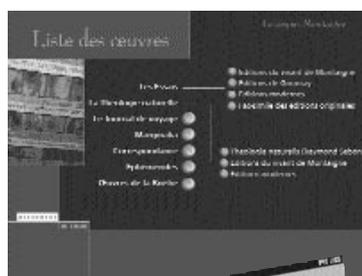


NON À LA BIBLIOTHÈQUE VIRTUELLE



Depuis que l'homme a commencé à produire des biens durables, il a été obligé de trier ce qu'il conservait et ce qu'il jetait. Aux bibliothèques est revenu le rôle de conserver au moins, au mieux de mettre à disposition les livres publiés. Et depuis qu'elles conservent, elles jettent, le manque de place vouant régulièrement une partie de la production au rebut.

L'émergence du support électronique va changer cela : pour la première fois, la réduction d'un document à une suite d'impulsions électriques abolit la frontière entre l'existant, le publiable et le conservé. Nous vivons aujourd'hui un temps où tout est « enregistré », et de fait publiable et conservable dès la conception.

Le tintamarre du village global

Cela ne laisse pas d'être affolant. Quand on imagine la masse d'informations produites et sauvegardées chaque minute par le village global, et qu'on se dit qu'un jour – demain – l'ensemble de ces informations sera simultanément disponible à tout un chacun, on imagine à quel point nous serons submergés. Nous verrons les flots de la mer Rouge se refermer sur nous et la route de la connaissance se fermer à jamais, comme ces fous que l'on retrouve morts chez eux au

milieu de monceaux d'emballages qu'ils n'ont pas eu le cœur de jeter.

Plus prosaïquement, à quoi serviront donc les bibliothèques, et les bibliothécaires, quand tout le savoir sera disponible et consultable chez soi ? En fait, le métier a vécu sa première révolution industrielle avec l'introduction de l'informatique dans les procédures de catalogage et de recherche bibliographique ; Chadwyck Healey France, devenue Bibliopolis depuis son rachat par les auteurs de cet article, y a pris sa part. Mais cette évolution technique n'est rien à côté de la formidable implosion qui va résulter de la numérisation des contenus eux-mêmes.

Les premières bases de données françaises en texte intégral

Aujourd'hui apparaissent sur le marché les premiers cédéroms de texte intégral. On peut en isoler trois types, tous trois tendant à l'exhaustivité.

Les bases lexicographiques. Tout le monde connaît Frantext, et son service original de consultation sur Minitel, qui préfigure sans doute un usage futur, effectué – et payé – à l'acte, des bases littéraires en ligne. Initiée et maintenue par le CNRS, cette base est alimentée par de nombreuses publications nouvelles. En revanche, sa consultation reste limi-

JEAN-PIERRE SAKOUN
JEAN-MICHEL OLLÉ

Bibliopolis - Le Catalogue des Lettres

tée aux occurrences d'un mot, sans accès direct au texte intégral.

Le corpus de textes. Les premières bases françaises en texte intégral ont vu le jour en 1996. Se démarquant des cédéroms culturels centrés autour d'une œuvre (souvent d'ailleurs proposée de façon partielle)¹, ces cédéroms se veulent sinon exhaustifs, du moins très représentatifs.

Acamedia a ainsi mis sur le marché un *Alexandre Dumas* qui propose soixante-six textes de l'auteur en texte intégral. Un *Victor Hugo*, présenté par Arborecence (Havas Éditions), rassemble peu de textes de l'auteur pourtant le plus prolifique de la littérature française, mais a le mérite de mettre l'œuvre en résonance avec son temps... et le cédérom en contact avec Internet, puisqu'il renvoie à des sites consacrés à l'auteur.

Le Catalogue des lettres, société sœur de Bibliopolis, a mis en vente cette année la première base littéraire française sur cédérom à vocation exhaustive : *Romanciers réalistes et naturalistes, 1820-1910*, soit trois cents romans en texte intégral comportant l'œuvre romanesque intégrale des grands auteurs de la période, de Balzac à Zola, et un choix représentatif des autres romanciers de la même mouvance.

