

**UNIVERSITE DE LA SORBONNE NOUVELLE**

**PARIS III**

**DOSSIER D'HABILITATION A DIRIGER DES RECHERCHES**

**Michel BERNARD**

**Études Littéraires Assistées par Ordinateur**

**REFLEXIONS SUR LES CONDITIONS D'EMERGENCE**

**D'UNE DISCIPLINE**

Sous la direction du professeur Henri BÉHAR

2003

# Introduction

Je me suis donné comme objectif, dans le présent dossier, de situer ma carrière et mon travail d'enseignant-chercheur dans le contexte de l'émergence d'une nouvelle discipline – ou d'un nouveau champ disciplinaire – que je propose d'appeler « Études Littéraires Assistées par Ordinateur » (ELAO)<sup>1</sup>. Comme candidat à l'habilitation à diriger des recherches, il m'importe avant tout de montrer comment les travaux auxquels je me suis consacré se situent dans un périmètre qui ne pourra être défini qu'en définissant en même temps ce qui le caractérise et ce qui le fonde scientifiquement. Je ne puis en effet, contrairement à la plupart des spécialistes de littérature, me rattacher à un siècle (je ne suis ni « dix-neuviémiste », ni « vingtiémiste », etc.), à un mouvement (je ne suis spécialiste ni du Romantisme, ni du Baroque, etc.), à un champ conceptuel (je suis pas un théoricien de la littérature, un généticien, ni un sociologue du roman ou un didacticien, etc.) L'ensemble des recherches que j'ai menées se rattachent, directement ou indirectement, à l'utilisation de l'informatique dans les études littéraires. Ce rapprochement d'un outil et d'un domaine scientifique n'est pas constitutif, en soi, d'une discipline, fût-elle ancillaire. L'utilisation de la machine à écrire ou de la fiche cartonnée par les spécialistes de littérature n'a jamais suscité de réflexion épistémologique, n'a pas

---

<sup>1</sup> . C'est Henri Béhar qui utilisa pour la première fois cette expression pour nommer son séminaire de 3<sup>e</sup> cycle à la Sorbonne Nouvelle (Paris 3). Je l'ai reprise dans le titre de mon *Introduction aux Etudes Littéraires Assistées par Ordinateur* (PUF, 1999). J'utiliserai le sigle ELAO pour les désigner dans la suite du mémoire.

engendré de domaine particulier de la connaissance. Il me faudra donc montrer en quoi l'apport de l'informatique aux études littéraires ne se borne pas à fournir de nouveaux outils aux chercheurs mais leur ouvre de nouvelles perspectives, leur offre la possibilité de définir de nouvelles voies.

J'essaierai de mener cette démonstration à travers l'examen de plusieurs domaines recouvrant à la fois des aspects des ELAO et des facettes de mon activité de chercheur. Je commencerai par montrer comment l'informatique a pu être introduite dans les études littéraires (recherche et formation), puis j'aborderai tout à tour les applications dans les domaines des bases de données, des hypertextes et de la lexicométrie. Je présenterai ensuite quelques travaux littéraires regroupés sous le nom de « lectures par le menu », où l'informatique est un moyen d'asseoir une nouvelle méthode critique.

L'aspect nécessairement personnel de ce dossier m'amènera à prendre comme exemple les travaux que j'ai moi-même menés, ou auxquels j'ai été associé, mais je les situerai systématiquement dans un contexte plus large, en montrant qu'ils complètent ou suivent d'autres entreprises de chercheurs du même domaine. De manière générale, on verra que les ELAO, contrairement à une certaine tradition des études littéraires, favorisent le travail d'équipe, la constitution de groupes et de réseaux, les œuvres collectives. Pour cette raison, je me référerai souvent aux travaux du Centre de recherche Hubert de Phalèse, créé en 1989 par Henri Béhar et que je dirige aujourd'hui. La collection Cap'Aggreg, qui vient de publier cette année son quinzième ouvrage (elle a débuté en 1991), a la particularité de proposer des ouvrages collectifs sous le pseudonyme de Hubert de Phalèse. Les noms des collaborateurs sont mentionnés, mais il s'agit d'une véritable œuvre commune, et non d'une série de contributions aboutées. Tous les textes publiés sont l'objet de relectures et de corrections collectives, au point qu'il est souvent impossible de dire qui est l'auteur de telle ou telle partie

d'un ouvrage. Je citerai dans ce dossier les ouvrages auxquels j'ai participé (douze sur quinze) mais je ne réclame pas la paternité exclusive des passages évoqués.

Par ailleurs, une de mes publications aura dans ce mémoire un statut un peu particulier. *L'Introduction aux Etudes Littéraires Assistées par Ordinateur*, que j'ai publiée aux PUF en 1999, recouvre en effet tous les aspects des ELAO qui seront envisagés ici. J'ai par conséquent choisi de ne pas réserver une section particulière à ce livre et d'y faire allusion en tant que de besoin, et donc de manière éparse. Cet ouvrage amorçait la réflexion épistémologique que je présente dans ce mémoire mais il visait surtout à donner des exemples concrets et des explications techniques, à fournir des indications pratiques sur un domaine encore méconnu. Le but de *L'Introduction aux Etudes Littéraires Assistées par Ordinateur* était de définir un champ disciplinaire en le nommant, en énumérant ses divers aspects et en montrant (en particulier par l'établissement d'une bibliographie) qu'il a une histoire, des spécialistes et des acquis. Je partirai donc du principe que ces démonstrations ont été faites (même si la rapidité de l'évolution des techniques rend très vite obsolète un discours sur l'informatique), que les références essentielles ont été fournies et qu'il ne s'agit plus à présent d'établir l'existence d'un champ disciplinaire mais de réfléchir aux conditions de son organisation scientifique et institutionnelle.

## **Remerciements**

Je tiens à remercier ici, parce que tous les travaux dont je vais parler sont, de quelque manière, collectifs, tous ceux qui ont participé, à un moment ou l'autre, aux activités du Centre de recherche Hubert de Phalèse, tous ceux qui, en organisant des colloques, en publiant des recueils, des revues, m'ont permis de m'exprimer, tous les élèves et étudiants, tous les collègues dont le contact m'a enrichi.

Je tiens surtout à remercier le professeur Henri Béhar pour une aide, un exemple et une amitié qui me soutiennent depuis près de vingt ans.

# **informatique et littérature : esprit de géométrie et esprit de finesse**

## **Naissance de la micro-informatique**

Sans vouloir donner un tour exagérément autobiographique à ce dossier, il m'apparaît néanmoins nécessaire de dire quelque chose de la démarche personnelle qui m'a amené à m'intéresser à l'informatique. C'est en effet une question qui m'a souvent été posée, et qui suppose toujours qu'il y aurait quelque incongruité à associer les deux spécialités. Y répondre permet d'ouvrir le débat sur cette incompatibilité alléguée et de commencer à tracer l'histoire d'une discipline. En effet, ma propre découverte de l'informatique coïncide avec l'apparition de ce que l'on appelait naguère la « micro-informatique ». C'est en 1978 que j'ai eu l'occasion de travailler pour la première fois avec un ordinateur. Dans le cadre d'un stage que l'École Normale Supérieure avait eu la judicieuse et surprenante idée de proposer aux nouveaux conscrits littéraires, j'ai pu m'initier à cette problématique encore toute neuve de l'utilisation de l'informatique dans les sciences humaines. Les ordinateurs étaient alors des machines fort peu commodes : pas d'écran (tous les résultats étaient imprimés), pas de lecteurs de disquettes (les données étaient entrées par le biais de fiches perforées). Les chercheurs étaient censés programmer eux-mêmes leurs requêtes (le PL/1 semblait alors le langage le plus adapté aux besoins des sciences humaines) mais ils étaient contraints, dans tous les cas, de passer par le truchement des informaticiens pour tout accès à la machine.

C'est dans ces années-là, sans que l'événement ait fait beaucoup de bruit, que les premiers micro-ordinateurs apparurent aux Etats-Unis (l'Apple II est de 1977). Ces nouvelles machines entraînent, on le sait, une véritable révolution dans l'usage de l'informatique. Il devenait possible d'utiliser seul une machine à la portée de la bourse d'un particulier, dotée de programmes relativement simples d'emploi. Les premiers programmes furent plutôt dédiés aux chiffres mais, bientôt, des « traitements de texte » permirent d'utiliser la puissance des ordinateurs pour traiter commodément du texte, pour l'écrire ou pour l'étudier (la première version de *Wordstar* a d'ailleurs vu le jour en cette année 1978). L'avènement d'une « micro-informatique » séduisit évidemment tous les chercheurs littéraires qui travaillaient déjà – au prix de quelles difficultés ! – avec les grands systèmes informatiques, mais il permit aussi à une nouvelle génération de littéraires – dont je suis – d'éviter d'emblée les complexités et les lourdeurs de l'informatique en temps partagé et donc de se faire de l'informatique une idée moins rébarbative.

## **Langue / langages**

Cette simplicité d'emploi était encore très relative, à l'aune de nos standards actuels. L'utilisateur devait en particulier se familiariser avec un ou plusieurs langages informatiques. C'est sur cet aspect de l'informatisation que j'ai eu tout d'abord à me pencher. Le premier article que j'ai publié, en 1985, porte sur le choix d'un langage de programmation pour les études et l'enseignement littéraire<sup>1</sup> : Basic ou LSE. Au moment où le plan « Informatique Pour Tous » avait procuré des ordinateurs à tous les établissements scolaires, le manque de programmes adaptés (en particulier à l'enseignement des Lettres) apparaissait de manière

---

<sup>1</sup>. « Enseigner le Français avec l'ordinateur », *Bulletin de l'Association des Professeurs de lettres*, n° 34, juin 1985, pp. 38-44.

particulièrement vive et il devenait urgent de développer des applications capables de manipuler et de favoriser l'étude des textes, et même de corpus de textes assez étendus, tels que ceux que l'on avait l'habitude de traiter dans les études littéraires.

Mes premières expériences de programmation datent de cette période. J'ai en particulier proposé aux Editions Larousse, qui avaient lancé en 1985 un « Concours de didacticiens » pour le cinquantième anniversaire des « Petits classiques », un logiciel baptisé *Ariane* et qui permettait d'explorer le texte de *Sylvie* de Gérard de Nerval (recherches de formes, tableaux et courbes d'occurrences, co-occurrences). Je ne gagnai qu'un lot de consolation<sup>1</sup> mais cet apprentissage de la programmation me fut très utile par la suite. D'autres chercheurs littéraires, à la même époque, suivaient la même route : une équipe de l'université de Toronto développait le logiciel *TACT* (1986-89), Etienne Brunet programmat *Hyperbase* (sorti en 1989) et Pierre Muller *PISTES* (commercialisé par le CNDP en 1989), deux logiciels de lexicométrie permettant de traiter, de manière simple et conviviale, des corpus de texte assez étendus.

On ne doit pas être surpris de voir des chercheurs littéraires se convertir à la programmation. Celle-ci est trop souvent référée aux mathématiques, alors que la manipulation d'un langage artificiel a au moins autant de liens avec la linguistique et l'étude des langues naturelles. J'ai

---

<sup>1</sup> . J. Le Merlus, responsable du concours, répondit à ma demande d'explications : « Par ailleurs, la librairie Larousse n'envisage pas, pour le moment, de créer, aux fins d'un traitement lexicométrique, des banques de données concernant les textes classiques. » (lettre du 6 février 1986). On sait que Larousse essaya de lancer quelques années plus tard (1994) une telle collection de « Petits classiques électroniques », sans grand succès d'ailleurs (seulement six titres parus en 1994).

tenté, dans un article de 1989, « Texte et programme »<sup>1</sup>, de tracer un parallèle entre ces deux types d'activité : écrire un programme informatique et écrire un texte littéraire. L'écriture d'un programme met en jeu un style, des choix, une forme de logique particulière au programmeur. Il n'est pas étranger à la littérature d'écrire un texte en tenant compte de contraintes formelles, d'écrire pour produire un effet, d'obéir de la logique, de lire un texte pour en retrouver les principes et en apprécier l'organisation. Toutes ces aptitudes du programmeur sont aussi celles qu'enseigne et que pratique le professeur de français.

### **Enseignement des Lettres / enseignement de l'informatique**

Je conclus cet article en émettant le vœu que soient incluses dans l'enseignement littéraire les connaissances nécessaires à un dialogue entre informaticiens et littéraires. S'il me semble en effet indispensable, en général, de permettre à des spécialistes de domaines différents de pouvoir se comprendre et travailler ensemble, cette nécessité est encore plus forte s'agissant de deux domaines qui ne se conçoivent pas dans l'isolement. De même que le langage est toujours expression d'un référent qui lui est extérieur, l'informatique manipule des informations qu'elle ne crée ni n'interprète. Le spécialiste du langage qu'est le littéraire, comme l'informaticien, sont donc par nature des hommes de dialogue, mettant leurs techniques au service des projets d'autrui. Cela se vérifie dans les métiers auxquels ouvrent ces deux disciplines. Les professeurs de français et les professeurs de Lettres n'ont pas comme raison d'être de former des spécialistes de langue et de littérature. Les programmes scolaires leur assignent la tâche d'enseigner à une génération à lire, écrire, penser. Ces compétences s'exerceront dans tous les domaines de la vie sociale. De même, un

---

<sup>1</sup> . « Texte et programme », *Bulletin de l'Association Enseignement Public et Informatique (EPI)*, n° 60, décembre 1990, pp. 163-170.

informaticien n'est pas un pur programmeur, et une de ses qualités professionnelles les plus appréciées consiste à savoir analyser les besoins de ses clients en comprenant leur mode de fonctionnement et leur façon de penser.

Il m'a donc semblé, très tôt, que ces deux spécialités devaient pouvoir être combinées dans les cursus et les programmes d'enseignement. J'ai pu expérimenter dans ce domaine tout au long de ma carrière d'enseignant. Après avoir reçu une formation en informatique pédagogique dans le cadre de l'Éducation nationale, d'abord en 1985 à l'occasion d'un stage « Informatique pour Tous », puis, surtout, au cours d'un stage dit « lourd » (trois mois) en 1986-87, j'ai enseigné pendant plusieurs années (1989-1993) l'option d'informatique en lycée. Il y avait peu d'enseignants de Lettres dans ce cas, et cette particularité m'a permis d'attirer dans ce cours des élèves de série littéraire qui ont pu s'initier au maniement de quelques logiciels et à des rudiments de programmation.

Parallèlement, j'étais recruté comme chargé de cours à Paris 3 pour y assurer un cours « Enseignement du français et informatique » (1989-91), financé par la MAFPEN et destiné à initier des étudiants de Lettres (DEUG 2 et licence) aux applications pédagogiques de l'informatique. Voici la présentation qui en était faite dans les brochures :

L'objectif de l'UV est de sensibiliser des enseignants ou de futurs enseignants aux possibilités de l'Enseignement Assisté par Ordinateur et également de les rendre capables d'utiliser l'ordinateur de manière autonome.

- Utilisation de logiciels de type professionnel dans le cadre scolaire: traitement de texte (WORD 5), gestionnaire de base de données (DBASE III+), tableur (EXCEL).
- Mise en œuvre et évaluation de didacticiels.
- Culture générale informatique: histoire, matériels, aspects sociaux, juridiques et institutionnels de l'informatique scolaire.
- Pédagogie et informatique: évaluation, objectifs, soutien,...
- L'apport de l'informatique à la recherche littéraire: lexicométrie, banques de données,...

Ce cours fut transformé à partir de 1991 en un cours baptisé « Littérature et informatique »,

que j'assure encore aujourd'hui, et qui a permis à plusieurs centaines d'étudiants de Lettres de découvrir les outils que l'informatique met à la disposition des chercheurs en littérature. Voici par exemple le sujet du premier examen (1992) :

### 1) Exercice pratique

A l'aide d'un logiciel de votre choix (les traitements de texte WORD ou WRITE sous Windows, le logiciel d'étude des textes PISTES), vous étudierez dans le texte de *Sido* de Colette (dont le chemin complet sur le disque dur est C:\PISTES\SIDO.TEX) l'usage de la couleur bleue.

En citant quelques contextes significatifs, vous expliquerez brièvement (une dizaine de lignes) quelles connotations l'auteur associe à cette teinte.

### 2) Développement

En introduisant un recueil d'études textuelles réalisées à l'aide de l'informatique, Pierre Muller écrivait en 1984 : « Nous ne nous cachons pas pour autant les limites de l'entreprise : toute analyse est nécessairement réductrice. Le travail mené au seul niveau du vocabulaire ne saurait suffire à construire une interprétation. » (*Des textes avec ou sans ordinateur, « Rencontres pédagogiques », n°3, INRP, 1984*).

En prenant comme exemple des situations abordées en cours, vous direz ce que vous pensez des scrupules de ce chercheur et des limites de la lexicométrie dans le domaine littéraire.

Il s'agit, on le voit, de combiner des savoir-faire pratiques (manipulation de l'ordinateur, utilisation de logiciels, mise en œuvre de bases de données,...) et une réflexion théorique sur l'emploi de ces outils et sur leur impact dans les études littéraires. Ce cours, accueillant des étudiants en deuxième année de DEUG et en licence, entrait dans le cadre d'un véritable cursus d'informatique pour littéraires, voulu par Henri Béhar et mis en place par l'UFR de Littérature et Linguistique Françaises et Latines, qui permettait à un étudiant se suivre en première année de DEUG un enseignement d'initiation à l'histoire littéraire par l'utilisation de la Banque de Données d'Histoire Littéraire (dont je parlerai plus loin), puis de suivre le cours optionnel « Littérature et informatique » et, en troisième cycle, de choisir le séminaire dirigé par Henri Béhar et portant sur la même problématique.

Par ailleurs, j'enseigne depuis 1990 dans le DESS « Lettres appliquées aux techniques éditoriales et à la rédaction professionnelle » à Paris 3 (module *Nouvelles technologies*). Il s'agit cette fois de donner à des étudiants généralement issus de filières littéraires les connaissances nécessaires à leur insertion professionnelle dans le monde de l'édition (traitement de texte, PAO, bases de données, etc.)

Comme maître de conférences à l'IUFM de Paris (1993-98), j'ai eu également l'occasion d'assurer des enseignements combinant littérature et informatique, en particulier un cours optionnel pour les élèves professeurs du secondaire intitulé « Enseignement du français assisté par ordinateur »<sup>1</sup>. Voici par exemple le programme des séances de la première année (1993-94) :

Première séance :

- Familiarisation avec le matériel : vocabulaire, mise en route, problèmes techniques fréquents, les supports de stockage, les standards, les matériels utilisés dans l'E.N., le B A BA du MS-DOS.
- Utilisation de base du traitement de texte : le clavier, frappe au kilomètre, mise en forme sommaire, sauvegarde, impression.

Deuxième séance :

- Usages pédagogiques du traitement de texte : brouillon électronique, rédaction en groupe, micro-édition, résoudre les problèmes de graphie, la correction en temps réel.
- Exercices sur traitement de texte : exercices à trous, test de Closure, texte à remettre en ordre, pastiche, recherches thématiques dans le texte intégral de *Poil de Carotte*.

Troisième séance :

- Les didacticiels : présentation de quelques didacticiels de grammaire, orthographe, vocabulaire, lecture rapide, entraînement à la mémorisation.

---

<sup>1</sup> . Voir « Internet et enseignement du français. Un mariage harmonieux », *Questions d'école*, n° 4, septembre 1997, p. 62.

- L'évaluation et la mise en œuvre des didacticiels : grille d'analyse (ergonomie, documentation, principes, présentation, traitement des réponses), fonction du didacticiel dans une progression pédagogique.

#### Quatrième séance :

- Approche lexicométrique de l'étude des textes : utilisation du logiciel PISTES (CNDP) sur le texte de *Voyage au bout de la nuit*, utilisation des index, des spécificités, des concordances, aperçu des tendances actuelles de la recherche, utilisations pédagogiques.

#### Cinquième séance :

- Les banques de données : les divers supports, panorama des ressources, modes d'accès, constitution d'une base de données.
- Un exemple de banque de données : la Banque de Données d'Histoire Littéraire.

#### Sixième séance :

- Les principes de l'EAO : historique, les modes d'apprentissage, une nouvelle relation enseignant-enseigné, perspectives (matériels, logiciels, intelligence artificielle).

#### Septième séance :

- D'autres matériels : le nano-réseau, le Macintosh, le CD-ROM, les réseaux locaux, la télématique, notions sur l'hypertexte et l'hypermédia.

#### Huitième séance :

- Les outils de productivité personnelle de l'enseignant : gestion des élèves, production de documents, exemples d'utilisation des outils professionnels (traitement de texte, tableur, gestionnaire de base de données, grapheur).

#### Neuvième séance :

- La langue et l'informatique : le français et l'informatique, problèmes terminologiques, orthographiques.
- Travaux pratiques sur les vérificateurs d'orthographe et les dictionnaires électroniques.

#### Dixième séance :

- Bilan : mise au point d'une séquence de français incluant l'ordinateur (en vue de l'évaluation du module), la place de l'informatique dans la discipline.

Cette longue expérience de l'enseignement de l'informatique à des littéraires<sup>1</sup> m'a permis de confirmer l'hypothèse de départ : les formations littéraires et informatiques peuvent fort bien se combiner. Le contenu de ces formations a beaucoup changé en quinze ans. Alors qu'il fallait, en 1989, montrer aux étudiants, dont certains approchaient d'un ordinateur pour la première fois, comment se servir du clavier et de la souris, leur expliquer la notion de « fichier » ou de « fenêtre », l'usage de l'informatique s'est considérablement développé dans la population et les étudiants d'aujourd'hui, nés au début des années 80, ont tous une culture informatique suffisante pour que l'on n'ait plus à passer par ces étapes de l'initiation.

Mais s'il n'y a plus aujourd'hui d'*a priori* par rapport à la machine, l'usage de l'informatique et son mode de raisonnement rebutent encore beaucoup de littéraires, qui acceptent d'utiliser un traitement de texte pour écrire mais qui ne voient pas l'intérêt d'aller plus loin dans l'intégration de l'informatique à leur travail de recherche.

### **Informathématiques ?**

Une des raisons de cette méfiance tient à la proximité supposée entre informatique et mathématiques. Il faut dire que l'organisation des études informatiques ne peut que conforter cette idée. Toutes les formations à l'informatique comportent, généralement au début, une forte proportion de mathématiques, ce qui a comme effet de dissuader tous ceux qui n'ont pas de solides préalables dans ce domaine. Les étudiants littéraires, généralement rétifs aux mathématiques, font un amalgame entre les « calculateurs électroniques » (premier nom des *computers*) et la science des nombres.

---

<sup>1</sup> . Voir « Internet et enseignement du français. Un mariage harmonieux », *Questions d'école*, n° 4, septembre 1997, p. 62.

Si l'on peut contester la nécessité d'avoir une formation avancée en mathématiques pour programmer les ordinateurs, il n'en est pas moins vrai que l'informatique requiert des compétences et une tournure d'esprit que, malheureusement, notre tradition pédagogique a cantonné, dans les dernières décennies, à l'enseignement scientifique. Tout le sens de l'action que je mène depuis plus de quinze ans est de montrer qu'un bon étudiant de Lettres est aussi apte qu'un bon étudiant scientifique à utiliser habilement un ordinateur, à écrire des programmes, à entrer dans la logique de la machine pour concevoir un algorithme et en juger les résultats, à manipuler des abstractions et un langage artificiel.

En revanche, l'application de l'informatique à certains domaines spécialisés requiert des connaissances mathématiques spécifiques. Je prendrai l'exemple des statistiques. Les littéraires, de manière générale, en usent peu et en usent mal. J'ai montré, à l'occasion d'une communication à un colloque de l'Institut International de Statistique en 1999<sup>1</sup>, que les statistiques mises en œuvre dans les études de métrique sont la plupart du temps indigentes et souvent erronées. Le calcul d'un pourcentage ou l'emploi d'une règle de trois semble souvent être le *summum* de l'outillage statistique dans les études littéraires. La lexicométrie, qui use de procédés bien plus complexes, paraît souvent hors de portée à des chercheurs en littérature qui abandonnent ce type d'approche à la linguistique quantitative, dont nombre de chercheurs ont une formation initiale de mathématiciens. Cette attitude est dommageable à une discipline qui a voulu naguère s'ériger en « science du littéraire » ou « science des textes et documents ». Même s'il n'est pas de science que du quantifiable, une approche scientifique requiert souvent

---

<sup>1</sup> . « Mesurer le mètre : poésie et statistiques », Actes de la 52<sup>e</sup> Session de l'Institut International de Statistique, 10-18 août 1999, Helsinki, *Bulletin de l'Institut International de Statistique*, Tome LVIII, 1999. Version en ligne : <<http://www.stat.fi/isi99/proceedings/arkisto/varasto/bern0811.pdf>>

des comptages efficaces et des statistiques sécurisées. Le refus *a priori* de tout chiffre renvoie fatalement les études littéraires à leur vieille réputation de subjectivité et d'approximation.

