

École Nationale Supérieure
des Sciences de l'Information
et des Bibliothèques

Diplôme de conservateur de bibliothèque

MÉMOIRE D'ÉTUDE



Structuration des collections et documents numérisés et Postes de Lecture Assistée par Ordinateur à la Bibliothèque nationale de France. L'exemple du *Manuel du Libraire et de l'amateur de livres* de Charles Brunet dans le contexte de la numérisation des éditions des imprimeurs lyonnais du XVI^{ème} siècle.

Éric Varon

Sous la direction de Monsieur Richard Bouché
Professeur à l'ENSSIB

BIBLIOTHEQUE DE L'ENSSIB



8109225

1996

1996

DCB

H6

Déposé le 06/01/96

**École Nationale Supérieure
des Sciences de l'Information
et des Bibliothèques**

Diplôme de conservateur de bibliothèque

Sous la direction de Monsieur Richard Bouché
Professeur à l'ENSSIB

Structuration des collections et documents numérisés et Postes de Lecture Assistée par Ordinateur à la Bibliothèque nationale de France. L'exemple du *Manuel du Libraire et de l'amateur de livres* de Charles Brunet dans le contexte de la numérisation des éditions des imprimeurs lyonnais du XVI^{ème} siècle.

Éric Varon



Stage effectué sous la direction de Monsieur Yannick Maignien, Chef de la Mission de Numérisation Scientifique, Direction du Développement Scientifique de la Recherche et des Réseaux, Bibliothèque nationale de France.

1996
DCB
46

1996

Abstract :

The constitution of an interdisciplinary collection of 100 000 digitalised works in bitmap or alphanumeric mode, and the reading station (CARE : Computer Aided Reading Environment) is a part of a more comprehensive Bibliothèque nationale de France program, that includes a vast collection of digitized books and linking it to servers providing textuals or bibliographical informations. The Lyon's Public Library provide a rich collection of books printing during the sixteenth centuy. The structured representation of texts in computer and standards for documents representation, such as SGML, are considered, in this context.

Résumé :

La constitution d'une collection interdisciplinaire en mode alphanumérique ou image et la station de lecture (Poste de Lecture Assistée par Ordinateur), sont les éléments d'un important programme de la Bibliothèque nationale de France, qui inclut une collection reliée à un serveur fournissant des textes ou des informations bibliographiques. La Bibliothèque municipale de Lyon fournit une riche collection d'ouvrages imprimés du XVI ème siècle. La représentation structurée des documents et le format SGML, sont utilisés dans ce contexte.

Mots clés :

France **Bibliothèque nationale
Poste de lecture assistée par ordinateur
Numérisation
Hypermédia

Title

Structuration of the collections and digitized works and the Computer Aided Reading Environment (CARE) at the Bibliothèque nationale de France. The exemple of the *Manuel du Libraire et de l'amateur de livres* by Charles Brunet in the context of the digitization of the Lyon's printers publising during the sixteenth century.

Je tiens tout d'abord à remercier Madame Marcelle Beaudiquez, responsable de la Direction du Développement Scientifique de la Recherche et des Réseaux de m'avoir accueilli comme stagiaire, Monsieur Yannick Maignien, Chef de la Mission de Numérisation Scientifique d'avoir accepté la direction de ce stage, ainsi que Monsieur Jean-Didier Wagneur et l'ensemble du personnel du Service de Numérisation. Mes remerciements s'adressent également à Monsieur Georges Perrin Chef du Service des Pôles associés, et à Monsieur Gérard Cathaly du service de Numérisation scientifique.

Table des matières :

Avant-propos.....	I-IV
Introduction.....	1
1. Un livre outil, interface textuelle, dans la perspective du PLAO ?.....	6
1.1. Une approche matérielle de la numérisation d'un corpus : l'exemple du Manuel de Brunet.....	12
1.2. La définition de type de document : principes généraux.....	12
1.3. La numérisation de l'Édition Lyonnaise du XVI ème siècle, le Manuel de Brunet et les stratégies de recherche au sein de la collection numérisée.....	14
1.4. Du texte numérisé aux liens entre notices.....	18
1.5. La collection numérisée.....	25
1.6. Favoriser la lecture/écriture assistée par ordinateur.....	26
-Conséquences des usages des fonds numérisés sur le choix des textes..	27
-Une bibliothèque cohérente.....	28
II) 2. Poste de Lecture Assistée par Ordinateur : définition.....	32
2.1. Première réalisation.....	33
2.2. Fonctions des Postes de Lecture Assistée par Ordinateur : les objectifs.....	35
2.3. Les fonctions centrales : la lecture/écriture assistée par ordinateur	37
2.4. L'annotation dynamique.....	38
2.4.1. La maquette d' AIS Berger-Levrault.....	43
a) La notion d'ULD, Unité Logique de Découpage.....	43
2.4.2. Critique de cette implémentation.....	50
3. Préfiguration des 3000 documents, liaison Poste de Lecture Assistée par Ordinateur et réseaux internationaux.....	52
3.1 Vers un nouvel encyclopédisme ?.....	59

Une nouvelle bibliographie dynamique : vers un réseau sémantique du document ?.....	60
Conclusion.....	65

Annexe 1. Numérisation des Éditions Lyonnaises du XVI ème siècle :

- Relevé des conclusions de la réunion du 10 juillet 1996
- Relevé des conclusions de la réunion du 20 septembre 1996
- Comité de pilotage "Numérisation", 19 septembre 1996

Annexe 2. Aspects techniques et financiers.....I-VI

Bibliographie.....I-IX

Avant-propos

Sens et objet de la recherche :

Notre recherche porte sur la constitution des collections numérisées, et le Poste de Lecture Assistée par Ordinateur (PLAO). La création d'une collection de cent mille documents à la Bibliothèque nationale de France, la mise en réseau d'une première collection de 3000 documents, la numérisation des fonds des Imprimeurs lyonnais du XVI^{ème} siècle posent des questions portant sur la structuration interne du document (balisage) et sur la structuration des corpus.

Ces formes de réédition des textes ouvrent de nouveaux chantiers pour le bibliographe : ceux de la structuration interne du document électronique, de la sélection et de la visibilité des éléments constituant le corpus, de la critique des sources avant la numérisation, et des études comparatives d'éditions. Pour ces dernières questions la relation base de données bibliographique/hypertexte avec les logiciels de lecture/écriture est essentielle, leur lieu de rencontre étant le PLAO.

La sélection de titres de la Librairie Lyonnaise du XVI^{ème} siècle provenant de la Bibliothèque municipale de Lyon, du fonds numérisé (D.D.S.R.) et de la réserve de la B.n.F (Direction de l'Imprimé et de l'Audiovisuel), nous paraît constituer une opération exemplaire, représentative de ces nouvelles questions. Parce qu'il s'agit d'une opération reposant sur la coopération de différentes Directions de la B.n.F. avec la Bibliothèque municipale de Lyon, la Direction du Livre et de la Lecture, des Universitaires spécialistes du XVI^{ème} siècle et parce qu'elle introduit la notion de D.T.D., Définition de Type de Document, norme précisant les conditions de description et d'utilisation des textes

numérisés en code ASCII.

