

Livre électronique et station de lecture assistée (1)

de Jacques Virbel

Centre national de recherche scientifique
et Institut de recherches en informatique de Toulouse (France)

et Yannick Maignien

Bibliothèque nationale de France

- [Le «Livre électronique» et sa lecture](#)
- [Quatre fonctionnalités de base de la «lecture assistée par ordinateur»](#)
- [Navigation, recherche, lecture](#)
- [Les Bibliothèques numérisées](#)

■ I - Le «Livre électronique» et sa lecture

La notion, et la formule même de «lecture assistée par ordinateur» du livre électronique est apparue en premier lieu, semble-t-il, à partir de 1988, à l'occasion des études et travaux entrepris dans le contexte initial de la Bibliothèque de France (2). Cette sorte de lecture se distingue très clairement de la consultation de bases de données textuelles, même si certains modèles du livre électronique présentent celui-ci comme une base de données particulière. Cette lecture se signale à l'attention par six propriétés fortes:

a) Le terme même de «livre électronique» possède déjà au moins quatre ou cinq sens, mais il s'agit encore le plus souvent d'*ouvrages électroniques* produits de la numérisation d'ouvrages sur support papier. Ce changement de support, qui touche entre autres à la perception et à la manipulation (au sens propre) modifie l'activité même de lecture par rapport à son contexte classique. Ici, l'assistance, en termes de «lecture assistée par ordinateur» (LAO) concerne la suppléance aux limitations et inconvénients de ce mode d'existence du texte (par ex.: surmonter la perte de perception de la dimension «épaisseur» et, par fait, celle des structurations typo-dispositionnelles à longue et moyenne portée). Le marquage ou balisage des structures logico-linguistiques des textes, introduites dans ou associées aux fichiers représentant ces textes, ainsi que diverses formes d'accès simulant des types de lectures de survol ou diagonales, peuvent assumer cette fonction.

b) Mais cette même assistance concerne aussi bien l'exploitation des possibilités nouvelles qu'offre ce mode, parmi lesquelles la prospection linguistique: inspection aussi détaillée que nécessaire du lexique,

application d'analyseurs syntaxiques, stylistiques, statistiques, ou autres, exploitation de connaissances linguistiques, elles-mêmes existant parfois sous forme de programmes informatiques. Ces possibilités, qui outrepassent généralement tout ce qui est envisageable «à la main», ouvrent la perspective de la pratique de tests d'hypothèses quasi-instantanés, et par le fait, celle d'une «lecture expérimentale».

c) Ce mode ouvre aussi des possibilités proprement infinies de dé/re-groupement/structuration, de sorte que la notion traditionnelle de *corpus*, au sens méthodologique (classes de documents pertinents) et technique (domaine d'application d'un programme donné) peut prendre la position d'une variable, et ainsi changer du tout au tout de statut, et contribuer aussi à cette allure expérimentale.

d) De plus, la numérisation de catalogues en même temps que celle des fonds crée une continuité elle aussi nouvelle, et la problématique de la caractérisation conceptuelle des fonds, ainsi que celle de la sélection des ouvrages (tout ou parties) pertinents pour une recherche s'en trouve entièrement redéfinie.

e) Une seconde continuité, tout aussi inédite, est créée entre la lecture et l'écriture, ici aussi par le fait de la continuité des supports (électroniques). La possibilité d'associer au texte des traces de réaction de lecture sous forme écrite (annotations, soulignements, attachement de commentaires, renvois automatiques, etc.) et l'utilisation de divers éditeurs (graphiques, hypertextuels, etc.) crée en effet un *continuum* entre lecture et écriture, tant au plan du travail individuel que collectif ou partagé.

f) Enfin, l'ensemble de ce contexte rend nécessaire la perception et la gestion des nombreux et différents produits de la lecture. On peut les appréhender comme une sorte de généralisation ou de mise au goût informatique par exemple des classiques tables analytiques, des cartes sémantiques ou encyclopédiques, ou encore de l'indexation individuelle, ou aussi des (presque déjà) classiques «historiques» ou «vues mosaïques» des systèmes hypertextuels.

■ II - Quatre fonctionnalités de base de la «lecture assistée par ordinateur» (LAO)

On conçoit que ces diverses possibilités ne sont viables que pour autant qu'un ensemble d'outils soient élaborés. Ils doivent concerner entre autres:

- la description des structures logico-linguistiques des textes,
- les descriptions encyclopédiques et indexicales des contenus,
- la modélisation d'activités de lecture (par ex.: typologies d'annotations),
- l'investigation linguistique fine sous les angles syntaxique, sémantique, stylistique, rhétorique, etc.,
- l'organisation, le classement et l'indexation individuels ou personnalisables,
- la perception de, et l'accès aux éléments opaques à un stade donné.