Pis encore, cette dichotomie entre ce qui relèverait de l'« esprit de géométrie » et ce qui appartiendrait à l'« esprit de finesse » aboutit à cette opposition fallacieuse que résume la boutade courante « sciences humaines et sociales *vs.* sciences inhumaines et associées » (ou, de l'autre point de vue, « sciences exactes et appliquées *vs.* sciences inexactes et inapplicables »...) Les deux branches de la connaissance ont à souffrir de cette étanchéité de leurs domaines, par l'effet de ce dénigrement réciproque mais aussi parce que le manque d'information sur les modes de raisonnement différents interdit le travail en commun et la réflexion épistémologique sur ses propres formes de pensée. Si l'on veut se référer à Pascal, qu'on le cite au moins en entier :

« Les géomètres qui ne sont que géomètres ont donc l'esprit droit, mais pourvu qu'on leur explique bien toutes choses par définitions et principes; autrement ils sont faux et insupportables, car ils ne sont droits que sur les principes bien éclaircis.

Et les fins qui ne sont que fins ne peuvent avoir la patience de descendre jusque dans les premiers principes des choses spéculatives et d'imagination, qu'ils n'ont jamais vues dans le monde, et tout à fait hors d'usage. »<sup>1</sup>

La monoculture, dans chacun des deux champs disciplinaires, donne des résultats étriqués, difficilement communicables. Il existe cependant des formations universitaires qui combinent les deux approches. J'ai fait en 2001, au Colloque « Computers, Literature and Philology

---

<sup>1</sup> . *Pensées*, édition Lafuma, 512-1.

(CLiP 2001) » une communication<sup>1</sup> qui présentait un panorama de toutes les formations de type « Littérature et informatique » en Europe. On en trouve un assez grand nombre mais elles sont aujourd'hui dispersées, parfois en perte de vitesse (le bi-DEUG « Lettres et Informatiques » de Paris VIII a par exemple disparu), souvent noyées dans des cursus du type « Informatique pour les sciences humaines ». Il serait intéressant de fédérer toutes ces initiatives au niveau européen, dans le cadre des réformes en cours, pour proposer par exemple un master « Littérature et informatique » qui permettrait aux étudiants d'acquérir des bases solides dans les deux domaines dans plusieurs universités européennes. Une autre initiative pédagogique consisterait à introduire dans le premier cycle de Lettres un enseignement de statistiques imité de celui qui est dispensé aux étudiants en histoire<sup>2</sup>.

### **Enseignement à distance**

Il faudrait ajouter à ces propositions une autre voie d'accès aux ressources informatiques pour les étudiants littéraires, et qui passe par les techniques de l'enseignement à distance. J'enseigne depuis 1999 à Télé 3, le service d'enseignement à distance de Paris 3, qui connaît une mutation profonde en raison du passage du stade de l'enseignement par correspondance à l'ère numérique. C'est ainsi que le cours de poétique que j'ai d'abord donné sous forme d'un polycopié envoyé aux étudiants est aujourd'hui disponible en ligne<sup>3</sup>, avec des avantages certains : possibilité de mise à jour en continu, illustrations en couleur, liens vers des

---

<sup>1</sup> . Colloque *Computers, Literature and Philology (CLiP 2001)*, « Philologien und Informationstechnologien », Gerhard-Mercator-Universität, Duisbourg, 6-9 septembre 2001. Communication : « Littérature et informatique à la Sorbonne-Nouvelle et dans quelques universités françaises ». A paraître.

<sup>2</sup> . Voir par exemple l'ouvrage de Pierre Saly, *Méthodes statistiques descriptives pour les historiens*, (Armand Colin, 1991) pour avoir une idée du contenu d'un tel enseignement.

<sup>3</sup> . < [http://www.tele3.univ-paris3.fr/lm/dlfftp1/ec\\_presentation.htm](http://www.tele3.univ-paris3.fr/lm/dlfftp1/ec_presentation.htm)>.

ressources en ligne, dialogue par courrier personnel ou forum, etc. J'assure aussi la mise à distance de mon cours « Littérature et informatique », sous la forme d'un enseignement consistant en un dialogue avec les étudiants, qui doivent effectuer un certain nombre de tâches (lectures, information, travaux de recherche) et m'envoyer les résultats pour évaluation.

Je ne m'appesantirai pas sur les aspects techniques et pédagogiques de cette expérience d'enseignement à distance mais il est important pour la question débattue ici d'analyser ce type de médiation pédagogique comme un moyen particulièrement efficace de familiarisation des étudiants littéraires à l'informatique. L'ordinateur et l'Internet se présentent ainsi comme de simples moyens de communication et des ressources documentaires. La pénétration de plus en plus grande de l'informatique dans les équipements et les pratiques quotidiennes, spécialement chez les plus jeunes, favorise une appropriation de l'ordinateur qui, dans un contexte comme celui de l'enseignement à distance, trouve immédiatement sa justification pratique. Même si les cours diffusés en ligne par Télé 3 ont aujourd'hui, fondamentalement, le même aspect que leurs référents de papier, le simple fait de les avoir numérisés et de les transmettre par Internet introduit chez les enseignants et chez les étudiants un rapport à l'informatique nouveau. Qu'il s'agisse en l'occurrence d'un enseignement de Lettres (Télé 3 permet de suivre tout le cursus jusqu'à la maîtrise de Lettres modernes) n'est évidemment pas indifférent à mon propos.

L'enseignement à distance, comme le traitement de texte ou le courrier électronique, sont des applications de l'informatique qui ne suscitent généralement pas de réserves de la part des littéraires, parce que leurs avantages et leur facilité d'utilisation sont assez importants pour vaincre les réticences et les *a priori*. De fait, on constate aujourd'hui un usage massif, dans les études littéraires, de ces moyens de communiquer. Un chercheur en littérature serait à l'heure actuelle particulièrement handicapé dans son travail s'il ne disposait pas d'une adresse

électronique et d'un traitement de texte. Nous avons vu, au cours des années 90, monter cette exigence de la part des éditeurs, des organisateurs de colloques, des services administratifs. L'enseignement à distance (qui suppose, d'ailleurs, la maîtrise des deux autres outils) concerne davantage l'enseignement que la recherche mais il entre aussi progressivement dans les habitudes de travail.

J'ai eu aussi une autre expérience intéressante à cet égard en assurant la coordination nationale des Journées d'Agrégation en Ligne<sup>1</sup> (depuis 2002). Il s'agit d'un réseau de sept universités française qui organisent chacune une journée d'étude sur un des points du programme de l'agrégation de Lettres. Les journées sont diffusées en visioconférence à destination de l'ensemble du réseau, ce qui permet aux étudiants et aux enseignants de suivre en direct les communications et les tables rondes, de dialoguer avec les intervenants et de leur poser des questions. Les enregistrements vidéo et les textes des communications sont ensuite stockés sur un serveur et sont disponibles *via* le Web. Une organisation de ce type, assez proche de la logique de l'enseignement à distance, est un moyen commode pour valoriser les efforts considérables (et bénévoles...) que consentent les organisateurs de journées d'agrégation et pour offrir des contenus de qualité à des étudiants dont certains sont éloignés des centres universitaires. On peut imaginer que ces techniques de visioconférence deviendront de moins en moins onéreuses et qu'elles faciliteront à l'avenir l'organisation des enseignements et des colloques scientifiques. Des expériences de ce type permettent aussi de familiariser les enseignants et les étudiants avec divers aspects du numérique.

---

<sup>1</sup> . On trouvera toutes les indications utiles sur cette organisation sur le site du Centre de recherches Hubert de Phalèse : < <http://www.cavi.univ-paris3.fr/phalese/>>.

Cependant, il ne s'agit là encore que d'un outil de communication, comme le téléphone ou la poste. Les ELAO ne se résument pas à cette culture informatique, de plus en plus diffusée mais qui ne constitue pas en soi une approche scientifique, et qui ne fournit pas aux littéraires d'outils spécifiques adaptés à leurs besoins. S'il est important de constater que l'usage quotidien de l'ordinateur ne peut que faciliter l'appropriation de l'informatique par les littéraires, il nous faut maintenant examiner ce qui relève proprement des ELAO, c'est-à-dire ce qui peut changer les méthodes et les concepts des études littéraires.

# Une base de données d'histoire littéraire

La Banque de Données d'Histoire Littéraire (BDHL) a été initiée par une équipe d'enseignants-chercheurs de l'université Paris 3 sur l'impulsion d'Henri Béhar<sup>1</sup>. Les premiers travaux remontent à 1985 et c'est un premier état déjà fonctionnel que j'ai découvert en rejoignant l'équipe de la BDHL en 1988. La BDHL avait été initialement développée sous le Système de gestion de base de données Multilog. Les données tenaient sur une disquette (360 Ko), le système sur une autre. Ma première contribution a consisté à écrire un programme *ad hoc* qui permettait de faire tenir l'ensemble de la BDHL sur une seule disquette avec un gain de temps de réponse substantiel. Cette version, baptisée « BDHL Junior », a longtemps été utilisée dans le cadre du cours d'initiation à l'histoire littéraire des étudiants de première année de DEUG de Paris 3 auquel j'ai fait allusion plus haut. Elle comportait, outre les fonctionnalités de recherche qui avaient déjà été implémentées depuis l'origine, des jeux permettant d'aborder les données de la banque sur le mode ludique, en particulier en mesurant combien 750 auteurs et 2 000 œuvres pouvaient dépasser les connaissances de quiconque.

---

<sup>1</sup> . Henri Béhar a rassemblé dans *La Littérature et son golem* (Honoré Champion, 1996) plusieurs articles qui racontent les débuts de la BDHL et exposent les principes qui y ont présidé (voir en particulier p. 21 et p. 31).

## Une nouvelle histoire littéraire

La BDHL, créée au moment où s'affirmait la nécessité d'une nouvelle histoire littéraire, voire même d'un retour à l'histoire littéraire<sup>1</sup>, était destinée à renouveler les pratiques dans ce domaine controversé des études littéraires. En s'affranchissant de la périodisation, mais aussi de la hiérarchisation des auteurs, des genres, des mouvements, l'histoire littéraire semblait pouvoir trouver ainsi des bases nouvelles, permettre des rapprochements inédits, se présenter comme une science. Dès le départ, la BDHL a été conçue à la fois comme un outil de recherche et comme un outil pédagogique. C'est d'ailleurs dans une revue destinée aux enseignants, la revue de l'EPI, que j'ai publié en 1988 un premier article<sup>2</sup> présentant la BDHL. Deux autres articles<sup>3</sup>, parus en 1992 et 1993, visent à montrer l'ensemble des possibilités de la BDHL, et surtout en quoi elle représente une tentative de repenser l'histoire littéraire.

En effet, la simple numérisation des données, et plus encore la possibilité de les manipuler, d'en dresser des statistiques, renouvelle la manière d'aborder l'histoire littéraire. La discipline imposée par l'informatisation oblige à s'interroger sur les objets manipulés. *Auteur, œuvre,*

---

<sup>1</sup> . Voir par exemple l'introduction du volume *L'Histoire littéraire aujourd'hui*, sous la direction d'Henri Béhar et Roger Fayolle, Armand Colin, 1990.

<sup>2</sup> . « La Banque de Données d'Histoire Littéraire », *Bulletin de l'Association Enseignement Public et Informatique (EPI)*, n° 51, septembre 1988, pp.172-177.

<sup>3</sup> . (avec H. Béhar et J.P. Goldenstein) : « La banque de données d'histoire littéraire. Principes, pédagogie, perspectives », 1992, *Texte* (Trinity College, University of Toronto), n°12.

« La Banque de Données d'Histoire Littéraire », in *Les Banques de données littéraires - Comparatistes et francophones*. Actes des Journées d'étude sur les Banques de Données Littéraires Francophones (Université de Limoges), 1993.

*titre, dates, thèmes, genre* : aucune de ces notions ne paraît évidente dès lors qu'il s'agit de les faire entrer dans les tables d'une base de données. C'est d'ailleurs là que se situe l'apport le plus net de l'informatique à l'ensemble des sciences humaines, qu'elle contraint à formaliser davantage ses concepts et ses procédures. C'est l'occasion d'une remise à plat, d'une refondation que la théorie littéraire avait amorcée depuis longtemps et que la constitution de bases de données permet de relancer dans un esprit pragmatique et concret.

## **Thématique**

C'est sur un de ces concepts de l'histoire littéraire, celui de *thème*, que j'ai fait porter ma thèse de doctorat (*Elaboration d'un thésaurus pour l'indexation thématique d'œuvres littéraires*<sup>1</sup>, publiée en 1994 chez Champion-Slatkine<sup>2</sup>). Dans le but de pouvoir coder les œuvres de la BDHL en indiquant leurs thèmes principaux, et permettre ainsi des requêtes sur les œuvres traitant de tel ou tel thème, j'ai entrepris de constituer un thésaurus d'environ 800 termes, répondant aux normes documentaires et facilitant le travail des indexeurs. Un tel travail suppose en amont une réflexion sur ce qu'il convient d'entendre par « thème littéraire », et suscite en aval des études sur les résultats d'une telle indexation. C'est par ce va-et-vient entre théorie et pratique que se caractérisent les ELAO. Rechercher, par exemple, les œuvres traitant de la jalousie suppose que l'on ait identifié les œuvres de la BDHL à l'aide de descripteurs thématiques, mais aussi, préalablement, que l'on ait défini ce qu'il faut entendre par « le thème de la jalousie ». Après une première vague d'indexation, on examinera le

---

<sup>1</sup> . Soutenue en juin 1992 à Paris 3, sous la direction de Henri Béhar.

<sup>2</sup> . *De quoi parle ce livre ? Elaboration d'un thésaurus pour l'indexation thématique d'œuvres littéraires*, Champion, 1994, 365 p.

résultat de la requête, quitte à préciser ensuite les définitions et à affiner l'indexation, et ainsi de suite.

Une des pistes explorées consistait à extraire automatiquement les mots clés thématiques du texte littéraire lui-même, en usant de méthodes mises au point par les documentalistes. Toutes les approches tentées ont échoué pour le texte littéraire, qui suppose une lecture et dont la thématique est beaucoup plus complexe que le sujet d'un texte documentaire. Cet échec illustre une autre particularité de la méthode expérimentale introduite dans les études littéraires par l'informatique. Il est en effet fréquent d'arriver à des impasses, de constater l'irréductibilité du texte littéraire aux procédures courantes dans le domaine documentaire. Les esprits chagrins ont même beau jeu de démontrer que l'informatique est inopérante dans le contexte de la littérature. Mais cette difficulté devient un atout et une source heuristique de premier ordre, dès lors que l'on s'affranchit du point de vue positiviste qui consisterait à attendre de l'informatique la résolution de questions qui ne ressortissent qu'à une réflexion épistémologique.

Les recherches relevant des ELAO aboutissent souvent à constater l'imperfection des outils informatiques pour analyser les faits littéraires. Mais cette impossibilité doit conduire, plutôt qu'à des anathèmes assez stériles (qui peuvent d'ailleurs, symétriquement, être le fait des informaticiens comme des littéraires), à la nécessité d'améliorer d'abord l'outillage informatique, et ensuite de penser le littéraire de manière plus adéquate. Ainsi, la confrontation entre la notion, si subjective, si polysémique, si floue, si *littéraire* en un mot, de « thème », et la rude logique d'une base de données m'a permis à la fois d'affiner la procédure de requête de la BDHL (élimination des synonymes, recours au thésaurus, recherches croisées, recherches de thèmes co-occurents) et de réfléchir sur la notion de « thème littéraire », de retrouver une logique parmi les multiples acceptions qu'on lui donne.

## Le point de vue des lecteurs

La plupart des travaux menés sur la BDHL retrouvent les positions théoriques de l'esthétique de la réception. Il ne s'agit pas d'un *a priori* mais du constat, maintes fois renouvelé, que seul ce point de vue sur le phénomène littéraire permet de sortir des apories où nous plongerions d'autres manières de penser. Le recours à l'informatique nous interdit en effet de concevoir la littérature autrement que comme un phénomène social dont l'historien ne peut que constater les manifestations. La question du « thème » en est un bon exemple. Il est en effet impossible de définir autrement ce concept que comme « ce que l'on entend habituellement par *thème* ». De même, un « grand auteur » est défini par l'opinion d'un lectorat donné et la reconnaissance des institutions, le nom qui le désigne est celui qu'on lui donne habituellement, le genre d'une œuvre est celui qui lui est généralement attribué, etc. Cette espèce de démagogie critique est tempérée par la démarche scientifique qui consiste à étudier l'opinion publique comme un fait objectif, mesurable.

Adopter le point de vue du lecteur revient à effectuer la troisième étape d'un mouvement dialectique qui a amené la critique à baser d'abord ses considérations et ses appréciations sur l'auteur, dont le génie était la pierre de touche de la valeur de son œuvre, puis sur le texte, dans lequel, par réaction, on s'est efforcé de chercher les principes immanents de la valeur littéraire. Mais ni les intentions de l'auteur ni les effets formels du texte ne permettent de résoudre les problèmes auxquels le concepteur d'une base de données se trouve confronté. Il est dans tous les cas plus sûr de considérer le phénomène littéraire comme une représentation, une construction dont on pourra étudier les caractéristiques et les variations. L'ouvrage inédit présenté avec ce dossier (*L'histoire littéraire au risque de l'informatique : la question du canon littéraire*) porte sur un des aspects de cette question, celui de la constitution du canon des « grands auteurs de la France ». À travers l'étude des manuels et ouvrages de références,

les variations du palmarès sont mises en évidence. Elles permettent de suivre l'évolution d'une *doxa* dont on pourra ensuite tenter de déduire la logique. C'est donc à travers le regard porté sur l'histoire littéraire que celle-ci est étudiée. Il s'agit d'une étude de deuxième degré, qui consiste à entériner le caractère de *phénomène* du fait littéraire<sup>1</sup> et à l'analyser en tant que tel, comme une construction et comme une représentation.

Cette démarche a aussi amené les concepteurs de la BDHL à proposer aux spécialistes d'autres littératures d'utiliser le modèle conceptuel et informatique de la BDHL pour réfléchir sur les différences d'approche de réalités diverses. Ainsi, à l'occasion de deux missions à l'université Laval (Québec) en avril 1995 et septembre 1996, j'ai eu l'occasion de travailler avec Clément Moisan et son équipe pour mettre au point une base de données de littérature québécoise. Ce projet n'a pas été entièrement réalisé, en raison de réductions budgétaires du côté québécois, sauf pour la partie immédiatement contemporaine, mais il a donné lieu à de fructueux échanges sur les différences interculturelles dans l'approche de nos histoires littéraires respectives. Les thèmes sélectionnés dans mon thésaurus, par exemple, se sont révélés dans bien des cas inutilisables pour indexer la littérature québécoise, ou les Québécois souhaitaient introduire dans la base des informations spécifiques : date d'immigration des auteurs, par exemple, ou pays d'origine, etc. Par ailleurs, un chercheur ivoirien, Bruno Gnaoulé-Oupoh, a proposé une adaptation de la BDHL au domaine littéraire africain<sup>2</sup>.

En dehors de ces applications à d'autres domaines francophones, une coopération (CAPES / COFECUB) est en cours avec une équipe brésilienne dirigée par Alckmar Dos Santos, professeur à l'Université Fédérale de Santa Catarina, pour adapter la BDHL à la littérature

---

<sup>1</sup> . Cf. Clément Moisan, *Le Phénomène de la littérature*, L'Hexagone, « Essais littéraires », 1996.

<sup>2</sup> . Son article est en ligne à l'adresse suivante : <<http://www.cavi.univ-paris3.fr/phalese/bdhli.htm>>.

brésilienne. J'ai également, lors d'une mission effectuée en 1998, noué des contacts avec une enseignante portugaise (professeur Herminia Amado Laurel, université d'Aveiro) qui a fait travailler ses étudiants sur les œuvres en français présentes dans les bibliothèques scolaires portugaises. Il s'agit cette fois de prendre en compte des regards étrangers sur la littérature française, pour évaluer son retentissement au fil des siècles et l'image qu'elle a eue.

### **Recherches en histoire littéraire avec la BDHL**

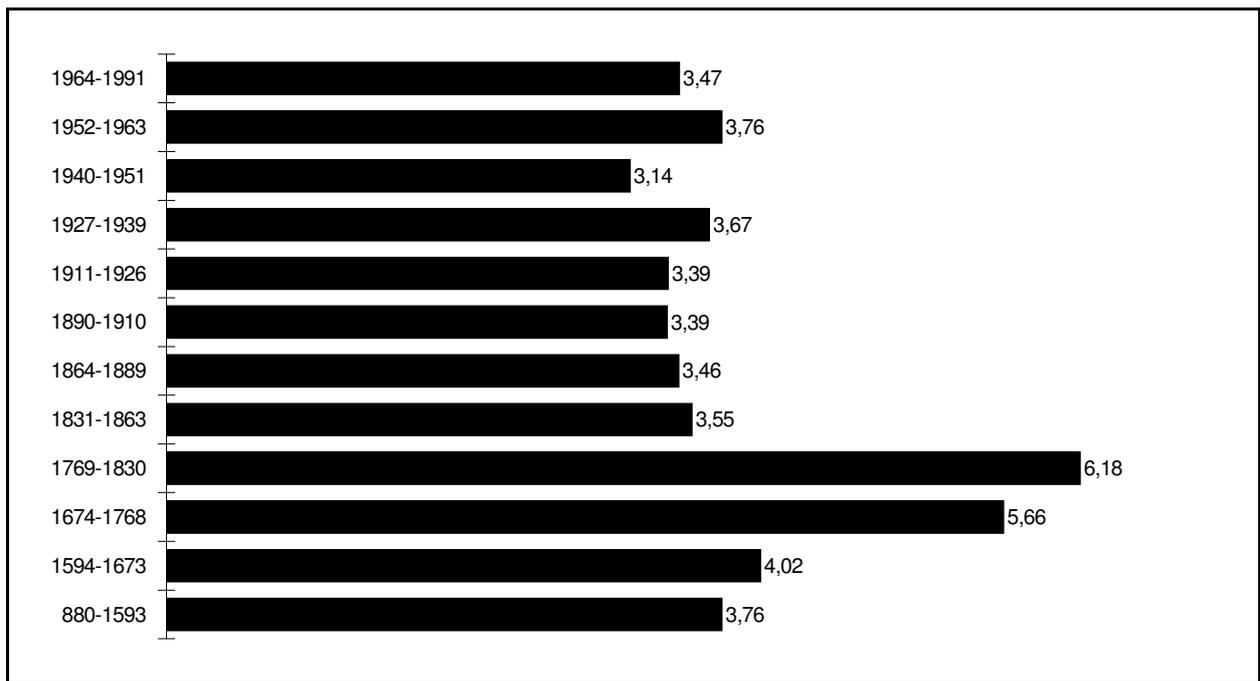
La BDHL a donné lieu également à des travaux de recherche plus spécifiques, qui prennent appui sur la BDHL pour fournir des aperçus statistiques et transversaux sur l'histoire littéraire. J'ai ainsi apporté une contribution à la titrologie dans une communication présentée lors d'un colloque organisé par la Maison française à Oxford en octobre 1993, « À juste titre. Une approche lexicométrique de la titrologie »<sup>1</sup>. Il s'agit d'étudier tous les titres des œuvres indexées dans la BDHL avec les méthodes de la lexicométrie. L'avantage de procéder à partir d'une base de données réside dans la possibilité de sérier les titres par période, par genre, par origine géographique des auteurs, etc. Il est possible de mettre en évidence des constantes, des caractéristiques des titres selon ces partitions. Une recherche de ce type a surtout la particularité de signaler des faits difficilement perceptibles autrement. C'est là une innovation des ELAO que de constituer ainsi en faits des mises en relation qu'un lecteur ne peut généralement pas établir.

