

# Le tour du propriétaire

Michel LENOBLE

VUILLEMIN (Alain) dir. : 1993, *Les banques de données littéraires, comparatistes et francophones* (Limoges : Presses de l'Université de Limoges et du Limousin) 276 pp. ISBN 2-910016-17-X. Prix : 125 FF. Distribué par le Secrétariat des PULIM, Faculté des Lettres et des Sciences Humaines, 39E, rue Camille Guérin, 87036 Limoges cedex, France.

Alain Vuillemin, professeur à l'Université de la Sorbonne Nouvelle (Paris III), nous propose un recueil d'articles réunis en plusieurs phases. L'idée originale de ce volume a germé lors de deux journées d'études consacrées à la « Bibliographie et documentation sur les littératures francophones », organisées par l'Association des Universités Partiellement ou Entièrement de Langue Française et l'Université des Réseaux d'Expression Française (AUPELF-UREF). L'ouvrage se divise en trois grandes parties, regroupant chacune des articles concernant la même problématique, à savoir : *Les enjeux* (pp. 17-64), *Les domaines* (pp. 65-200) et *Les perspectives* (pp. 201-276). Chaque article comporte soit des notes infrapaginales, soit des références bibliographiques minimales.

Les 26 articles de ce volume sont précédés d'un *Avant-propos* et d'une *Introduction* précisant le contexte dans lequel il faut situer cet ouvrage. Jean-Claude VAREILLE (pp. 9-10) y voit une occasion unique de réfléchir aux bouleversements des conditions de la recherche en littérature et en linguistique depuis la venue de l'ordinateur dans ces domaines de sciences humaines. Jacques FONTANILLE (pp. 11-16) décèle deux changements majeurs par rapport au passé : la modification du « contrat de lecture » et la dématérialisation de l'objet de recherche, bien plus essentiels d'ailleurs, que l'évolution du paysage de la recherche qui prend désormais un tour tout à fait concret depuis l'avènement des équipes de recherche, des réseaux et d'une nouvelle communauté scientifique.

La première partie, *Les enjeux*, s'ouvre par un article de Jacques VIRBEL présentant (pp. 19-28) les fonctionnalités de base des stations de lecture de la future Bibliothèque de France (BdF). Outre un environnement de recherche informatisé pour les recherches bibliographiques, la saisie de textes, la rédaction de textes, la consultation de dictionnaires

---

✉ Département de Littérature Comparée; Université de Montréal; C.P. 6128; Succ. « A »; Montréal (Québec); CANADA H3C 3J7.  
E-mail : lenoble@m.ere.umontreal.ca

---

en accès direct, les liens avec la communauté scientifique et les connexions avec les autres bibliothèques, des stations de lecture assistée par ordinateur sont mises à la disposition des visiteurs de la BdF. On est en droit de se demander si ces fonctionnalités offertes aux chercheurs ont vraiment leur place dans une bibliothèque, alors que les connaissances requises pour travailler efficacement avec ces nouvelles stations informatisées sont loin de faire partie du bagage courant de formation de tous les chercheurs en Lettres. Sans se faire critique du projet de la BdF, Virbel situe cependant très bien les enjeux fondamentaux dissimulés derrière ces nouveautés techniques; il essaie de voir *in concreto* les modalités de réalisation du travail d'annotation et de recherche qu'effectue tout chercheur sur son corpus d'étude. Il indique lesquelles de ces pratiques seraient informatisables dans un futur rapproché.

Jean-Louis JOUBERT (pp. 29–32) coordonne le projet de bibliographie systématique, exhaustive, informatisée et consultable en ligne des littératures francophones. Malgré les problèmes inhérents à la définition des limites du corpus, à l'élaboration des normes communes de publication et aux différences de matériels et de logiciels, la bibliographie en accès direct sera située sur un serveur de l'Université de Limoges, exploitée via les réseaux de télécommunication (TRANSPAC) et le système *Spirit*. L'UREF mettra dans les pays du Sud des points SYFED (SYstème Francophone d'Édition et de Diffusion) à la disposition des universités et des chercheurs. Diffusion papier et support informatique continueront à cohabiter pour un temps encore. Quelques publications papier viennent déjà d'être publiées : *Bibliographie de littérature maghrébine (1980–1990)*, *Bibliographie des études littéraires haïtiennes*. Viendront bientôt une *Bibliographie libanaise* et une *Bibliographie courante de l'Afrique et du monde négro-africain*.