C'est que le livre électronique ne constitue pas une simple substitution de support (comme la photocopie, ou les micro-formes), mais redéfinit les procédures d'accès au texte, et par là-même, suscite de nouvelles pratiques, certaines reconduisant dans ce contexte des pratiques «classiques» de lecture, d'autres découvrant des possibilités inédites directement inspirées par ce mode d'existence du livre. Ces pratiques doivent être au moins en partie anticipées afin d'intervenir dans la conception du livre électronique, et dans celle de l'environnement de lecture associé; et elles sont certainement très contrastées selon différents contextes de lecture. Sans naturellement préjuger des recherches et développements en cours, on peut néanmoins faire l'hypothèse que certaines fonctionnalités ne pourront pas ne pas se retrouver dans tous les cas de figure, parce qu'elles semblent en quelque sorte consubstantielles à la notion même de

lecture du livre électronique. En se référant aux quelques remarques précédentes, on peut par exemple citer le groupe de fonctionnalités que nous avons déjà proposé ailleurs de désigner par la formule «complexe MAPS»:

- Le Marquage ou balisage: l'association au texte de diverses marques dénotant explicitement les unités de sa structure logico-linguistique (phrases, paragraphes, chapitres; actes, scènes; définitions, théorèmes, démonstrations; etc.). Il s'agit ici de la structuration-auteur (souvent réinterprétée par l'éditeur ou le maquettiste), par rapport à cette sorte de structuration-lecteur en quoi consiste une part de l'annotation.
- L'Annotation qui permet justement à un lecteur particulier d'associer à des passages d'un texte, des caractérisations et/ou des commentaires propres. Le point important est ici que l'annotation électronique peut réaliser l'«idéal» que l'annotation traditionnelle ne satisfait que médiocrement, à savoir la (re)lecture du livre guidée par, et à travers, les annotations. Dans une acception ambitieuse, les annotations peuvent aussi être typées, composées, etc., et d'un autre côté communiquées ou échangées.
- La Prospection, par quoi on peut entendre l'ensemble des possibilités offertes pour réaliser des investigations fines du texte (lorsque celui-ci est en mode alpha-numérique), en termes lexicaux, syntaxiques, sémantiques, stylistiques, etc. La prospection gagne naturellement à s'appuyer sur des connaissances linguistiques variées et peut de surcroît exploiter diverses méthodes formelles (booléennes, statistiques, etc.)
- La Structuration, ou organisation, ou classification, de (segments ou unités de) textes, c'est-à-dire la composition d'entités textuelles issues de textes divers en de nouveaux ensembles (corpus), arguments d'opérations textuelles. Cette fonctionnalité prend en compte le fait que dans l'univers numérisé, toutes sortes d'opérations de groupement/dégroupement de textes sont relativement aisées à exprimer et à exploiter. Elle doit reposer sur un système de catalogage, et au-delà, de représentations de connaissances encyclopédiques.

Il n'est pas déraisonnable d'imaginer déjà ici des solutions techniques pour l'expression et la représentation de ces fonctionnalités (par exemple, il est possible que dans certains cas un balisage de type SGML ou TEI soit suffisant pour le marquage et l'annotation; d'autres cas plus complexes exigeront une représentation hypertextuelle, etc.) Mais un point important est à relever: ces quatre fonctionnalités entretiennent de très nombreuses interdépendances, aussi bien au niveau de la conception, qu'ensuite, à celui de leur mise en œuvre. En particulier, chacun des types d'unités qui leur sont propres (unités structurales du marquage, passages d'annotation, etc.) sont mutuellement inter-définissables. Ainsi par exemple, un passage de tel type (unité d'annotation) peut être déterminé par propriété de comprendre telles formes (repérables par prospection) dans le cadre d'un même paragraphe (unité de marquage), au sein d'un texte donné, ou d'un certain corpus (unité d'organisation). On peut alors faire l'hypothèse que ce qui peut contraster les divers usages évoqués ci-dessus (en termes de types d'ouvrages, de publics lecteurs, d'objectifs de lecture, de disciplines, etc.) concerne justement les exigences de contenu et les formes des combinaisons de ces quatre fonctionnalités (naturellement entres autres, et sans préjuger du fait que certaines configurations de lecture ne les exigent pas toutes).

■ III - Navigation, recherche, lecture

La station de lecture assistée par ordinateur peut ouvrir à d'autres fonctionnalités qui couvrent en fait, à l'avenir, l'accès à l'ensemble des fonctions génériques d'une bibliothèque dont le but est de classer, organiser et signaler les connaissances dont les documents sont porteurs. On peut appeler cette approche

méta-documentaire, dans la mesure où elle n'est plus seulement lecture d'un document, mais lecture d'outils de signalement et même exploration des classes d'appartenance disciplinaires de ces collections. Les termes de «recherche d'information» et de «navigation» sont alors appropriés. Le premier renvoie à une fonction de pré-sélection des documents, et permet de se constituer un corpus «personnel»; le second autorise l'approche raisonnée de grands nombres de documents hétérogènes, et de situer ce corpus «personnel» dans des ensembles de connaissances.