Il s'agit du point de vue éditorial de rassembler sur le même support des textes autrefois proches que l'Histoire et le Temps ont séparés. Dans la République des lettres, tous les textes sont frères. Les auteurs d'une époque se lisent et s'interrogent, leurs textes se répondent. Ce dialogue entre tous les textes d'une période se perd ensuite et finit par se réduire, au gré des aménagements de collections, à de pauvres tête-à-tête : Balzac contre Zola, ou plutôt *La Comédie humaine* versus *Les Rougon-Macquart*, au pire *Le Père Goriot* contre *Germinal*. Plus que la cathédrale engloutie chère à



**ON PEUT PLUS
RAISONNABLEMENT
PRÉDIRE
QUE RESTERONT
SUR CÉDÉROMS
TOUTES LES BASES
QUI DEMANDENT
DES CONSULTATIONS
LONGUES
ET CONSTANTES ET
QUE MIGRERONT
VERS LE RÉSEAU
TOUT CE QUI A
BESOIN DE MISE À
JOUR RAPIDE ET
RELÈVE
D'INTERROGATIONS
PLUS PONCTUELLES**

Michel Serres, c'est la richesse des débats qui s'y tenaient que nous cherchons à exhumer.

La somme érudite. Bibliopolis et la librairie Honoré Champion publient, sous la direction de Claude Blum, un *Montaigne* où se mêlent et se répondent sur le même disque toutes les éditions, diplomatiques et établies, de toutes les œuvres publiées par lui de son vivant, mais aussi par sa fille d'adoption, Mme de Gournay, après sa mort, ainsi que toutes les grandes éditions critiques parues du XVI^e siècle à nos jours, et aussi une édition complète de ses œuvres en anglais, italien, allemand, espagnol, le facsimilé de toutes les premières éditions et de l'exemplaire de Bordeaux. Le cédérom trouve là tout son sens : lui seul permet de rapprocher ces textes et de recréer cette communauté culturelle autour d'une époque ou d'un auteur.

Même si les grands prêtres de la nouveauté annoncent sa fin au profit d'Internet, on peut plus raisonnablement prédire que resteront sur cédéroms, pour des raisons de vitesse d'accès, toutes les bases qui demandent des consultations longues et constantes et que migreront vers le réseau tout ce qui a besoin de mise à jour rapide et relève d'interrogations plus ponctuelles.

C'est pourquoi, à Bibliopolis, nous avons fait migrer notre logiciel vers une interface de type Internet Explorer ou Netscape, même s'il nous semble prématuré aujourd'hui de porter nos bases sur Internet.

Les besoins des bibliothécaires...

Du texte électronique, certes, mais pour quoi faire ? demandent les bibliothécaires.

Le premier argument qui plaide en sa faveur est qu'il permet indiscutablement une augmentation du fonds de la bibliothèque bien supérieure à ce que des budgets toujours serrés et toujours aléatoires permettent quand il s'agit de papier. Même à leur prix de vente actuel, l'acquisition de trois cents romans reproduits sur un cédérom à 9 000 F revient incomparable-

1. Voir le *Madame Bovary* de Ubisoft, qui, malgré d'indiscutables qualités pédagogiques, ne propose pas l'intégralité du roman.

ment moins cher que les trois cents volumes papier correspondants, fusent-ils tous édités en livre de poche, ce qui est loin d'être le cas. Et un cédérom, c'est dix mètres de rayonnage en moins.

Le second relève de l'offre culturelle fondamentale : les cédéroms de littérature présentent des corpus complets de textes aux lecteurs, aux enseignants, aux étudiants, aux chercheurs, et la possibilité de lancer des requêtes susceptibles de satisfaire aussi bien une recherche extrêmement pointue que la préparation d'un cours... ou d'une dissertation.

Enfin le support électronique même, solide et peu encombrant, permet de modifier les pratiques de prêt, de conserver en lecture la forme papier plus fragile, et de faire ainsi de substantielles économies de maintenance du fonds.

... et des lecteurs

Du texte électronique, certes, mais pour faire quoi ? demandent les lecteurs.

