. La représentation littéraire de la réalité est gouvernée par les règles de l'idiolecte textuel." (1972)

toponymes, ni, comme l'étudie Cl. Quémar (p. 85), le "phénomène d'attraction lexicale par homophonie ou par assonance (et/ou par allitération)". Bref, on le constate avec Rastier : le sens de tels passages réside bel et bien dans les modifications et/ou conservations du signifié verbal. Quelles que soient la rêverie poétique de Marcel sur ces substances concrètes, et l'ontologie idéaliste du narrateur, pour qui l'*essence* d'un lieu est contenue dans les mots qui définissent le toponyme, dont ils imitent les sonorités (mimologiques), le sens linguistique des noms est constitué de leurs qualifications en contexte, sans se réduire à une réalité mentale (par le renvoi à l'esprit) ou à une réalité matérielle (par le renvoi aux choses). Voilà donc en quoi le sens est propre au texte, dans une situation d'interprétation globale, comme l'explique encore Rastier (1994 a).

II. Thématique et statistique lexicale

Avec le recueil collectif *L'analyse thématique des données textuelles* (Didier, 1995), la redéfinition du concept de "thème littéraire", celui des sentiments, s'est opérée dans un corpus de 350 œuvres romanesques, sur la période 1830-1970 ¹¹. Elle a donc requis l'assistance informatique pour cerner ses occurrences pertinentes, sans se départir de sa domination herméneutique : "pour atteindre ses objectifs, la thématique doit guider l'analyse lexicale, puis interpréter ses résultats qui sans cela resteraient inutilisables pour une sémantique textuelle. L'analyse lexicale, dont la statistique est un auxiliaire, ne propose pas d'elle-même des indices à l'analyse thématique. Les logiciels d'interrogation imposent certaines démarches, mais ne proposent rien. Ils servent à confirmer ou infirmer des hypothèses, et tout dépend donc de la stratégie d'interprétation." (*ibid.* repris dans Rastier, 2001, p. 191) ¹²

¹¹ Ainsi par exemple E. Bourion, étudiant le thème de la peur, faisait déboucher la recherche informatisée, de corrélats, sur l'enquête sémantique, laquelle organise l'exposé par les « axes » ainsi trouvés : "nous avons retenu, parmi les cooccurrents sélectionnés par le test probabiliste, seulement ceux qui étaient attestés chez des auteurs différents, dans différents sous-corpus (autour de *frayeur*, *épouvantable*, *faire peur*, etc.), avec un score statistique de 4 et une fréquence minimale de 4. Dans cet ensemble d'associations, nous avons étudié les cooccurrents sélectionnés, en contexte, pour repérer les isotopies sur lesquelles ils sont indexés."

¹² "Pour progresser, l'analyse thématique doit donc dépasser l'analyse lexicale. Elle en est cependant tributaire, pour des raisons qui tiennent à l'état de l'art [...]. Le mot à partir duquel peut commencer la recherche thématique n'en est pas l'objet, à la différence d'un mot-vedette qui ferait l'objet d'une recherche lexicographique. On va certes chercher, en utilisant les moyens d'assistance informatisés, d'autres mots et expressions qui sont cooccurrents. Une fois interprétés, les cooccurrents pour lesquels on aura identifié une relation sémantique seront considérés comme des *corrélats*, c'est-à-dire comme des lexicalisations complémentaires de la même molécule sémique. Le réseau des corrélats relie les manifestations lexicales du thème.