Il faut en effet ajouter que cette information méta-documentaire ouvre sur la possibilité (ou la nécessité) de «représenter» ces classes ou ces liens, donc de «traduire», dans le cadre de l'assistance informatisée, ces informations en «graphes» ou en schémas simulant, éventuellement spatialement, topographiquement, l'organisation des connaissances. Cette schématisation synthétique, sinon simplifiée, permet le repérage de trajets complexes d'exploration dans des fonds numériques importants, facilité qui se substitue donc à une mémorisation humaine incapable à elle seule d'organiser sans de tels repères le champs documentaire.

La philosophie, avec l'encyclopédisme des Lumières, s'est déjà posé ce problème de *relations* à optimiser entre organisation des connaissances d'une part et efficacité de leur représentation et de leur classement à des fins éditoriales d'autre part. Le «désordre» des connaissances comme chemin de lecture est en fait de rigueur. Chacun en fait l'expérience, la lecture savante est rarement linéaire. Il ne peut en être autrement vu la richesse des données et la multiplicité des intérêts humains. D'Alembert dans le *Discours préliminaire de l'Encyclopédie* relève ce nécessaire et irréductible «désordre»:

Le système général des sciences et des arts est une espèce de labyrinthe, de chemin tortueux, où l'esprit s'engage sans trop connaître la route qu'il doit tenir. Pressé par ses besoins et par ceux du corps auquel il est uni, il étudie d'abord les premiers objets qui se présentent à lui; pénètre le plus avant qu'il peut dans la connaissance de ces objets; rencontre bientôt les difficultés qui l'arrêtent, et soit par l'espérance ou même par le désespoir de les vaincre, se jette dans une nouvelle route; revient ensuite sur ses pas; franchit quelquefois les premières barrières pour en rencontrer de nouvelles; et passant rapidement d'un objet à un autre fait sur chacun de ces objets à différents intervalles et comme par secousses, une suite d'opérations dont la discontinuité est en effet nécessaire de la génération même de ses idées. Mais ce désordre, tout philosophique qu'il est de la part de l'esprit, défigurerait, ou plutôt anéantirait entièrement un arbre encyclopédique dans lequel on voudrait le représenter. (p. 111)

Une bibliothèque générale, et même jusqu'à un certain point une bibliothèque spécialisée, est à l'image de ce «désordre labyrinthe» qui a inspiré ensuite Borges: l'océan des connaissances est à la mesure de l'infinité des possibilités d'écriture et d'énonciation. La bibliothèque de Babel est la bibliothèque de toutes les bibliothèques, de toutes les variations et combinatoires linguistiques possibles. Ou, pour rester dans la métaphore marine: «L'univers n'est qu'un vaste Océan, sur la surface duquel nous apercevons quelques îles, plus ou moins grandes, dont la liaison avec le continent nous est cachée.» (D'Alembert, p. 114). Il ne faut donc pas oublier que la cartographie est un artefact, un substitut et une représentation abstraite, aussi parce que le réel est pour l'essentiel caché, non encore «lié», ce qui est particulièrement vrai dans l'univers hyperdocumentaire.

Le paradoxe du système informatique, dans l'assistance au lecteur-chercheur d'information, reste bien qu'il doit respecter cette indépassable richesse et hétérogénéité des données, et pour autant offrir une assistance par l'utilisation de l'ensemble des outils logiques de représentation et d'organisation. Le premier ordre reste bien «généalogique», au sens heuristique ou de construction successive de

connaissances. Le second, d'exposition, est encyclopédique.

A fortiori dans le domaine de fonds électroniques, la visibilité de ce parcours devient cruciale de par le nombre immense de documents consultés, appelés à l'écran. Cette possibilité n'a de limites que celles de la bibliothèque qui contient ces fonds, ou même l'ampleur du réseau où virtuellement on sait (ou espère) pouvoir trouver l'information. Par ailleurs, il a déjà été souligné combien l'appel à l'écran fait disparaître ou du moins nivelle les distinctions éditoriales et matérielles dont procèdent les ouvrages. La fonction de cartographie renvoie donc aux impératifs de repérage et de contextualisation, de modélisation, de simplification et de synthèse, de variation des points de vue ou des échelles, de guide d'exploration et de mnémotechnique.

Qu'est-ce que la navigation? Cette notion est largement héritée de la pratique en bibliothèque, réactualisée par l'apparition de l'informatique: un parcours de fichier bibliographique est déjà en soi une navigation. Les renvois de notice à notice, les rebonds de vedette à vedette, d'indice de classification à un autre, sont autant de parcours, combinables éventuellement, superposables et cumulables. Ce travail est rendu possible par la disposition physique des données, leur structure commune. Aujourd'hui la numérisation de documents et leur mise à disposition à travers des serveurs étend encore les acceptions de cette navigation. Cependant il reste à établir les conditions de cette navigation: l'univers des références qui décrivent les documents est structuré, les documents numérisés sont structurés. Quel lien peut-il y avoir entre ces deux volets de la mise à disposition documentaire? La notion de navigation va devoir s'élargir pour prendre en compte tous les niveaux auxquels elle peut s'effectuer; on naviguera aussi bien entre des notices descriptives, des extraits significatifs de documents - eux-mêmes structurés -, plusieurs documents, ou encore dans un même document dont les passages mis en rapport par l'intervention du lecteur dans son travail vont devenir autant de structures pertinentes.