Un ouvrage de référence bibliographique comme le *Manuel* de Brunet peut être considéré à la fois comme une base de données et une interface textuelle conduisant à d'autres textes d'un corpus numérisé. Il s'agit donc d'un *livre outil*, texte qui une fois numérisé donne accès à d'autres livres, rendant possible l'émergence d'une *bibliographie dynamique*, reposant sur des *réseaux sémantiques de documents*.

D.F. McKenzie (1) introduit ainsi le concept de bibliographie dans l'optique anglo-saxonne en citant Sir Walter Greg :

"L'objet des recherches du bibliographe est une série de papiers ou de parchemins couverts de certains signes écrits ou imprimés. Il ne s'intéresse à ces signes que comme autant de traces arbitraires. Leur sens n'est pas son affaire" (1) (...) Une étude récente de Ross Atkinson (2) vient étayer cette thèse en s'appuyant sur des travaux du sémioticien américain, C.S. Peirce. On peut par exemple affirmer que les signes qui constituent un tel livre, tels que le bibliographe doit les lire n'ont de valeur qu' iconique ou indicielle.

En bref, les signes iconiques sont ceux qui font intervenir un rapport de similitude ; ils représentent un objet, à peu près de la même façon qu'un objet représente son modèle.

L'étude comparative des éditions pratiquée par L'ITEM-CNRS se situe dans cette nouvelle bibliographie induite par la numérisation des textes et se sépare de la Gestion Électronique de Documents. La notion d'édition critique et comparative prend un sens différent quand il s'agit du texte contemporain. Une grande bibliographie classique comme le *Manuel* de Brunet nous permet d'approcher l'univers de

(1) D.F. McKenzie, *La bibliographie et la sociologie des textes*, Préface de Roger Chartier, Paris : Éditions du Cercle de la Librairie, 1991.

(2) Ross Atkinson, *An Application of Semiotics to the Definition of Bibliography*, *Studies in bibliography*, 33, 1980, pp. 54-73.

l'érudition. On peut suivre des ouvrages à travers des ventes célèbres, évaluer leurs cotations, retrouver les usages anciens qui faisaient citer un auteur par les formes latines déclinées de son nom (Estienne, Stephanus, ou bien Cicéron, Cicero, etc.). Cet univers de l'érudition est une aide précieuse pour les chercheurs qui doivent faire abstraction de leurs conceptions actuelles pour retrouver des manières de penser et de dire, des sensibilités propres au XVI^{ème} siècle, inscrites dans la matérialité des textes et dans les formes d'expression et de formulation des idées, c'est également une source d'interrogation en regard de l'encodage et de la normalisation d'un texte présentant des formes irrégulières (typographies avec des polices de caractères diverses, présentation évoluant au fil des années, la publication s'étant étendue entre 1810 et 1860).

Les grands répertoires bibliographiques (Brunet), peuvent être croisés avec les documents contenus dans la base Frantext.

Le texte numérisé n'est plus simplement un livre digitalisé, mais un outil de lecture tout à fait nouveau permettant un nouvel encyclopédisme. Si l'information arrive rapidement depuis de nombreuses sources, il importe de choisir parmi la masse informationnelle la part, qui intéressera un utilisateur précis à un moment précis. La recherche au sein d'une masse presque illimitée de données demande une fonction d'assistance intelligente. La bibliographie peut-elle avec d'autres outils, apporter sa contribution à la constitution de ce nouvel encyclopédisme ?

Les idées qui sous-tendent la notion de nouvelle bibliographie sont la complémentarité, l'édition critique, les bibliographies iconiques, descriptives et analytiques. La question des acquisitions ne saurait être

posée en terme de documents singuliers, mais renvoie à une problématique plus structurelle, l'idée de collection numérisée, ensemble vivant dont la complexité dépasse la simple addition des documents qui la compose. Rassembler toutes les connaissances, les mettre en ordre, les rendre accessibles et les diffuser, telles sont les composantes de toute entreprise encyclopédique qui, du même coup, favorise la circulation entre les différentes disciplines et décloisonne le savoir. Comme le note Edgar Morin : "Les savoirs qui liés permettraient la connaissance de la connaissance sont disjoints et morcelés". (1)

Considérant les répertoires bibliographiques numérisés en mode texte comme des bases de données, il serait nécessaire d'appliquer les méthodes relationnelles. Toutefois nous nous situons dans un univers de connaissance érudite, c'est à dire non normalisée : c'est pourquoi il est utile de mettre en avant les méthodes comparatistes de la philologie : mise en perspective d'édition, repérage des gloses, etc. Tout ce travail s'appuie sur l'hypertexte et non plus exclusivement sur la recherche exacte par extraction de requête : c'est au cours du parcours de lecture et de comparaison que l'étincelle du sens jaillit. Nous tenterons de souligner comment une interface de navigation portant sur une collection numérisée en mode image et en mode texte doit utiliser ces deux méthodes d'investigation complémentaires et croisées, pouvant être complétées par des outils de recherche des occurrences d'un texte. Le lieu privilégié de ces croisements est le Poste de Lecture Assistée par Ordinateur relayé par Internet (réseau Renater).

(1) Edgar Morin , La Méthode 4. Les idées, Paris : Éditions du Seuil, 1991.

Introduction

Le lecteur qui vient consulter les fonds d'une bibliothèque est en quête d'une mémoire, totale, universelle, portant sur tous les champs de la connaissance.

À la Bibliothèque nationale de France, lieu de relations privilégiées avec les savoirs et de travail dans un environnement virtuel, en réseau, le texte numérisé et les Postes de Lecture Assistée par Ordinateur (PLAO) ont été conçus, comme les moyens d'assurer une communication maximale des collections sous forme de documents électroniques, tout en permettant une consultation à distance:

“Du point de vue de la communication, deux grandes nouveautés doivent marquer l'ouverture de la Grande Bibliothèque : la création sur place de postes de travail équipés d'instruments d'aide à la recherche, à la lecture et à la rédaction, et surtout la possibilité de consulter à distance des catalogues et des documents numérisés”(1).

Une bibliothèque numérique exige-t-elle une structuration particulière ? La structuration des fonds est-elle conçue en relation avec la structuration du document numérique ? Ces deux questions étant posées dans la perspective d'un programme dont la finalité est d'offrir, à travers le Poste de

(1) Patrice Cahart, Michel Melot *Propositions pour une grande bibliothèque : rapport au Premier ministre*, Paris : La Documentation française, 1989. Collection des rapports officiels..

Lecture Assistée par Ordinateur (PLAO) (1), une plus large accessibilité aux documents en favorisant la simultanéité des consultations et l'interrogation à distance.

Une bibliothèque numérique exige-t-elle une structuration particulière ?

Au cours de notre stage à la Bibliothèque nationale de France effectué au Service de la Numérisation scientifique sous la direction de Monsieur Yannick Maignien (2), stage complété par le colloque *Hypertextes littéraires et éditions savantes* (3), et la participation aux travaux portant sur la numérisation du *Manuel du libraire et de l'amateur de livres* de Jacques Charles Brunet (1780-1867) deux principes se sont imposés:

-la question des acquisitions ne saurait être posée en terme de documents singuliers, mais renvoie à une problématique plus structurelle : l'idée de collection numérisée, ensemble vivant dont la complexité dépasse la simple addition des documents qui la compose.