Prenons l'exemple de ce constat : le XIX<sup>e</sup> siècle affectionne les titres commençant par *les*. Seule la machine peut repérer de tels faits. Est-ce à dire qu'il ne s'agit pas d'un *phénomène*,

---

<sup>1</sup> . « À juste titre : a lexicometric Approach to the Study of Titles », *Literary and Linguistic Computing*, vol. 10, n°2, 1995, pp. 135-141. Publié en français dans la revue en ligne *Lexicometrica* (n°1) : <<http://www.cavi.univ-paris3.fr/lexicometrica/article/numero1/bernard.htm>>.

puisque un lecteur humain ne peut s'en rendre compte, et donc que cela ne mérite pas l'étude, en vertu des principes énoncés plus haut ? On peut répondre dans un premier temps que les études littéraires, dans leur ensemble, ont justement comme projet d'exacerber les capacités d'attention du lecteur banal, et donc de mettre en évidence des faits qu'une lecture non professionnelle ne détecte pas toujours. Les aperçus de la narratologie ou de la psychocritique, par exemple, peuvent surprendre le lecteur amateur, voire l'irriter, parce qu'elles pointent dans le texte des faits qu'il n'y avait pas lus. Il en est de même pour les ELAO, qui appliquent au texte des méthodes spécialisées, éloignées de l'usage courant. D'autre part, un *phénomène* littéraire, pour n'être pas consciemment perçu par le lecteur, n'en est pas moins opérant. L'usage de l'article défini pluriel dans les titres du XIX<sup>e</sup> siècle est peut-être indicatif d'une tendance à présenter des fresques sociales (*Les Misérables*, *Les Rougon-Macquart*, *Les Maîtres-sonneurs*, etc.), ou des séries (*Les Orientales*, *Les Châtiments*, *Les Chimères*, etc.), à faire nombre dans l'annonce quasiment commerciale que constitue un titre, à suggérer que le livre contient tout un monde, décline plusieurs possibilités. Ces impressions conférées par les titres sont bien éprouvées par le lecteur, mais sans qu'il se rende forcément compte qu'elles sont produites par l'article. Beaucoup de faits mis en évidence par les procédés de la lexicométrie ou par l'analyse statistique des bases de données sont de cet ordre. Le lecteur cultivé reconnaîtra instinctivement un titre du XVIII<sup>e</sup> siècle, ou du Moyen-Âge, mais sans savoir dire ce qui lui permet cette reconnaissance. Les phénomènes étudiés par les ELAO sont donc parfois imperceptibles à la lecture, mais l'hypothèse peut être formulée qu'ils agissent néanmoins sur les lecteurs. La même discussion pourrait être menée en ce qui concerne les intentions de l'auteur, qui peut user de procédés mis en évidence par l'ordinateur mais sans les formuler de manière consciente. Il en est ainsi de la longueur des titres (en nombre de mots moyen) :



Il est évident qu’aucun auteur n’avait conscience d’une norme dans ce domaine, mais le fait est que des modes existent, et que les statistiques les mettent en évidence.

Un autre travail a été mené sur la base de l’indexation thématique de la BDHL. Il consistait à étudier le thème des *sentiments* dans les romans entre 1830 et 1970<sup>1</sup>. Ce travail met en exergue une autre potentialité des ELAO, qui permettent de travailler sur de très gros corpus, ce qui est inenvisageable autrement. Les 534 romans recensés par la BDHL pour cette période peuvent être étudiés du point de vue des mots clés que l’on a sélectionnés comme désignant des sentiments. La méthode qui préside à tous les travaux de la BDHL a été suivie : pour définir ce qu’est un « sentiment », j’ai compilé plusieurs listes qui se présentent comme des

---

<sup>1</sup> . H. Béhar et M. Bernard : « La nébuleuse des sentiments », *L’Analyse thématique des données textuelles. L’exemple des sentiments*, dirigé par François Rastier, publié par Eveline Martin, Didier érudition, 1995, pp. 53-84.

listes de sentiments, et j'ai conservé les 25 termes le plus fréquemment cités : amour, admiration, affection, jalousie, orgueil, etc. Les statistiques portent ensuite sur la présence diversifiée de ces mots clés dans l'indexation des romans du corpus, et pointent sur des faits qui peuvent intéresser l'histoire littéraire : les sentiments co-occurent le plus fréquemment, par exemple, sont amour/amitié et amour/jalousie, beaucoup plus fréquemment qu'amitié/jalousie ou amour/angoisse, etc. De tels constats sont relatifs à la qualité de l'indexation et au choix des œuvres dans la BDHL, mais ils ont une valeur heuristique certaine, en signalant des pistes de réflexion qui peuvent être ensuite suivies de manière plus précise.

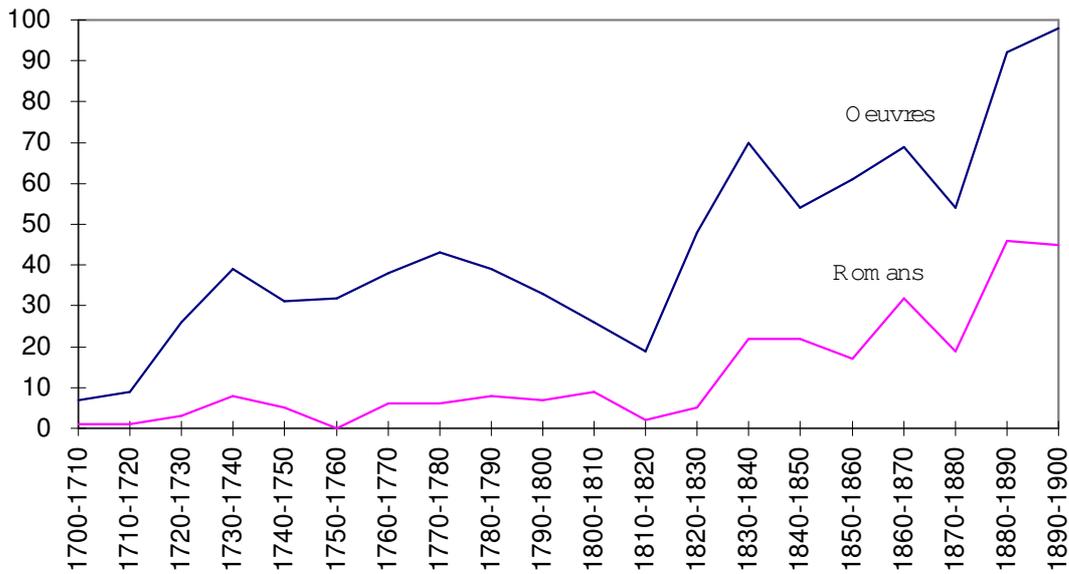
Ces deux points sont à souligner, du point de vue méthodologique qui est le mien dans le cadre de ce dossier :

- Les ELAO ont souvent comme résultat de signaler des pistes de recherche, plutôt que d'effectuer l'ensemble d'une étude. En traitant de vastes collections de données, l'ordinateur et les statistiques mettent en évidence des faits quantitatifs qu'une étude qualitative plus fine doit permettre de cerner de plus près. De même que l'archéologie aérienne ne dispense pas les archéologues de monter des expéditions pour aller se rendre compte sur place de ce qui a été repéré au sol, l'informatique se contente souvent de désigner certaines particularités des corpus à la curiosité de chercheurs qui peuvent ainsi gagner du temps en concentrant leurs efforts sur les zones les plus prometteuses. La BDHL peut ainsi suggérer à qui voudrait étudier les sentiments dans les romans des XIX<sup>e</sup> et XX<sup>e</sup> siècle une liste d'œuvre où ces thèmes sont particulièrement nombreux, ou bien des œuvres dans lesquelles on trouve des co-occurrences de sentiments particuliers, ou signaler la corrélation entre telle particularité biographique et la présence de tel sentiment dans l'œuvre, etc.

- La qualité des résultats obtenus dans le traitement statistique d'une base de données est éminemment liée à la qualité de la numérisation, de la saisie, aux conventions qui ont présidé à l'indexation de la base, aux choix qui ont été effectués. Le chercheur doit être très critique et très exigeant à cet égard. Le prestige de l'informatique est souvent tel que l'on prend pour argent comptant tout ce qui est présenté sous son patronage, et les journaux (ou les revues savantes, hélas...) sont remplis de graphiques illusoire ou de calculs scabreux. Pour prendre un exemple cité dans l'article sur la titrologie, il ne faut pas s'émerveiller que le mot « prononcée » soit un des termes spécifiques des titres du XVII<sup>e</sup> siècle, et encore moins échafauder là-dessus des hypothèses sur le goût du Classicisme pour le langage : l'explication réside tout simplement dans la présence dans le corpus des titres de sept « oraisons funèbres » de Bossuet, toutes « prononcées en l'église de... » Ce type d'artéfact est très commun dans le traitement de bases de données textuelles, et il faut s'en méfier. De même, il ne faut jamais oublier que tous les chiffres obtenus à partir du traitement des données de la BDHL sont relatifs à la sélection opérée au départ et au choix de n'intégrer que des œuvres qui figuraient dans un grand nombre de manuels. On ne peut s'étonner, par exemple, que les œuvres des dix dernières années y soient mal représentées.

Mais, en prenant ces précautions, la BDHL peut se révéler un outil irremplaçable pour l'étude des grands corpus. J'ai ainsi proposé, dans le cadre d'une journée d'étude intitulée « De nouveaux outils pour la recherche littéraire », organisée par Nathalie Ferrand à l'École Normale Supérieure en janvier 1995, une étude encore plus large sur « le roman des XVIII<sup>e</sup> et

XIX<sup>e</sup> siècles »<sup>1</sup>. Une étude de ce genre a également pour objet de constater des variations de grande ampleur sur le long terme. Voici par exemple un graphique qui représente le nombre de romans et la production complète sur toute la période 1700-1900, à partir des données de la BDHL :

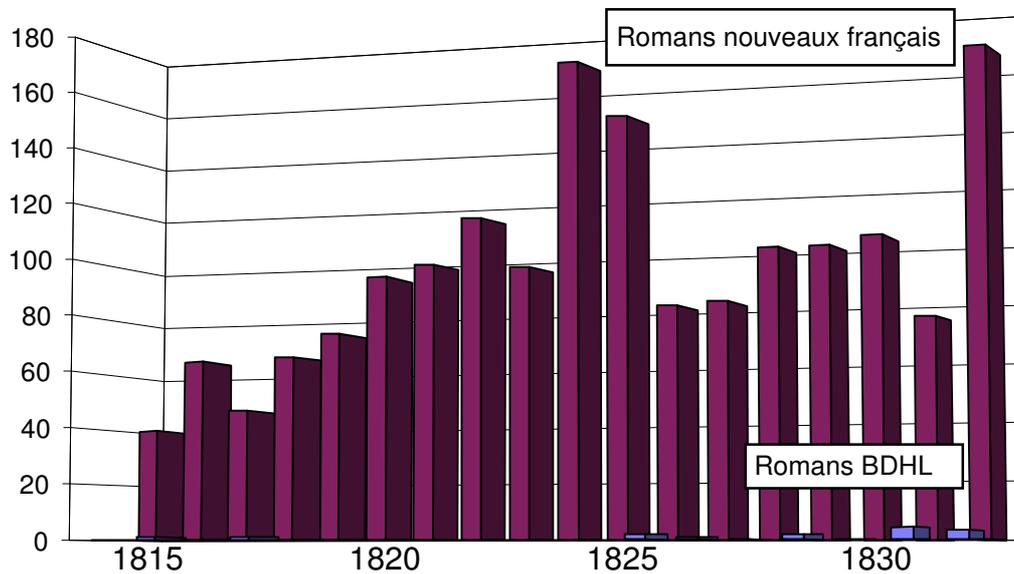


On retrouve là une notion connue, celle de l'expansion de la production romanesque aux XIX<sup>e</sup> et XX<sup>e</sup> siècles (ce qui a motivé le découpage chronologique de cette étude). Cependant, la comparaison des deux courbes rappelle que la croissance du roman se fait dans le cadre d'une augmentation de l'édition dans son ensemble d'une part, et d'autre part que la BDHL, comme toutes les histoires littéraires, retient d'autant plus d'œuvres qu'elles sont plus proches de nous. Pour visualiser ces deux faits, on peut utiliser tout d'abord un graphique qui montre ce

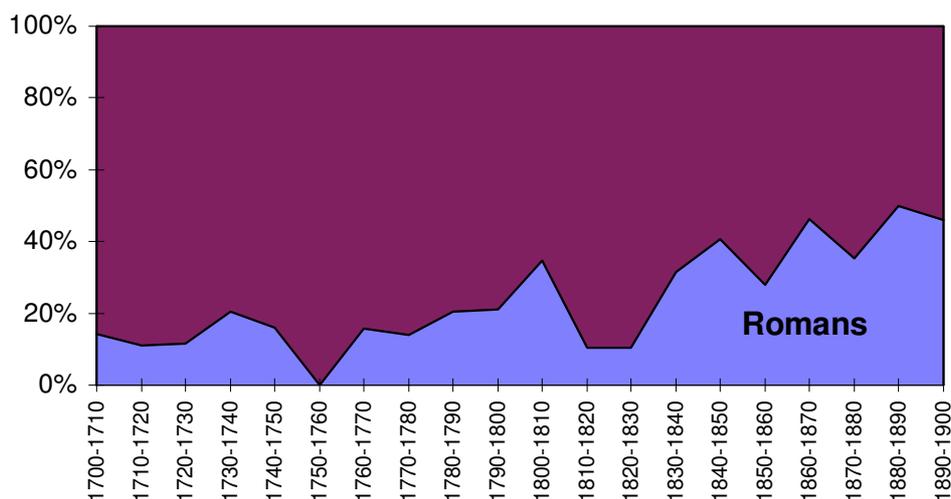
---

<sup>1</sup>. « Du *Diable boiteux* au *Jardin des supplices*, le roman dans la Banque de Données d'Histoire Littéraire », in *Banques de données et hypertextes pour l'étude du roman*, Nathalie Ferrand dir., PUF « Écritures électroniques », 1997.

que représentent les romans retenus dans la BDHL par rapport à l'ensemble de la production entre 1815 et 1832, soit 14 romans sur 1 717 :



La BDHL, n'enregistrant que la littérature consacrée, ne peut donc se prétendre représentative de l'ensemble de la production éditoriale. Par ailleurs, il est possible de visualiser la proportion des romans dans l'ensemble des œuvres retenues par la BDHL :



Ce graphique évite les défauts que présentait la courbe affichée plus haut. Elle montre qu'il y a bien une croissance du roman au détriment des autres genres, mais cette croissance est

moins impérieuse que ce que laissait croire la courbe, et elle présente des irrégularités (une chute, par exemple, dans les années 1810-1830).

J'ai déjà dit avec quelle prudence il fallait considérer les résultats statistiques. Je voudrais à présent insister sur la dangereuse séduction de leur représentation graphique. Alors qu'il est instinctivement méfiant devant des tableaux de chiffres, le chercheur littéraire est souvent moins suspicieux devant des graphiques qu'il croit pouvoir lire et interpréter plus aisément. La comparaison que je viens de faire entre les deux courbes ci-dessus est illustrative des précautions à adopter pour jauger la justesse d'un graphique. Lanson mettait déjà en garde contre la séduction des chiffres, en son époque de positivisme :

« Défions-nous des *chiffres*. Le chiffre ne fait pas disparaître ce qu'il y a de flottant ou de flou dans l'impression : il le déguise. Quiconque sait un peu écrire trouvera dans le langage ordinaire les moyens de rendre sensibles les nuances sans lesquelles il n'y a pas chez nous d'exactitude, et ces nuances ne se chiffrent pas.

Désabusons-nous des *courbes*, dont nous faisons le symbole du développement des idées littéraires. Mais elles y supposent, ou elles y introduisent : 1° *l'unité*; 2° *la continuité*. Or il y a des mouvements qui éclatent, comme des épidémies, en plusieurs lieux à la fois, et des genres qui naissent deux ou trois fois, avant de vivre. La *courbe* donne donc souvent une représentation inexacte des faits. »<sup>1</sup>

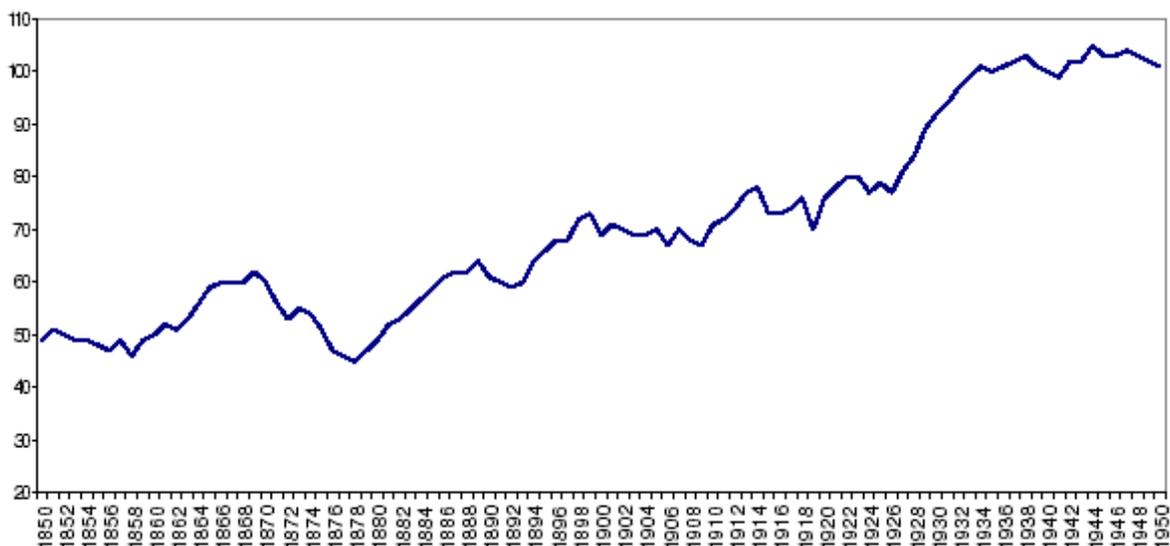
Disons plutôt que les graphiques sont des moyens rhétoriques pour convaincre de la justesse d'une analyse, plus que des preuves irréfutables de l'existence d'un fait. J'ai donné dans mon

---

<sup>1</sup> . LANSON (Gustave) : *Essais de méthode, de critique, et d'histoire littéraire*, L. Peyre éd., Hachette, 1965, « Méthode de l'histoire littéraire », p. 41.

*Introduction aux Études Littéraires Assistées par Ordinateur* (p. 58-59) quelques exemples d'illusions d'optique fréquentes dans la lecture des graphiques. Disons, de manière générale, que la prudence du chercheur doit être grande dans ce domaine et qu'il faudra, plus que jamais, aller vérifier la justesse des faits qui sont suggérés par les graphiques.

Une autre étude d'ensemble<sup>1</sup>, présentée au congrès de la SELF XX en 2001, interroge la notion de « siècle littéraire », en l'occurrence le XX<sup>e</sup>. Peut-on trouver dans les données et les courbes de la BDHL ou d'autres bases de données des éléments statistiques qui justifient les découpages séculaires proposés par l'histoire littéraire ? Si l'on se base sur une donnée comme le nombre d'écrivains en activité (en vie et ayant déjà publié une œuvre importante), d'après la BDHL, pour chaque année entre 1850 et 1950, on obtient un graphique de ce type :



---

<sup>1</sup> . « Quand commence le XX<sup>e</sup> siècle littéraire ? Essai de réponse statistique », *Le Temps des Lettres. Quelles périodisations pour l'histoire de la littérature française du XX<sup>e</sup> siècle ?*, Michèle Touret et Francine Dugast-Portes éd., Presses Universitaires de Rennes, « Interférences », 2001, p. 195-209.

On distingue bien quelques légers effets de génération sur cette courbe, par exemple en 1878 ou pendant la guerre de 1914-1918, mais rien ne permet de fixer avec sûreté une date limite entre XIX<sup>e</sup> et XX<sup>e</sup> siècles. Une étude de ce genre permet surtout de réfléchir sur les critères qui devraient justifier une coupure séculaire autour de laquelle il serait possible de classer, dans les manuels, les écrivains qui appartiennent à l'un ou l'autre siècle. Il existe cependant des écrivains dont la vie se situe exactement à cheval sur les deux périodes (par exemple Anatole France, Pierre Loti ou Alfred Jarry). Il n'est pas impossible de supposer que ces écrivains inclassables sont, pour cette raison, minimisés par l'histoire littéraire, alors qu'à l'inverse on constate que nos plus grands écrivains ont des vies qui se confondent à peu près avec les siècles qu'ils sont supposés représenter.

Ce type de détour méthodologique et de parcours de recherche, qui consiste à essayer de répondre à une question posée à la BDHL et d'en arriver au terme de l'étude à remettre en cause les termes de la question, est caractéristique des travaux présentés dans ce chapitre. Il peut paraître choquant de dire que nos grands écrivains ont, plus que du génie, des dates de naissance et de décès faciles à loger dans un manuel de littérature découpé par siècles, mais si l'on considère les dix noms le plus souvent cités en tête de palmarès, la chose est évidente : Voltaire (1694-1778), Molière (1622-1673), Jean-Jacques Rousseau (1712-1778), Victor Hugo (1802-1885), Jean Racine (1639-1699), Pierre Corneille (1606-1684), Denis Diderot (1713-1784), Honoré de Balzac (1799-1850), Emile Zola (1840-1902), Montaigne (1533-1592). L'apport d'une base de données est précisément de permettre ce genre de découverte en distanciant le matériau littéraire, si chargé d'affectivité. En cherchant des justifications à la périodisation séculaire, nous avons donc trouvé – à l'inverse – des raisons de penser que l'*a priori* séculaire modifiait notre vision de l'histoire littéraire (c'est-à-dire ce qu'enregistre la BDHL) au point de ne pas y intégrer ceux des écrivains qui n'entrent pas dans le moule.

## Etudes d'œuvres à l'aide de la BDHL

Un autre emploi de la BDHL consiste à éclairer une œuvre donnée en la contextualisant à l'aide de l'ensemble des données. On en lira des exemples dans les volumes de la collection Cap'Aggreg. Chaque fois, une œuvre est confrontée, par chacune de ses caractéristiques, à l'image de l'histoire littéraire que constitue la BDHL. Par exemple, il est possible de dresser la liste<sup>1</sup> des poètes symbolistes contemporains d'Apollinaire (né en 1880) :

Stéphane MALLARME	1842-1898	SAINT-POL ROUX	1861-1940
Paul VERLAINE	1844-1896	Maurice MAETERLINCK	1862-1949
Joris-Karl HUYSMANS	1848-1907	René GHIL	1862-1925
Arthur RIMBAUD	1854-1891	Henri de REGNIER	1864-1936
Émile VERHAEREN	1855-1916	Gustave LE ROUGE	1867-1938
Jean MOREAS	1856-1910	Paul CLAUDEL	1868-1955
Rémy de GOURMONT	1858-1915	André GIDE	1869-1951
Albert SAMAIN	1858-1900	Paul FORT	1872-1960
Josephin PELADAN	1859-1918	O. V. de L. MILOSZ	1877-1939
Jules LAFORGUE	1860-1887	Pierre-Jean JOUVE	1887-1976

On pourrait, sur le même modèle, fournir la liste des œuvres publiées en 1913 comme *Alcools*, des événements littéraires contemporains, des œuvres publiées au Mercure de France, etc. Un traitement plus complexe, basé sur l'indexation thématique, permet de fournir une liste d'œuvres indexées avec les mêmes mots thématiques que l'œuvre étudiée. Voici par exemple le résultat obtenu pour *La Condition humaine*<sup>2</sup> :

---

<sup>1</sup> . Hubert de Phalèse : *Quintessence d'Alcools. Le recueil d'Apollinaire à travers les nouvelles technologies*, Nizet, 1996, p. 15.

<sup>2</sup> . Hubert de Phalèse : *Les Voix de la Condition humaine. La Condition humaine d'André Malraux à travers les nouvelles technologies*, Nizet, 1995, p. 22.