Une des «couches documentaires» privilégiée est bien celle des données bibliographiques, contenues dans les bases informatisées (par exemple l'ensemble des bases sciences humaines du CNRS), dans la mesure ou la référence «auteur» ou «matière» peut se croiser avec la localisation de ces informations dans les documents eux-mêmes (bibliographies de fin de monographie ou de revues critiques, citations, table des matières, index, etc.) À l'avenir, ces «couches documentaires», intermédiaires entre le catalogue et le document, devront faire l'objet d'une approche normalisée (de la même façon que le format InterMarc fait autorité pour le catalogue) afin de permettre des procédures de navigation en réseau indépendantes de l'origine des sources.

Les procédures à envisager seront sûrement itératives, au sens où, par ces approches successives, par «couches», le but sera de «cerner» l'information recherchée non pas de façon prédictive, pré-établie par une indexation ou une classification *a priori*, mais bien par voisinage «intéressant» le domaine d'étude. Se repérer, localiser: savoir où l'on est, d'où l'on vient, où l'on va. Cela suppose de pouvoir prendre des repères: est-on en vue des côtes d'un continent disciplinaire? Doit-on au contraire chercher des azimuts dans un océan indifférencié de documents, mais alors quel est le système astronomique, la carte céleste qui permet cette référence? Force est d'avancer qu'une bibliothèque c'est d'abord l'ensemble des repères fournis au lecteur pour accéder au document. (Par exemple la BN a des catalogues «Auteurs», des catalogues par périodes historiques, mais pas des catalogues «Matières» systématiques). Un univers de simulation doit, dans l'espace électronique, se substituer à ces repères. Cet environnement doit être mi-sensible, analogique, mi-intelligible, numérique, en un sens comme l'est une carte qui est une modélisation simplifiée du réel, mais contient des outils de calcul (repérage orthonormé, rectification entre nord géographique et nord magnétique, etc.)

Ces outils («information vizualizer») sont actuellement mis en œuvre par un certain nombre de sociétés

américaines dont Rank Xerox Parc à titre expérimental. Citons à titre d'exemples: *The Butterfly application*, *The GAIA protocol*, *The Protofoil interface*, *Perspective wall*, *Cone tree*, *The Hyperbolic tree browser*, *The Scatter Gatter loop*, ou documents lents (3).

Les systèmes à l'œuvre dans le domaine bibliothéconomique, et qui sont déjà un mode de structuration, sont variés. Classifications décimales Dewey, CDU, qui découpent l'espace du savoir en diverses classes elles-mêmes partagées selon des points de vue chronologiques, géographiques... De même les langages d'indexation matière fournissent-ils une description par sujet dans un vocabulaire contrôlé et organisé: LCSH, Rameau... Ces listes de vedettes matières, ces thésaurus sont organisés et maintenus, en fonction de l'évolution respective des disciplines (voir l'importance croissante des sciences dans ces évolutions) et selon la connaissance plus approfondie de certains domaines. C'est cette organisation méta-documentaire (dans le sens où elle trace un cadre *a priori*) qui produit une cartographie permettant de «situer» tout document, mais aussi de cartographier un parcours de recherche, d'étayer une modélisation en profil de recherche.

Là encore, il est fondamental de reprendre la démarche de D'Alembert dans le *Discours préliminaire de l'Encyclopédie* (démarche qui elle-même théorisait les questions encyclopédiques systématiques et éditoriales abordées par Bacon et Chambers). Au contraire de l'ordre «généalogique» et hétérogène dont nous partions plus haut (et qui est aussi pour une bibliothèque la succession du Dépôt légal ou l'accumulation des acquisitions successives de documents):

L'ordre encyclopédique de nos connaissances [...] consiste à les rassembler dans le plus petit espace possible, et à placer, pour ainsi dire, le philosophe au dessus de ce vaste labyrinthe dans un point de vue élevé d'où il puisse à la fois apercevoir les sciences et les arts principaux; voir d'un coup d'œil les objets de ses spéculations, et les opérations qu'il peut faire sur ces objets; distinguer les branches générales des connaissances humaines, les points qui les séparent ou qui les unissent; et entrevoir même les routes secrètes qui les rapprochent.