Or aujourd'hui "l'herméneutique numérique", telle que la conçoit Mayaffre (2002) veut éviter les écueils du réductionnisme : "Lorsque que le littéraire avait affaire à un grand corpus, il le lisait avec un questionnement précis, une hypothèse de travail établie, sans quoi il se perdait dans cette lecture. Pourtant, le danger de ces questionnements *a priori* ou exogènes est double et doublement cruel." D'abord par le simple fait "de projeter ses hypothèses de travail sur le texte", avec "le risque de toujours finir par trouver ce que l'on cherche (c'est-à-dire au fond ce que l'on sait déjà ou croit savoir)". Ensuite, parce "qu'une hypothèse de lecture – quand bien même se trouverait-t-elle être pertinente – en oblitère toujours d'autres, plus pertinentes peut-être." Voilà pourquoi il est nécessaire que "la lecture, la compréhension, le questionnement, puis l'interprétation des textes gagnent aujourd'hui en rigueur pour sortir du tout-subjectif. Dans un retournement spectaculaire, la démarche inductive complète la démarche déductive, le positivisme se substitue au constructionnisme." Car, "de l'*hypothético-déductif* en vigueur nous passons à un *positivisme-inductif* original. La démarche épistémologique face au texte se trouve donc inversée : là où traditionnellement le chercheur interrogeait le texte sur la base d'hypothèses de travail construites, c'est le texte qui interroge le chercheur sans tabou et sans *a priori*. Par une lecture différente (hypertextuelle plutôt que linéaire, nous l'avons vu, mais aussi paradigmatique plutôt que syntagmatique, quantitative plutôt que qualitative), l'ordinateur voit autre chose pour déranger nos certitudes et élargir l'horizon étroit de nos modes (aux deux sens du terme) d'interrogation." Ce développement paraît très convaincant, en particulier pour qui a recours à des logiciels de lexicométrie. Certes, devant un écran de concordances lexicales attestées en contexte, "se laisser interpellé par lui pour découvrir ses éléments saillants que l'on ignore devient passionnant", Toutefois le *positivisme-inductif* est un leurre, ou plutôt une solution de facilité. Rastier (2001, p. 96) prévient que "la notion de donnée invite à la prudence. Du seul fait qu'une hypothèse a présidé à son recueil, une donnée est *donnée* non à l'observateur, mais par l'observateur. Par exemple, toute acception est liée à un genre et à un discours : pour la rendre interprétable et transformer son occurrence en donnée, il importe donc, de restituer cet entour." Ainsi, par exemple, quand dans la littérature régionale, le logiciel souligne la fréquence des comparaisons de la nature avec un joyau (H. Russell, *Souvenirs d'un montagnard* : "Où trouverait-on un site plus romantique que le sommet de l'orgueilleuse colline couronnée par les Bains de Vénasque ? Sapins en haut, sapins en bas, rochers partout : et sur l'autre rive de l'Essera, qui coule à 200 mètres plus bas comme un torrent d'émeraude et d'aigue-marine, [...] Je

Mais il faut pouvoir discerner les meilleurs points d'entrée dans ce réseau : la "vedette" n'est qu'un de ces points d'entrée, présumé lexicaliser synthétiquement le thème que l'on cherche à décrire." (2001, pp. 206-7)

regrettait ces grandes vagues de *saphir*, ces obélisques de glace, ces chaos verts et azurés, qui sont une des merveilles des Pyrénées." etc.), l'interrogation de banques textuelles dépassant cette monographie en relativise l'originalité, car elle révèle combien le genre narratif contemporain de Russell (de Balzac, Gautier, Sand à Verne et Zola) n'était pas avare de cette pierrerie esthétisante, telle qu'elle était chantée aussi dans le corpus poétique, remontant du romantisme à la pétrification précieuse de type pétrarquiste, pour le blason de la femme aimée ("Puisque tu veux dompter les siècles tout-perdants Par le rare portrait de ses grâces divines, Frise de *chrysolits* ses tempes *ivoirines*, Fais de *corail* sa lèvre et de *perle* ses dents : Fais ses yeux de *crystal* y plaçant au-dedans Un cercle de *saphirs* et d'*émeraudes* fines, Puis musse dans ces ronds les embûches mutines De mille amours taillés sur deux *rubis* ardents. Fais d'*albâtre* son sein, etc."). Bref, le relevé d'occurrences – des "données" lexicales – ne constitue un indice qu'en étant rapporté à un genre. Il suppose un tri, pour ne retenir que les données qui concordent avec cette topique littéraire, laquelle contredit donc la pure démarche inductive. Dès sa lecture dans le concordancier, le nom de joyau ainsi thématiqué perd sa "restriction positiviste", celle-là même, dans les sciences du langage, à laquelle "s'en tient le courant scientifique" et "la tradition logico-grammaticale" (Rastier, 2001, p. 5).

Ce mythe de la démarche inductive, fondée sur des données *endogènes* fournies par un ordinateur qui balayerait le corpus de façon autonome ¹³, perdure aujourd'hui chez Adam, qui dénigre "le concept encore trop rudimentaire d'isotopie. J.-M. Viprey montre que la co-occurrence ou la collocation (chez Halliday et Hasan ou chez Harris) est un facteur primordial de textualité" (2006, p. 5), ces marques dont Rastier avait déjà souligné les insuffisances pour expliquer la cohésion sémantique d'un texte : "La notion de texture proposée par Halliday et Hasan montre bien les limites d'une approche linguistique discrètement positiviste" (1994 b). Ainsi Viprey, prenant l'exemple d'un roman balzacien, propose d'y étudier le thème financier de la vente en laissant le logiciel imposer une liste de vocables formant un ensemble "micro-distributionnel", dont la co-occurrence interne est réglée par la méthode statistique de "l'analyse factorielle". Il explicite : "Les relations lexicales ainsi cartographiées (cartographie des vocables les plus occurrents; la proximité entre les items figure leur parenté de profils collocatifs), j'ai proposé de les désigner sous le terme global d'isotropie." Soit un terme qui de l'aveu même de son créateur "n'a pas été choisi au hasard", mais pour entrer, sinon "en compétition",