-la nécessité d'articuler de l'information non formalisée, souvent du texte, avec des configurations structurales exprimant des figures de contenu, c'est-à-dire des schèmes formels de représentation des connaissances. La constitution du fonds numérisé ne peut être dissociée des finalités qui lui sont assignées et qui conditionnent les choix qui sont faits. En effet, la démultiplication des accès ne serait rien si elle ne

(1) Poste de Lecture Assistée par ordinateur conçu dès 1989 par Jean Gattegno, Alain Giffard, Bernard Stiegler

(2) Monsieur Yannick Maignien, DDSR, Chef de la Mission scientifique de numérisation

(3) Programme Patrimoine écrit (CNRS-ITEM) : Colloque hypertextes littéraires et éditions savantes 12-14 septembre 1996, ENS rue d'Ulm, coordination Monsieur Jean-Louis Lebrave.

reposait sur l'ergonomie d'un système susceptible de répondre à une recherche d'efficacité, obtenue par la lecture assistée par ordinateur.

La structuration des fonds est-elle conçue en relation avec la structuration du document numérique ?

La réalisation concrète de l'hypertexte exige une structuration du document. La norme SGML permet de décrire la structure des documents, qui est ensuite interprétée par un logiciel.

Pour illustrer l'articulation de notre problématique nous nous appuierons sur l'exemple concret de la numérisation en cours du *Manuel du libraire et de l'amateur de livres* de Brunet (1).

Dans le cadre de la Direction du Développement Scientifique et des Réseaux, le Secteur des Pôles associés (2) et la Mission Scientifique de Numérisation mènent une politique de numérisation des collections lyonnaises du XVI^{ème} siècle, en étroite collaboration avec la Bibliothèque municipale de Lyon (3).

Messieurs Maignien et Wagneur (4) ont souhaité qu'une étude soit menée pour examiner les possibilités de structuration des notices de l'ouvrage de Brunet en format SGML dans l'hypothèse d'une numérisation en mode texte. En effet une structuration des divers éléments de la notice permettrait aux futurs utilisateurs un accès plus souple aux

(1) Voir en annexe les comptes rendus des réunions des 10/7/96 et 20/9/96.

(2) Monsieur Georges Perrin, DDSR, Chef du Secteur des Pôles associés

(3) Voir en annexe les comptes rendus des réunions des 10/7/96 et 20/9/96.

(4) Monsieur Jean-Didier Wagneur, DDSR, adjoint de Monsieur Maignien à la Mission Scientifique de numérisation

éléments d'information contenus dans chaque étude bibliographique, tout en favorisant des processus d'indexation informatique, plus élaborés et plus fins par rapport à une numérisation en mode image, n'autorisant pas le balisage du texte par des marqueurs (tag). Le projet *Text Encoding Initiative* (1), permet de rattacher cette structuration à une réflexion collective sur le document électronique.

L'étude pour la définition d'une D.T.D. (Définition de type de document), dans le cadre de l'opération de numérisation de l'ouvrage bibliographique de Jacques Charles Brunet pose dans la pratique, la problématique des connaissances sur les connaissances : structurer un tel répertoire en balise SGML ayant du sens pour l'utilisateur, ne peut se faire que par un travail en commun avec les spécialistes du XVI^{ème} siècle qui définiront leurs attentes.

L'utilisation souhaitée est la recherche bibliographique rétrospective par auteurs et titres des ouvrages anonymes. Il est donc nécessaire de définir une structure qui permette d'isoler les vedettes et les éléments les plus significatifs. Il faut alors appliquer la notion de représentation du contenu du document d'une part, l'accès à l'information de l'autre. Cet accès repose sur des *noeuds* et des *liens*. Les possibilités de liage sont multiples. Il s'agit du lien qui existe entre un mot et une remarque, du lien entre un texte et un commentaire de ce texte ou son résumé, entre un texte et une référence bibliographique, une remarque historique, un relevé de

(1) Cahiers Gutenberg n° 24 (spécial TEI) - juin 1996.

catalogue des ventes. Tous ces textes constituent avec leurs liens une base de données. Comme l'a noté Monsieur Jean-Pierre Balpe (1):

Toute syntaxe se déploie sur deux axes : l'axe des commutations ou axe paradigmatique qui définit des classes d'équivalences, l'axe de permutations ou axe syntagmatique, qui détermine des relations d'ordre.

Chaque syntaxe particulière n'est rien d'autre qu'un ensemble spécifique de définition des classes et des relations d'ordre.

Dans les hyperdocuments, l'axe paradigmatique concerne la constitution de groupements d'unités d'informations.

L'axe syntagmatique concerne les relations gouvernées par les valences.

Il est tout à fait possible, en examinant quelques cas, d'esquisser une "hyper-grammaire" (Englebart, 1984 ; Codd 1985), d'essayer de tracer quelques-unes des lois générales permettant de maîtriser de façon relativement abstraite les modes de circulation dans les hyperdocuments.

Un texte comme le *Manuel* de Brunet peut être considéré à la fois comme une base de données et une interface textuelle conduisant à d'autres textes d'un corpus numérisé. Il s'agit donc d'un *livre outil*, texte qui une fois numérisé donne accès à d'autres livres, rendant possible l'émergence d'une *bibliographie dynamique*, reposant sur des *réseaux sémantiques de documents*. L'instruction d'une recherche

(1) Jean-Pierre Balpe, *Hyperdocuments hypertextes et hypermedias*, Paris : Eyrolles, 1989

suppose une cartographie des ouvrages sélectionnés au regard de nombreux critères intellectuels croisés : appartenance à un ensemble thématique, chronologique, contexte de production, place dans l'histoire du livre ancien.

1. Un livre outil, interface textuelle, dans la perspective du PLAO ?

Jacques-Charles Brunet, *Manuel du libraire et de l'amateur de livres* de Jacques Charles Brunet (1780-1867), est l'oeuvre de toute une vie comparable à l'oeuvre de Quérard (1). Brunet publie la première édition en 1810 et travaille sur cette publication pendant cinquante ans, jusqu'à la 5^{ème} édition publiée en 6 volumes et que nous utilisons actuellement.

Avec cet ouvrage Brunet souhaitait décrire les "livres rares, précieux, singuliers et aussi les ouvrages les plus estimés en tous genres qui ont paru tant dans les langues anciennes que dans les principales langues modernes depuis l'origine de l'imprimerie jusqu'à nos jours" (1470-1860). C'est donc un ouvrage essentiel pour la connaissance du livre ancien, dont les collections sont dispersées pour de multiples raisons.

Il s'agit d'une bibliographie générale, internationale, rétrospective, sélective et commentée qui retient environ **47 500** ouvrages présentés dans l'ordre alphabétique des auteurs et des titres anonymes avec "l'histoire des différentes éditions qui en ont été faites; des renseignements nécessaires pour reconnaître les contrefaçons et collationner les anciens livres [avec] une concordance des prix

(1) Quérard, Joseph, 1797-1865, 1^{ère} édition publiée de 1827 à 1839 en 10 volumes.

auxquels une partie de ces objets ont été portés dans les ventes publiques en France, Angleterre et ailleurs, depuis près d'un siècle ..." Les descriptions sont extrêmement minutieuses (1).