Et D'Alembert de préciser ce qu'est cet ordre encyclopédique et représentatif, dans un sens tout à fait identique aux recherches mises en œuvre dans les outils cités plus haut tels que les «visualiseurs d'information»:

C'est une espèce de mappemonde qui doit montrer les principaux pays, leur position et leur dépendance mutuelle, le chemin en ligne droite qu'il y a de l'un à l'autre; chemin souvent coupé par mille obstacles, qui ne peuvent être connus dans chaque pays que des habitants ou des voyageurs, et qui ne sauraient être montrés que dans des cartes particulières fort détaillées. Ces cartes particulières seront les différents articles de l'*Encyclopédie*, et l'Arbre ou Système figuré en sera la Mappemonde.

On sait que ces «Cartes» ou «Arbres» du savoir, en plus de leur rôle mnémotechnique, étaient déjà proposés dans les «Prospectus» rédigés par Diderot pour promouvoir de façon synthétique l'ampleur de l'aventure éditoriale.

Cet ordre «encyclopédique» est arbitraire par convention, donc relatif à l'intentionnalité théorique et au «point de vue» qui y préside et non à l'ordre objectif des documents rassemblés:

Mais comme dans les cartes générales du globe que nous habitons, les objets sont plus ou

moins rapprochés, et présentent un coup d'œil différent selon le point de vue où l'œil est placé par le géographe qui construit la carte, de même la forme de l'arbre encyclopédique dépendra du point de vue où l'on se mettra pour envisager l'univers littéraire. On peut donc imaginer autant de systèmes différents de la connaissance humaine que de mappemondes de différentes projections; et chacun de ces systèmes pourra même avoir, à l'exclusion des autres, quelque avantage particulier. Il n'est guère de savants qui ne placent volontiers au centre de toutes les sciences celle dont ils s'occupent, à peu près comme les premiers hommes se plaçaient au centre du monde, persuadés que l'univers était fait pour eux.

Cette «intentionnalité» de lecture comme «point de vue» est une notion essentielle: il ne s'agit pas d'avoir une arborescence de classification «objective», mais bien de repérer *cette* lecture et non une autre, *cette* recherche et non une autre. Naviguer est à chaque fois une pratique différente en fonction des types de lecture engagée. Là encore la métaphore peut être filée: la navigation «aux instruments» pour un paquebot n'est pas la même que celle «à l'estime» pour un bateau de plaisance. D'autres auteurs ont déjà souligné ces aspects et ces références à la cartographie dynamique des connaissances (par exemple *Atlas* de Michel Serre, les cinécartes de Pierre Lévy, etc.)

Dans le domaine de l'édition électronique, les possibilités de conformation et d'adaptation au type de lecteur et à son intérêt, sont poussées très loin, au point de n'avoir plus aucun rapport avec l'édition papier. Exemple: le «*personal journal*» présente une information formatée en fonction de l'intérêt du lecteur (4). Cette personnalisation de l'information est la pierre d'achoppement de la navigation électronique. Comment au sein d'un ensemble diversement structuré d'informations documentaires, aller chercher la «bonne information» requise par un profil de lecture? Inversement: comment, à partir des requêtes de lecture capter et construire un profil dont la définition sera une variable de navigation dans les fonds?

D'Alembert voit bien aussi que l'ordre encyclopédique tient sa hiérarchie au nombre des liaisons, au sens hyperdocumentaire du terme comme dans Xanadu, que les connaissances ou documents entretiennent entre eux, mais en souligne en même temps l'arbitraire logique, - ce ne peut être qu'outil relatif et précaire devant la richesse insondable des données:

Quoi qu'il en soit, celui de tous les arbres encyclopédiques qui offrirait le plus grand nombre de liaisons et de rapports entre les sciences, mériterait sans doute d'être préféré. Mais peut-on se flatter de le saisir? [...] Mais souvent tel objet qui par une ou plusieurs de ses propriétés a été placé dans une classe, tient à une autre classe par d'autres propriétés, et aurait pu tout aussi bien y avoir sa place. Il reste donc nécessairement de l'arbitraire dans la division générale.

C'est d'ailleurs, pour D'Alembert, nécessairement un compromis qu'il faut trouver entre système encyclopédique et «ordre généalogique», afin d'avoir un produit éditorial à peu près satisfaisant: choisir une division qui nous a paru satisfaire tout à la fois le plus qu'il est possible à l'ordre encyclopédique de nos connaissances et à leur «ordre généalogique»; de la même façon que, pour être efficace et synthétique, le lecteur sur une station informatisée devra avoir un compromis acceptable entre représentation classificatoire des fonds documentaires et ordre généalogique de sa recherche ou bibliothèque personnelles. Ajoutons enfin que D'Alembert justifie longuement le rapport entre rangement alphabétique des articles dans l'édition, et au contraire le système des «renvois» (accompagnant chaque article) qui donne toute sa force implicite à l'ordre encyclopédique, on dirait maintenant à l'hypermédia.