¹³ A tel point que Mayaffre affirme aujourd'hui (2008, p. 94) que "*endogène* est sans doute le mot-clef de la linguistique de corpus, ainsi peut-on envisager une statistique endogène, une stylistique endogène (Viprey), une lexicologie endogène, une sémantique endogène, etc." Soit un nouvel avatar de l'immanence.

du moins "en coopération avec le terme et la notion d'isotopie" de Rastier (2006, p. 78) Ainsi dans *Le Père Goriot* l'isotropie du vocable *vendre* a pour "attirance" immédiatement visible le groupe *mère, fille, seulement, pauvre, argent, robe*. Or quand dans son évaluation du modèle Viprey conclut plus généralement que de telles "lignes saillantes de la collocation lexicale seront des interprétants de la structure séquentielle beaucoup plus féconds que la simple répétition" (p. 82), et qu'il plaide lui-même "pour le passage du *discontinu* logico-grammatical (pour reprendre les termes de Rastier) au *continu* sémantique", il fait passer au second plan l'épreuve des contextes d'attestation des co-occurrences lexicales¹⁴ qui sont précisément à même de déterminer leur relation thématique.

Cette priorité du niveau lexical-statistique sur le niveau textuel sémantique s'appuie sur cet argument théorique : "Même si l'on entend que le sème isotopant (et le sémème isotopé) s'actualisent dans le cadre d'une sélection, contrainte par le texte (ce qui et notamment le thème majeur de *Sens et textualité*), ces catégories n'en restent pas moins exogènes et projetées, à partir de descriptions du lexique, sur le vocabulaire. Comme telles utiles et nécessaires, elles n'en sont pas pour autant suffisantes pour décrire, modéliser et grammatiser la réticulation matérielle d'un texte singulier. Les catégories de l'*isotopie* ne sauraient jamais se substituer à la formalisation descriptive des singularités sémantiques du texte à l'étude, dont la collocation généralisée nous semble une phase cruciale. Descriptive, cette formalisation n'en est pas moins une opération interprétative première, celle où le texte s'interprète en quelque sorte lui-même, se déploie, se restructure et objective les lignes de force de son vocabulaire." (p. 78) A travers cette auto-génération, le recours à "la configuration *isotropique*" fait ressurgir le mythe positiviste selon lequel "le statut des *unités du texte* est endogène" (p. 81), elles qui procèdent par "champs qui interagissent" (p. 82), et qui se présentent sous forme de vocables décontextualisés.

Mais pourquoi cette "réticulation" (tabulaire et non-linéaire) devait-elle reléguer à l'exogène la constitution de réseaux associatifs fondés sur la notion d'isotopie, laquelle réclame un minimum de linéarisation (de "séquentialité", dit Viprey), du fait qu'elle est un phénomène syntagmatique ? Pour reprendre l'exemple balzacien, un simple regard sur les lexèmes *mère, fille* unit leur sème inhérent /filiation féminine/ à /finance/ inhérent à *pauvre, argent*, via le sème /parure féminine/ de *robe*¹⁵. Un tel contenu sémantique, croyons-nous, relève aussi

¹⁴ Et ce, malgré sa copie d'écran des "concordances du verbe *vendre*" (p. 85) qui laisserait croire le contraire.

¹⁵ Relation inter-sémique qui n'apparaîtra qu'au crible des contextes, dont les deux suivants du *Père Goriot* sont représentatifs : "Je conçois alors qu'Anastasia se fasse faire une *robe* lamée, et veuille attirer sur elle tous les regards chez Mme de Beauséant, en y paraissant

"d'un processus endogène" (p. 82) ¹⁶, dans la mesure où il ne s'agit nullement de "catégories projetées", mais d'une sélection-actualisation à partir des seuls passages du roman – qui la font varier et dont elle dépend.

En sorte que la métaphore géographique ne doit pas leurrer : se laisser guider par "carte" statistique, dite "atlas de l'hypertexte", qui permet de visualiser les rapports de voisinage graphique témoignant de la structuration lexicale, désorienté, dès lors qu'elle est découplée de la relation sémantique contextuelle, qui, selon nous est seule apte à donner un fondement à la notion d'*environnement thématique*, entendue comme ces collocations lexicales statistiquement justifiées.