Avec **14** thèmes communs, la plus proche est *Les Conquérants* (roman d'André Malraux, 1928) avec : *action, colonialisme, combat, communisme, destin, engagement, héroïsme, histoire, liberté, mort, révolution, torture, violence, pouvoir.*

Puis, toujours du même auteur, vient *L'Espoir* (roman, 1937) avec **11** thèmes communs : *action, combat, communisme, fraternité, héroïsme, liberté, mort, peuple, révolution, solidarité, violence.*

Avec **9** thèmes communs viennent ensuite *La Peste*, roman d'Albert Camus (1947) : *absurde, action, combat, destin, engagement, héroïsme, histoire, mort, solidarité* ; et *L'État de siège* du même (théâtre, 1948) : *absurde, combat, destin, engagement, héroïsme, liberté, mort, sacrifice, pouvoir.*

Avec **8** thèmes : *Les Justes*, du même dramaturge : *engagement, héroïsme, homicide, liberté, mort, révolution, sacrifice, violence.*

Avec **7** thèmes communs on trouve *Quatrevingt-Treize*, roman de Victor Hugo (1874) : *fanatisme, histoire, liberté, mort, peuple, révolution, violence* ; *L'Insurgé* de Jules Vallès (1886) : *engagement, héroïsme, histoire, homicide, liberté, peuple, révolution* ; *L'Homme révolté*, essai de Camus (1951) : *absurde, héroïsme, histoire, homicide, liberté, révolution, violence.*

Pour **6** thèmes communs, les rapprochements commencent à devenir plus hasardeux, encore qu'on puisse y trouver une justification tenant non pas au genre mais à la problématique du roman : Montaigne, *Essais* (1580) : *action, fanatisme, histoire, liberté, mort, violence* ; Lamennais, *Paroles d'un croyant* (essai, 1834) : *combat, fraternité, liberté, peuple, violence, pouvoir* ; Romain Rolland, *Le 14 Juillet* (drame, 1902) : *héroïsme, liberté, mort, révolution, sacrifice, violence* ; et, du même, *L'Âme enchantée* (roman, 1934) avec *destin, héroïsme, histoire, mort, révolution, sacrifice* ; Jules Romain, *Les Hommes de bonne volonté* (roman, 1932-1946) : *argent, communisme, histoire, liberté, révolution, solidarité* ; Sartre, *Les Mouches* (drame, 1943) : *action, angoisse, engagement, liberté, solidarité, violence* et *Les Mains sales* (drame, 1948) avec : *action, communisme, engagement, liberté, mort, violence* ; Camus, *Caligula* (drame, 1944) : *absurde, destin, homicide, liberté, mort, pouvoir* ; Nizan, *Aden Arabie* (1931) : *angoisse, colonialisme, combat, communisme, liberté, mort* ; et enfin de Malraux les *Antimémoires* (1967) qui tiennent autant de l'autobiographie que du roman, et l'on ne sera pas surpris d'y retrouver certaines constantes : *action, engagement, fraternité, histoire, violence, pouvoir.*

Les rapprochements ainsi obtenus suggèrent d'autres lectures, réfèrent le roman étudié à certaines traditions de notre littérature, le situent dans des paradigmes auxquels on n'aurait pas pensé spontanément. Le rapprochement avec Camus, par exemple, peut susciter des parallèles intéressants sur la manière différenciée de traiter des mêmes thèmes. Est-il besoin de répéter, cependant, que ce type de résultat est profondément dépendant de l'indexation pratiquée et du choix des œuvres indexées ?

La notion mise en œuvre ici est celle de *réseau*. Elle est aujourd'hui fondamentale dans la pensée de l'informatique, a imprégné fortement la sociologie et peut apporter des éclairages

pertinents en littérature. L'histoire littéraire traditionnelle est tributaire, comme la plupart des sciences mises au point au XIX<sup>e</sup> siècle, d'une vision hiérarchique du domaine de connaissance. L'ensemble de l'histoire littéraire, comme dans les tableaux de classification de l'histoire naturelle, est divisée en siècles, puis en générations, en mouvements, en genres. On trouve de fait beaucoup de tableaux de classification dans les histoires littéraires<sup>1</sup>, et surtout une image très compartimentée du *continuum* temporel. L'utilisation d'une base de données décloisonne la masse des informations et permet dès lors d'établir des connexions que la structure arborescente masquait. C'est ce qui apparaît lorsque l'on relie un point du réseau, ici une œuvre particulière, à l'ensemble des autres points. Chaque œuvre va se trouver raccordée, par exemple, par sa date de publication, par son auteur, par ses thèmes, par son genre, par les mots de son titre, par la profession de son auteur, etc. à d'autres points du réseau, alors que dans une vision classique de l'histoire littéraire, on l'aurait seulement envisagée dans la case où elle aurait été rangée préalablement, et on l'aurait alors évaluée en fonction de sa représentativité par rapport à cette catégorie. C'est là un des critères de valorisation les plus fréquents de la critique de type lansonien : le chef-d'œuvre, ou le grand auteur, est représentatif de son temps, d'un mouvement, d'une tendance esthétique, voire d'un phénomène de société. Cette perception de l'ensemble documentaire couvert par l'histoire littéraire est tributaire du support papier utilisé jusqu'à aujourd'hui. Le support numérique permet, au contraire, une vision réticulaire des données, dans laquelle chaque point peut à son tour devenir un centre, pour les besoins de l'étude.

---

<sup>1</sup> . Un des sommets du genre est constitué par le manuel de Bornecque, *La France et sa littérature. Guide complet dans le cadre de la civilisation mondiale*. (Les Editions IAC ; Lyon, deux tomes, 1953 et 1957).

## Une navigation hypertextuelle dans l'histoire littéraire

Le mode de lecture de données organisées en réseau ne peut être le même que celui que l'on adopte face à un livre. Même si les manuels d'histoire littéraire comportent le plus souvent des outils de navigation non-séquentiels, comme les index, les tables des matières, les codes de couleur des pages, les tableaux synoptiques, etc., ils invitent à une lecture linéaire, qui trace un parcours chronologique à l'allure de récit séquentiel. Cette disposition induit des erreurs d'appréciation, comme celle par exemple qui fait croire à la lecture du Lagarde et Michard que le Symbolisme succède au Naturalisme, et serait donc une réaction contre les excès de celui-ci, parce que les deux chapitres sont présentés dans cet ordre. L'emploi de la BDHL fait au contraire apparaître les synchronies de cette moitié du XIX<sup>e</sup> siècle : *Germinal* est de la même année (1885) que les *Complaintes* de Laforgue, etc.

Le mode de parcours le plus adapté aux données réticulaires est la navigation hypertextuelle. On appelle « hypertexte » un ensemble documentaire dans lequel on a marqué les relations sémantiques par des liens informatiques permettant de se déplacer d'un point à un autre. Le succès du Web a popularisé cette notion et diffusé les compétences qui permettent d'en tirer parti. Les programmes d'interrogation que j'ai conçus pour l'interrogation de la BDHL en ligne<sup>1</sup> présentent de l'histoire littéraire une vision hypertextuelle. À partir d'une page quelconque, des liens permettent de rayonner vers toutes les informations analogues dans l'ensemble de la base. Voici un exemple, celui de la fiche sur Zola :

---

<sup>1</sup> . Ils sont tous deux accessibles à partir du site du Centre de recherche Hubert de Phalèse : <<http://www.cavi.univ-paris3.fr/phalese/hubert1.htm>>. Le premier, plus ancien, est écrit pour ColdFusion (j'en ai assuré l'analyse et la direction de projet, pas la programmation), le second est écrit en PHP (j'ai cette fois assuré seul l'ensemble du projet).

**Zola (Emile)** : De son vrai nom Emile Edouard Charles Antoine Zola. Né en 1840 (Paris), mort en 1902 (Paris) à l'âge de 62 ans. Sa première oeuvre a été publiée en 1868. Il avait 28 ans. Il a bénéficié des sources de revenus suivants : [Ecrivain](#), [Journaliste](#), [Responsable de la publicité chez Hachette](#).

Emile Zola a écrit notamment : *Les Contes à Ninon* (1864), *Thérèse Raquin* (1868), *La Fortune des Rougon* (1871), *Les Rougon-Macquart* (1871-1893), *La Curée* (1871-1872), *Le Ventre de Paris* (1873), *La Conquête de Plassans* (1874), *La Faute de l'abbé Mouret* (1875), *L'Assommoir* (1876-1877), *Nana* (1880), *Pot-Bouille* (1882), *Au Bonheur des dames* (1883), *Germinal* (1885), *L'Oeuvre* (1886), *La Bête humaine* (1890), *L'Argent* (1891), *La Débâcle* (1892), *Le Docteur Pascal* (1893), *Les Trois Villes* (1894-1898).

Il a été membre des groupes suivants : [Naturalisme](#) (1867-1902), [Socialisme](#) (1889-).

Ses contemporains les plus proches (moins de cinq ans de différence) sont :

Nom	Prénom	Année de naissance	Année de décès
<a href="#">Reider</a>	Paul	1835	1923
<a href="#">Clément</a>	Jean-Baptiste	1836	1903
<a href="#">Becque</a>	Henri	1837	1899
<a href="#">Villiers de l'Isle-Adam</a>	Auguste	1838	1889

[Raccourci vers rechercheoeuvres.php?datepub=1873](#)  Internet

Un clic de souris sur n'importe quel mot en bleu de cette page permettra d'obtenir d'autres informations : un clic sur le prénom fournit la liste de tous les écrivains se prénommant « Émile », l'intitulé de la profession ouvre une liste des écrivains qui ont exercé des métiers semblables, un clic sur la date de naissance ou de décès ouvrira une liste des écrivains nés ou morts ces années-là, un clic sur la date d'une des œuvres permettra de lire la liste des œuvres publiées la même année, etc. Ce parcours, s'il présente des risques de désorientation bien connus des spécialistes de l'hypertexte, permet néanmoins de se forger une vision assez neuve de l'histoire littéraire, en traçant des parentés nouvelles entre les œuvres, les auteurs. Il peut sembler insolite, par exemple, de regrouper des écrivains par leurs prénoms mais, outre que les prénoms marquent les générations, il n'est pas extravagant de penser qu'un écrivain peut être secrètement attiré par un autre écrivain portant le même prénom que lui. Plus classiquement, la comparaison d'une œuvre avec toutes les œuvres relevant du même genre permet de la situer chronologiquement et thématiquement dans sa catégorie générique.

J'ai insisté sur l'intérêt de ce type de navigation informationnelle dans un article<sup>1</sup> publié dans la revue *éc/arts* en 2000 : l'utilité d'une base de données tient moins à la masse des informations qu'elle contient qu'à la possibilité de les relier entre elles. Un des modes de cette mise en relation est l'établissement de statistiques. On a vu plus haut que les statistiques permettent d'obtenir des aperçus synoptiques sur de vastes corpus ; elles permettent aussi de situer un point du réseau dans un ensemble plus large. Dans un article<sup>2</sup> publié dans la revue de statistiques *Student*, dans un numéro consacré à l'emploi des statistiques dans les domaines artistiques, j'ai tenté de montrer comment l'existence d'une base de données dans un domaine comme l'histoire littéraire et les traitements statistiques qu'on lui applique pouvaient permettre d'une part de repérer des événements et d'autre part de situer les données dans un contexte plus ou moins large. Il s'agit là de la BDHL mais aussi de toutes les autres bases de données qui existent aujourd'hui dans le domaine littéraire. Voici par exemple un graphique<sup>3</sup> obtenu à partir de la base *Opale-Plus*, c'est-à-dire le catalogue électronique de la Bibliothèque Nationale de France. Il représente la production éditoriale et la production littéraire pour la période 1850-1950 :

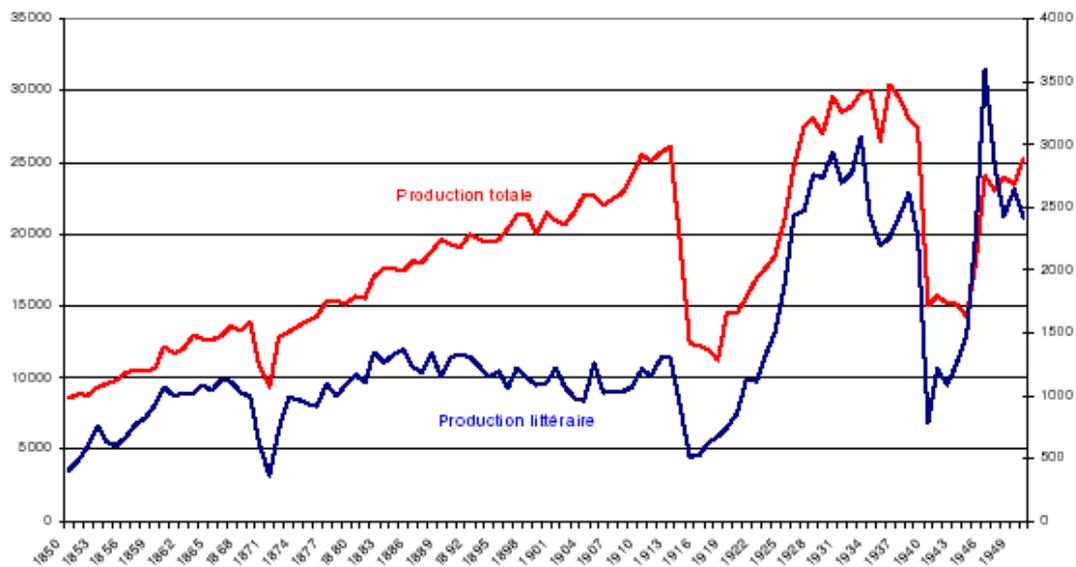
---

<sup>1</sup> . « Une banque de données d'histoire littéraire sur l'Internet », *éc/arts*, n°2, 2000, p. 371-373.

<sup>2</sup> . « Statistiques et histoire littéraire », *Student*, vol. 4, N°1, Juin 2001, « Special issue on Statistics in Art », Presses Académiques Neuchâtel, p. 5-11.

<sup>3</sup> . Publié dans « Quand commence le XX<sup>e</sup> siècle littéraire ? Essai de réponse statistique », *Le Temps des Lettres. Quelles périodisations pour l'histoire de la littérature française du XX<sup>e</sup> siècle ?*, Michèle Touret et Francine Dugast-Portes éd., Presses Universitaires de Rennes, « Interférences », 2001, p. 195-209.

## Dépôt légal 1850-1950



Les deux courbes sont assez parallèles, sauf dans une période (1870-1914) qui correspond à peu près à la période du Naturalisme et du Symbolisme. Dans ce laps de temps, le nombre de titres littéraires stagne, voire régresse, alors que l'ensemble de la production s'envole. Il est évident qu'un fait de ce genre peut intéresser une histoire littéraire générale mais peut aussi permettre de situer une œuvre particulière dans un contexte limité à quelques années. La navigation hypertextuelle peut ainsi se doubler de changements d'échelle, particulièrement délicats à gérer dans toutes les disciplines historiques.

Deux articles<sup>1</sup> publiés dans la revue *Histoires littéraires* font état des possibilités ouvertes à l'histoire littéraire par *Opale-Plus* mais aussi par *Gallica*, la banque de données textuelles de

---

<sup>1</sup>. « *BN-OPALE-PLUS* en ligne : un outil précieux pour l'histoire littéraire », *Histoires littéraires*, Avril-mai-juin 2000, n°2, p. 115-118.

Collectif (avec Jean-Pierre Goldenstein) : « *Gallica*, une très grande bibliothèque en ligne », *Histoires littéraires*, 2002, n°9, « Chronique de l'@ », p. 145-150.

la BNF. Il s'agit de banques de données très différentes de la BDHL dans leur conception, puisqu'elles visent à l'exhaustivité alors que la BDHL ne représente qu'une sélection, celle qui est pratiquée par l'histoire littéraire. Ces banques, qui préfigurent explicitement une future bibliothèque virtuelle, permettent dès aujourd'hui d'effectuer des recherches inenvisageables par les moyens traditionnels. Outre les études statistiques illustrées plus haut, la recherche en plein texte dans le catalogue de la BNF permet de formuler des requêtes jusque là impossibles, comme par exemple de chercher toutes les œuvres publiées par un éditeur, mais aussi tous les titres contenant un mot ou une expression. Pour les œuvres anciennes, qui ne sont pas indexées par thèmes, et dans le titre desquelles figurent souvent des indications variées, ce dispositif permet des recherches très intéressantes. Il est même à prévoir que ce catalogue électronique devrait, dans les années qui viennent, susciter des découvertes importantes. *Gallica* est d'ores et déjà une banque de données d'une grande utilité, même si beaucoup des ouvrages qu'elle contient ne sont disponibles que sous forme d'images. On peut, par exemple, balayer toutes les tables des matières du XIX<sup>e</sup> siècle de *La Revue des deux-mondes* à la recherche d'un mot, d'un nom. Là encore, cette possibilité nouvelle devrait permettre, par exemple, la découverte d'inédits encore inconnus.

L'existence de ces deux banques de données, et d'un grand nombre d'autres ressources de ce type disponibles en ligne ou sous forme de cédéroms<sup>1</sup>, change en profondeur le travail de la recherche en histoire littéraire. Elle l'accélère, tout d'abord, en permettant d'obtenir très rapidement des informations qui nécessitaient jusqu'alors de longues manipulations livresques, difficulté opérationnelle sur laquelle ont buté toutes les tentatives d'histoire

---

<sup>1</sup> . On trouvera un inventaire (un peu daté désormais) et une typologie (toujours valable) de ces ressources dans *Introduction aux Études Littéraires Assistées par Ordinateur*, p. 89 et sq.

littéraire scientifique selon le vœu de Lanson. Mais, surtout, elle permet d'accéder à des informations impossibles à obtenir autrement. Cette possibilité est d'autant plus déterminante que la vitesse de croissance de nos bibliothèques augmente très rapidement et que la documentation de plus en plus riche qu'elles renferment deviendrait inutilisable sans un outil qui permette de la gérer et de l'explorer.

Cependant, la mise en œuvre de ces outils documentaires pose une nouvelle fois la question de la formation des littéraires. Un outil informatique puissant n'est pas forcément facile d'utilisation, quoi qu'en disent les éditeurs de logiciels, et le chercheur ne pourra en tirer parti que s'il a appris à s'en servir. Le plus important dans ce domaine n'est pas de connaître les particularités de telle application mais d'avoir assimilé la logique générale de l'interrogation des bases de données : opérateurs booléens, champs, index, mots clés, troncatures, etc. Ayant acquis ces préalables, un chercheur deviendra autonome et saura s'adapter à une interface et à une base de données particulières.

### **Production de bases de données**

Plus encore, il aura la capacité intellectuelle et opérationnelle d'envisager la création de bases de données. En effet, j'ai présenté jusqu'à maintenant des exemples d'utilisation des bases de données, dans lesquelles le chercheur n'est qu'un consommateur d'information, mais il faut aussi affirmer que, comme dans d'autres disciplines scientifiques, les chercheurs en littérature ont tout intérêt à constituer des bases de données. À tout le moins, on ne voit pas quel avantage aurait aujourd'hui l'antique boîte à fiches sur une base de données informatisée à l'aide d'un simple traitement de texte, qui permet par exemple de retrouver immédiatement une information située n'importe où dans une fiche. Mais, pour aller plus loin, la constitution de bases de données structurées, gérées par des logiciels spécialisés, peut fournir un outil de

recherche extrêmement efficace, parce qu'il permet une approche statistique ou l'interfaçage avec d'autres bases de données. Cette possibilité d'interconnexion est un des principes constitutifs de la BDHL, qui peut être croisée avec d'autres bases de données. Un chercheur qui aurait par exemple constitué une base d'adaptations cinématographiques d'œuvres littéraires pourrait aisément récupérer les données de la BDHL pour vérifier, par exemple, si les différents genres ou mouvements littéraires sont également traités par le cinéma.

La création d'une base de données peut être également une forme de publication de résultats de recherche. C'est une pratique courante chez nos collègues terminologues ou lexicologues, mais bien des recherches littéraires pourraient donner lieu, entre autres résultats, à des bases de données. La pratique commence à prendre de l'ampleur dans le domaine bibliographique, et l'on trouve sur l'Internet un grand nombre de bibliographies spécialisées régulièrement mises à jour. Cette démarche pourrait être appliquée à toutes les recherches qui comportent des relevés, des dénombrements, l'établissement de listes, la mise en évidence de séries. Si je prends comme exemple les titres des vingt dernières thèses déposées au fichier central<sup>1</sup>, soit :

1. La découverte de soi dans l'œuvre romanesque de Michel Butor
2. L'ironie dans le théâtre français du vingtième siècle
3. Le roman policier français au début du XX<sup>e</sup> siècle
4. L'édition littéraire en Bourgogne au dix-neuvième siècle
5. Louis-Ferdinand Céline et ses images
6. La lecture identitaire des œuvres romanesques antillaises : étude de réception critique
7. La construction d'une représentation de soi dans le manuscrit de la *Vie de Henry Brulard* de Stendhal : une approche génétique

---

<sup>1</sup>. Thèses déposées entre mai et août 2003.

8. La réception du théâtre de Jean Cocteau en Allemagne au XX<sup>e</sup> siècle
9. Le merveilleux dans les romans de chevalerie de la fin du Moyen-Âge à la renaissance (XIV<sup>e</sup> siècle-XVI<sup>e</sup> siècle)
10. La relation amoureuse dans l'œuvre de Patrick Modiano
11. L'histoire dans l'œuvre de Jean Giono : de la vision romanesque de l'histoire au mythe personnel
12. Les écrits marginaux d'Aragon , une explication du mouvement et de sa transformation
13. Le kaléidoscope à travers l'écriture de quatre poètes femmes : Anne-Marie Albiach, Gabrielle Althen, Marie-Claire Bancquart et Esther Tellerman
14. L'autobiographie féminine au XVIII<sup>e</sup> siècle. Exemple : *L'Histoire de Madame de Montbrillant* de madame d'Épinay
15. L'autoportrait d'un nihiliste occidental : la littérature de Michel Houellebecq (1958-) de 1991 à 2001
16. La poursuite d'un idéal, le livre unique de Baudelaire à Mallarmé, Pierre-Jean Jouve et Du Bouchet
17. Entre mémoires apocryphes et mémoires apocryphes en science-fiction : la science-fiction marque-t-elle la différence ?
18. Un problème du faire artistique : la répétition. Approche poétique / poétique de l'œuvre de Michel Butor
19. Intimité et vécu social dans l'œuvre d'Annie Ernaux
20. Jean-Jacques Rousseau et la tentation de l'espace bucolique

Il me semble que, dans ce que l'on voudra bien considérer comme un échantillon des sujets récents de la recherche en littérature française, plusieurs se prêteraient fort bien à la publication d'une base de données utile à d'autres chercheurs : 3 (bibliographie de romans

policiers), 4 (éditeurs et éditions bourguignons), 6 et 7 (articles critiques). D'autres auraient avantage (si j'interprète bien les titres) à s'appuyer sur des bases de données : 2 (passages ironiques dans les pièces dépouillées), 5 (répertoire d'images avec leurs caractéristiques), 7 (liste de corrections et ajouts), 9 (répertoire des indices de merveilleux dans le corpus), 18 (relevé des effets de répétition chez Butor). Peut-être d'autres sujets, dont il est difficile de comprendre le propos, relèveraient-ils de cette démarche. Tous, cela va sans dire, ont intérêt à organiser en base de données leur bibliographie.

Enfin, la constitution d'une base de données, comme j'ai tenté de le montrer à propos de la BDHL, invite à repenser le domaine sur lequel on travaille. Cette salutaire mise en perspective suffirait, me semble-t-il, à justifier la constitution d'une base de données, ou tout au moins à la concevoir. Elle est en effet de nature à pointer les faiblesses théoriques ou conceptuelles d'un projet de recherche, à lever des ambiguïtés, à préciser une problématique. De ce point de vue, même l'échec dans la constitution d'une base de données pourrait avoir un effet bénéfique, si cette tentative amène par exemple à assumer en connaissance de cause le caractère flou et indéfini d'une notion ou d'un objet de recherche.