Nous avons cité longuement D'Alembert car, à notre connaissance, avec Leibniz, c'est l'auteur qui a

poussé le plus loin la formalisation de ce problème complexe qu'est l'organisation des connaissances, comme compromis acceptable entre la richesse infinie des connaissances et leur accumulation d'une part, et la simplicité et l'efficacité d'outils de manipulation d'autre part. Mais ce sont bien là des fonctionnalités que tente de maîtriser l'informatique documentaire actuelle. Entre l'efficacité des représentations synthétiques et l'appauvrissement réducteur de la formalisation, l'informatique documentaire actuelle navigue entre les mêmes écueils.

Les notions de cartographie ou d'arbre que l'on retrouve dans les outils visuels informatisés sont donc des métaphores qui doivent résulter d'un côté en des indications bibliographiques que la bibliothèque virtuelle est capable de communiquer, de l'autre en des domaines de recherche singuliers liés au parcours du lecteur/chercheur (5). Au fond, la numérisation des documents repose tout à la fois la question de structuration physique de l'écrit et celle de la structuration logique des connaissances. Le «format» codex se trouve bouleversé de ce double point de vue; il en est donc de même pour l'espace qui le rassemble, la bibliothèque, au profit d'une entité nouvelle, virtuelle, qui reste à penser.

■ IV - Les Bibliothèques numérisées

De même que le livre électronique n'est pas pensable (ou du moins simplement pas examinable) sans station de lecture, cette dernière ne peut exister pleinement qu'en relation avec sa source naturelle, la bibliothèque numérisée. Et de même qu'une bibliothèque traditionnelle n'est pas qu'un simple rassemblement d'ouvrages, une bibliothèque numérisée n'est pas qu'un simple gisement de fichiers. *A fortiori* si de telles bibliothèques peuvent être mises en réseau. Ici, les questions se présentent donc comme beaucoup plus ouvertes.

Dans le contexte classique, les fonds d'ouvrages d'une bibliothèque sont accessibles via des catalogues, qui constituent un outil spécifique d'accès, mais aussi un lieu de convergence d'éléments très divers:

- c'est une plate-forme de rencontre entre la visée «bibliothécaire» et «lectorale» du document (pour ne prendre que les acteurs les plus immédiatement concernés, bibliothécaire et lecteur);
- la constitution des fonds (ou de sous-ensembles de tels fonds) repose en principe sur une certaine logique (ou une «économie») qui tend à associer à un ouvrage, à côté de son sens «intrinsèque» (?), en principe indépendant de sa localisation, une «valeur» supplémentaire (métaphoriquement, au sens saussurien) qui elle dépend de cette localisation, *i.e.* de l'appartenance de l'ouvrage en question à un fonds donné, et donc aussi de l'environnement de sa référence, qui autorise mais aussi limite, moyennant tel ou tel mode de navigation dans ce catalogue, telles ou telles possibilités d'extension de recherche de références par voisinages thématique, chronologique, auctorial, etc.;
- de la même manière, toute la partie indexicale des catalogues (mots-clés, vedettes, etc.) résulte d'une composition complexe de renvois à des organisations conceptuelles génériques ou générales des savoirs, et de la spécificité ou spécialisation des fonds considérés; et elle ne peut que varier d'un fonds à un autre. Une part importante du savoir-faire d'un bibliothécaire, dans son rôle d'accueil du public, résulte d'une bonne appréhension de ces données, rapportée aux requêtes des utilisateurs, lesquelles peuvent évoluer dynamiquement.

La numérisation tant des fonds d'ouvrages que des catalogues renouvelle tous ces aspects dans des proportions dont on n'a pas encore pris la mesure de l'amplitude et de la profondeur. Pour l'essentiel:

a) Un nouvel usage des catalogues numérisés, qui deviennent «cherchables» de différentes manières, sans

qu'il soit clairement acquis que leurs structures et leurs contenus traditionnels soient les plus pertinents ou les plus efficaces.

b) La possibilité d'inter-connexion des catalogues (et par le fait, des fonds) par union, renvois ou autres relations, qui peut engendrer des pertes d'orientation, et multiplier les facteurs de relativités référentielles signalés ci-dessus.

c) Une nouvelle relation fonds-catalogues, par le fait que les rubriques des catalogues peuvent être étendues autant que de besoin. Par exemple, on peut envisager d'enrichir un article de catalogue relatif à un ouvrage donné par la recopie, dans des rubriques dédiées, de fragments significatifs de l'ouvrage, tels que tables de matières et sommaires, résumés (auctoriaux ou non), index, quatrièmes de couverture, etc. Ces extraits significatifs ont d'abord pour vocation de donner une image plus précise de l'ouvrage dans la perspective de l'évaluation de sa pertinence pour le lecteur. Mais ils peuvent eux-mêmes être recherchables, et constituer une sorte de base de données bibliographiques, éditoriales et encyclopédiques particulière. On rencontre alors ici aussi des problèmes de cohérence, cohésion, consistance, complétude, de telles bases de connaissances. Il n'apparaît pas que cette nouvelle forme de consultation, intermédiaire entre celle des contenus des catalogues, et celle des ouvrages proprement dits soit simple à définir, mais qu'elle constitue au contraire un enjeu peut-être majeur du versant culturel des «autoroutes de l'information».