Cette conception cartographique trouve un autre écho chez M. Kastberg (2008) dans un premier temps de son enquête sur le traitement statistique des données textuelles par le très performant logiciel *Hyperbase* (version 8.0 lemmatisée, 2008). L'auteur se propose "non seulement d'extraire du corpus les items les plus fréquents, mais de rechercher la corrélation entre deux items, autrement dit la « proxémie » ¹⁷. Ce traitement des co-occurrences est depuis longtemps un des enjeux majeurs de la lexicométrie. De plus, dans les recherches littéraires il s'avère un outil prestigieux, notamment dans l'extraction des réseaux isotopiques, ou isotropiques." ¹⁸

Elle obtient ainsi, pour le corpus Le Clézio, une liste de paires, classés par ordre décroissant. Or dans ces contextes *a minima* ¹⁹, les associations sont si doxales, qu'elles ne révèlent, *in fine*, que la banalité de leur relation métonymique :

1098.64 nu pied

385.74 fond vallée

285.90 arbre feuille

234.69 ciel étoile

dans tout son éclat et avec ses diamants." (p. 238) Antithétique de l'incipit : "Son jupon de laine tricotée, qui dépasse sa première jupe faite avec une vieille *robe*, et dont la ouate s'échappe par les fentes de l'étoffe lézardée, résume le salon, la salle à manger, le jardinet, annonce la cuisine et fait pressentir les pensionnaires." (p. 55)

¹⁶ Il s'agit là de ce que Mayaffre appelle des "isotopies endogènes au corpus" (2008, p. 100). En revanche, il semble que Viprey relègue à l'exogène les isotopies *afférentes*, par exemple celles qui forment le triplet dysphorique "liquidation, trahison, prostitution" qu'il donne pour le sens contextuel de *vendre*, distinctes du "commerce proprement dit" (2006, p. 80).

¹⁷ Elle précise : "L'étude automatique des collocations et de la micro-distribution des termes nous incite à aborder l'extraction automatique des univers sémantiques et des collocations par des biais différents : d'un côté par l'extraction d'un univers thématique, gravitant autour d'un mot-pôle, et de l'autre, par le recensement des co-occurents et de séquences d'items."

¹⁸ Soit une conciliation avouée des travaux de Rastier et Viprey.

¹⁹ Cf. Mayaffre (2008, p. 95) : "la forme minimale du contexte d'un terme, nécessaire à sa compréhension-interprétation, n'est pas le syntagme ou la phrase mais la co-occurrence".

821.02 femme homme	372.49 mer vague	282.52 cœur poitrine	226.95 colline pierre
763.43 <i>ciel nuage</i>	368.83 nord ouest	268.35 dieu seigneur	224.95 nuage poussière
657.52 enfant femme	365.53 eau goutte	267.55 escalier marche	224.84 jardin maison
623.35 œil visage	346.98 bras jambe	260.09 maison toit	224.50 étoile nuit
525.63 mer vent	344.30 larme œil	258.99 mer oiseau	223.97 <i>ciel horizon</i>
525.36 <i>ciel soleil</i>	305.51 bruit moteur	249.77 montagne vallée	223.83 horizon mer
502.21 lumière soleil	303.70 nord sud	248.33 <i>ciel lumière</i>	222.26 front joue
479.39 feuille papier	300.94 bruit vent	245.32 corps intérieur	215.56 horizon ligne
476.56 cheveu visage	294.13 <i>ciel terre</i>	243.26 pied sol	214.72 horizon soleil
437.05 mère père	293.21 rue ville	239.85 canne champ	210.39 cheveu œil
412.23 <i>ciel mer</i>	290.10 cri oiseau	236.20 arbre branche	203.47 doigt main

Une seconde méthode ²⁰ est alors requise pour pallier une telle fragmentation :

"Mais, on peut se rapprocher d'une association, d'un microcosme co-occurentiel d'une autre manière. Prenons un item lexical, *ciel*, un des lemmes les plus fréquents du corpus leclézien. Grâce à *Hyperbase* nous faisons une extraction du contexte immédiat, ici le paragraphe, qui entoure les 2949 occurrences du lemme *ciel*. L'extraction automatique du contexte d'un item lexical permet la création d'un sous-corpus qui est soumis à un calcul de spécificité, puisqu'on ne recherche plus une relation entre un mot et un texte, mais une relation privilégiée entre les mots eux-mêmes." ²¹

"Choisissons par exemple l'item *ciel*, reflétant bien une thématique chère à notre auteur ; voici les concurrents les plus proches du pôle : *ciel, nuage, soleil, mer, terre, lumière, étoile, horizon, oiseau, vent, couleur, eau, avion, montagne, lune, espace, centre, fumée, éclair, nuit, femme, colline, fleuve, vallée, toit, étendue, arbre, plaine, ligne, pluie, vague, sable, pierre, herbe, main, désert, lueur, gens, droit, fond, dune, aile, brume*".