Cette partie alphabétique est complétée par une "Table méthodique" qui suit le classement des libraires en usage depuis le 17^{ème} siècle, que Brunet a largement contribué à faire connaître et qui a été appliqué dans de nombreux répertoires jusqu'à l'adoption de la CDU. Cette table en 5 divisions (théologie, jurisprudence, sciences et arts, belles lettres, histoire) avec des subdivisions très détaillées reprend, en une séquence numérique continue de 31872 numéros, tous les livres cités dans la 1^{ère} partie, mais dans un signalement très abrégé.

-Utilisation du *Manuel* :

-Parmi les ouvrages en toutes langues on trouve des ouvrages français et le *Manuel* peut être utilisé comme bibliographie rétrospective "nationale" sélective pour la période antérieure à la *France littéraire* de Quérard (2), soit de 1470 à 1699 environ.

-la recherche peut être faite au titre pour les ouvrages anonymes, recherche que ne permettent pas *La France littéraire* et *La Littérature française contemporaine* dont le classement ne concerne que les auteurs.

-une recherche documentaire par grand domaine peut être faite grâce à la "Table méthodique", à condition de pouvoir

(1) Pour une description complète du *Manuel* de Brunet, voir l'ouvrage de Madame Marcelle Beaudiquez Directrice de la DDSR, *Guide de bibliographie générale : méthodologie et pratique*. Nouvelle éd. revue et mise à jour. München ; New York ; Paris : Saur, 1989

(2) Opus cité p.6

replacer le sujet recherché dans un cadre qui n'est plus familier à notre conception moderne du partage des connaissances.

-inversement, à partir d'un ouvrage donné, figurant dans la partie alphabétique, il est possible de retrouver d'autres ouvrages se rapportant au même domaine de la connaissance: à la fin de chaque notice, un numéro entre crochets carrés renvoie à la section correspondante de la Table "méthodique".

Il est à noter que la plupart des fonds anciens des bibliothèques municipales ont été classés selon le "Système des libraires" adopté par Brunet pour la présentation des notices de son *Manuel*. Les magasins de livres sont alors dispersés en cinq parties : Théologie, Jurisprudence, Sciences et Arts, Belles lettres, Histoire. Aujourd'hui la numérisation permet de contourner l'appartenance arborescente à une classification, pour pratiquer l'intertextualité multiforme dans laquelle peut s'inscrire un ouvrage. Il s'agit alors d'organiser un instrument de mise en relation de corpus dispersés.

L'utilisation souhaitée est : la recherche bibliographique rétrospective par auteurs et titres des ouvrages anonymes. Il est donc essentiel de trouver une structure qui permette d'isoler les vedettes et les éléments les plus significatifs. On peut distinguer les textes à statut iconique (textes numérisés en mode image), des indices et symboles qui caractérisent les corpus saisis en mode texte très structurés, utilisés comme *livres outils*. Des parcours heuristiques de recherche sont alors possibles. Cet aspect doit être mis en perspective avec

l'idée que la numérisation est corrélée au fait majeur du découplage entre conservation d'un document, et lieu ou temps de consultation. Ainsi sont ouvertes les possibilités de communication, de traitement à distance, de complémentarité au sein d'un corpus de documents, d'origines et de supports différents. La numérisation permet la mise en commun de corpus physiquement distants, l'échange de documents partagés avec les Pôles associés qui disposeront de leurs fonds spécifiques et de terminaux d'accès communicant avec l'ensemble des collections. Pour que ces corpus soient utilisés au mieux, il semble nécessaire de travailler à la fois avec l'idée du statut des types de textes numérisés (*livres apportant des connaissances sur les livres*), de la structuration des documents suivant leurs usages (mode image, mode texte), des réseaux sémantiques qui relieront alors ces textes (un exemple est donné dans la deuxième partie sous le titre "proposition de D.T.D.")..

Cette idée d'utiliser des livres outils, comme moyen de recherche heuristique rejoint les préoccupations du projet européen HPB (Hand Printed Book), qui réunit un consortium de bibliothèques européennes de recherche, dans le but de mettre sur pied une base de données informatisée, portant sur les collections de livres anciens en Europe. Monsieur Dupuigrenet-Desroussilles (1), responsable de ce projet, montrait récemment, en prélude au colloque de Roanne (1er et 2 octobre 1996), comment la création de cette base

(1) Monsieur François Dupuigrenet-Desroussilles, est Directeur de l'École nationale supérieure des Sciences de l'information et des Bibliothèques depuis 1995, après avoir assuré la Direction du service de l'inventaire général au département des imprimés de la Bibliothèque nationale de France.

reposait sur le constat que les collections de livres anciens européens sont dispersées pour de multiples raisons : " on produit au 16 e siècle à Venise pour Paris, à Paris pour l'Angleterre, à Lyon pour l'Espagne etc. Elles le sont aussi pour des raisons moins pacifiques, à la suite de saisies, prises de guerre et confiscations diverses..."

Tous ces aléas de l'histoire ont pour conséquence que nous pouvons trouver des fonds importants de livres français par exemple en Russie, parce qu'il y a eu de très importants achats faits pendant la période révolutionnaire. C'est aussi le cas en Allemagne, par exemple à Göttingen. Quant à elle, la Bibliothèque nationale de France compte certainement une des collections d'ouvrages de littérature italiens du 16 e siècle parmi les plus importantes du monde"⁽¹⁾. Tout en mettant en oeuvre une politique de description et d'identification des collections, il serait souhaitable de redonner vie et éclat aux contenus des collections de livres anciens. Dans ce contexte la numérisation en mode texte d'outils bibliographiques mis en perspective avec des collections d'oeuvres en mode image, permettrait de reconstituer des collections virtuelles d'ouvrages dispersés physiquement, mais reliés par l'histoire et l'appartenance à une culture commune. La numérisation serait alors un outil permettant de lutter contre les connotations symboliques attachées à la notion de patrimoine écrit, collections parfois décrites comme figées, ayant fini leur vie. La création de collections numérisées structurées autour d'outils

(1) Les migrations du patrimoine écrit, ces livres venus du bout du monde ..., Entretien avec François Dupuigrenet-Desrousilles directeur de l'ENSSIB, Colloque national de Roanne, 1 er et 2 octobre 1996, Actualité Rhône-Alpes du livre numéro 116, septembre 1996, supplément à Livres-hebdo et Livres de France.

bibliographiques d'investigation permettrait d'inventer (au sens de découvrir), de sauver de l'oubli des oeuvres dispersées, qui retrouveraient leur statut d'objets vivants, avec toute une histoire que l'on pourrait remettre en perspective avec notre univers mental d'aujourd'hui. Parce que notre culture est structurée par ce patrimoine dont nous gardons des traces mnésiques et linguistiques. Par l'élaboration d'un document de substitution accessible, organisé au sein de la collection numérique, ne pourrait-on rejoindre la préoccupation des spécialistes du livre ancien : "l'un des buts que doivent se fixer les personnes chargées de leur conservation, c'est de continuer à alimenter cette histoire et ne pas les conserver simplement avec un syndrome du coffre-fort". (1) La spécificité d'une collection numérisée est de pouvoir rassembler ce qui est dispersé, de permettre au lecteur de sélectionner un corpus, non en fonction de son appartenance à une classification disciplinaire, ou déterminée par sa localisation physique, dans l'un ou l'autre des départements thématiques de la bibliothèque, mais au regard de nombreux critères intellectuels tels que chronologie, les contextes de production, l'histoire du livre, l'interdisciplinarité des auteurs, etc... La notion de nouvel encyclopédisme virtuel, cartographie des savoirs, remédiant à la dispersion physique des volumes, est alors posée.