d) La possibilité d'un autre enrichissement, l'expression (plus aisée que dans les catalogues de facture classique) de diverses relations inter-textuelles (par ex.: est /comporte une traduction de X, un commentaire de X, une réplique à X, etc.), certainement très intéressante pour certaines recherches, pose néanmoins le problème de la formalisation de ce type de relations (ex.: quel est le statut d'une traduction de traduction? d'un commentaire de résumé?)

e) Une nouvelle relation bibliothèque-lecteur, sur deux plans:

- la nécessité d'envisager la numérisation d'ouvrages comme quelque chose de différent, et de plus qu'un simple transfert de support mais plutôt comme une sorte de préparation de document, où l'on doit anticiper les usages ultérieurs pour intégrer dans le document délivré les dispositifs autorisant ces usages. La question se pose tout particulièrement à propos des marquages de toutes sortes de structures textuelles et linguistiques, bien au-delà des habituelles questions d'import-export de documents, ainsi qu'à propos de l'application d'outils de lecture assistée par ordinateur;
- une autre anticipation nécessaire est celle du besoin pour le lecteur de constituer des bibliothèques personnelles (qu'il s'agisse d'un espace de travail personnel alloué au sein du fonds, comme par exemple dans le modèle Xanadu, ou d'une recopie en local de sous-ensembles de fonds ainsi détachés de leurs sites comme dans le modèle élaboré par la BNF. A nouveau ici, le mode de préparation doit s'appuyer sur des connaissances des principes d'organisation personnalisée chez les utilisateurs pour que, au minimum, il n'y ait pas d'incompatibilité de principe (sauf au prix d'un retraitement d'ensemble) avec ceux qui sont, parfois implicitement, introduits lors de la structuration des fonds. Par exemple ces principes seront très certainement différents selon que l'on prête à l'utilisateur le besoin d'utiliser un simple «bureau» à la manière des ordinateurs Mcintosh, une organisation hypertextuelle, ou une véritable base de données textuelles structurée en mode SGML. Sur ces questions se pose directement le rapport entre les mémoires externes et les processus de mémorisation interne.

f) Dans le contexte traditionnel, la relation œuvre-ouvrage (physique) est très variée eu égard aux aspects de «conditionnement» constitué par le livre en tant qu'objet physiquement autonome: une œuvre sur plusieurs volumes, en nombre entier ou non; plusieurs œuvres dans un seul volume, comme dans le cas des anthologies, etc. Le mode numérisé permet d'envisager une adéquation plus étroite entre une unité de livraison, et une œuvre ou même une partie d'œuvre quand elle correspond aux besoins du lecteur (un

seul chapitre de tel roman, un seul acte de telle pièce, etc.) Mais il se pose alors des problèmes de limite, de granularité des objets délivrables, en relation avec leur identification cataloguale, et leur adressage (qui ne peut pas, par exemple, reposer en général sur des tables des matières de facture classique). Plus largement encore, la numérisation, et les possibilités de groupement et dégroupement qu'elle autorise techniquement pose des questions non triviales de référence entre par exemple «œuvre» et «texte», «original», «exemplaire» et «substitut», toutes questions centrales dans une théorie formelle de la référence symbolique.

Le livre électronique, les stations spécialisées en autorisant la lecture, dans un sens en partie renouvelé du mot, les fonds numérisés d'où procèdent ces livres enfin, et les modalités techniques et intellectuelles de leur pénétration et, en dernier ressort, de leur intelligence: nous avons tenté sommairement d'en indiquer quelques interdépendances technico-logico-culturelles, au moment où le numérique commence à déferler sur la civilisation du livre.

■ Notes

1 - Ce texte est une version révisée de l'article intitulé «Le Livre électronique et le concept de "station de lecture assistée par ordinateur"» paru dans *Littérature, informatique, lecture*, textes réunis par Alain Vuillemin et Michel Lenoble, Limoges, Presses universitaires de Limoges, 1999. Il est publié ici avec l'aimable autorisation d'Alain Vuillemin et de Michel Lenoble, des Pulim de l'Université de Limoges (France) et avec le concours du Certel de l'Université d'Artois.

2 - Les aperçus présentés ici prennent leur source, entre autres, dans une collaboration entretenue avec Th. Cloarec et J. Pendjé (Bibliothèque nationale de France), G. Auffret (Advanced Information Systems), et O. Mazhoud et E. Pascual (IRIT), dans le cadre du projet européen MEMORIA (CEE. Action MLAP. DG XIII).

3 - Voir: *Communications of the ACM*, avril 1995, vol. 38, no 4 : «Rich interaction in the digital library» (article de l'équipe Rank Xerox de Palo Alto), pp. 29-39.