En découle "une représentation graphique des liens préférentiels qui tissent un réseau de ce mot choisi pour pôle, sous forme de graphe". L'annexe ci-dessous donne ces copies d'écran pour un autre mot pôle, *hydre* dans les deux corpus en prose et en vers de Hugo, traité par *Hyperbase* (figures 1 et 2). Ces graphes servent au traitement du thème du reptile mythologique chez cet auteur, dans une optique comparative au sein de deux genres distincts. Certes elle fait

²⁰ Après celle-ci, par laquelle, comme l'explique Brunet (2007, p. 9), "la recherche sur les associations s'appuie sur le tableau des co-occurrences, dont la fonction *corrélats* a fourni une vue d'ensemble, sous forme d'analyse factorielle." Ces couples de vocables ont ainsi donné d'autres pistes lexicales (outre les *spécificités* statistiques) au lecteur qui n'aurait pas de choix *a priori* se son thème. Toutes ces méthodes offertes par le logiciel organisent, en le retardant, le retour au texte (comme le dit Mayaffre, 2008) ; aussi fiabilisent-elles l'enquête.

²¹ Plus exactement, "Nous pouvons ensuite effectuer le même calcul que tout à l'heure, celui concernant les co-occurents, cette fois-ci en prenant le sous-corpus de l'environnement thématique du mot-pôle comme base pour le calcul."

ressortir des lexicalisations identiques (mais avec des scores différents) ou différentes (par exemple, le co-occurent *Michel*, individualisé, devient l'anonyme et multiple *anges*). Mais ces vocables, comme ces "liens co-occurentiels" dont parle Kastberg, n'acquièrent de sens qu'une fois resitués dans leurs contextes verbaux respectifs. Nous persistons à croire que cette seconde opération, dans l'ordre des priorités, doit avoir la primauté, *in fine*, pour mériter le qualificatif de "thématique", au-delà de son acception statistique dans la commande du logiciel.

Certes on a là une base fiable à partir de laquelle peut commencer l'interrogation des contextes d'attestations de ces substantifs en réseau. Les associations lexicales étant justifiées par les calculs, peut débiter l'enquête qualitative sur la nature des corrélations textuelles. Ainsi pour reprendre l'exemple de l'ouvrage numérisé de Russell que l'on a cité, le précieux *environnement thématique* que fournit *Hyperbase* (par ordre hiérarchique décroissant), pour le même pôle *ciel*, consiste en une liste, dans laquelle la normalité des relations métonymique sollicite moins l'intérêt du lecteur que l'originalité de certains corrélats, le premier étant par exemple "perçant"²² :

*Ecart*²³ *Corpus Mot*

4.76 20 FUMÉE

3.56 15 FLÈCHE

²² L'intérêt majeur du logiciel est qu'il fournit d'un clic, *via* un concordancier, tous les contextes d'attestation. Pour Russell, ceux-ci explicitent la relation violente entre pic terrestre et surface céleste, selon le topos d'un glaive angélique : "le fier Mont-Perdu, couvert de neige, et la pointe nue, mais orgueilleuse aussi du Cotieilla, *perçant l'ardente vapeur du ciel* d'Espagne, comme un énorme Vésuve éteint, ou le génie morose des sables." "Je revoyais à l'horizon, dans une clarté extraordinaire, les Pyrénées *perçant de leurs stérilités neigeuses l'azur* presque tropical du ciel : je souffrais d'être si bas". "Enfin à l'Est, *perçant le ciel comme un poignard*, la pyramide dentelée des Corvetas éblouit toute la région par sa blancheur." "Sans doute on n'y voit pas, comme dans les Alpes, des aiguilles aériennes, des flèches vertigineuses, dominant leurs voisines de mille mètres, des pyramides et des chaos de glace qui semblent des Icebergs échoués au milieu de l'Europe, des obélisques de quatre mille mètres, *perçant le ciel comme des paratonnerres ou des poignards*." "Emblèmes sacrés de l'innocence, *perçant les nues*, et dominant la terre entière, ces grandes masses blanches symbolisent nos élans, nos ascensions morales vers l'Infini, vers l'Idéal, autrement dit vers Dieu." "*Perçant à peine un nuage* crépusculaire et jaune, avec lequel se confondaient au loin des plages bourbeuses, une espèce de soleil apparaissait vaguement dans le brouillard et la fumée qui s'échappait de longues cheminées grises et solitaires, dont on cherchait en vain la base." Une seule occurrence transfère l'acuité à la vision : "C'est une image de l'infini, car *l'œil le plus perçant* ne saurait distinguer, dans une immensité si indécise, le point de contact entre la terre et le ciel."