(1) Les migrations du patrimoine écrit, ces livres venus du bout du monde ..., Entretien avec François Dupuigrenet-Desrousilles directeur de l'ENSSIB, Colloque national de Roanne, 1 er et 2 octobre 1996, opus cité (cf : p.10)

1.1. Une approche matérielle de la numérisation d'un corpus : l'exemple du *Manuel* de Brunet

Le *Manuel* comprend 40 000 notices, 7449 pages, 72 lignes par page sur deux colonnes (chaque colonne est numérotée). Du point de vue de la richesse du contenu, il s'agit d'une base de données importante : Brunet s'était donné pour but d'éliminer par absorption toutes les autres bibliographies de son époque, et dans ses différentes préfaces il énumère les travaux dont il a tiré parti, "en extrayant la substance de ce qu'ils contiennent", et les rendant inutiles. La numérisation et le balisage d'un tel corpus reprennent l'oeuvre de Brunet avec de nouveaux moyens.

1.2. La définition de type de document : principes généraux

Une définition structurée du document permet de décrire l'organisation logique des documents, par exemple la collection des *Que sais-je ?*. La T.E.I ⁽¹⁾ va plus loin car elle propose non seulement de s'adresser à une classe d'ouvrages, mais même à des textes aussi variés que des pièces de théâtre, des oeuvres poétiques, des textes juridiques. Cette approche a été faite par des philologues qui ont construit un inventaire - une sorte de flore au sens de Buffon - des divers éléments pouvant constituer un document littéraire, cet inventaire constituant une "base tag sets". Chaque balise (tag) est utilisée pour définir un élément logique.

Une D.T.D. définit :

(1) Text Encoding Initiative : système de codage permettant l'échange et la circulation de documents électroniques en mode texte. Il s'agit d'un projet international né en novembre 1987 au Vassar College (New York) et soutenu par la Commission Européenne DG XIII...

- les noms des éléments autorisés ;
- la fréquence d'apparition des éléments ;
- l'ordre d'apparition des éléments ;
- la présence obligatoire ou non du balisage, par exemple les balisages de début ou de fin ;
- les contenus des éléments c'est-à-dire les noms des éléments autorisés à apparaître à l'intérieur d'autres, jusqu'au niveau de données *caractères* ;
- les attributs de balises et leurs valeurs par défaut ;
- les conventions d'éditions qui peuvent être exploitées.

Les linguistes comme Noam Chomsky (1) décrivent les règles grammaticales par des balises du type :

<Phrase> = <GN> <GV>

<GN> = <ART> <ADJ>* <Nom>

<GV> = <Verbe><ADJ> | ...

<Art> = le | la | un | ...

Avec cette grammaire et en parenthésant les éléments internes, une phrase comme : "le célèbre *Manuel* de Brunet qui englobe totalement les bibliographies", s'écrit :

<Phrase>

<GN><Art>Le</Art><Adj>célèbre</Adj><Nom>*Manuel* de Brunet</Nom><pronomrelatif>qui</pronomrelatif></GN><GV><Verbe>englobe</Verbe><Adv>totalement</Adv></GV><GN><Article>les</Article><Nom>bibliographies</Nom></GN></Phrase>

Une telle notation a été employée pour les langages de définition de documents structurés (SGML, HTML), par

 (1) né en 1928 à Philadelphie, professeur au Massachusetts Institute of Technology, il est l'un des plus célèbres linguistes contemporains, proposant une nouvelle définition de la nature formelle du langage et de la grammaire générative et des automates de reconnaissance..

exemple :

<roman> = <chapitre>*

<chapitre> = <titre><paragraphe>*

<paragraphe> = <texte>

<titre> = <texte>

indique qu'un roman est formé de chapitres, qu'un chapitre est formé d'un titre et d'un nombre quelconque de paragraphes, et qu'un paragraphe c'est du texte.

1.3. La numérisation de l'Édition Lyonnaise du XVI^{ème} siècle, le *Manuel* de Brunet et les stratégies de recherche au sein de la collection numérisée:

Le volume énorme des masses textuelles envisagées dans le cadre du projet de numérisation, ou de l'offre d'information sur des réseaux ouverts comme Internet posent la question de la recherche d'information pertinente, qui devra reposer sur des instruments de navigation (index), des outils bibliographiques à numériser en mode texte. La question de l'instruction de la stratégie de recherche heuristique est liée à la structuration interne de ces documents indispensables pour une utilisation en réseau des collections (1). Nous avons avancé l'hypothèse du *livre-outil*, notamment le *Manuel du libraire et de l'amateur de livres* de Brunet, saisi en mode texte suivant une D.T.D., soigneusement organisée pour mettre en valeur les aspects informatifs du texte. Nous tenterons, pour illustrer cette utilisation, de présenter un parcours heuristique reliant les notices *La Boétie, Xénophon*

(1) cette question a été débattue au cours de la réunion du 11/07/96. Les participants ont été unanimes pour demander la numérisation d'ouvrages bibliographiques en mode texte et notamment le *Manuel* de Brunet. : cf. Compte rendu en annexe.

et *Montaigne*. Si le lien La Boétie Montaigne semble évident, le détour par Xénophon, l'est beaucoup moins. C'est pourquoi hypertexte et base de données sont deux ressources complémentaires.

Ce parcours pourrait être effectué à partir d'une notice ainsi structurée, (les balises SGML ont été ajoutées au texte original) :

<notice>

<vr>La BOETIE (Etienne de)</vr>

<biblio> Historique description du solitaire et sauvage pays de Medoc (dans le Bordelois), par feu M. de La Boétie, conseiller du roy, *Bordeaux, Millanges, 1593, in-12.*</biblio>

<num>[24686]</num>

<p>Nous n'avons pas encore eu occasion de voir cette description fort rare, et à laquelle, selon le Nouveau Le-Long, n°2230, on a joint quelques vers de La Boétie, qui ne se trouvent pas dans l'édition donnée par Michel Montaigne des oeuvres de son ami.