Claire PANIJEL souligne (pp. 33–44) que l'offre mondiale d'information électronique est d'environ 8 000 banques de données, ce qui représentait, en 1990, un marché de 9 milliards de dollars. La part de marché détenue par les banques de données en Sciences humaines et sociales représente moins de 10 % des banques disponibles et la littérature n'occupe qu'une faible partie de cette fraction. Elle note également l'émergence de produits mixtes, comportant non seulement une bibliographie, le texte intégral, des traductions, mais aussi des atlas, des photographies, une encyclopédie et des logiciels permettant des analyses linguistiques. Panijel liste pour les littératures francophones les banques de données pertinentes : catalogues de bibliothèques ou librairies, les répertoires de thèses, les répertoires d'organismes de documentation ou de recherche, les banques de textes intégraux et les répertoires bibliographiques.

Après un court article (pp. 45–48) à saveur bibliométrique portant sur la francophonie et dû à la plume de Jacques FAULE, Claude CAZALÉ-BÉRARD (pp. 49–56) s'interroge sur le devenir du texte et l'évolution du travail de recherche en Lettres à l'ère électronique. Elle s'intéresse principalement au texte à des fins de critique textuelle et note, à cet égard, une redéfinition du concept de texte, qu'elle envisage dans une évolution diachronique constante. Nous sommes, selon elle, à l'ère du texte comme procès, dont les différents états constituent les éléments d'une banque de données. Elle suggère une nouvelle économie du travail de recherche, selon laquelle il s'agirait d'abord de refaire l'histoire du texte (génétique) avant d'en faire une étude sociologique (avatars de la circulation et de la diffusion) et enfin, de voir les aspects du texte qui ont principalement retenu l'attention des critiques. Elle ébauche à peine le travail de recherche sur le contenu même des textes et ne semble pas réaliser que très peu de chercheurs sont réellement

familiers avec le matériel, les techniques et les méthodologies particulières aux traitements automatiques des textes littéraires.

Henri BÉHAR donne, de façon succincte (pp. 57–64), quelques exemples de travaux exécutables avec des logiciels et des CD-ROM sur le poste de travail du littéraire. En conclusion, il réaffirme (p. 64) très justement les trois fonctions de l'ordinateur dans les recherches en littérature, à savoir la fonction d'instrument de validation des hypothèses, la fonction documentaire et informative, et la fonction heuristique. Il est, lui aussi, convaincu que le travail de recherche à l'aide de l'ordinateur est le point de départ obligé de toute recherche et que rien ne peut être dit qui ne soit vérifié et même quantifié.

La deuxième partie du présent ouvrage, *Les domaines*, rassemble une série d'articles traçant plus les contours de la francophonie qu'ils ne traitent de développements informatiques concrets, ou n'abordent de problèmes épistémologiques ou méthodologiques s'y rattachant.

Jacques CHEVRIER dresse un bilan général de la production littéraire francophone (pp. 67–77), discute le concept et la réalité de la francophonie, et présente un relevé des productions par aires géographiques. Bizarrement, il ne retient des auteurs francophones de Belgique que les auteurs d'origine flamande! Les auteurs wallons semblent totalement méconnus.

Charles BONN (pp. 79–95) et Fériel KACHOUKH (pp. 97–102 et pp. 103–107) font, dans trois articles différents, état du relevé bibliographique du domaine maghrébin et du projet LIMAG de bases de données bibliographiques des littératures maghrébines en langue française. En plus des livres et des articles de littérature seconde, ce projet inclut également un répertoire international des chercheurs; le tout est géré par le SGBDR *Paradox*. Le projet LIMAG a une triple mission : documentaire, de traitement de l'information et de restitution/diffusion sur supports informatique et imprimé. Vu la décentralisation des opérations de collecte de l'information bibliographique, l'équipe responsable fait face à des problèmes d'uniformisation de la saisie des données et de leur indexation, ainsi qu'à des problèmes d'intégrité des fichiers de la base de données. C'est là leur principal souci.

Dans ses deux articles (pp. 109–112 et 113–120), Virginia COULON présente la banque de données des littératures africaines (LITAF) répertoriant la production littéraire en langue française pour l'Afrique sub-saharienne. Cette banque est gérée par le logiciel *Cds-Isis* de l'UNESCO. Suite à cette expérience, Coulon livre ses conseils pratiques à l'intention de futurs développeurs de banques de données du même genre, et ce, particulièrement en ce qui a trait à la planification, l'analyse et l'évaluation des moyens matériels et financiers requis.

Maguy GRASSIN (pp. 121–128), après avoir tracé un portrait de la situation de la langue et de la culture louisianaises, évoque la façon de procéder retenue pour constituer la base de données bibliographiques « Langue et Culture en Louisiane ». Il s'agit en fait dans un premier temps de saisir par traitement de texte les répertoires existants, pour ensuite y ajouter les nouvelles entrées au fur et à mesure des publications. *Spirit* sera probablement utilisé pour son exploitation publique.