# Hypertextualités

L'hypertexte, envisagé dans le chapitre précédent comme un outil documentaire, est aussi un média éditorial et un vecteur de création littéraire. Cette nouvelle dimension du texte ne peut rester en dehors du champ des études littéraires, qui ont, dans les dernières années, intégré dans leur démarche des points de vue médiologiques déjà bien installés en génétique, en bibliologie, en bibliométrie, en stemmatologie, etc. Les études littéraires, bien qu'elles aient concentré leurs efforts autour de la notion de *texte* immatériel, se préoccupent néanmoins des supports et des médiations par lesquelles le texte parvient à ses lecteurs. L'avènement d'un nouveau média ne peut être dédaigné par des spécialistes du texte, et le regard des littéraires sur les potentialités de l'écriture hypertextuelle sera déterminant pour son évolution.

## Edition hypertextuelle

Une partie de mes travaux dans ce domaine porte sur l'édition hypertextuelle de textes littéraires. J'ai développé un modèle d'édition critique hypertextuelle dans le cadre du projet FALMER<sup>1</sup>, dont j'ai présenté les grandes lignes au congrès de l'ALLC-ACH à Glasgow en juillet 2000. La première réalisation dans ce domaine est l'édition des œuvres complètes de

---

<sup>1</sup> . Voir « The FALMER project : an Electronic Critical Edition », *Literary and Linguistic Computing*, vol. 16, n°1, 2001, p. 179-187 et « Pour un Ducasse électronique », *Les Lecteurs de Lauréamont*, J.-J. Lefrère et M. Pierssens éd., actes du quatrième colloque international sur Lauréamont, Montréal, 5-7 octobre 1997, Du Lérot éditeur, pp. 53-61.

Lautréamont / Ducasse<sup>1</sup>, mise en ligne en 1997. J'ai développé pour cette édition une interface qui permet d'en faire un hypertexte intégral. Il faut entendre par là que chaque mot du texte est un « bouton » hypertextuel qui permet d'accéder à une page où figurent toutes les occurrences du mot dans le reste du texte et, dans un grand nombre de cas, une notice explicative comparable à l'apparat critique des éditions littéraires. Il est possible également de consulter les variantes du texte, une bibliographie, une biographie, en bref tous les compléments habituels présents dans les éditions critiques, mais aussi des outils plus rares, comme un index complet de tous les mots du texte. Cette même interface a été utilisée pour une édition en ligne des *Complaintes* de Laforgue<sup>2</sup>, et une édition de la *Légende des siècles* est actuellement en chantier. Je passerai rapidement ici sur les avantages économiques et matériels de cette forme d'édition de textes rares<sup>3</sup>, pour m'étendre un peu plus sur l'apport de l'hypertexte à l'édition de textes littéraires.

Ce dispositif, s'il ne fait que systématiser le principe de la note ou du renvoi, induit en réalité, par son caractère massif, une autre approche lexicale du texte littéraire. On ne lit pas de la même manière un texte sur papier et un hypertexte constellé d'invites à la navigation transversale. La linéarité propre au *codex*, essentielle au bon fonctionnement de la plupart de nos procédés d'écriture (toutes les formes du récit, mais aussi l'architecture des recueils poétiques, les effets de prolepse et d'analepse, la temporalité théâtrale, etc.) se trouve ici détruite par les possibilités de navigation réticulaire. Cela peut convenir à des textes fragmentaires, éventuellement à certains recueils poétiques, mais se révèle suffisamment gênant dans le cas du roman ou du théâtre pour que ce type d'édition soit réservé à l'étude

---

<sup>1</sup> . Disponible à partir du site Web du Centre de recherches Hubert de Phalèse.

<sup>2</sup> . *Ibid.*

<sup>3</sup> . Voir à cet égard *Introduction aux Études Littéraires Assistées par Ordinateur*, p. 143.

d'un texte, plutôt qu'à sa lecture première. Même si la pratique de l'anthologie nous a habitués à lire les textes sous forme de fragments, il est évident que l'œuvre littéraire est initialement destinée à être lue intégralement et séquentiellement. La question de l'ergonomie des écrans et autres affichages informatiques est un faux débat, qui sera réglé tôt ou tard par l'évolution des techniques. La question centrale est celle de l'inadaptation fondamentale d'un texte littéraire conçu pour le papier au média hypertextuel. On peut certes mimer le support papier sur un écran d'ordinateur, comme le font la plupart des interfaces actuellement disponibles (certaines vont même jusqu'à faire entendre le froissement de la page qui tourne<sup>1</sup>...), mais cette transposition ne peut être, à l'évidence, qu'une étape transitoire dans l'avènement de nouveaux modes de lecture, de même que les incunables ont imité un moment les manuscrits pour s'orienter ensuite vers des mises en page plus adaptées à l'impression.

Les conséquences de ce choc culturel peuvent être déterminantes pour l'avenir de la littérature. Dans une communication<sup>2</sup> présentée à Bologne en 1998 dans le cadre d'un colloque sur les « Hypertextes de création », je me suis attaché à montrer que le scénario le plus probable était celui d'une disparition d'une bonne partie de notre patrimoine littéraire à

---

<sup>1</sup> . Voir ici par exemple : < <http://www.crea-clic.com/crea-clic/livres.htm>>.

<sup>2</sup> . « Lectures hypertextuelles et textes littéraires traditionnels : transitions », communication au colloque « Iperesti creativi : modifiche della scrittura e nuove tecnologie », 15 mai 1998, Bologne, texte en italien sur le site de *Bollettino '900* : <http://www.unibo.it/boll900/numeri/2001-ii/W-bol/Bernard/Bernardtesto.html>. J'ai également présenté des communications allant dans le même sens à Paris X, en 2000 (Journée d'étude « Recherche Assistée par Ordinateur. Représentations », organisée par Françoise Deconinck-Brossard, Université Paris X, 25 mars 2000. Communication « Textes numérisés : le lit de Procuste ») et à Munich, en 2001 (Journées d'étude « Medientechniken von Literatur und Sprache: Kino - Hörspiel - Hypertext - Multimedia », organisées par Jochen Mecke, université de Regensburg, Munich, 8-10 octobre 2001, communication « Lectures littéraires, de la page à l'écran ». A paraître.).

l'occasion de ce changement de support, comparable au passage de la littérature orale à la littérature écrite. Certaines œuvres seront adaptées au nouveau média, et donc profondément altérées, d'autres ne pourront être restructurées et disparaîtront. Il est évident que ce processus sera fort lent (de même que la littérature orale n'est pas tout à fait morte aujourd'hui) mais il semble néanmoins que le mouvement de disparition du livre est aussi irréversible que l'a été en son temps la disparition du *volumen* au profit du *codex* ou celle du manuscrit à l'arrivée de l'imprimerie. Le rôle d'un chercheur dans un tel contexte n'est d'ailleurs pas d'alimenter fantasmes, nostalgies et prédictions eschatologiques mais d'observer scientifiquement les évolutions très profondes qui transforment aujourd'hui le monde éditorial.

Dans le domaine des éditions critiques, matériau de base des études littéraires, l'apport du média hypertextuel peut néanmoins se révéler extrêmement bénéfique, parce qu'il permet d'environner le texte d'un appareil critique très développé et d'en faciliter la lecture. L'édition de Lautréamont / Ducasse, en particulier, a permis un véritable travail d'équipe autour du texte, aussi bien dans la collaboration avec les chercheurs ducassiens que dans la participation de dizaines d'étudiants qui ont eu à apporter leur pierre à l'édifice. Chaque fiche lexicale peut en effet être complétée par un collaborateur, et être mise en ligne au fur et à mesure. Cela donne une souplesse incomparable à l'édition savante et permet de transformer l'édition critique en une sorte de colloque permanent, une œuvre collective en perpétuel devenir. Ces possibilités pourraient modifier en profondeur nos modes de travail actuels. Il serait possible par exemple que des communautés de chercheurs se retrouvent, de manière virtuelle, autour d'un site Web qui présenterait en permanence le dernier état des connaissances sur un sujet (un auteur, par exemple). C'est déjà là le projet des revues spécialisées actuelles (celles, par exemple, des *Sociétés des amis*), mais le nouveau média permettrait une consultation et une mise à jour infiniment plus commodes. L'utilisation de la grande Toile du Web étend

d'ailleurs le travail collaboratif de manière inouïe. Ainsi, dans l'édition des œuvres de La Fontaine / Ducasse, il est possible de passer d'une allusion à *Phèdre* dans les *Poésies* à une note d'aspect traditionnel qui explique la référence mais aussi, au choix du lecteur, à la citation complète du passage évoqué ou même, sur un autre site, à la lecture intégrale de la pièce de Racine. Il est ainsi possible de lier tout noyau informationnel à l'ensemble du contenu du Web, dans l'esprit réticulaire qui peut conférer à tout point le statut de centre.

J'ai pu également expérimenter cette disposition à mettre en œuvre ce que Pierre Lévy appelle l'« intelligence collective » du réseau<sup>1</sup> en assurant les fonctions de chef de projet du site<sup>2</sup> consacré au bicentenaire de Victor Hugo par le Ministère de l'Éducation Nationale en 2002. Il s'agissait non seulement de coordonner les services de l'Éducation Nationale et ceux des autres ministères concernés pour offrir des informations substantielles, mais surtout de permettre à l'ensemble des acteurs du monde éducatif de faire vivre ce bicentenaire en organisant des manifestations diverses, en collaborant au site, en échangeant des informations. L'utilisation d'une liste de diffusion, que Hubert de Phalèse a déjà expérimentée depuis plusieurs années avec LITOR<sup>3</sup>, a permis de créer un véritable réseau de professionnels de la recherche, de la documentation et de l'enseignement, mais aussi d'amateurs de l'œuvre de Hugo.

---

<sup>1</sup> . Pierre Lévy, *L'intelligence collective. Pour une anthropologie du cyberspace*, La Découverte, coll. « Science et société », 1994.

<sup>2</sup> . <<http://www.victorhugo.education.fr>>. J'ai publié une brève présentation de ce site : « victorhugo.education », *Les dossiers de l'ingénierie éducative*, n°37, décembre 2001, p. 65.

<sup>3</sup> . Liste de diffusion sur l'informatique et la littérature, animée par le centre de recherche Hubert de Phalèse (<<http://www.cavi.univ-paris3.fr/phalese/litor1.htm>>).

Par ailleurs, les informations et documents proposés par le site lui-même jouent eux aussi la carte du réseau, en ce sens qu'ils renvoient systématiquement à d'autres sites, officiels ou privés. Une enseignante, Danielle Girard, a notamment tenu à jour une vaste collection d'adresses de sites Web offrant des ressources hugoliennes. Cette architecture en réseau permet de valoriser toutes les initiatives qui ont été prises lors de l'année du bicentenaire : l'INaLF a mis en ligne une trentaine de textes de Hugo provenant de Frantext, la BNF a numérisé pour *Gallica* des *Œuvres complètes* et organisé une exposition qui a eu son pendant sur le Web, le Ministère de la Culture a créé une base de données d'événements liés au bicentenaire, etc.

Une célébration de ce genre relève plus de la « vie littéraire », au sens large, que de la recherche universitaire mais cette expérience m'a donné une idée précise de ce que peut être le rôle de l'Internet dans la constitution de réseaux collaboratifs institutionnels. La structure hypertextuelle donnée à l'information sur le Web a comme pendant une organisation humaine elle aussi fondée moins sur des hiérarchies, telles que celles qui existent encore dans l'administration, mais sur un réseau de compétences et d'intérêts. Il ne s'agit pas, cependant, d'une Thélème éthérée et utopique mais d'une forme d'organisation bien réelle, avec ses problèmes de financement, de pouvoir, de décision, de partage des tâches, ses dates butoirs et ses contraintes techniques. L'émergence de ce type d'organisation face aux structures pyramidales encore dominantes aujourd'hui est similaire au changement qui a été introduit dans le fonctionnement des institutions de recherche médiévales par l'invention de l'imprimerie. À l'apparition d'un nouveau média correspondent inévitablement des modifications structurelles dans les institutions qui créent et qui diffusent l'information, en même temps qu'une mutation dans les structures de cette information.

## **Hypertextes de création**

Si l'hypertexte est, en effet, pour la lecture première des textes littéraires, un média inadapté, qu'il vaut mieux réserver à des fins d'étude et de recherche, de nouvelles formes d'écriture littéraire peuvent tirer parti de ses potentialités et s'adapter à ses contraintes. Une autre partie de mon travail sur l'hypertexte a consisté à m'interroger sur les nouvelles écritures hypertextuelles. Si l'on part de l'analogie entre l'émergence du nouveau média électronique et d'autres événements technologiques du passé, somme toute de moindre ampleur, telles l'invention de l'imprimerie, l'apparition des rotatives, de la machine à écrire, etc., il est possible de prédire qu'à un nouveau média correspondront nécessairement de nouvelles formes d'écriture. De ce point de vue, il serait suicidaire pour les chercheurs littéraires de ne se concevoir que comme des spécialistes du livre en papier. Cette vision restrictive de l'histoire de la littérature, qui les coupe déjà en grande partie de domaines comme celui de la littérature orale ou de genres mixtes comme la bande dessinée, les programmes de jeux informatiques ou les arts vivants, risque, dans le contexte d'une disparition progressive de l'édition traditionnelle, de rendre caduc leur domaine d'investigation. Cette nécessaire adaptation passe par la prise en compte de nouveaux genres littéraires, qui n'en sont aujourd'hui qu'au stade des balbutiements et des prototypes, mais qui permettent de penser les contraintes structurelles du nouveau média.

Dans un article<sup>1</sup> de 1993 pour la revue *Texte* (Toronto), j'ai avancé l'idée que l'hypertexte dotait le langage d'une troisième dimension, en me fondant sur l'analogie suivante :

---

<sup>1</sup> . « Hypertexte : la troisième dimension du langage », *Texte* (Trinity College, University of Toronto), 1993, n°13/14, pp. 5-20. Version en ligne : <[http://www.chass.utoronto.ca/french/litera/Revue\\_Texte/bernard.PDF](http://www.chass.utoronto.ca/french/litera/Revue_Texte/bernard.PDF)>

Discours	Figure géométrique	Dimensions
cri, borborygme	point	0
discours oral	droite	1
discours écrit, texte	surface	2
hypertexte	volume	3

De même que le langage a mis des siècles à explorer les deux dimensions que lui conférait, avec le passage à l'écrit, la feuille de papier, il faudra beaucoup de temps aux écrivains pour inventorier les possibilités de cette troisième dimension qui leur est offerte.

J'ai étudié, pour en donner une idée, quelques prototypes d'hypertextes littéraires disponibles actuellement<sup>1</sup>. Ils requièrent, à l'évidence, des stratégies d'écriture très différentes de celles que nous connaissons aujourd'hui. Voici quelques modifications lourdes de conséquences :

- Un lecteur peut se déplacer à sa guise dans l'hypertexte, sans que l'auteur puisse prévoir tous les itinéraires possibles. L'auteur ne peut plus s'appuyer sur un déroulement chronologique, que le lecteur pourrait modifier à sa guise.
- On ne peut lire intégralement un hypertexte, et il y a donc des secteurs qui peuvent rester inconnus, et il est impossible de le résumer.
- Le lecteur peut participer à la production du texte, ne serait-ce que par ses choix, mais aussi dans certains cas en écrivant lui-même (les auteurs américains parlent de « wreader »).

Toutes ces innovations requièrent une adaptation des pratiques d'écriture mais aussi de nouvelles compétences de lecture. L'expérience montre que les lecteurs sont aujourd'hui déroutés par cette géographie textuelle réticulaire, qu'ils n'ont pas acquis les savoir-faire adaptés à la construction de sens à partir d'un hypertexte. Si l'on met face à face l'inexpérience des lecteurs et celle des auteurs qui ne peuvent que tâtonner à la recherche d'une nouvelle écriture, on décrit une situation qui n'évoluera que lentement, à l'échelle d'une génération. Quand la pratique du Web aura suffisamment familiarisé le public avec l'hypertexte, il sera sans doute possible qu'une littérature adaptée au nouveau support émerge et trouve son assise. Les problèmes ne sont d'ailleurs pas seulement lexiques mais aussi commerciaux, juridiques, techniques. Pour cette raison, le devoir de veille qu'assume la recherche littéraire dans ce domaine doit se partager avec les spécialistes d'autres disciplines.

Il y a cependant, dans ce domaine comme dans tous ceux que j'aurai à évoquer dans ce dossier, une spécificité du littéraire, qu'il importe de discerner. Cette tâche est plus aisée au voisinage d'autres disciplines, surtout quand elle est aussi éloignée que l'informatique. Dans l'étude de l'hypertexte, il est facile de distinguer ce que peut apporter l'informaticien à la connaissance d'un système technique de gestion des données et ce que peut en dire le littéraire, qui se préoccupera plus de la construction du sens et des procédés d'écriture. De ce point de vue, les ELAO ne perdent pas les études littéraires en les diluant dans la technicité, mais au contraire permettent d'en affirmer les propriétés et d'en établir la légitimité.

---

<sup>1</sup> . Voir aussi : « Lire l'hypertexte », *Littérature et informatique. La littérature générée par ordinateur*, Textes réunis par A. Vuillemin et M. Lenoble, Artois Presses Université, 1995, pp. 313-325. Version en ligne : [http://www.bibliotheque.refer.org/littinfo/4\\_bernar.htm](http://www.bibliotheque.refer.org/littinfo/4_bernar.htm)

# Lexicométrie

Cette revendication d'une spécificité du littéraire est particulièrement importante dans le domaine de la lexicométrie, c'est-à-dire des statistiques sur le vocabulaire. Cette discipline, en effet, a été créée, est essentiellement pratiquée et illustrée par des linguistes. Les travaux de linguistiques quantitatives ont maintenant près d'un demi-siècle d'existence, ils sont basés sur des concepts et des procédures bien établis, profitent régulièrement des progrès de l'informatique. L'usage de ces techniques par des littéraires n'a rien de naturel, même si des pionniers comme Charles Muller ou Etienne Brunet ont montré tout le parti que l'on pouvait en tirer.

La première difficulté tient à la représentation que nous nous faisons du texte littéraire. Nous avons tendance, en littérature, à refuser de considérer le texte d'une œuvre comme la simple séquence des signes qui le composent. Il reste encore quelque chose des origines exégétiques de notre discipline dans la révérence que nous conservons à l'égard du chef-d'œuvre, dans la perception duquel nous laissons entrer un grand nombre de connaissances extérieures au simple texte : biographie, histoire littéraire, intertextualité, génétique, etc. L'épisode structuraliste n'a pas fait table rase de cette posture, et il y a toujours une répugnance instinctive à prendre en considération la pure matérialité du texte littéraire.

C'est cependant sous cette forme élémentaire que l'œuvre littéraire peut devenir objet de lexicométrie. Comme nous l'avons vu plus haut pour la constitution d'une base de données en histoire littéraire, l'ordinateur ne peut manipuler que des informations réduites à la forme

binaire. En l'occurrence, l'œuvre la plus géniale, la plus émouvante, la plus importante pour l'histoire de la pensée ne sera au final qu'une succession de signes dont certains sont identifiés comme caractères et d'autres comme séparateurs. C'est à ce prix que nous pourrions obtenir des statistiques et – éventuellement – mieux lire l'œuvre traitée.

## **Exploration des bases textuelles**

Le premier avantage de ces traitements est de permettre une exploration rapide des corpus. On est ici plutôt dans l'ordre des fonctions documentaires, à la limite de la lexicométrie, mais il s'agit, historiquement, des premières applications, et des plus souvent mises en œuvre. Il y a très longtemps, en effet, que les spécialistes du texte produisent et utilisent des concordances, permettant de retrouver tous les contextes des mots d'un texte, classés par ordre alphabétique. Le nom d'Hubert de Phalèse, que nous avons choisi pour notre Centre de recherches, est celui d'un bénédictin du XVII<sup>e</sup> siècle qui a publié une concordance de la Bible, énorme volume<sup>1</sup> compilé par des centaines de moines qui a été utilisé pendant très longtemps par les théologiens et les prédicateurs. C'est en hommage au travail de ces précurseurs que nous avons choisi ce nom, et aussi pour montrer que l'usage de concordances pour lire un texte n'a rien de révolutionnaire, même si l'usage de l'informatique constitue évidemment un saut qualitatif important dans cette pratique. Notons au passage que la première application de l'informatique à l'étude des textes a consisté à compiler une concordance de la *Somme* de Thomas d'Aquin (Roberto Busa, en 1949).

---

<sup>1</sup> . *Sacrorum Bibliorum vulgatae editionis concordantiae Hugonis, cardinalis... emendatae primum a Francisco Luca, nunc denuo... expurgatae ac locupletatae cura et studio... Huberti Phalesii,... Editio novissima...* , Lugduni sumptibus J.-B. de Ville, 1677.

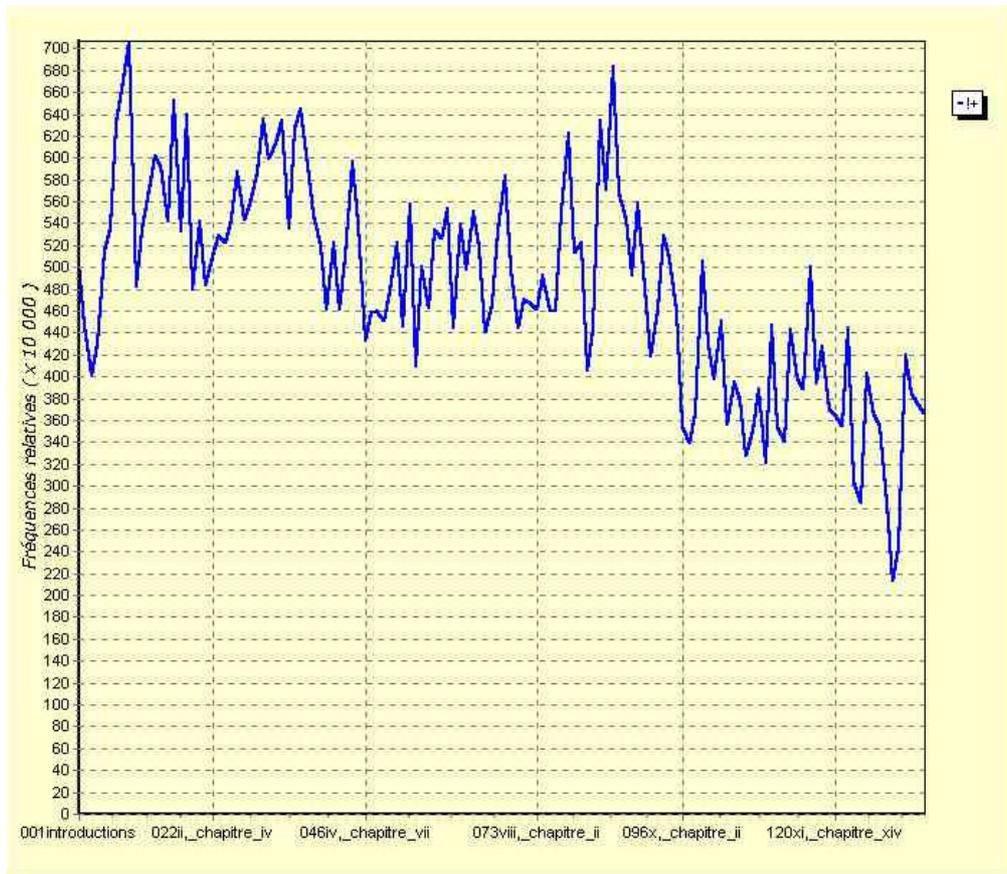
Un des résultats de cet usage des concordances peut être trouvé dans les volumes de la collection Cap'Agreg sous la forme de « fiches thématiques », auxquelles j'ai contribué avec les autres membres du Centre de recherche. Il est possible par exemple d'établir une synthèse sur le système des couleurs dans une œuvre, comme dans l'étude<sup>1</sup> que j'ai présentée lors d'une journée d'agrégation, en 1999, sur «Le système des couleurs dans *Corinne*». L'ordinateur fournit tous les contextes des mots qui désignent des couleurs, et l'on peut ensuite tenter de synthétiser ces données en traçant un tableau général des dominantes chromatiques de l'œuvre. Il est bien évident que le rôle de la machine est ici purement fonctionnel, et qu'elle permet seulement d'effectuer des recherches qui auraient, sinon, coûté beaucoup de temps et occasionné des erreurs. En particulier, il est impossible à la machine d'accéder à une notion thématique comme celle de « couleur », et le chercheur devra donc établir lui-même la liste des formes<sup>1</sup> référant à des couleurs. Il est toujours possible, pour ce faire, de s'appuyer sur des listes ou des dictionnaires électroniques, mais la méthode la plus efficace consiste à lire la totalité de l'index alphabétique en y repérant les mots qui relèvent du champ lexical étudié.