<p>M. Léon Feugère, après avoir donné <biblio>*Étienne de la Boétie, études sur sa vie et ses ouvrages, Paris, Labitte, 1845, in-8°,* </biblio> a publié les <biblio>*Oeuvres complètes* de cet ami de Montaigne, *avec ses notes, Paris, Delalin, 1846, in-12*</biblio> <num>[19956]</num>, volume contenant tous les opuscules connus de La Boétie, y compris les 29 sonnets insérés dans les diverses éditions des essais de Montaigne, mais non pas la *Description du Médoc* qui est introuvable. L'écrit le plus remarquable de La Boétie, <biblio>*La Servitude volontaire ou le Contre un,* a été imprimée pour la première fois dans les *Mémoires de l'Etat de la France sous Charles IX*

(par Simon Goulart), dont il est paru trois éditions de 1576 à 1578. Coste l'a insérée dans les éditions de Montaigne qu'il a données en 1727 et depuis, c'est ce qu'ont fait également presque tous les éditeurs des Essais, qui sont venus après Coste; mais la meilleure édition que l'on ait de cet éloquent morceau de la prose est celle qui est imprimée à la suite de la Notice bio-bibliographique de la Boétie...; par le Dr J. -F. Payen, Paris, typographie de F. Didot, 1853, in-8. de 146 pp., avec une vue du Castelet de La Boétie, et un fac-similé de son écriture. M. Payen y a donné pour la première fois le vrai texte de l'auteur d'après un manuscrit contemporain et authentique. Il a été tiré en Gr. Pap. vingt-cinq exemplaires de cette excellente notice.

Parmi les éditions de la *Servitude volontaire*, on doit distinguer celles de Paris, Daubrée, 1835, in-8 et in-18 avec une préface analytique et apologétique de Fr. de Lammenais; et une autre sous ce titre :

DE LA SERVITUDE volontaire ou le contr'un, par Étienne de la Boétie, ouvrage publié en l'an 1549, et transcrit en langage moderne, pour être plus à la portée d'un chacun voire des moins aisés, par Adolphe Reschalet (anagramme de Charles Teste). Bruxelles et Paris, chez les marchands de nouveautés 1836, in-18, avec ou sans errata. L'édition n'a pas été mise en vente. (Notice de M. Payen, p.58.)

- Ménagerie, voy. Xénophon.

La notice Xénophon contient la mention suivante :

<notice>

.....

<p>-La ménagerie de Xénophon ; les règles de mariage de Plutarque ; lettres de consolation de Plutarque à sa femme, trad. du grec en françois par Estienne de la Boëtie, ensemble quelques vers latins et françois de son invention ; item un discours sur la mort dudit seigneur de la Boëtie par M. de Montaigne. Paris, Frederic Morel, 1571, ou 1572, in-8.

<note>On trouve difficilement ce volume complet, c'est-à-dire avec les vers françois, qui, bien qu'indiqués sur le titre daté de 1571, et réunis au volume dont on a en même temps réimprimé le titre avec la date de 1572. Vendu ainsi complet, 15 fr. d'Hangard, et 20 fr. en 1814 ; 42 fr. 50 c. Nodier.

L'édition de *Paris, Cl. Morel, 1600*, pet. in-8, sous le titre de *Mesnageries d'Aristote et de Xénophon...traduites de grec en François par feu Estienne de la Boëtie...et mises en lumière avec quelques vers françois et latin du dict La Boëtie par Michel sieur de Montaigne*, est une réimpression, faite page par page, de celle de 1571-1572.</note>

Dans ce contexte on pourrait retenir comme base de description les éléments suivants :

L'unité retenue serait la notice d'auteur ou d'anonyme, chaque notice commençant par une chaîne de caractères en majuscules d'imprimerie accentuées, le changement de notice est signalé par une nouvelle chaîne de caractères du même type, la notice comprend des paragraphes et un numéro unique entre crochets carrés.

Les délimiteurs :

La syntaxe concrète est définie par les délimiteurs

< ouverture d'une balise

</ ouverture d'une balise terminant une instruction

-- début et fin de commentaire

< ! ouverture de déclaration de balisage

< ? ouverture d'instruction de traitement

La déclaration de balisage commence la définition d' un type de document :

```
<!DOCTYPE notice SYSTEM "notice.dtd" [
```

```
< ! ELEMENT notice -(vr/nt/p+/biblio/com/num/renv/note  
/vente) ]>
```

vr = vedette retenue (auteur ou anonyme);

nt=nom ou titre ;

p+ = paragraphe, le + indique qu'il y en a généralement plusieurs tandis que vedette et numéro sont uniques pour chaque notice ; com = commentaires ;

biblio = bibliographie en italique et abrégée dans une notice;

num = numéro entre crochets carrés unique pour chaque notice et permettant d'utiliser la table systématique ;

renv = renvoi

note = note

vente = vente (des indications sur des ventes célèbres de livres tels La Vallière ou Pixierécourt permettent à l'historien du livre de suivre le parcours d'un texte à travers des éditions multiples).

1.4. Du texte numérisé aux liens entre notices

Des liens peuvent être établis entre des éléments bibliographiques cachés dans les notices, comme dans notre exemple le texte de Montaigne à propos de La Boétie mentionné dans la description d'une traduction du grec établie par La Boétie. Aux possibilités d'atteindre sélectivement telle ou telle entrée, la numérisation ajoute des capacités relationnelles, qui transforment le texte en une base de données structurée. La notion de réseau sémantique

du document intervient quand on peut croiser les éléments internes du texte balisé avec les descriptions normalisées en format InterMarc qui décrivent les attributs externes du document. L'écran du Poste de Lecture Assistée par Ordinateur devient alors le lieu de rencontre de ces deux techniques de l'informatique : système de gestion de base de données et hypertexte, qui permettent au lecteur savant de confronter les éditions, de critiquer les sources d'information dans la tradition de la philologie. Les idées de constitution des collections, de catalogage scientifique qui forment la trame de la bibliothéconomie peuvent rencontrer les méthodes de l'Internet : adresses URL, mise en réseau, hypertexte (http), tout en offrant des garanties quant à la qualité des documents proposés, des possibilités d'annotation et de sauvegarde du travail de lecture/écriture. Un document de la collection des imprimeurs lyonnais sera étudié au moyen d'un réseau sémantique comprenant :

-la description faite par les catalogueurs par exemple :

Agrippa de Nettesheim, Henri Corneille [Agrippa von Nettesheim, Heinrich Cornelius]

devient : L 100 \$w.1..b.freAgrippa\$mHenri Corneille

Vedette principale personne physique (notice GEAC n'affichant pas le numéro d'identification de la notice d'autorité)

-la notice du *Manuel* de Brunet avec balise SGML :

<vr>AGRIPPA (Henr.-Corn.)</vr> vedette retenue

La notice InterMarc mentionne le fait qu'il s'agit d'un document électronique :

300 \$aDate d'éd. du microfilm provenant d'un catalogue

d'éditeur \$a Réunit 2 vol.\$aRéunit : "De occulta philosophia" ; "Geomantia" ; "Elementa magica Petri de Abano philosophi" ; "Ex libro Iohan. Triteni octo quaestionum, libellus" ; "Artem brevem Raymondi Lulii commentaria" ;

L'origine de l'édition sur microfiches utilisée pour la numérisation :

324 1\$kNum. BNF de l'éd. de \$bCambridge(Mass.)

\$cOmnisys\$d[ca 1990]\$eFrench books before 1601\$

v.121.2\$g1 microfilm\$kReprod. de l'éd. de \$bLyon

\$cBeringos fratres\$d1600

peut être mise en relation avec les descriptions de Brunet décrivant le recueil dans l'édition lyonnaise :

Lugduni, per Beringos fratres, 2 tom. en 3 vol. pet. in-8 12 à 18 fr. [8863]

Le traité De occulta philosophia est le principal ouvrage de ce recueil. Il fut d'abord publié à Anvers, chez Graphe, en 1531 (D. Clément dit 1530), in-4. Celle de Cologne, 1533, in-fol, qui a été réimpr. sous la même date, renferme trois livres (...) Le 4 e livre, ajouté aux éditions faites après la mort d'Agrippa n'est pas de lui.