Jacques LEMARIGNIER (pp. 129–134) fait une présentation générale rapide de l'utilisation de la banque de textes littéraires français *Frantext* à la Bibliothèque Publique d'Information (BPI). Il montre combien il est facile, sur un corpus de 2 500 textes,

d'effectuer des recherches de citation, de documenter l'histoire d'un mot ou de son utilisation, de préparer une étude de thème.

Étienne BRUNET, à partir de la même base, réalise une étude (pp. 135–145) brillante et raffinée de la « raison ». Son article est l'exemple type de travail statistique sur *Frantext* accompagné de commentaires très lucides venant d'une très longue expérience de recherche dans le domaine, et très conscients des limites de la méthode et des interprétations que l'on peut en tirer.

Éveline CADUC propose une recherche (pp. 147–156) sur le « vert » chez Saint-John Perse dont les données statistiques ont été établies par le logiciel *Hyperbase* d'Étienne Brunet. Elle met l'accent sur l'aspect récursivité et similarité des composants phonétiques, lexicales, syntaxiques, métriques et connotatives du texte.

Pierre BRUNEL (pp. 157–159) et surtout Alain VUILLEMIN (pp. 161–176) nous donnent un historique du projet *Ariel-Littéral* de banque de bibliographie comparatiste et francophone de Paris-Sorbonne (Paris IV). Cette base présente l'originalité d'inclure un bref résumé, sorte d'approximation conceptuelle du contenu, pour chacune des entrées et de permettre une recherche, par le biais de *Spirit*, de ces informations en texte intégral. Il peut en ressortir, par exemple, une analyse fine de la composition, de la structure et des clivages du vocabulaire critique. Dans un avenir rapproché, Vuillemin projette de développer les possibilités d'interrogation multilingue de la base *Ariel*.

Frédéric FOUSSIER (pp. 177–194) fait une présentation détaillée de *Spirit* (Système Probabiliste d'Indexation et de Recherche d'Information Textuelle), lequel travaille à partir des résumés en texte intégral qui accompagnent les références bibliographiques de la base *Ariel*. Il détermine le pouvoir informationnel des entrées bibliographiques en étudiant les fréquences et les fonctions des lemmes de leur résumé. *Spirit* permet l'interrogation en langage libre, l'indexation automatique et l'interrogation multilingue. Néanmoins, la combinaison *Spirit-Ariel* pâtit de deux inconvénients : l'un, lié aux défauts inhérents à tout système de dépistage d'information, à savoir la rigidité des analyseurs et des dictionnaires, et l'autre, venant du fait que le travail est effectué sur des textes seconds (résumés produits manuellement) plutôt que sur les textes intégraux des ouvrages référencés.

Cette deuxième section du livre se termine par un court article (pp. 195–200) de Marie Anne POLO DE BEAULIEU à propos du projet de banque de données des *exempla* médiévaux recensant et les recueils d'*exempla* et les *exempla* eux-mêmes.

La dernière partie de l'ouvrage, consacrée aux *Perspectives*, débute par deux articles (pp. 203–220 et 221–227) portant sur la maquette d'une banque de données bilingue en littérature comparée : *El Hadj*. Tout d'abord, Frédéric FOUSSIER essaie d'appliquer des techniques de recherche documentaire à des travaux d'analyse littéraire au prix de raccourcis et de simplification théoriques inquiétantes. Le titre de son article nous annonce « une édition de texte à distance », alors qu'en fait, il s'agit d'une simple opération de téléchargement. Foussier affirme aussi (p. 213) que le logiciel *Spirit* est : « [...] un instrument inédit pour explorer l'architecture thématique secrète d'un récit », car dit-il : « [...] *Spirit* épargne à l'utilisateur le soin de faire varier sa question ». Il n'y a là rien d'inédit puisque c'est déjà pratique courante depuis que les concordanciers existent et que l'on effectue des recherches sur les racines et les sèmes. Foussier suggère enfin (pp. 217–219) une série de nouvelles fonctionnalités prévues pour *Spirit* telles que : des

statistiques lexicales, des outils d'assistance à la traduction, l'étude du rythme des phrases. Si l'on en juge par la description succincte qu'il nous donne de ces fonctionnalités, on ne peut s'empêcher de noter qu'elles ont déjà fait l'objet de développements beaucoup plus poussés par ailleurs<sup>1</sup>.