²³ Il s'agit de l'*écart réduit* (ce "test probabiliste" dont parle Bourion ci-dessus) ainsi expliqué par E. Brunet (2006, p. 31) : "La distribution d'un mot est rarement régulière à travers un corpus et des écarts s'y observent entre la fréquence d'un mot observée dans un

19.62	120	BLEU	4.74	10	VOYAGEUR	3.55	29	BLEUES
8.55	193	TERRE	4.69	488	SOLEIL	3.45	38	INFINI
8.30	34	AZUR	4.69	15	TEINTES	3.39	16	GÉANTS
8.15	195	NUAGES	4.69	15	BLANCHEURS	3.38	163	QUEL
7.59	13	LIMPIDE	4.30	44	BRISE	3.38	118	FEU
7.53	50	PUR	4.29	23	ORIENT	3.38	23	DOMINANT
7.02	8	PERÇANT	4.19	12	INNOCENCE	3.36	57	ROUGES
6.61	142	HORIZON	4.13	31	CROIT	3.26	24	IDÉAL
5.95	7	VOS	4.07	55	BLANCHEUR	3.24	254	NEIGES
5.95	7	PROFILANT	4.01	3005	DU	3.21	41	NEIGEUSES
5.81	109	LUMIÈRE	3.96	13	PURES	3.19	112	PLAINES
5.76	11	SÉRÉNITÉ	3.96	13	POLAIRE	3.18	70	CIMES
5.54	358	SOUS	3.96	13	ENDORT	3.16	113	DESSUS
5.53	4576	LES	3.79	20	LEVANT	3.14	25	SPLENDEURS
5.48	8	PURETÉ	3.75	14	BRILLANTES	3.14	25	ENTENDRE
5.15	42	NUAGE	3.70	120	ÂME	3.12	103	OCÉAN
5.08	9	COMMENÇAIENT	3.66	28	CRÊTES	3.08	18	POUSSIÈRE
4.98	51	ÉTOILES	3.64	21	IMMENSITÉ	3.07	34	LUEURS
4.88	76	NOIR	3.64	21	GRACIEUSE	3.03	26	ÉTERNELLE
4.76	20	PYRAMIDES	3.56	15	PÂLE	3.01	291	NATURE

Si l'on ne peut que souscrire à la conclusion de M. Kastberg (2008) :

"Ces différentes manières d'envisager les associations et les environnements thématiques à partir d'un mot-pôle à l'intérieur d'un corpus permettent non seulement de comparer différents calculs statistiques, mais fournissent aussi des preuves quant à la solidité de ces analyses. Les résultats sont en effet souvent très semblables et la complétude des différentes études constitue une base solide pour l'étude des isotopies et des collocations dans un texte. [...] Néanmoins, si le traitement statistique des occurrences d'un corpus permet d'en faire une description formelle, le traitement statistique co-occurentiel et de la proxémie d'items lexicaux nous fait franchir un palier important, permettant de basculer vers la sémantique et la phraséologie. C'est pourquoi cette technique s'avère constituer un atout précieux pour l'analyse sémantique d'un texte."

c'est en précisant aussitôt qu'au-delà de la phraséologie²⁴, l'enquête thématique sur les contextes d'attestation constitue l'essentiel d'une approche qui n'accepte les qualificatifs d'*immanente* et d'*endogène* que dans le cadre de la sémantique interprétative. C'est à cette condition que le "continu" du sens textuel dont parle Viprey ne peut être atteint.

Références

- Adam, J.-M. (2006) "Autour du concept de *texte*. Pour un dialogue des disciplines de l'analyse des données textuelles", *J.A.D.T.*
- Barthes, R. (1967) "Proust et les Noms", *Nouveaux essais critiques*, Seuil, Points [1980]
- Barthes, R. (1970) *S/Z*, Seuil, Points [1991].

texte et la fréquence théorique qu'on était en droit d'attendre, vu la proportion du texte dans l'ensemble."

²⁴ Doxale ou propre à un auteur, telle, chez Russell, la structure *perçant* + *COD céleste*.