Pour lire ? Non, chacun sait qu'il est difficile de lire sur écran². Pour travailler. Rassemblées sur une ou plusieurs galettes, ces bases de texte intégral seront interrogeables, comparables, rapprochables, découpables, recopiables, déformables, recomposables, à condition d'être interrogées par des logiciels d'indexation et de recherche suffisamment puissants³.

Le logiciel Trevi, ex-Caravan, mis au point par Chadwyck Healey France et repris et développé par Bibliopolis, est actuellement un des quatre meilleurs

2. Chaque fois que nous faisons cette affirmation, nous sommes contredits dans nos démonstrations, sur nos stands, par nombre de personnes qui, au hasard d'une recherche lexicographique, « tombent » sur un texte de Balzac qu'ils ne connaissaient pas, comme cela leur arrivait au hasard d'un rayonnage, et se mettent à lire le texte qui défile sur l'écran.

3. On comprendra que cela dépasse largement les initiatives, fort respectables au demeurant, d'internautes mettant sur le réseau, en vrac et sans garantie, des textes littéraires « interrogés » par des *browsers*.

L'ami américain

Les bibliothèques américaines achètent beaucoup de textes électroniques, à des prix plus élevés que ceux pratiqués en Europe. Elles ne sont pas pour autant aussi riches que nous nous complaisons à l'affirmer. Elles n'en ont pas moins pris le virage du texte électronique depuis quelques années.

De fait, la présence anglo-américaine est massive dans l'offre globale de littérature électronique. La guerre culturelle n'aura pas lieu sur les réseaux et les logiciels, durablement américains, mais sur les contenus. Il suffit pour s'en rendre compte de mettre face à face la puissance des éditeurs électroniques américains, de Chadwyck Healey à UMI, la floraison de textes gratuits fédérés par le projet Gutenberg, qui recense et organise les textes de littérature mis sur le *net* par les chercheurs, la multiplication de forums consacrés aux auteurs, et le nombre de publications érudites qui commencent à migrer vers une forme exclusivement électronique, ce qui est considéré aux États-Unis comme inéluctable.

Cette avance se traduit également dans l'utilisation d'Internet. Pour la première fois, cette année, les bibliothèques américaines ont dépensé plus d'argent sur Internet qu'en cédérom dans leurs achats de littérature électronique. Pour la première fois le *Library Journal* (septembre 1997) ouvre le dossier des droits littéraires sur Internet. On s'achemine vers un accord qui pourrait

faire date, il est vrai facilité par la conception anglo-saxonne du *copyright*. Toutes les études relèvent par ailleurs que la coexistence de la même offre sur cédérom et en ligne ne nuit pas au support cédérom, qui continue à être préféré à Internet par nombre d'établissements outre-Atlantique. A des prix élevés : les produits proposés y sont en moyenne une fois et demie plus cher que les produits français. Mais nombre de bibliothèques s'organisent en consortium pour acheter les produits et en partager l'exploitation. Cette pratique reste fort étrangère au milieu des bibliothèques françaises. Pour suivre la laborieuse émergence d'un consortium d'humanités anglo-américaines à l'initiative de quelques enseignants de la Sorbonne, nous savons, et nous regrettons, que de telles pratiques se heurtent à beaucoup de difficultés administratives et financières, souvent surestimées par les bibliothécaires eux-mêmes. Il s'agit là certainement d'une réticence culturelle que les associations, l'Agence bibliographique de l'enseignement supérieur, la Fédération française de coopération entre bibliothèques devraient prendre à cœur de lever dans les années à venir. L'avenir électronique des bibliothèques est sans doute quelque part entre Internet, Intranet, et des réseaux spécifiques entre bibliothèques.

logiciels mondiaux de recherche lexicographique. Mis au point pour exploiter les bases de la BnF, il a été adapté par nos équipes à l'exploration des textes longs et donne des résultats prodigieux en terme de recherche et de combinaison des critères et des opérateurs.

Nous en sommes aujourd'hui aux balbutiements de la recherche littéraire assistée par ordinateur. C'est ici que les bibliothécaires ont un rôle majeur d'initiation à ces méthodes, eux qui sont la plupart du temps plus avancés dans ces domaines que les lecteurs qui fréquentent leurs établissements. D'autant que l'exploration de ces bases littéraires devrait produire, au même titre qu'au siècle dernier la généralisation des manuels, de nou-

velles pratiques pédagogiques que nous sommes loin d'avoir entièrement explorées aujourd'hui⁴.