Alain VUILLEMIN trace l'ébauche d'une banque de textes bilingue ou multilingue permettant les interrogations multilingues venant préfigurer les modèles futurs d'édition de textes littéraires. Vuillemin y défend, de façon fort intéressante, une vision du texte électronique comme objet « infiniment variable et pluriel » dont les « états imprimés n'en seraient que des aspects particuliers [...] liés aux contraintes d'une technologie, celle de la conception et de l'impression des livres ». Les « états informatisés en révéleraient au contraire les virtualités potentielles, voire son mouvement intérieur, son dynamisme. »

Bernard GICQUEL suggère (pp. 229–238) dans son article une conception de la stylistique « rationnelle et technologique pour donner prise à automatisation ». Il s'intéresse particulièrement aux opérations menant du discours littéraire au discours scientifique à son sujet; c'est là que, selon lui, s'effectue le travail de recherche en stylistique. C'est, à notre avis, une approche originale des problématiques littéraires qui permettra de faire avancer l'automatisation de la recherche en littérature à condition d'étudier les modalités de la transition entre l'objet et le discours scientifique sur l'objet.

Jean-Pierre BALPE (pp. 239–251) nous livre une excellente typologie des générateurs de textes littéraires et des éléments permettant de distinguer les différents types de textes générés. La littérature générée par ordinateur (LGO) revêt en fait un double intérêt puisque, d'une part, l'exercice de la génération vient enrichir par ses résultats la pratique de l'analyse et que, d'autre part, la LGO elle-même sert de repoussoir pour mieux faire apparaître les contours du littéraire, de l'écriture, de la lecture et des droits d'auteur.

Le présent ouvrage se termine par deux articles, l'un de Michel BERNARD (pp. 253–269) est consacré à la banque de données d'histoire littéraire rassemblant des auteurs, des œuvres, des événements historiques et littéraires ainsi que des institutions (académie, revues, prix, mouvements). Elle permet des interrogations croisées faisant ressortir des réponses présentant des caractéristiques communes. Bernard souhaite pouvoir, grâce à ce type de bases de données, renouveler la réflexion en histoire littéraire (objet, méthode, périodisation), faciliter la recherche documentaire et fournir des outils pédagogiques.

*Les banques de données littéraires, comparatistes et francophones* se termine par l'article de Jérôme BASCHET qui nous donne une description (pp. 271–276) de la base de données iconographiques sur vidéodisques des manuscrits de la bibliothèque Vaticane. Le projet prévoit la prise de 400 000 clichés d'enluminures, miniatures, lettres historiées et décors secondaires auxquels sont couplées des fiches documentaires avec les références ainsi qu'une série d'index décrivant leur contenu (thème, lieu, noms propres, objet, légende, etc.). Le SGBD prévoit des interrogations croisées.

---

<sup>1</sup> Voir, par exemple, au sujet de l'étude du rythme de la phrase, l'étude fondamentale de Brunet : BRUNET (Étienne) : 1985, « La phrase de Zola », *La critique littéraire et l'ordinateur / Literary criticism and the computer*, B. Derval et M. Lenoble, dirs (Montréal : Derval & Lenoble), pp. 111–157.

\*  
\*   \*   \*

L'ouvrage proposé par Alain Vuillemin est en soi une véritable mine d'informations concernant les divers projets bibliographiques et littéraires de la francophonie. C'est une photographie instantanée de la situation actuelle, à la fois inquiétante et pleine de promesses. Inquiétante, parce qu'elle donne à voir les zones où peu est réalisé (Wallonie) et où beaucoup n'est encore que projet, parce qu'elle cerne les manques criants de ressources et de standardisation, et qu'elle souligne le caractère souvent artisanal des réalisations. Pleine de promesses, car elle fait état d'un foisonnement de projets et d'une volonté de rattraper le temps perdu.

En plus de nous faire faire le tour du propriétaire, Vuillemin ouvre, à quelques reprises, des perspectives de réflexion stimulantes quant à l'avenir de l'automatisation de plus en plus poussée des recherches en littérature. Deux angles d'attaque nous paraissent tout à fait porteurs : l'étude du travail *in concreto* des chercheurs (Virbel) et l'étude du métadiscours sur le littéraire (Gicquel), afin d'en découvrir les modalités et les finalités, resserrent petit à petit le domaine résistant à une plus grande automatisation.

Outre deux anglicismes inattendus en pareil contexte (*scanner* [numériseur] et *login* [code d'accès]), nous devons souligner que cette publication est bien plus informative — c'est là son but — que théorique ou réflexive; elle est cependant indispensable à qui s'intéresse à la francophonie, au développement des banques de données et aux industries de l'information et de la langue.