Les capacités de recherche de l'ordinateur peuvent aussi être utilisées pour représenter, à l'aide de graphiques, les occurrences de formes dans le corpus étudié. Les logiciels de lexicométrie proposent de plus en plus de fonctionnalités de ce type. Il est ainsi possible, à

---

<sup>1</sup> . « Le système des couleurs dans *Corinne* », Journée d'agrégation Hubert de Phalèse, « Corinne et Ségalen », Université Paris III, 6 novembre 1999, mis en ligne : <<http://www.cavi.univ-paris3.fr/phalese/agreg2000/couleurs.htm>>. Voir aussi « Le mot pouvoir », *Analyses et réflexions sur Zola. La Fortune des Rougon. Figures du pouvoir*, Edition Marketing. Ellipses, 1994, pp.66-68.

l'aide du logiciel Lexico3<sup>2</sup>, de tracer la courbe représentant le nombre de ponctuations dans chaque chapitre de *Gil Blas*<sup>3</sup> :



La courbe montre que la longueur des phrases croît tout au long du roman. Une autre application de ce type consiste à représenter sur un graphique les occurrences de noms de

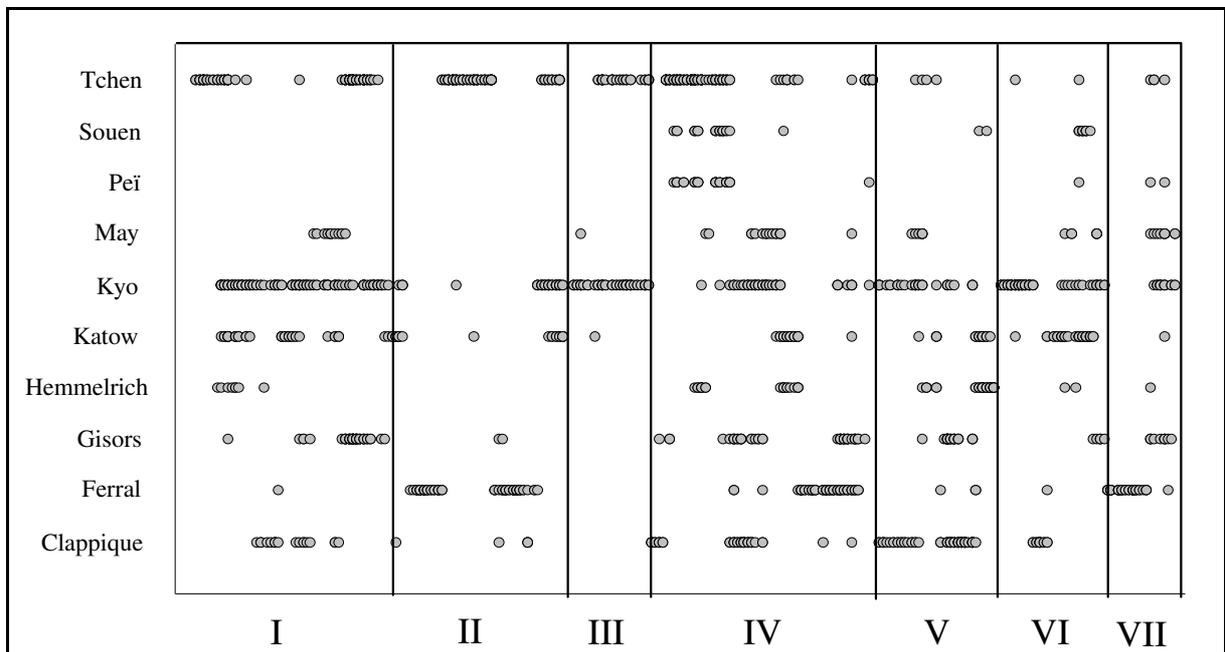
---

<sup>1</sup> . On sait que ce terme a en lexicométrie un sens précis. Une « forme » ou « forme graphique » est une série de caractères placés entre deux séparateurs. C'est dans ce sens que j'utilise le mot quand il est question de lexicométrie.

<sup>2</sup> . Dû à André Salem et à son équipe de l'université Paris 3.

<sup>3</sup> . *Les Bons Contes et les bons mots de Gil Blas*, Nizet, 2001, p. 74.

personnages dans un roman ou une pièce de théâtre, ce qui permet d'obtenir une synopsis des présences et absences simultanées. Voici un exemple<sup>1</sup>, celui de *La Condition humaine* :



Ce tableau permet de voir quels sont les personnages principaux, quels personnages apparaissent ensemble, lesquels ne se croisent jamais, ou apparaissent épisodiquement. Ce type de cartographie textuelle rend beaucoup de services parce qu'elle permet de saisir rapidement des relations et des phénomènes textuels complexes. Elle demande néanmoins, comme je l'ai dit plus haut à propos de la BDHL, une grande prudence et un retour systématique au texte pour vérifier ce que font apparaître les graphiques (dans l'exemple proposé, chaque point correspond à une occurrence du nom du personnage, ce qui peut également renvoyer aux discours sur un personnage, et pas forcément à sa présence comme acteur dans le récit).

<sup>1</sup> . *Les Voix de la Condition humaine. La Condition humaine d'André Malraux à travers les nouvelles technologies*, Nizet, 1995, p. 60.

Ce type de méthode permet aussi d'explorer des corpus beaucoup plus vastes, constitués en bases de données textuelles. La plus importante est Frantext. Comme membre d'un laboratoire rattaché au CNRS et à l'INaLF, j'ai suivi de près les travaux qui ont suivi l'achèvement du Trésor de la Langue Française. Lors du colloque « Mots - Textes - Dictionnaires », CNRS-INaLF, (Nancy, juin 1990), l'équipe Hubert de Phalèse avait déjà plaidé pour une meilleure adaptation de Frantext aux besoins spécifiques des littéraires<sup>1</sup>. En effet, Frantext a été conçue au départ en vue des travaux lexicographiques de l'INaLF et les principes de saisie ne sont pas toujours convenables pour les littéraires (choix des textes numérisés, absence des titres de poèmes, citations supprimées, textes tronqués, etc.) La situation s'est beaucoup améliorée aujourd'hui et Frantext est un outil de premier ordre pour les recherches littéraires. On trouvera encore dans les volumes de la collection Cap'Agreg des exemples d'utilisation de Frantext dans le cadre d'études littéraires. J'en cite quelques-uns :

- Recherche des origines du titre d'Aragon *Les Voyageurs de l'impériale*. Frantext fournit des contextes littéraires de *l'impériale*, siège de voiture ou allusion à l'Empire colonial, mais aussi de l'expression entière « voyageurs de l'impériale »<sup>2</sup>.
- Fortune d'Apollinaire, à travers les citations de son nom dans Frantext<sup>3</sup>.
- Le thème de la Chine et de Shanghai dans la littérature française avant *La Condition humaine*<sup>1</sup>.

---

<sup>1</sup> . Collectif (avec H. Béhar, L. Pouzadoux et P. Rebollar) : « Utilisation de Frantext en recherche littéraire » in *Frantext - Autour d'une base de données textuelles - Témoignages d'utilisateurs et voies nouvelles*, Didier Erudition, 1992, pp. 162-166.

<sup>2</sup> . *En voiture avec Les Voyageurs de l'Impériale d'Aragon*, Nizet, 2001, p. 42.

<sup>3</sup> . *Quintessence d'Alcools. Le recueil d'Apollinaire à travers les nouvelles technologies*, Nizet, 1996, p. 24.

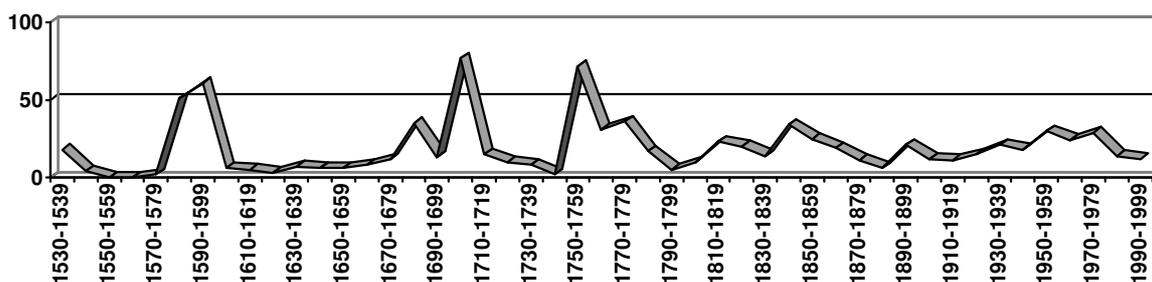
L'usage de la base de données est ici incontournable, dans la mesure où il procure des résultats pratiquement impossibles à obtenir autrement. Personne ne peut lire les 3 700 textes que contient Frantext à la recherche de toutes les occurrences du mot « Chine ». Les études thématiques les plus poussées portent sur des corpus de deux ou trois centaines d'œuvres, ce qui est déjà considérable, et le plus souvent ne recourent qu'à des citations indirectes, ce qui présente le risque de toujours suivre les mêmes ornières, d'épouser des idées reçues. L'usage de Frantext offre l'opportunité de véritables découvertes ponctuelles mais aussi la possibilité d'établir de vastes synthèses.

Lors d'un colloque sur « La France et la Pologne. Histoire, mythes, représentations », à Lyon, en 1998, j'ai présenté une communication<sup>2</sup> sur l'image de la Pologne et des Polonais dans la littérature française du XVI<sup>e</sup> au XX<sup>e</sup> siècle. Il s'agit donc d'embrasser ici un corpus considérable. Ce genre de travail s'apparente à la fois aux fiches thématiques évoquées plus haut, en ce sens que les quelque 3 500 contextes fournis par la base doivent être classés selon une typologie qui permettra de dire quels types de propos sont tenus sur la Pologne. Il relève aussi des études statistiques évoquées plus haut à propos de la BDHL. Il est possible, par exemple, de tracer un graphique indiquant comment ces occurrences de mots liés à la Pologne apparaissent dans le corpus tout au long de la période :

---

<sup>1</sup> . *Les Voix de la Condition humaine. La Condition humaine d'André Malraux à travers les nouvelles technologies*, Nizet, 1995, p. 27.

<sup>2</sup> . « L'usage du mot Pologne dans les textes littéraires français (XVI<sup>e</sup>-XX<sup>e</sup> siècles) : étude thématique et lexicométrique » in *La France et la Pologne. Histoire, mythes, représentations*, Françoise Lavocat éd., Presses Universitaires de Lyon, 2000, p. 29-45.



On note ainsi des périodes d'intérêt plus marqué pour la Pologne, que l'Histoire permet souvent d'expliquer. Toutes les précautions d'usage formulées plus haut sont encore valables ici. Je n'en donnerai qu'un exemple, un peu plus spécifique : une « polonaise » n'est pas forcément une femme originaire de Pologne mais aussi un morceau de musique, un manteau, etc. On peut considérer que cela relève toujours de la thématique polonaise, mais il faudra être attentif à ces cas d'homographie dès lors qu'il s'agit d'étudier des corpus textuels.

### Spécificités d'un texte

En dehors de ces applications documentaires, très utiles mais assez frustes du point de vue statistique, la lexicométrie permet surtout d'envisager de comparer les différentes parties d'un texte, ou un texte avec un corpus plus grand, en mettant en évidence les formes particulièrement fréquentes ou particulièrement rares. On utilise pour cela divers outils statistiques. Un des plus efficaces est le calcul des spécificités, mis au point par Pierre Lafon. On lira un exemple d'application de ce calcul dans le volume<sup>1</sup> de la collection Cap'Agreg consacré à *Corinne* de M<sup>me</sup> De Staël. Un tableau (p. 31) fournit les formes spécifiques de chacun des vingt livres dont se compose l'œuvre, ce qui permet de faire apparaître non seulement les dominantes thématiques, souvent aisément identifiables, mais aussi des particularités stylistiques moins visibles à la lecture.

---

<sup>1</sup>. *Corinne à la page*, Nizet, 1999.

Le calcul des spécificités peut être appliqué non seulement aux différentes parties d'un même texte mais aussi à des corpus plus étendus, qui servent de référence pour juger des caractéristiques d'un texte donné. J'ai publié des études de ce genre sur plusieurs auteurs, Saint-John Perse<sup>1</sup>, Nathalie Sarraute<sup>2</sup>, mais aussi, dans la collection Cap'Agreg, sur M<sup>me</sup> de Staël<sup>3</sup>, Laforgue<sup>4</sup>, Lesage<sup>5</sup>, Aragon<sup>6</sup>, Claude Simon<sup>7</sup>, Beckett<sup>8</sup>, Apollinaire<sup>9</sup> et Malraux<sup>10</sup>. J'ai mis au point une procédure qui permet d'extraire de Frantext les index des corpus à étudier puis, à l'aide d'un programme que j'ai écrit, de calculer les spécificités de toutes les formes du texte pivot. Il suffit ensuite de les classer pour obtenir la liste des formes spécifiques d'un texte. Voici par exemple les formes plus fréquentes chez Sarraute que dans les œuvres contemporaines :

---

<sup>1</sup> . « Informatique et littérature », in *Postérités de Saint-John Perse*, Actes du colloque de Nice 4, 5 et 6 mai 2000, textes réunis et présentés par Eveline Caduc, ILF-CNRS « Bases, corpus et Langage », Association des Amis de la Fondation Saint-John Perse, Publications de la Faculté des Lettres Arts et Sciences Humaines de Nice, 2002, pp. 285-295. Version en ligne : <<http://www.up.univ-mrs.fr/~wperse/bernard.htm>>.

<sup>2</sup> . « “Mes mots à moi” : aperçus lexicométriques sur l'œuvre de Nathalie Sarraute », in *Nathalie Sarraute. Du tropisme à la phrase*, textes réunis et présentés par Agnès Fontvieille et Philippe Wahl, Presses Universitaires de Lyon, 2003, pp. 59-69.

<sup>3</sup> . *Corinne à la page*, Nizet, 1999, p. 29.

<sup>4</sup> . *La Forgerie des Complaintes de Jules Laforgue*, Nizet, 2000, p. 50.

<sup>5</sup> . *Les Bons Contes et les bons mots de Gil Blas*, Nizet, 2001, p. 68.

<sup>6</sup> . *En voiture avec Les Voyageurs de l'Impériale d'Aragon*, Nizet, 2001, p. 77.

<sup>7</sup> . *Code de La Route des Flandres. Examen du roman de Claude Simon*, Nizet, 1997, p. 46.

<sup>8</sup> . *Beckett à la lettre. En attendant Godot, Fin de partie*, Nizet, 1998, p. 37.

<sup>9</sup> . *Quintessence d'Alcools. Le recueil d'Apollinaire à travers les nouvelles technologies*, Nizet, 1996, p. 30.

<sup>10</sup> . *Les Voix de la Condition humaine. La Condition humaine d'André Malraux à travers les nouvelles technologies*, Nizet, 1995, p. 43.

est, elle, Véra, Alain, ce, », «, [virgule], a, c', tu, Maman, Lemaire, Guimier, cela, mais, Germaine, sent, qu', ça, Gisèle, je, moi, ne, sais, tout, là, [points de suspension], faut, Kolia, Maine, Berthe, quelque, bien, Micha, chose, très, Gacha, quand, air, n', ces, mots, peut, rien, si, sait, eux, importe, petit, tante, pas, aussitôt, oh, ils, oui, ah, cette, lui, maintenant, Jardot, allons, Papa, eh, fauteuils, Lili, tss, sensation, ha, peu, moments

et celles qui y sont plus rares, voire absentes :

[point d'exclamation], [point virgule], fit, trois, fut, car, nuit, hommes, eut, Julien, demanda, Olivier, heures, M, ni, dix, guerre, mer, ville, chef, jours, eût, mit, voiture, soir, Simon, matin, prit, saint, parce, avaient, regarda, Marie, mort, quatre, femmes, route, moins, lorsqu', amour, Jean, ensuite, fût, leva, Henri, demi, vingt, Philippe, eau, corps, dieu, ayant, midi, faisait, allait, huit, cependant, feu, cent, noir

Cette dernière liste est plus surprenante dans le cadre des études littéraires, qui n'utilisaient guère, avant l'arrivée de l'informatique, cette notion de « forme rare » dans un texte. On a parfois l'impression qu'un auteur répugne à certaines expressions, ou qu'il fuit certaines réalités, on peut même aller jusqu'à supposer des tabous, des interdits ou des phobies, mais en faire la preuve par un repérage des formes rares dans le texte est très difficile sans ordinateur. Il est pourtant tout aussi révélateur sur l'écriture et l'imaginaire d'un auteur de repérer ses évitements que de mettre en évidence ses tics et ses mots préférés, qui sont souvent plus faciles à repérer à la lecture. L'informatique a même permis la création du concept de *nullax*, terme qui désigne une forme totalement absente du corpus alors que sa fréquence dans le corpus de référence est importante. Le *nullax* est souvent l'indice d'un refus marqué, d'un interdit absolu, qui peut être d'ordre stylistique, thématique, psychologique, esthétique, etc.

Ces résultats bruts ne sont évidemment pas directement utilisables, et il faut prendre des précautions méthodologiques pour les exploiter. La plus importante concerne le choix du corpus de référence. En effet, dire qu'une forme est abondante ou rare dans un texte n'a de signification que par rapport à un point de comparaison, qui est ici le corpus de référence qui permet de calculer les spécificités. Pour obtenir le résultat qui vient d'être cité, le vocabulaire

de quatre œuvres de Nathalie Sarraute, *l'Ère du soupçon* (1956), *Le Planétarium* (1959), *Pour un Oui ou pour un Non* (1982) et *Enfance* (1983), a été comparé à celui d'un ensemble de 225 romans et pièces de théâtre de la seconde moitié du XX<sup>e</sup> siècle. Il s'agit d'un corpus littéraire, et suffisamment proche des textes sarrautiens pour ne pas poser de problèmes de diachronie. Mais ces précautions peuvent être jugées insuffisantes, et certaines spécificités n'apparaître peut-être qu'en raison de la présence dans le corpus de référence d'œuvres très différentes (pour prendre un exemple caricatural, « Julien » apparaît comme une spécificité négative chez Sarraute parce que c'est le prénom d'un personnage central dans *La Maison des autres* et dans *Le Cœur des vivants* de Bernard Clavel, où ce mot apparaît plus de mille fois).

Il est donc très important de bien choisir le corpus de référence. J'ai développé ce point dans une communication<sup>1</sup> de 2000 sur « Le vocabulaire spécifique d'une œuvre », où j'ai montré l'effet d'un changement de corpus de référence sur le calcul des spécificités de *À La Recherche du temps perdu*. Selon que le corpus est constitué d'œuvres postérieures ou antérieures, que l'on y intègre ou pas les œuvres poétiques ou théâtrales, les spécificités sont différentes :

---

<sup>1</sup> . Colloque du groupe Sémantique des textes, « Corpus littéraires - Recueil et numérisation, analyses assistées, didactique », organisé par François Rastier et Franck Neveu, 20-21 octobre 2000, Université Paris VII. Communication : « Le vocabulaire spécifique d'une œuvre ». A paraître.

Formes spécifiques de la <i>Recherche</i> par rapport à :						
Romans 1913-1922	Romans 1850-1950	Romans 1850-1922	Romans 1913-1950	Tout 1913- 1922	Romans d'Anatole France	Romans de Barrès
que qu' avait duchesse avais était être même chez eût de grand' qui elle princesse si mais baron ne duc	que qu' avais duchesse ( ) princesse grand' me qui même mais si eût être pas avait pour chez où	que qu' avais duchesse ( ) me pas même mais grand' princesse être qui si j' eût je ou pour	que duchesse qu' avais princesse qui grand' duc eût baron avait de si chez me ( ) être car où	avait était elle qu' avais que eût loup duchesse me grand' princesse baron si chez pas être ne faisait pour	que qu' même ( ) être avais si avait comme pas nous où pour mais eût cela qui duchesse faire	avait était pas elle avais que qu' n' on si comme cela ne aussi chez dire princesse faire non ou

La problématique abordée ici n'est pas spécifiquement informatique. Elle ressortit au débat sur les notions de « norme », d'« écart », de « style ». On ne peut en effet fonder les études stylistiques que sur la notion d'un écart entre la langue d'un auteur et une entité, d'ailleurs difficile à définir, qui serait la langue courante. Rien n'est plus flou que cette notion, et l'informatique, une fois de plus, ne fait que pointer une zone d'ombre des études littéraires. On ne peut en effet signaler, par exemple, l'emploi abondant d'adjectifs qualificatifs par un écrivain sans supposer, le plus souvent tacitement, qu'il existe une norme quant à la proportion d'adjectifs dans un texte littéraire. Or, cette norme n'est jamais précisée, il n'existe pas d'échelle de fréquence, comme celle de Dubois-Buyse pour la langue courante, qui permettrait de dire si l'emploi d'une forme est fréquent ou pas dans un texte de Corneille, par exemple. Les indications de ce type qui figurent dans le *Trésor de la Langue Française* sont elles-mêmes calculées à partir de Frantext, sans précision générique autre que littéraire /

technique.

Comme nous l'avons vu dans les applications en histoire littéraire, les ELAO sont donc amenées, parce que la logique des machines est contraignante, à se poser des questions épistémologiques que les littéraires auraient dû se poser préalablement pour asseoir leur discipline sur des bases scientifiques. Ce n'est d'ailleurs pas le rôle des chercheurs spécialistes des ELAO que de répondre à ces questions mais celui de l'ensemble de la communauté scientifique. En revanche, l'informatique permet de poser les questions de théorie littéraire de manière très pragmatique, ce qui peut aider à les aborder sur des bases plus sûres. On conçoit par exemple qu'il est plus facile de répondre à la question « À quel corpus doit-on comparer le *Recherche* ? » qu'à la même question posée sous la forme « Quelle est la langue littéraire courante à l'époque de Proust ? ». En particulier, la première question peut être abordée dans le cadre d'une démarche expérimentale. Il est possible de tester plusieurs corpus de référence pour juger des résultats et en déduire des règles plus générales.

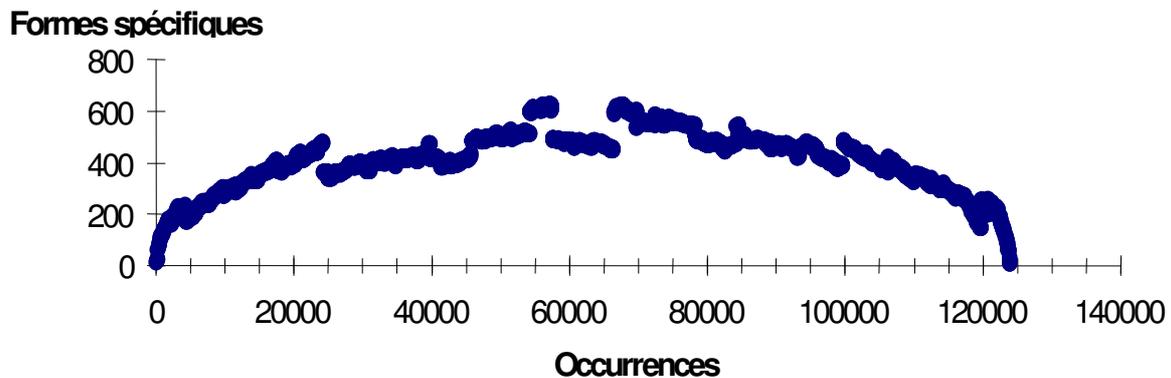
C'est un des apports majeurs de l'informatique aux études littéraires que d'y introduire cette approche expérimentale. Rien n'est plus simple, par exemple, de vérifier *a priori* si l'œuvre de Stendhal contient beaucoup d'occurrences de « rouge » et de « noir ». Une telle recherche, sans ordinateur, aurait coûté tant de peine qu'aucun chercheur ne s'y serait lancé à tout hasard. Il est ainsi possible de vérifier des hypothèses avant de se consacrer à l'étude à proprement dit, qui passe évidemment par une lecture approfondie des textes, que le recours à la machine ne saurait remplacer.