4 - Cf. *Wall Street Journal*, article du *Monde* du 24 avril 1995 et cahier «Multimédia» de *Libération* du 28 avril 1995 «Quotidiens branchés».

5 - Pour d'autres références «philosophiques», voir Y. Maignien «La Constitution de la collection numérisée de la Bibliothèque nationale de France: vers un nouvel encyclopédisme?» in *Literary and Linguistic Computing* et «La Bibliothèque virtuelle ou de l'Ars memoria à Xanadu», *Bulletin bibliographique de France*.

■ Références bibliographiques

«Rich interaction in the digital library» (article de l'équipe Rank Xerox de Palo Alto), *Communications of the ACM*, vol. 38, no 4, avril 1995, p. 29-39.

Chahuneau F., Lécluse Ch., Stiegler B., Virbel J., «Prototyping the Ultimate Tool for Scholarly Qualitative Research on Texts», in *8e conférence annuelle du «New Oxford English Dictionary»*, Waterloo, 18-20 octobre 1992.

Charles M., *Rhétorique de la lecture*, Paris, Seuil, 1977.

Chouchan D., «Du livre à l'ordinateur», in *La Recherche*, Paris, n° 8, 1990, p. 96-98.

D'Alembert et Diderot, *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers, I. Discours préliminaire* de D'Alembert, Paris, Garnier Flammarion, 1986.

Delany P., Landow G. (éd.), *Hypermedia and Literary Studies*, Cambridge, MIT Press, 1990.

Genette G., *L'Oeuvre de l'art*, Paris, Seuil, 1994.

Goodman N., *Languages of Art, an Approach to the Theory of Symbols*, Hockett, 1968.

Gross M., «Les Banques de données du LADL. Analyse automatique du français», in *Colloque INL'91*, Nantes, 23-24 janvier 1991, p. 361-386.

Iser W., *L'Acte de lecture*, Mardaga, 1976.

Jauss H., *Pour une esthétique de la réception*, Paris, Gallimard, 1978.

Jones W., «"As We May Think"? Psychological Considerations in the Design of a Personal Filing System», in Guindon R. (éd.), *Cognitive Science and its Applications for Human Computer Interaction*, Lawrence Erlbaum Associates Publishers, 1988, p. 235-288.

Lebrave J.-L., «Hypertextes-mémoires-écriture», in *Genesis*, Paris, no 5, 1994, p. 9-24.

Maignien Y., «La Constitution de la collection numérisée de la Bibliothèque nationale de France: vers un nouvel encyclopédisme?», in *Literary and Linguistic Computing*, vol. 10, no 1, 1995, Oxford University Press.

Maignien Y., «La Bibliothèque virtuelle, ou de l'Ars memoria à Xanadu», in *Bulletin bibliographique de France*, 40, 2, 1995, p. 8-17.

Nielsen J., «Online Documentation and Reader Annotation», in *International Conference on Work with Display Units*, Stockholm, 12-15 mai 1986.

Pascual E., «Système de description en langage naturel de l'architecture de documents», *Convention IA 91*, Paris, 14-17 janvier 1991.

Rossi P., *Clavis Universalls*, Editions Jérôme Millon, 1993.

Searle, J. A., «Taxonomy of Illocutory Speech Acts», in K. Gunderson (éd.), *Language, Mind and Knowledge*, University of Minesota Press, 1975, p. 334-336.

Sperberg-McQueen G.M., Lou Burnard (éd.), *Guidelines for the Encoding and Interchange of Machine-Readable Texts*, ACH-ACL-ALLC, 1994.

Spiro R., Bruce B., Brewer W., *Theoretical Issues in Reading Comprehension*, Lawrence Erlbaum Associates Publishers, 1980.

Stiegler B., «Machines à écrire et matières à penser», in *Genesis*, Paris, no 5, 1994, p. 25-50.

Virbel J., «Annotation dynamique et lecture expérimentale: vers une nouvelle glose?», in *Littérature (Informatique et Littérature)*, Paris, Larousse, no 96, 1994, p. 91-105.

Virbel J., «Reading and Managing Texts on the "Bibliothèque nationale de France" Station», in *Colloque national sur l'écrit et le document* (CNED), Nancy, 6-7 juillet 1992. Actes in: *Bigre*, no 80, p. 192-199.

Virbel J., «La Lecture assistée par ordinateur et la station de lecture de la Bibliothèque nationale de France», in Vuillemin A. (éd.), *Les Banques de données littéraires, comparatistes et francophones*, Limoges, PULim, 1993, p. 19-28.

Yates F., *L'Art de la mémoire*, Paris, Gallimard, 1966.

2003 (1999)

Voir dans l'encyclopédie de *l'Astrolabe*:

[L'Avenir de la lecture interactive](#)

[Critique de la lecture informatique](#)

[L'Edition électronique](#)