Brunet conclut que l'édition lyonnaise est la meilleure :

Mais la meilleure édition est facile à reconnaître, parce qu'il s'y trouve au verso du titre un portrait d'Agrippa gravé en bois, qui n'est pas aux deux autres. Remarquez de plus que la bonne édition de ce premier volume est la plus complète ; elle a 799 p., tandis que la réimpression en italique n'en a que 677, et il y manque six ouvrages.

Les acquéreurs de la Bibliothèque nationale de France ont fait le choix de numériser cet auteur d'après les microfilms réalisés à Cambridge Massachusetts, eux-mêmes réalisés

d'après l'édition lyonnaise. Cette donnée bibliographique est signalée par les catalogueurs (zone 324 d'Intermarc : note de reproduction en fac-similé). Le chercheur pourra consulter le Manuel de Brunet, pour connaître les différentes éditions et éventuellement faire des annotations personnelles, à partir de ses propres travaux sur d'autres éditions. Cette activité de *servuction* (1), service produit par la synergie entre agent du service public et utilisateur, rejoint les fonctions assignées à la bibliographie par Monsieur Roger Chartier : " Loin d'être un savoir cantonné et auxiliaire, voué à l'inventaire et à l'interprétation de données formelles mises au service de l'édition de textes, la bibliographie ainsi redéfinie devient une discipline essentielle pour reconstituer comment une communauté donne forme et sens à ses expériences les plus fondamentales à partir de textes multiples qu'elle reçoit, produit et s'approprie". (2)

A la B.n.F. la station de lecture/écriture personnelle est étudiée de manière à augmenter la lisibilité. Elle offre un choix d'écrans adaptés au lecteur (taille, résolution, couleur), l'offre commerciale dans ce domaine se diversifiant (écrans de grande taille, 19 à 22 pouces, résolution de 300 dpi) destinés à la publication assistée par ordinateur. Le type, la vitesse du défilement à l'écran et son contrôle, la création d'outils de maniement logique et physique d'écrans, de fenêtres multitâches (plusieurs programmes pouvant être exécutés parallèlement), en font un outil ergonomique pour la navigation textuelle et l'annotation.

(1) Jean-Michel Salaün, Le marketing des bibliothèques, Paris : Éditions du Cercle de la librairie, 1992.

(2) Roger Chartier, préface à l'ouvrage de D.F. McKenzie : La bibliographie et la sociologie des textes, Paris Éditions du Cercle de la Librairie, 1991.

Messieurs Jean-Louis Lebrave et Bernard Cerquiglini ont montré comment la philologie bénéficie des techniques de l'hypertexte, par l'utilisation d'une disposition "écranique", autorisant l'affichage en parallèle des différentes versions d'un texte ancien, permettant de structurer des groupes de copies, en *familles*, reliées par des liens de type hypertextuel. *Tout le travail de l'érudition philologique consiste, en remontant dans le temps, à reconstituer le texte original à partir de ces copies plus ou moins fautives.* La matérialité de l'écran de l'ordinateur permet la comparaison minutieuse - des textes : *La comparaison entre manuscrits joue un rôle crucial. Elle débouche sur la notion de "lieux variants". Étant donné un texte dont on suppose que sous sa forme abstraite et idéale, il est achevé, cohérent, et stable, mais qu'il nous a été transmis de manière instable, on l'organise en une séquence de fragments invariants et de lieux invariants, fragments apparentés, mais différents les uns des autres selon l'exemplaire matériel du texte* (1).

Les stations de lecture/écriture proposent des fonctions d'édition, de publication. Ce qui implique des fonctions de traitement de texte sophistiquées, incluant des outils de PAO, ouvrant des possibilités de préparer des rapports, des conférences, des supports de cours ("transparents" destinés aux rétroprojecteurs), diapositives (traitement de la photographie par ordinateur), bandes vidéo, bases de données personnelles, processeurs d'idées.

(1) Jean-Louis Lebrave . *L'hypertexte et l'avant texte, Texte et ordinateur : les mutations du lire-écrire.* Actes du colloque interdisciplinaire tenu à l'université Paris X Nanterre, 6, 7, 8 juin 1990, Éditions de l'Espace Européen 1991.

Pour utiliser au mieux son propre travail, des fonctions d'impression, d'expédition vers son ordinateur à domicile par modem, de fax, peuvent être jointes. Opérer une meilleure visibilité sur de grands corpus de connaissance, tel est ce que permet l'ordinateur : grande recension de textes via l'indexation, l'annotation, le classement de ces connaissances et la possibilité de réactiver de façon automatique le capital d'un travail entrepris de longue haleine.

Cependant la qualité technique ne serait que de peu d'utilité, si elle ne s'accompagnait d'un riche environnement de lecture, soit des dictionnaires mono et multilingues, des encyclopédies, des dictionnaires de synonymes, analogiques, étymologiques, des thésaurus, des bases de données bibliographiques (BN-Opale, puis Catalogue Collectif de France), des répertoires bibliographiques numérisés tels le *Manuel* de Brunet, les tables des matières et index.

Ces stations de travail doivent également permettre la communication entre des utilisateurs différents (forums, bulletins électroniques), avec l'administration de la Bibliothèque et différents serveurs d'autres bibliothèques.

Un second groupe de fonctions qui implémentent la lecture assistée par ordinateur et l'édition représente le cœur du système. La définition de ces fonctions repose sur une analyse détaillée de la lecture de recherche afin d'en transposer les caractéristiques au sein du dispositif informatisé. C'est une lecture de fréquentation d'appropriation qui, dans son principe, et malgré quelques analogies de surface ne revient pas à la consultation d'une base de données textuelles. C'est en effet une lecture qui

s'inscrit dans le long terme, qui constitue un parcours individualisé. C'est une lecture attentive, d'inspection et d'observation scrutative des contenus, des formes, du lexique mais c'est aussi dans le même temps une lecture d'exploration. C'est enfin une lecture qui engendre une intense activité de rédaction. Elle a un rythme lié à la recherche qui conduit à s'arrêter, mettre en ordre, explorer, plus ou moins minutieusement, avec des fluctuations, des moments de cristallisation et de relecture. La préservation et l'organisation des traces de ce processus sont essentielles pour comprendre le parcours individuel de recherche.