- Bourion, E. (1995) "Le réseau associatif de la peur", in *L'analyse thématique des données textuelles*, Didier, en ligne à www.revue-texto.fr
- Brunet, E. et Kastberg-Sjöblom, M. (2000) "La thématique. Essai de repérage automatique dans l'œuvre d'un écrivain", *J.A.D.T.*
- Brunet, E. (2006) Le logiciel *Hyperbase* – Manuel d'utilisation.
- Brunet, E. (2007) "Fréquences et séquences. Mise en œuvre dans *Hyperbase*", *Lexicometrica : Topographie et topologie textuelles*.
- Collot, M. (1988) "Le thème selon la critique thématique", *Communications* 47.
- Compagnon, A. (1992) "Ce qu'on ne peut plus dire de Proust", *Littérature* 88.
- De Biasi, P.-M. (1987) "L'avant-texte", article mis en ligne le 18 janvier 2007, disponible sur : <http://www.item.ens.fr>.
- Debray-Genette, R. (1976) "Genèse des aubépines", *Poétique* 25, repris dans le collectif *Recherche de Proust*, Seuil, Points [1980].
- Debray-Genette, R. (1977) "Génétique et poétique : esquisse de méthode", *Littérature* 28.
- De Lattre, A. (1984) *Le personnage proustien*, Corti.
- Genette, G. (1966) "Structuralisme et critique littéraire", *Figures I*, Seuil, Points [1981].
- Genette, G. (1976) *Mimologiques*, Seuil.
- Genette, G. (2001) "Peut-on parler d'une critique immanente ?", *Poétique* 126.
- Greimas, A.-J. (1966) *Sémantique structurale*, PUF.
- Greimas, A.-J. et Courtés, J. (1979) *Dictionnaire raisonné de la théorie du langage*, Hachette.
- Jenny, L. (2002) "Hypertexte et Genèse", *Littérature* 125.
- Kastberg-Sjöblom, M. (2008) "Attirance thématique : fréquences ou séquences ? Extraction des isotopies sémantiques d'un corpus textuel", revue électronique *Texte et corpus*, n°3 (Actes des Journées de la linguistique de Corpus 2007).
- Marantz, E. (1989) "La genèse du personnage proustien", *B.S.A.M.P.* 39.
- Mayaffre, D. (2002) "L'Herméneutique numérique", en ligne à www.astrolabe.fr.
- Mayaffre D. (2008) "L'entrelacement lexical des textes, co-occurrences et lexicométrie", revue électronique *Texte et corpus*, n°3.
- Milly, J. (1974) "Sur quelques noms proustiens" *Littérature* 14.
- Quémar, Cl. (1979) *Essais de critique génétique*, Flammarion.
- Rastier, F. (1987, rééd. 1996) *Sémantique interprétative*, PUF.
- Rastier, F. (1989) *Sens et textualité*, Hachette, rééd. électr. à www.revue-texto.fr.
- Rastier, F. (1992) "Thématique et génétique – L'exemple de *Hérodiade*", *Poétique* 90.
- Rastier, F. (1994 a) "Sur l'immanentisme en sémantique", *Cahiers de Linguistique Française* 15, en ligne à www.revue-texto.fr
- Rastier, F. (1994 b) "La macrosémantique", in *Sémantique pour l'analyse*, Masson, et en ligne à www.revue-texto.fr
- Rastier, F. (1997) "Parcours génétiques et appropriation des sources", in *Texte(s) et intertexte(s)*.
- Rastier, F. (2001) *Arts et sciences du texte*, PUF, notamment le chapitre "Thématique et topique", qui reprend "La sémantique des thèmes ou le voyage sentimental", paru in *L'analyse thématique des données textuelles*, Didier, 1995, et en ligne à www.revue-texto.fr
- Rastier, F. (2006) "Formes sémantiques et textualité", *Langages* 163.
- Richard, J.-P. (1974) *Proust et le monde sensible*, Seuil.
- Richard, J.-P. (1975) "Table ronde", *Cahiers Marcel Proust*.
- Riffaterre, M. (1972) "Le poème comme représentation", *Poétique* 9.
- Viprey, J.-M. (2006) "Structure non-séquentielle des textes", *Langages* 163.

Annexe

Environnement thématique du mot-pôle *Hydre* dans deux corpus de Hugo.

Présentation par ordre hiérarchique décroissant ²⁵

Corpus de poésie versifiée (Hydre 64 occ.)

<i>Ecart</i>	<i>Corpus</i>	<i>Mot</i>
3.85	12	anarchie
3.79	13	écailles
3.63	16	Michel
3.60	112	chaos
3.58	17	ébauche
3.15	29	dragon
3.13	107	vingt
3.12	30	triple
3.08	162	contre
2.96	177	obscur
2.90	39	île
2.88	40	delà
2.78	264	fut
2.73	146	eût
2.68	50	rampe
2.65	52	monstres
2.59	55	terreur
2.58	56	fureur
2.51	60	globe
2.49	388	sort
2.44	397	vie
2.44	65	plaines
2.37	70	ver
2.35	127	terrible
2.33	130	horrible
2.33	129	cri
2.30	75	argent
2.29	134	obscur
2.22	141	ouvre
2.21	82	fumée
2.14	298	encor
2.10	227	oiseau
2.00	167	lieu

Corpus de prose (Hydre 38 occ.)