4. Nous travaillons actuellement, en coopération avec des entreprises de génie logiciel, dans quatre directions : l'introduction du langage naturel, qui permet un accès plus convivial, donc plus aisément partageable par tous, aux requêtes de recherche ; l'introduction de l'intelligence artificielle, qui permet au logiciel de s'enrichir et d'affiner ses recherches au fur à mesure qu'il intègre les paramètres ou les habitudes du chercheur ; la mise au point de thésaurus particuliers qui permettra au chercheur d'explorer des notions et des concepts ; enfin, la production de cartographies du vocabulaire des textes, qui permettent de repérer des thématiques, en mettant en évidence des rapprochements ou au contraire des isolats.

Cette relation nouvelle doit être à double sens. En effet, pour la première fois, les bibliothécaires vont devoir acquérir des sommes de textes trop importantes pour qu'ils puissent en évaluer eux-mêmes la pertinence. Il faudra bien qu'alors les usagers jouent un rôle de prescripteurs. Il y a là source d'une nouvelle forme d'échange qui devrait peu à peu trouver ses marques.

Pourquoi pas les éditeurs papier ?

On peut s'étonner, à voir les avantages multiples du support électronique pour les corpus d'œuvres littéraires, de ne pas voir les éditeurs se précipiter sur ce créneau et offrir à tous les publics plus de cédéroms de littérature.

Un certain nombre de raisons expliquent ce relatif désintérêt, la plus importante restant économique. Numériser de la littérature dans de bonnes conditions coûte très cher. Pour rentabiliser auprès du grand public un cédérom fabriqué pour un coût oscillant entre un et deux millions de francs et vendu à moins de 400 F, il faut en vendre au moins 20 000 exemplaires.

NOUS EN SOMMES AUJOURD'HUI AUX BALBUTIEMENTS DE LA RECHERCHE LITTÉRAIRE ASSISTÉE PAR ORDINATEUR

Or sans même parler de la qualité de l'offre, l'état du parc actuel fait que le nombre moyen de ventes d'un cédérom culturel ne saurait dépasser 1 500 exemplaires, sauf rares exceptions.

De fait, les acheteurs naturels de cédéroms sont, et restent, les bibliothèques. Elles seules sont intéressées par les grandes masses de textes organisées en corpus thématiques... et ont les moyens de les acquérir. Nous avons quant à nous mis au point des méthodes qui permettent de publier des ouvrages de haute érudition comme le *Montaigne* pour des prix proches de 25 000 F, et des sommes de textes comme *Romaniers réalistes et naturalistes* pour des sommes inférieures à 10 000 F. Ces prix restent très inférieurs à ce que produisent nos amis anglo-saxons.



Du virtuel au réel, il n'y a qu'un pas

De fait, au-delà de la nouveauté du média, la littérature électronique soulève bon nombre des problèmes de positionnement auxquels bibliothèques et bibliothécaires vont être confrontés dans les années à venir.

Dans ce monde où, pour la première fois, les capacités de stockage seront à la hauteur des capacités de production de l'information, c'est-à-dire quasiment infinies, les gestionnaires de l'information vont avoir un rôle d'orientation et de prescription de plus en plus important. Et la nécessité de leur expertise va suivre la courbe de

croissance de la production de textes nouveaux, auxquels s'ajoutera la conversion rétrospective des textes anciens.

LES ACHETEURS NATURELS DE CÉDÉROMS SONT, ET RESTENT, LES BIBLIOTHÈQUES

Non, la bibliothèque ne doit pas être virtuelle. Placée au cœur des échanges culturels et gérant l'extraordinaire foisonnement d'informations qui nous attend, recentrée sur les services au lecteur, lieu d'échanges, de culture, et d'organisation des chemins de la connaissance, elle doit fortement s'ancrer dans la réalité des besoins de ses usagers. C'est ainsi qu'elle exploitera au mieux ses fonds, qu'ils soient virtuels ou physiques.

Septembre 1997