### **Découpage automatique des corpus**

Cette démarche expérimentale peut être illustrée par les travaux que j'ai menés sur le

découpage automatique des corpus. Je viens de dire comment le calcul des spécificités permettait de signaler les mots spécifiques (positifs ou négatifs) d'une partie d'un corpus par rapport à un ensemble, mais il s'agit dans ce cas de parties déjà délimitées, soit par l'auteur, soit par l'éditeur, soit par le chercheur dans une œuvre (chapitres, parties, actes, volumes, récit/discours, etc.) On constate que, selon le découpage adopté, le nombre de spécificités peut varier considérablement dans l'ensemble du texte. Il est donc tentant de demander à un programme de découper lui-même le texte pour optimiser le nombre de spécificités obtenues. Cela permettra d'isoler des parties du texte qui usent d'un vocabulaire très particulier.

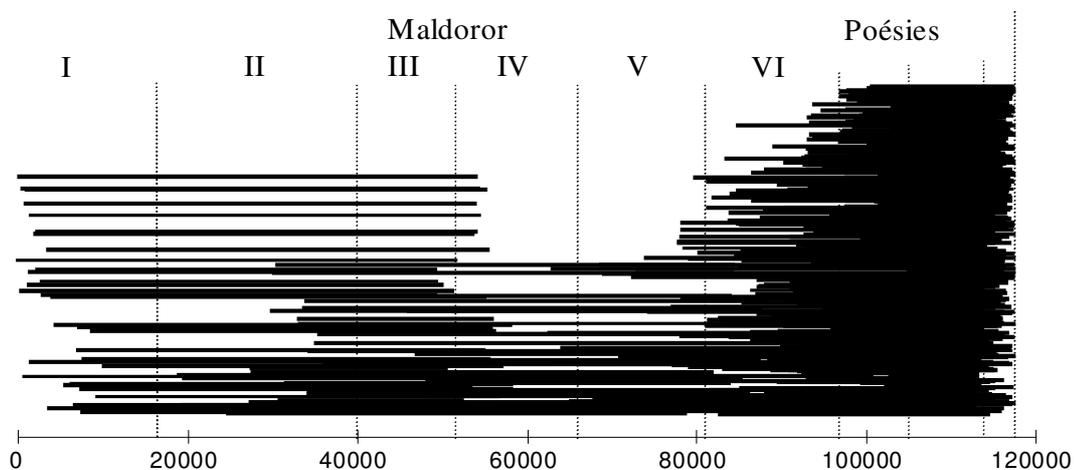
J'ai conçu un programme qui permet de faire ces opérations et propose des découpages du texte. Le principe en a été exposé<sup>1</sup> lors des Journées internationales d'Analyse statistique des Données Textuelles de Nice, en 1998. En prenant le texte de *La Route des Flandres* de Claude Simon comme exemple (parce qu'il s'agit justement d'un texte sans partitions de l'auteur, et assez difficile à délimiter), j'ai obtenu le schéma suivant :



Il faut entendre que si l'on place une coupure à un point du texte porté en abscisse (exprimé en nombre d'occurrences), on obtient le nombre de formes spécifiques indiqué en ordonnée.

L'allure générale de la courbe est normale : si l'on calculait les spécificités des deux premières pages, ou des deux dernières, on n'obtiendrait que fort peu de termes spécifiques, et le plus grand nombre de formes spécifiques est obtenu en coupant à peu près le corpus en son milieu. En revanche, ce sont les multiples irrégularités de la courbe qui permettent de faire des observations intéressantes. Chacun des décrochages constatés correspond à une mutation brusque du vocabulaire du texte, et délimite des portions de texte présentant de fortes particularités lexicales. Comme dans toutes les applications de ce genre, il faut bien entendu aller vérifier le texte pour y chercher les causes stylistiques et thématiques de ces ruptures de vocabulaire.

Un autre traitement issu des mêmes calculs permet par exemple de montrer les partitions les plus riches en formes spécifiques dans les œuvres de Lautréamont / Ducasse :




---

<sup>1</sup> . «Découpage automatique d'un corpus par la calcul des spécificités», *Actes des 4èmes Journées internationales d'Analyse statistique des Données Textuelles*, textes réunis par S. Mellet, UPRESA, Université de Nice, 1998, p. 125-134.

Chacun des traits horizontaux correspond à une partition du texte qui génère un grand nombre de formes spécifiques (classées de haut en bas par ordre décroissant du nombre de formes spécifiques). On voit nettement que les *Poésies* (I et II) utilisent un vocabulaire particulier, ainsi que, à un moindre degré, les trois premiers *Chants de Maldoror*. On voit aussi la parenté possible entre le dernier des six *Chants* et les *Poésies*. On trouvera des études du même type, dans la collection Cap'Agreg, à propos d'Apollinaire<sup>1</sup>, de Laforgue<sup>2</sup> et de Mme de Staël<sup>3</sup>.

### **Une lexicométrie littéraire ?**

Les exemples utilisés dans ce chapitre montrent que l'usage de la lexicométrie peut fournir aux études littéraires de solides points d'appui. Mais son emploi sur des textes littéraires demande un surcroît de précautions :

- La langue littéraire est plus complexe que celle des textes documentaires ou techniques. L'emploi poétique du vocabulaire, l'usage de figures comme la métaphore ou l'ironie, la versification, la répugnance à la répétition, les archaïsmes, etc. rendent assez délicates la mise en œuvre et l'interprétation des procédures de la lexicométrie.
- Les études littéraires portent fréquemment sur de petits corpus, rendant moins pertinentes les statistiques, plus à l'aise sur les grands nombres.
- Les études thématiques ou sémantique sont difficiles à mener avec des méthodes quantitatives, qui ne peuvent prendre en compte que des formes graphiques, et non des unités de sens.

---

<sup>1</sup> . *Quintessence d'Alcools. Le recueil d'Apollinaire à travers les nouvelles technologies*, Nizet, 1996, p. 36.

<sup>2</sup> . *La Forgerie des Complaintes de Jules Laforgue*, Nizet, 2000, p. 47.

<sup>3</sup> . *Corinne à la page*, Nizet, 1999, p. 35.

- Les phénomènes sonores, si importants en stylistique et en métrique, échappent à l'étude d'un corpus codé selon le code orthographique. Une étude phonétique devra passer par un recodage du texte en alphabet phonétique.
- Les études stylistiques sont souvent plus attentives au détail de l'expression qu'à des phénomènes plus étendus, mis en évidence par la lexicométrie.

Il y a donc bien une pratique proprement littéraire de la lexicométrie, même si les outils employés sont les mêmes que dans d'autres disciplines (ils nécessitent cependant quelques adaptations). Les difficultés naissent de cette différence que notait déjà Lanson à son époque<sup>1</sup> : il n'est de science que du général, alors que les études littéraires portent sur le particulier, l'original, le détail. Cette antinomie peut être résolue en distinguant des échelles d'observation. De même que, dans les sciences expérimentales, les phénomènes observés sont différents selon que l'on se situe au niveau atomique, microscopique ou macroscopique, l'ordinateur, parce qu'il médiatise notre lecture du texte, nous permet d'accéder à d'autres échelles d'observation. Il fait dès lors émerger d'autres types de faits que ceux que nous avons l'habitude d'interpréter. Mais il faut se rappeler que les différents niveaux d'observation se rapportent à une même réalité, et qu'ils entretiennent des relations de cause à effet qu'il nous appartient de mettre en évidence.

### **De quelques erreurs d'appréciation**

L'emploi de la capacité de calcul des ordinateurs peut aussi servir à dissiper quelques erreurs et à démonter certaines constructions de l'esprit qui n'ont de logiques que l'apparence. Sur ce

---

<sup>1</sup> . LANSON G., *Méthode de l'histoire littéraire*, Paris, Les Belles Lettres, 1925 ; repris dans *Essais de méthode, de critique, et d'histoire littéraire*, L. Peyre (éd.), Paris, Hachette, 1965.

terrain, l'ordinateur est un contrôleur redoutable et à peu près infaillible. Je passerai rapidement sur le fait qu'à l'heure actuelle aucun chercheur littéraire ne devrait se permettre des affirmations péremptoires sur la fréquence d'un mot dans un texte, sur sa présence ou son absence, sur l'impossibilité de trouver l'origine d'un intertexte sans avoir vérifié dans Frantext ou d'autres bases de données l'exactitude de ses dires. De même, il n'est plus permis aujourd'hui d'ignorer les bases de données bibliographiques et les catalogues électroniques, les ressources de l'Internet ou de dictionnaires et encyclopédies sur cédérom. Un sottisier, sur ce point, serait aussi inutile que malveillant.

En revanche, l'informatique peut être employée pour évaluer certaines théories échafaudées par des écrivains ou des chercheurs et dont on peut essayer de vérifier la véracité de manière expérimentale.

### Les anagrammes de Tzara

Un des travaux auxquels je me suis livré dans ce domaine porte sur une théorie qui a occupé Tzara à la fin de sa vie. Il avait cru déceler dans les œuvres de Villon des anagrammes qui auraient permis au poète de dissimuler dans son texte des allusions à son entourage, à ses aventures, tout un roman que Tzara se faisait fort de dévoiler dans un livre qui n'a jamais vu le jour mais dont le manuscrit et les brouillons ont été publiés. J'ai montré dans un article<sup>1</sup> paru dans *Romania* en 1995 que sa démonstration était fautive et qu'elle reposait sur une méconnaissance des règles statistiques.

---

<sup>1</sup> . « Le Secret de Villon à l'épreuve de l'ordinateur - Tzara et les anagrammes », *Romania*, n°449-450, tome 113, 1992-1995, pp. 242-252.

Je n'entre pas ici dans le détail de la démonstration, que l'on pourra lire dans l'article, mais je voudrais insister sur l'apport méthodologique de l'informatique dans ce cas. L'erreur de Tzara consiste à inférer de quelques constatations ponctuelles l'existence de tout un système crypté dans les œuvres de Villon. Ce type de démarche intellectuelle est assez fréquente dans notre domaine pour que l'on s'attarde un peu sur sa logique. Il n'est pas rare, en effet, qu'un critique s' imagine avoir trouvé dans un texte littéraire des motifs qui y seraient cachés par l'auteur pour communiquer à quelques initiés un sens caché au plus grand nombre : romans à clé, procédés cabalistes, anagrammes, acrostiches, jeux de mots, rébus, palindromes, équivoques, les procédés sont nombreux qui permettent de coder le texte littéraire.

Certains de ces codages échapperont toujours à la vérification de l'ordinateur, qui ne nous permettra jamais de dire si les écrivains étudiés par Jean-Paul Weber ont bien bâti leur univers imaginaire sur les traumatismes enfantins qu'il a cru identifier, ou si *Les Liaisons dangereuses* sont bien un portrait de la bonne société grenobloise. En revanche, tout ce qui est formalisable, comme la recherche d'anagrammes, peut fort bien être vérifié par l'ordinateur. Il suffit de concevoir un programme adapté et l'ordinateur peut refaire les calculs et les repérages avec une grande sûreté.

### Les anagrammes de Saussure

Un autre exemple de ce type de travaux est la recherche que j'ai menée avec le Professeur Francis Gandon sur les anagrammes de Saussure, fameuses depuis que Jean Starobinski leur a consacré un livre<sup>1</sup>. Pour publier les dossiers manuscrits de Saussure sur les anagrammes, Francis Gandon souhaitait tester les solutions proposées par Saussure comme je l'avais fait pour Tzara. J'ai conçu à cet effet un programme qui permet de rechercher les anagrammes

d'un mot donné dans un texte (en l'occurrence le *De natura rerum*). Le programme permet de respecter toutes les règles formelles que Saussure a progressivement élaborées. Le résultat a été publié par Francis Gandon dans son édition<sup>2</sup>. Il ne conclut pas à une pure erreur de Saussure, mais maintient que certains des constats de celui-ci sont troublants. Quoi qu'il en soit, l'informatique offre la possibilité de vérifier les intuitions de Saussure, de démontrer qu'il triche parfois dans certaines démonstrations, ou au contraire passe à côté de phénomènes qui auraient validé ses thèses.

Il existe bien d'autres procédés de ce genre, procédant de la littérature, du jeu ou de l'occultisme, qui pourraient être soumis à l'épreuve de l'informatique.

---

<sup>1</sup> . Jean Starobinski, *Les mots sous les mots ; les anagrammes de Ferdinand de Saussure*, Gallimard, 1971.

<sup>2</sup> . Francis Gandon : *De Dangereux édifices. Saussure lecteur de Lucrèce. Les cahiers d'anagrammes consacrés au De rerum natura*, Editions Peeters, Louvain-Paris, « Bibliothèque de l'Information Grammaticale », 2002.

Voir en particulier p. 397.

## Lectures par le menu

L'un des reproches les plus fréquents adressés aux ELAO tient à la présentation de ses résultats. Les tableaux de chiffres, les équations, les graphiques, les détails techniques qui émaillent le plus souvent les publications dans ce domaine rebutent les littéraires, qui ont l'habitude de présentations plus rédigées, plus textuelles. La collection Cap'Agreg a tenu compte de cette critique au cours de son évolution, en masquant de plus en plus les démonstrations les plus techniques, en rejetant en notes ou sur le Web les explications techniques, en rédigeant davantage les données d'abord présentées sous forme de tableaux. Cela ne nuit en rien au sérieux de la démarche mais assure une meilleure lisibilité. Tout cet appareil technique se justifiait dans la jeunesse des ELAO, où les chercheurs accordaient autant d'attention à des méthodes encore controversées et mal connues qu'aux résultats produits, où le public pouvait encore subir l'ascendant d'une informatique encore prestigieuse parce que nouvelle, mais il devient inutile aujourd'hui.

La série d'études que je vais présenter dans ce chapitre procèdent de la même démarche. Il s'agit d'études littéraires qui seraient impossibles à mener sans le recours à l'informatique mais dans la présentation desquelles toute référence aux procédés mis en œuvre a été gommée. Rien n'est plus naturel : aucun critique, après tout, ne s'aviserait de donner des indications sur la marque de sa machine à écrire ou sur la manière dont il a rempli ses fiches cartonnées, et personne ne se soucie de le savoir. Il n'y a pas de raison pour que l'outil informatique soit plus indiscret et nécessite des explications oiseuses, en dehors d'études

spécialisées écrites pour des spécialistes. Il m'a semblé en outre que la méthode des ELAO paraîtrait plus légitime si elle était jugée sur ses résultats littéraires, validés comme tels. Pour cette raison, les articles que je vais présenter ici ont été proposés à des revues littéraires connues, et non à des revues spécialisées en informatique. Les comités de lecture ont donc évalué des articles littéraires, sur les critères propres à la discipline, sans avoir à prendre en compte l'adéquation de la méthode informatique aux résultats obtenus.

### **Le souci du détail**

Le titre sous lequel j'ai rangé les articles de la série des « lectures par le menu » signifie tout à la fois que ces lectures sont assistées par ordinateur (utilisation des « menus » des logiciels) mais surtout qu'elles se préoccupent de détails du texte. Cette voie des études critiques, inaugurée par les *microlectures* de Jean-Pierre Richard<sup>1</sup>, a suscité plusieurs travaux récents, dont *Lectures du détail* de Naomi Schor (Nathan, 1994), *Pouvoir de l'infime* (textes recueillis par Luc Rassin et François Schuerewegen, Presses Universitaires de Vincennes, 1997) et *Le Détail* (textes d'un colloque de 1997 réunis par Liliane Louvel, publications de la Licorne, UFR Langues Littératures Poitiers, 1999). J'ajouterai le livre du critique d'art Daniel Arasse, *Le Détail. Pour une histoire rapprochée de la peinture* (Flammarion, « Champs », 1996, première édition en 1992), dont les analyses peuvent être transposées à la littérature.

Dans chacune de ces études, il s'agit de partir d'un détail du texte, qui a retenu l'attention à la lecture, et de s'interroger sur les raisons de cette mise en valeur par la mémoration<sup>2</sup> de ce qui aurait dû passer inaperçu. Le détail, dans l'esthétique classique, est fait pour disparaître dans l'ensemble, et l'attention au détail est systématiquement dévalorisée. Néanmoins, les

---

<sup>1</sup> . Jean-Pierre Richard, *Microlectures*, Paris, Seuil, 1979.

<sup>2</sup> . « Évocation d'un souvenir fixé plus ou moins longtemps auparavant » (*TLF*).

approches psychanalytiques du texte littéraire ont permis de revaloriser le détail en lui donnant une signification plus forte, en ce sens que le détail est révélateur de ce qui est caché par le texte. La méthode que j'ai suivie est plutôt basée sur l'analyse des réactions du lecteur, dans le but d'expliquer pourquoi un détail peut attirer l'attention. Le mode d'explication est plus littéraire que psychanalytique, en ce sens qu'il renvoie moins à l'inconscient du lecteur ou du texte qu'à la mise en évidence de l'importance du détail dans l'œuvre. Ce qui est caractérisé initialement comme détail, et par là-même taxé d'insignifiance, est souvent l'indice d'axes majeurs du texte, qui sont trahis par le caractère insolite ou symbolique du détail.

### **Revue de détail**

Je commencerai par résumer brièvement le principe de chacune des études qui relèvent de cette démarche en disant pour quelles raisons (attention au détail et usage de l'informatique) elles sont des « Lectures par le menu » :

« Les *je/nous* de Victor Hugo. Pluriel de modestie et pluriel de majesté dans *Les Misérables* », *Op. Cit. Revue de littératures française et comparée*, Publications de l'Université de Pau, n°3, novembre 1994, pp. 129-136.

Le détail analysé ici est l'emploi par Hugo du pronom « nous » dans *Les Misérables*. Il s'agit donc d'une étude stylistique, sur la manière dont Hugo utilise le « nous » pour se désigner lui-même, comme auteur/narrateur. L'usage de l'informatique est rendu indispensable par la grande masse textuelle du roman, dans laquelle l'ordinateur peut chercher avec précision les occurrences de « nous » (qu'il faut bien entendu trier ensuite pour isoler les pluriels de majesté) ou d'expressions comme « l'auteur de ce livre ». À partir des 570 occurrences de « nous », il est possible de tracer un portrait intellectuel du narrateur et, au-delà, de définir le

projet de Hugo comme conteur et comme rhéteur. L'emploi du « nous » est en effet un procédé qui aide Hugo, en confondant lecteur et narrateur, à impliquer, à convaincre, à faire de la narration un long plaidoyer pour « les misérables ».

« Sylvie ou le pourpre proustien », *Bulletin Marcel Proust*, 1996, n°46, pp. 72-108.

Cette étude part d'une appréciation critique de Proust, qui affirme dans *Contre Sainte-Beuve*, à propos de *Sylvie*, de Gérard de Nerval : « La couleur de *Sylvie*, c'est une couleur pourpre, d'une rose pourpre en velours pourpre ou violacé ». Une vérification à l'aide de l'ordinateur montre qu'en réalité il n'y a dans *Sylvie* qu'une seule allusion à la couleur pourpre : « Au sortir du bois, nous rencontrâmes de grandes touffes de digitale pourprée [...] ». L'étude essaie d'expliquer comment Proust en est arrivé à surestimer cette notation, sans doute en raison de la très forte connotation affective chez lui du pourpre et de la digitale. L'article comporte donc deux parties : il est d'abord montré que la phrase qui a attiré l'attention de Proust a des résonances dans toute l'œuvre de Nerval, et une étude du thème dans les œuvres de Proust montre ensuite que son interprétation de l'œuvre de Nerval révèle également une thématique profonde de son écriture. Il s'agit donc ici de « lire une lecture », ou d'interpréter un souvenir de lecture pour mieux comprendre à la fois l'œuvre lue et l'imaginaire du lecteur.

« Les Misérables ou les ciseaux à froid », in *Victor Hugo 4 - Science et technique, La Revue des Lettres modernes*, Minard, 1999, pp. 97-115.

Le détail est ici un minuscule motif narratif qui a attiré mon attention à la lecture : à deux reprises, dans les *Misérables*, Hugo mentionne un outil, le « ciseau à froid ». Dans les deux cas, Jean Valjean est sauvé grâce à cet outil. L'étude, en partant de cette remarque et de cette coïncidence, essaie de montrer que le poids symbolique de cet objet renvoie à des thématiques profondes du roman et, de manière plus large, à des nœuds de l'imaginaire hugolien.

L'ordinateur a été mis à contribution pour explorer rapidement l'immense œuvre de Hugo afin d'y retrouver d'autres apparitions de ces thèmes. Frantext permet également d'aller chercher des attestations dans son énorme corpus, pour retrouver, par exemple, un extrait de *Rocamboles* où Ponson du Terrail rappelle que le ciseau à froid est un instrument de cambrioleur.

« *Madame Bovary ou le danger des sucreries* », *Romantisme*, n°103, 1999, pp. 41-51.

Le point de départ est ici une brève remarque de Flaubert lors de la scène du bal au château de la Vaubeysard, où Emma Bovary découvre avec émerveillement le luxe aristocratique : « Le sucre en poudre même lui parut plus blanc et plus fin qu'ailleurs. » Cette remarque incidente est référée au reste du roman, où apparaissent d'autres images du sucre et de la blancheur, qui tissent un motif thématique cohérent, lié au destin tragique d'Emma. Il est également montré que ce qui pouvait passer initialement pour une marque d'ironie de la part du narrateur n'est qu'une notation psychologique et concrète qui ne marque en rien la crédulité d'Emma mais au contraire son appréciation précise et réaliste de son environnement. L'ordinateur permet, ici encore, d'inventorier rapidement les occurrences de certains motifs dans l'œuvre de Flaubert mais aussi dans la littérature contemporaine.

« *Le jardin des Plantes ou l'hypotypose de la place Monge* », *Frontières de la fiction*, Alexandre Gefen et René Audet éd., Editions Nota Bene, Presses Universitaires de Bordeaux, 2001, p. 375-395.

Claude Simon, dans son roman *Le Jardin des plantes*, se livre à une description très précise d'un lieu qu'il ne nomme pas et qu'il est cependant possible d'identifier comme étant la place Monge, dans le cinquième arrondissement de Paris. Ce procédé est interrogé dans l'ensemble

de l'œuvre de Simon. Ce qui semblait être un détail découvert fortuitement se révèle être une manière d'écrire révélatrice du projet poétique de Claude Simon. L'ordinateur permet de retrouver des détails matériels dans l'ensemble de l'œuvre, à la recherche des indices de cette référentialité problématique.

« Les chefs d'œuvre de la langue française sont les discours de distribution pour les lycées », *Les Poésies d'Isidore Ducasse, Actes du cinquième colloque international sur Lautréamont, Marseille, 13-15 octobre 2000*, Jean-Jacques Lefrère et Michel Pierssens éd., *Cahiers Lautréamont*, livraisons LV et LVI, Du Lérot, éditeur, 2001, p. 111-124.

La phrase prise comme titre est une citation des *Poésies* de Ducasse, qui semble ici se livrer à une parodie ou écrire une boutade. L'étude tente de montrer que, bien au contraire, les discours de distribution des prix que Ducasse a pu écouter sont écrits dans un style et avec des intentions qui correspondent largement avec le projet littéraire et moral qui s'exprime dans les *Poésies*. L'ordinateur a été surtout employé ici pour explorer les bases de données bibliographiques et retrouver ainsi la trace des discours de distribution des prix dont Ducasse avait pu s'inspirer.

« Eugène Mutiaux, le parrain de Proust », *Ce que je ne sais pas, Cinquième colloque des Invalides*, Jean-Jacques Lefrère et Michel Pierssens éd., Du Lérot éditeur, 2002, p. 81-84.

Cette communication décrit une recherche en gestation<sup>1</sup> sur un personnage méconnu et qui semble cependant capital pour la connaissance de Proust : son parrain, Eugène Mutiaux. Mutiaux était collectionneur d'art, japonais en particulier, et il est envisageable qu'il ait eu une influence déterminante sur l'éducation artistique de Proust. Par ailleurs, le thème du baptême est très important dans la *Recherche* et la relation difficile de Proust à sa judaïté confère à son parrain un rôle sans doute central dans son histoire personnelle et littéraire. J'ai été mis sur la piste de ce personnage par une découverte faite lors d'une recherche informatique sur les mots absents de la *Recherche*. La forme « parrain » n'y est jamais utilisée, ce qui a attiré mon attention.