<i>Ecart</i>	<i>Corpus</i>	<i>Mot</i>
7.62	16	encor
5.47	130	chaos
5.47	24	onde
5.14	175	obscur
5.04	1033	ombre
5.02	462	sort
4.77	16	écailles
4.58	160	oiseau
4.56	22	cieux
4.55	75	aile
4.50	24	ver
4.27	34	monts
4.16	40	anarchie
4.04	48	rampe
4.01	50	airain
3.93	56	dragon
3.92	3963	sous
3.90	58	ébauche
3.90	58	astres
3.87	61	plaines
3.86	978	ciel
3.85	166	gouffre
3.81	66	triple
3.73	190	flamme
3.72	75	azur
3.71	76	fauve
3.69	197	obscur
3.67	201	affreux
3.64	83	monstres
3.59	619	voit
3.58	90	cendre
3.57	92	flots
3.57	91	fureur
3.55	94	anges

Les associations que Hugo privilégie dans son corpus poésie Présentation par ordre hiérarchique décroissant des dix premières

<i>test</i>	<i>mot 1</i>	<i>mot 2</i>
9005.92	abîme	ciel
22.29	étoile	voile
20.04	nuit	ombre
19.15	jour	nuit
19.14	nombre	ombre
18.92	âme	flamme
16.24	enfant	mère

²⁵ Avec une élimination des déterminants et de la ponctuation, pour ne laisser que plus de place aux candidats lexicaux privilégiés. Le mot-pôle est aussi restreint au singulier.

16.15	bruit	nuit
16.14	nu	pied
15.34	flot	mer

La première association de notre thème de monstruosité reptilienne n'apparaît que bien plus loin dans le classement, avec le score suivant : "5.57 chaos hydre"

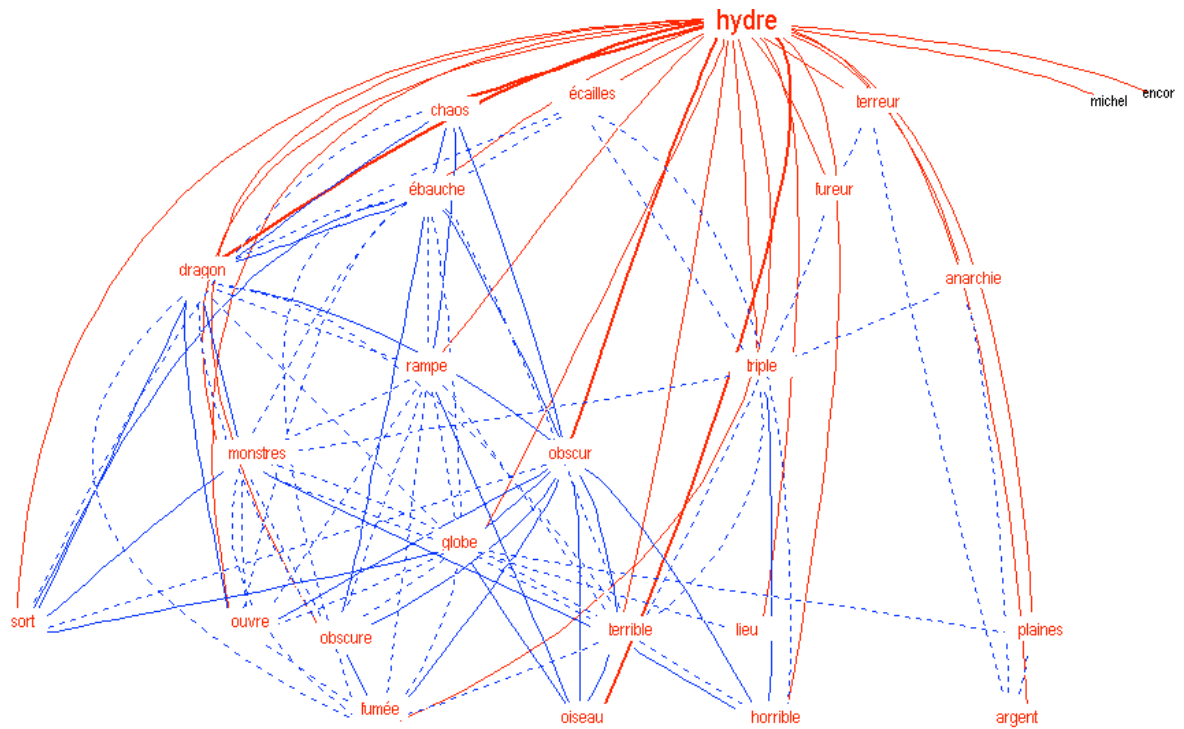


Figure 1 – Graphe du mot-pôle dans le corpus poésie