### **Une méthode critique**

Au-delà des acquis de chacune de ces études, je voudrais insister à la fois sur la méthode critique qu'elles délimitent et sur l'apport de l'informatique à l'invention de cette démarche. Il s'agit en effet, sous la forme de plusieurs travaux portant sur des œuvres diverses, de proposer une méthode critique qui consiste à retrouver la cohérence thématique d'une œuvre en l'examinant depuis un infime détail. Cette conception fractale de l'œuvre littéraire, qui voit dans chaque détail d'un texte une image complète de sa totalité, est aussi liée à la notion de réseau, déjà évoquée à plusieurs reprises ici. Faire momentanément d'un point du texte le centre de celui-ci, et montrer que l'on peut accéder depuis cette porte d'entrée à l'ensemble

---

<sup>1</sup> . J'en avais déjà annoncé le programme à Tokyo, lors d'une conférence dans le cadre du congrès de la Société Japonaise de Langue et Littérature Françaises, en juin 2001.

des thématiques, des problématiques et de la poétique d'une œuvre revient à voir dans l'œuvre littéraire un réseau de significations et de formes.

Aucune des « lectures par le menu », qui ne sont que des articles, n'a encore suffisamment d'ampleur pour montrer qu'en étendant le réseau plus loin, il serait possible de rendre compte non pas de quelques motifs seulement mais de l'ensemble de l'œuvre. Ce serait là l'objet d'un livre entier, que je nourris le projet d'écrire un jour. Les études que j'ai publiées jusqu'à maintenant sont des essais et des présentations d'une méthode critique qui pourrait être aisément mise en œuvre à une plus grande échelle.

L'apport de l'informatique à une lecture en réseau est évidemment indispensable, moins peut-être sous l'angle pratique de l'exploration des textes que par l'image qu'elle donne de l'œuvre littéraire. J'ai déjà parlé plus haut de l'impact de l'hypertexte sur la lecture des textes, dont il casse la structure séquentielle en établissant des liens insolites d'un mot du texte vers des passages qu'aucun dispositif textuel n'indique. C'est, à l'évidence, ce type de parcours qui suggère des lectures comme celles que j'ai proposées dans cette série d'articles. Lire une remarque de Flaubert sur la blancheur du sucre et aller vérifier toutes les occurrences de ce mot dans le reste du roman est une démarche très éloignée des lectures traditionnelles. Ce parcours paradigmatique est bien entendu celui des études thématiques, mais il porte habituellement sur des motifs plus visibles, d'apparence moins anecdotique. On étudiera ainsi le thème de l'ennui, celui de la féminité, voire le personnage du médecin ou du paysan dans *Madame Bovary*, mais pas le motif du sucre.

La différence entre ces deux approches me semble provenir des outils employés. Une étude thématique faite de manière traditionnelle, par des relectures répétées du roman et l'établissement de fiches, ne peut entrer dans un niveau de détail aussi fin qu'une étude

effectuée à l'aide de l'ordinateur. La première raison, que j'ai déjà eu l'occasion d'évoquer, tient à l'amorce de l'étude. On ne se lance pas dans une étude longue et contraignante à tout hasard ; on veut être sûr dès le début que le thème sur lequel on va travailler est réellement présent dans l'œuvre, et même qu'il a une importance suffisante pour justifier une étude d'une certaine ampleur. En revanche, les recherches informatiques permettent d'expérimenter préalablement une hypothèse de travail. Il est évident que je ne me suis pas engagé dans l'étude du motif du sucre sans avoir préalablement vérifié qu'il avait des échos nombreux dans l'œuvre, que le réseau était assez dense pour offrir un matériau d'étude solide. C'est donc, de ce point de vue, le perfectionnement des moyens de recherche qui permet ce genre d'étude, même si les meilleurs critiques (je pense par exemple à Georges Poulet ou à Jean-Pierre Richard) peuvent rivaliser par leur extraordinaire mémoire et leur culture étendue avec ce type de technique. Je ne suis pas sûr, néanmoins, que la tête la plus pleine et la mieux faite que l'on voudra imaginer puisse retrouver instantanément, de mémoire, deux citations de la *Correspondance* de Flaubert où le sucre en poudre est associé à des images érotiques.

Cette aisance à se mouvoir dans l'épaisseur des textes n'est pas destinée à fournir une prothèse de culture littéraire mais une manière plus rapide de confronter les œuvres, ce qui est après tout la mission des spécialistes du texte depuis les premiers exégètes. Quand ceux-ci mirent au point des dispositifs matériels, comme la roue à livres, ou textuels, comme les concordances, ils cherchaient déjà à accélérer le processus de mise en relation entre des points éloignés de la masse documentaire qui entretiennent cependant des rapports éclairants. Dans les lectures par le menu, il ne s'agit pas de s'en tenir à l'examen du détail pour lui-même, mais de le mettre en réseau avec tout ce qui peut s'y rapporter dans le reste du texte. L'objet de l'étude est d'ailleurs moins le détail, qui n'est qu'une amorce, mais les relations qui le constituent en centre d'un graphe.

C'est de ce point de vue que les lectures par le menu relèvent de l'esthétique de la réception. Le choix du détail initial, qu'il soit le fait d'un écrivain (Proust), d'une recherche informatique, d'un lecteur anonyme ou de moi-même, est interrogé comme s'il faisait partie de l'œuvre, parce qu'il est un phénomène de lecture. Il est à noter cependant qu'il n'y a rien de psychanalytique ou de personnel dans ma démarche : l'objet de l'étude n'est pas de montrer ce que la mémoration du détail dévoile du lecteur, mais ce qu'elle apprend sur le texte. Il est toutefois évident que lorsque le lecteur est Marcel Proust, la lecture devient amorce d'écriture et prolonge l'étude littéraire sur l'axe intertextuel.

Une objection doit cependant être envisagée. Est-ce que tout point du texte ne pourrait pas devenir un détail révélateur ? Si l'on étend la notion de réseau à ses dernières conséquences, on peut en effet supposer que n'importe quelle phrase d'un texte pourrait être étudiée selon les modalités des lectures par le menu. Cette idée, tout d'abord, n'a rien qui doive effaroucher ou qui puisse jeter le doute sur la méthode exposée ici. Il serait effectivement possible d'appliquer à tout syntagme la technique qui consiste à rechercher toutes ses occurrences dans le texte et à en commenter les résonances. Cependant, l'expérience montre que tous les points d'entrée ne donnent pas des résultats probants. Il est des phrases clés dans une œuvre, des passages qui, parce qu'ils sont disposés en des points remarquables (*incipit*, *desinit*, début de chapitre, d'alinéa, etc.), ou qu'ils sont désignés par des procédés syntaxiques ou rhétoriques (répétition, anaphore, paronomase, etc.), ou que leur thématique éveille des échos dans l'ensemble du texte, favorisent la mémoration et deviennent exemplaires. J'ai souvent pu vérifier que tout lecteur conserve d'une œuvre littéraire un certain nombre de souvenirs assez vifs et qui semblent souvent peu représentatifs de l'œuvre dans son ensemble. L'attention au détail étant systématiquement dévalorisée dans notre culture critique (comme le montrent Daniel Arasse et Naomi Schor), les lecteurs ont souvent quelque réticence à évoquer ces

souvenirs. Une des ambitions des lectures par le menu est de permettre une revalorisation de ce type de rapport au texte littéraire.

### **Le rapport à l'objet**

On aura remarqué que les détails étudiés dans les lectures par le menu correspondent dans plusieurs cas à des objets : sucre, ciseau à froid, place de Paris, digitale. Il ne me semble pas que le caractère matériel de ces motifs soit dû au hasard. La première cause est d'ordre pratique. Il est en effet plus aisé de travailler sur un thème par le biais d'une recherche d'occurrences dans une base de données textuelles quand ce thème est réduit à une réalité désignée par un nombre de termes limité. Il y a évidemment moins de synonymes de « ciseau à froid » ou même de « pourpre » que de « jalousie » ou d'« exil ». Le travail est donc plus facile à mener dans le premier cas puisque l'on peut être à peu près sûr qu'une recherche ramènera tous les contextes pertinents.

Mais il y a entre l'usage de l'informatique et l'attention au statut de l'objet dans le texte littéraire une parenté plus profonde, qui tient, me semble-t-il, à la médiation d'un outil technologique dans la relation du lecteur à l'œuvre. Lire à l'aide d'un ordinateur n'est pas une opération qui va de soi, et qui n'est pas sans conséquence sur la perception du texte. Outre les nouvelles pratiques de lecture qu'elle induit (l'habitude d'une lecture hypertextuelle influe sur les comportements de lecture séquentielle, de même qu'un adepte du *zapping* supporte difficilement de suivre un film du début à la fin), l'irruption de l'ordinateur dans le champ de la lecture littéraire, qui renvoie le texte à sa matérialité (comme nous l'avons vu dans le chapitre sur la lexicométrie), et à une matérialité que l'accoutumance au livre et la désincarnation de la notion même de *texte* nous avait fait oublier, nous fait reconsidérer le statut de l'*objet textuel* (on me permettra d'user de cette amphibologie pour désigner à la fois

le texte entendu comme objet et l'objet figuré dans le texte).

# Conclusion

Au terme de ce parcours des pôles qui délimitent à la fois le périmètre de mon expérience de chercheur et celui des Etudes Littéraires Assistées par Ordinateur, je voudrais tenter de fournir, en guise de conclusion, une réponse, au moins provisoire, à la question posée par ce dossier (et par ma candidature à l'habilitation à diriger des recherches) : Les ELAO sont-elles aujourd'hui une discipline à part entière ?

## **Une discipline ancillaire ?**

La première manière de concevoir le rôle des ELAO consiste à en faire une discipline ancillaire des études littéraires. Il s'agirait ainsi de lui donner le même statut que d'autres compétences auxiliaires utiles au chercheur en littérature, comme l'histoire du livre, la génétique, la paléographie, la linguistique, etc. Il peut s'agir par ailleurs de sciences parfaitement constituées en domaine mais qui ne sont envisagées par le chercheur que comme des outils pour l'aider dans un travail à visée purement littéraire. De ce point de vue, les ELAO ne seraient pas un champ scientifique par eux-mêmes, mais le point de contact entre deux disciplines, l'informatique et les études littéraires. Cette conception interdisciplinaire et utilitaire des ELAO n'a rien d'infamant, et suffirait à justifier leur existence. De même que certains physiciens se sont spécialisés dans la microscopie électronique et travaillent pour que leurs collègues aient à leur disposition des instruments toujours plus performants, certains spécialistes de littérature seraient chargés de vérifier, de sélectionner, d'adapter des outils

informatiques à l'usage de leurs collègues, qu'ils aideraient à mettre en œuvre une technologie et des procédures étrangères à leur discipline.

Tout contact interdisciplinaire, comme ceux qui existent entre littérature et histoire, littérature et linguistique, littérature et psychanalyse, littérature et sociologie, etc., suppose ainsi la formation d'un corps de passeurs, sur l'une et l'autre rive, capables d'introduire dans leur champ d'origine certains des acquis et des outils en usage dans l'autre, auquel ils se sont initiés selon diverses modalités. Leur rôle consiste à traduire en termes acceptables des visions souvent très différentes de domaines du réel eux-mêmes très éloignés. L'extrême spécialisation de la science aujourd'hui, et le compartimentage qui en résulte, nécessitent toujours plus de truchements de ce type. Le rôle des ELAO, dans ce cas, se conçoit dans un cadre assez étroit, et il ne nécessite que la formation de quelques spécialistes, dans le cadre de filières restreintes. Je dirais que c'est à peu près leur situation actuelle. Leur utilité est suffisamment reconnue pour que la plupart des universités littéraires se préoccupent de donner à leurs étudiants une formation minimum dans ce domaine. Il ne s'agit souvent que d'une formation bureautique et documentaire, mais elle fait partie des objectifs affichés, admis à la fois par le corps professoral, les étudiants et le Ministère.

Le personnel enseignant assurant ces enseignements peut aller d'informaticiens ouverts aux problématiques des sciences humaines et assurant des cours d'initiation aux logiciels de base à des littéraires de formation curieux de ces questions techniques. Il suffirait, pour faire vivre des ELAO entendus ainsi, de s'assurer que certains enseignants recrutés en Lettres disposent d'une formation suffisante sur ces questions. Cette formation elle-même, et la recherche nécessaire dans ce domaine, pourraient être encadrées par un petit nombre de professeurs (je ne crois pas qu'il y en ait plus de deux ou trois en activité en France aujourd'hui) qui se spécialiseraient, au moins partiellement, dans ce champ disciplinaire.

## Un champ disciplinaire ?

Mais un tel scénario ne me semble pas propre à donner à l'informatique toute la place qu'elle devrait avoir dans les études littéraires. J'ai déjà exposé plus haut les deux raisons pour lesquelles je ne pense pas que les ELAO puissent être cantonnées à un rôle auxiliaire :

- Elles sont devenues des outils indispensables à toute recherche littéraire, et il est dangereux de laisser croire qu'elles ne sont qu'un outil parmi d'autres à la disposition du chercheur, ce qui sous-entendrait que celui-ci aurait le choix de ne pas s'en servir. J'ai essayé de montrer que, bien au contraire, contrairement à d'autres disciplines comme l'histoire ou la sociologie, l'informatique est incontournable et ne relève plus aujourd'hui d'un choix méthodologique. La notion de « discipline ancillaire » n'a de sens que dans le cas où les outils que fournit la discipline ne sont employés que par certains chercheurs, travaillant dans des domaines spécifiques. Tout le sens de ce dossier est de montrer qu'il n'en est rien : tout chercheur littéraire, quel que soit le sujet sur lequel il se penche, a intérêt à utiliser un outil informatique.
- L'informatique n'est pas une technique parmi d'autres pour le littéraire, parce qu'elle touche au cœur de sa discipline : le maniement de l'information. De même, la révolution informatique que connaît aujourd'hui notre société n'est en rien comparable à l'arrivée d'autres technologies, comme dans le domaine des transports ou de l'énergie. Il s'agit d'une révolution informationnelle, qui modifie en profondeur les modes de production et les modes de pensée. L'impact, sur une science des textes, d'une technologie capable de manipuler les textes ne peut être même comparée avec l'influence qui fut celle, par exemple, de l'histoire sur les études littéraires à la fin du XIX<sup>e</sup> siècle. Comme j'ai essayé de le montrer, c'est la substance même de nos supports de travail qui est en train de

changer avec l'essor de la numérisation.

Pour ces deux raisons, d'ailleurs complémentaires, cantonner les ELAO à une position auxiliaire reviendrait à confondre l'essentiel et l'accessoire. La notion de « champ disciplinaire », malgré ou en raison de la part d'ambiguïté que contient l'expression, me semble mieux correspondre avec ce que devraient être les ELAO que celle de « discipline ». Les ELAO sont un « champ disciplinaire » en ce sens qu'elles sont au confluent de deux disciplines, mais aussi en ce sens qu'elles occupent une place à l'intérieur d'une discipline, qui est celle des études littéraires, qu'elles ne prétendent ni remplacer ni concurrencer.

Mais les ELAO ne pourront être un champ disciplinaire à part entière que si elles parviennent à ne plus être pure collection d'outils hétéroclites empruntés à la panoplie de l'informaticien. Il faut bien dire cependant qu'il y a eu beaucoup de bricolage conceptuel dans ce domaine. Avec le ravissement que suscitent les nouveautés, des chercheurs se sont emparé un peu au hasard de gadgets à la mode pour les appliquer à leur domaine d'étude, obtenant des résultats au moins douteux. Si les ELAO ne consistent qu'en un catalogue d'applications, il faut effectivement leur laisser le statut de discipline ancillaire : elles ne sont un objet ni d'enseignement (mais de formation) ni de recherche (mais de veille technologique). L'accession au rang de champ disciplinaire suppose une réflexion épistémologique qui fonde les ELAO en les dotant d'une définition, de contours stables, d'un corps de principes et de précautions méthodologiques.

### **Spécificité de l'approche littéraire**

C'est à cette tâche que je me suis employé, comme j'ai tenté de la montrer dans ce dossier. J'ai voulu mettre en évidence qu'au-delà des aides ponctuelles que l'informatique pouvait apporter au chercheur littéraire (et qui ne sont d'ailleurs pas négligeables en elles-mêmes),

elle lui donnait surtout l'opportunité de repenser sa discipline et ses pratiques. Il faut surtout se garder de penser l'irruption de l'informatique dans le champ littéraire comme l'annexion, par une logique scientifique triomphante, d'une contrée barbare et obscurantiste. C'est à un mouvement exactement inverse que nous assistons aujourd'hui dans le monde scientifique. Les sciences dites « exactes » sont en pleine phase de doute (théories « post-modernes », chaos, logiques floues, problèmes éthiques et déontologiques), en recherche de sens. Il serait paradoxal qu'au moment où physiciens et mathématiciens se tournent vers les logiques floues, ou possibilistes<sup>1</sup>, les spécialistes de littérature, sans doute les plus à même d'évoluer dans un univers conceptuel de ce type, fait uniquement de langage, sans concepts stables, aux limites indéfinies, abandonnent leurs savoir-faire au profit d'un scientisme appauvri.

C'est bien à un dialogue d'égal à égal qu'incitent les ELAO, entre deux disciplines qui ont des choses à apprendre l'une de l'autre. L'informatique, en effet, revenue des errements millénaristes et des mythes de l'« intelligence artificielle », tente elle aussi d'intégrer les notions de logique floue au fonctionnement des ordinateurs. Des questions comme la traduction automatique, le traitement automatique des langues naturelles, l'indexation automatique des documents lui posent encore des problèmes insolubles. On peut imaginer que des spécialistes de littérature ont quelque chose à apprendre à des informaticiens sur la manière dont un texte véhicule une signification.

Mais c'est surtout à penser les spécificités de notre discipline que les ELAO doivent nous aider. J'ai exposé dans ce dossier quelques-uns des aspects de cette réflexion : des concepts comme ceux de *texte*, de *thème*, d'*histoire littéraire*, d'*auteur*, des pratiques comme la lecture,

---

<sup>1</sup> . Voir par exemple les *Eléments de logique floue* de Louis Gacôgne (Hermes Science Publications, 1997).

l'édition sont remises en question par l'usage de l'informatique. Notre discipline, vue de l'extérieur, a un aspect à peu près unique : nous écrivons des textes pour expliquer des textes. La limite entre ces deux types de textes n'est pas totalement fixée, et nous n'avons aucun autre critère que l'assentiment de nos lecteurs pour prouver que nos explications sont justes. Si l'on ajoute à cela que les textes littéraires que nous étudions sont volontiers autotéliques et que critiques et écrivains sont parfois les mêmes personnes, que la légitimité sociale de nos recherches est sans cesse remise en question, nous aurons fait le portrait d'une discipline paradoxale, et qui tire sa force de la précarité même de ses fondations. Nous ne manipulons en effet que du symbolique, et c'est bien là ce qui donne de la valeur à nos travaux. Si l'État a choisi de faire enseigner la langue nationale par des professeurs de Lettres, c'est bien en raison de la très forte charge symbolique des connaissances que nous sommes chargés de transmettre.

Ces éléments constitutifs de notre discipline sont inaccessibles à la logique informatique. Mais c'est justement à son contact qu'ils sont mis en valeur. De même que l'on réfléchit mieux à sa condition nationale en franchissant les frontières ou en accueillant des étrangers, on comprend mieux la discipline que l'on pratique en la confrontant à une discipline qui questionne jusqu'aux fondements même de nos pratiques.

### **Projets de recherche**

C'est à ce dialogue disciplinaire, sous ses différents aspects, que je compte me consacrer dans les années à venir :

- Création de formations mixtes de littérature et d'informatique, pour permettre à des équipes pluridisciplinaires, peut-être européennes, d'assurer un enseignement de haut niveau dans les deux domaines.

- Développement de l'enseignement à distance, sous toutes ses formes, y compris la diffusion de contenus de niveau recherche.
- Animation d'une équipe de recherche, si possible interdisciplinaire, qui permette de tester, de développer, d'adapter des applications informatiques et de réfléchir sur les résultats obtenus.
- Enrichissement de la BDHL (développement des coopérations internationales, en particulier) et aide à la création de bases de données. Je souhaite également prolonger le travail sur la question de l'établissement du canon littéraire en réfléchissant aux autres concepts sur lesquels repose la pratique de l'histoire littéraire : auteur, œuvre, périodes, mouvements, etc.
- Approfondissement de la question des statistiques en littérature, sans doute sous la forme d'un manuel de statistiques à l'usage des littéraires.
- Amélioration des interfaces hypertextuelles pour l'édition d'œuvres littéraires.
- Écriture d'une « étude par le menu » de plus grande ampleur (biographie ou étude littéraire)

# Table des matières

<i>Introduction</i> .....	2
<i>informatique et littérature : esprit de géométrie et esprit de finesse</i> .....	6
Naissance de la micro-informatique .....	6
Langue / langages .....	7
Enseignement des Lettres / enseignement de l'informatique .....	9
Informathématiques ? .....	14
Enseignement à distance.....	17
<i>Une base de données d'histoire littéraire</i> .....	21
Une nouvelle histoire littéraire.....	22
Thématique.....	23
Le point de vue des lecteurs .....	25
Recherches en histoire littéraire avec la BDHL .....	27
Etudes d'œuvres à l'aide de la BDHL .....	37
Une navigation hypertextuelle dans l'histoire littéraire .....	40
Production de bases de données.....	45
<i>Hypertextualités</i> .....	49
Edition hypertextuelle .....	49

<b>Hypertextes de création</b> .....	<b>54</b>
<b><i>Lexicométrie</i></b> .....	<b>58</b>
<b>Exploration des bases textuelles</b> .....	<b>59</b>
<b>Spécificités d'un texte</b> .....	<b>65</b>
<b>Découpage automatique des corpus</b> .....	<b>70</b>
<b>Une lexicométrie littéraire ?</b> .....	<b>73</b>
<b>De quelques erreurs d'appréciation</b> .....	<b>74</b>
Les anagrammes de Tzara .....	75
Les anagrammes de Saussure .....	76
<b><i>Lectures par le menu</i></b> .....	<b>78</b>
<b>Le souci du détail</b> .....	<b>79</b>
<b>Revue de détail</b> .....	<b>80</b>
« Les <i>je/nous</i> de Victor Hugo. Pluriel de modestie et pluriel de majesté dans <i>Les Misérables</i> », <i>Op. Cit. Revue de littératures française et comparée</i> , Publications de l'Université de Pau, n°3, novembre 1994, pp. 129-136.....	80
« Sylvie ou le pourpre proustien », <i>Bulletin Marcel Proust</i> , 1996, n°46, pp. 72-108.....	81
« Les Misérables ou les ciseaux à froid », in <i>Victor Hugo 4 - Science et technique, La Revue des Lettres modernes</i> , Minard, 1999, pp. 97-115.....	81
« <i>Madame Bovary</i> ou le danger des sucreries », <i>Romantisme</i> , n°103, 1999, pp. 41-51. ....	82
« <i>Le jardin des Plantes</i> ou l'hypotypose de la place Monge », <i>Frontières de la fiction</i> , Alexandre Gefen et René Audet éd., Editions Nota Bene, Presses Universitaires de Bordeaux, 2001, p. 375-395.....	82
« Les chefs d'œuvre de la langue française sont les discours de distribution pour les lycées », <i>Les Poésies d'Isidore Ducasse, Actes du cinquième colloque international sur Lautréamont, Marseille, 13-15 octobre 2000</i> , Jean-Jacques Lefrère et Michel Pierssens éd., <i>Cahiers Lautréamont</i> , livraisons LV et LVI, Du Lérot, éditeur, 2001, p. 111-124.....	83
« Eugène Mutiaux, le parrain de Proust », <i>Ce que je ne sais pas, Cinquième colloque des Invalides</i> , Jean-	

Jacques Lefrère et Michel Piessens éd., Du Lérot éditeur, 2002, p. 81-84.....	84
<b>Une méthode critique.....</b>	<b>84</b>
<b>Le rapport à l'objet.....</b>	<b>88</b>
<b><i>Conclusion</i> .....</b>	<b>90</b>
<b>Une discipline ancillaire ? .....</b>	<b>90</b>
<b>Un champ disciplinaire ? .....</b>	<b>92</b>
<b>Spécificité de l'approche littéraire.....</b>	<b>93</b>
<b>Projets de recherche .....</b>	<b>95</b>
<b><i>Table des matières</i> .....</b>	<b>97</b>