Cette recherche rejoint une problématique beaucoup plus vaste, comme le montrent à la fois l'ampleur des débats et le nombre des articles publiés. Pour Bernard Stiegler, il s'agit de *la maïeutique d'un nouvel instrument spirituel*. Lire signifie déjà écrire (et l'on pourrait rapprocher cette idée de ce que dit Umberto Eco dans *Lector in fabula* (1) : lire n'est pas un acte neutre, il se noue entre le lecteur et le texte une série de relations complexes, de stratégies singulières qui, le plus souvent modifient sensiblement la nature de l'écrit original). La lecture est le travail d'une *impression* (Bernard Stiegler). Souligner, entourer un ou des mots, ficher, rapprocher, corréler, corner une page (pratique déconseillée avec les livres de bibliothèque), glisser un signet parfois lui-même annoté, un "post-it", commenter, résumer, faire la synthèse de plusieurs textes, schématiser, élaborer un plan, rédiger : il n'est pas possible de distinguer où, parmi ces

(1) Eco, Umberto *Lector in fabula ou la Coopération interprétative dans les textes narratifs*, Paris : 1985.

opérations, commence l'écriture, où cesse la lecture. Plus la lecture devient "savante", plus on ne lit qu'en inscrivant, on n'écrit qu'en déchiffrant, écrire, c'est toujours lire et relire, citer, montrer, enchevêtrer, réagencer ses lectures antérieures, y compris en les oubliant. C'est aussi préparer ses lectures à venir, *s'apprêter à lire le nouveau en cheminant par le fouillis textualisé des mots, en capillarissant son corpus, en y tramant ses réseaux*, (Bernard Stiegler) (1).

1.5. La collection numérisée :

L'originalité du programme de numérisation de la Bibliothèque consiste à réaliser simultanément une collection électronique évaluée aujourd'hui aux alentours de 100 000 livres ou documents et à développer une station de lecture informatique permettant le travail de recherche approfondie sur les textes structurés (annotation, copie, création de fichiers, archivages, etc.).

La numérisation n'est pas conçue comme un support de substitution qui permettrait d'éviter la communication des livres dans leur forme d'origine (elle pourra faciliter la consultation par le plus grand nombre, d'ouvrages fragiles ou rares). Elle a pour ambition de favoriser des pratiques de lecture complémentaires : sur support papier d'une part, et sur support numérique d'autre part. Il faut également noter que pour ce qui concerne le support informatique le lecteur aura à sa disposition deux types de documents, une copie en mode image qui est la représentation fidèle de l'original (mise en

(1) Bernard Stiegler : Machines à écrire et matières à penser, Revue Genesis numéro 5, 1994.

page, typographie, iconographie) et une version en mode alphanumérique désolidarisée du modèle traditionnel (et ouvert éventuellement à toutes les fonctionnalités de l'hypertexte).

La lecture de document sur support électronique, va-t-elle modifier les pratiques de lecture attachées au livre sous la forme de codex, tel que l'a connu notre culture depuis cinq siècles ?

Elle changera en tout cas les pratiques institutionnelles actuelles puisque la localisation des documents numérisés sera indifférente au lecteur qui pourra les consulter à distance, obtenir des copies électroniques lui permettant de travailler chez lui ou dans la salle de lecture de la bibliothèque ou dans celle d'un "pôle associé".

"Ce qui caractérise cette évolution c'est à la fois la nouvelle forme de l'ambivalence entre système de signes et perception sensible, et le nouveau rapport au temps" (1).

1.6. Favoriser la lecture/écriture assistée par ordinateur

Des terminaux de lecture (" Poste de lecture simple", PLS) permettront d'accéder à ces documents, mais un seuil nouveau est atteint dès lors que le terminal de travail informatique ("Poste de Lecture Assistée par Ordinateur", PLAO) permettra une valeur ajoutée sur laquelle nous reviendrons.

(1) Yannick Maignien : l'oeuvre d'art à l'époque de sa reproduction numérisée, Paris : Bulletin des Bibliothèques de France, tome 41 numéro 1, 1995

La politique de numérisation intègre en effet le pouvoir de traitement informatique dont le lecteur dispose, grâce à de puissantes stations de travail.

Le poste de travail permet d'accéder à l'information : à partir du catalogue, de choisir et de charger les ouvrages numérisés, de se connecter au réseau de bases de données en ligne, grâce à une architecture ouverte du système, compatible avec Internet, ainsi qu'avec des serveurs de CD-ROM.

La collection numérisée a pour finalité le travail des chercheurs. De ce point de vue, l'objectif est de constituer une *bibliothèque idéale*.

Cette bibliothèque idéale est la résultante de deux directions, de deux ordres de préoccupations distinctes, mais étroitement corrélées aux caractéristiques de la numérisation :

- un de ces axes sera qualifié de bibliothèque (relativement) immatérielle
- l'autre de bibliothèque cohérente

Conséquences des usages des fonds numériques sur le choix des textes

Plus que les caractéristiques du médium électronique (ses coûts, ses contraintes, et possibilités), ce sont les usages attendus de ce fonds numérisé qui induisent des critères de sélection des ouvrages.

Le premier critère de choix porte sur l'édition scientifique qui propose l'appareil critique le plus important, notamment les éditions critiques des œuvres reconnues par les spécialistes comme essentielles, tout en offrant aussi quelques variantes si l'établissement du texte pose des problèmes de source. Cela

renvoie à la possibilité de travailler sur différentes éditions d'une même œuvre.

Le deuxième groupe est composé de la littérature critique (après une sélection exigeante) dans la mesure où celle-ci est considérée quasiment comme une œuvre. Les travaux critiques qui ont fait date historiquement seront prioritaires. La décision de retenir telle ou telle édition d'œuvres complètes constitue en soi un choix de commentaire critique.

Enfin, le dernier groupe comprend des traductions. En premier lieu des traductions anciennes (par exemple les traductions des textes philosophiques de l'époque médiévale au XIXème jamais rééditées), en second lieu des traductions plus modernes.

L'ordre des choix varient en fonction de l'auteur, de l'époque considérée et de ses spécificités. Les traductions latines des textes grecs, par exemple celles de Marcile Ficin qui jouent un rôle essentiel pour la Renaissance seront prioritaires et considérées comme une œuvre.

Une bibliothèque cohérente

Ce critère renvoie à la constitution du fonds lui-même. Là encore, la finalité du projet est fondatrice : il s'agit de réunir un fonds correspondant aux attentes des chercheurs de chaque discipline et ouvrant des passerelles interdisciplinaires. Cette idée entraîne le recours à des équipes de spécialistes formant des commissions ou réfléchissant à titre individuel. Les commissions ont été constituées sous la direction de Messieurs

Miquel et Chartier. En ce qui concerne les *grands lecteurs* on peut citer Messieurs Chistian Jacob, Jacques Roubaud, Patrick Bazin, et les experts Messieurs Bernard Stiegler, Jacques Virbel.

Ces chercheurs ont été mis à contribution pour l'évaluation des logiciels de lecture/écriture par l'expérimentation des stations de travail. De ces travaux résultent :

- la conception des différentes maquettes de PLAO ;
- la définition de l'assistance au travail intellectuel (la *maïeutique d'un nouvel instrument spirituel* ⁽¹⁾, pour le philosophe Bernard Stiegler.

Il convient de préciser que les disciplines qui président à la réalisation du programme de numérisation sont la littérature, l'histoire, les sciences politiques, juridiques, économiques, la philosophie, l'histoire des sciences, l'ethnologie, l'anthropologie, la sociologie, la linguistique.

Cette collection structurée est pensée en relation avec les préoccupations bibliothéconomiques :

La conservation de collections repose pour l'essentiel sur la restauration et le microfilmage des documents.

La numérisation, même si elle n'est pas d'abord une technique de conservation, permettra en tant que technique de préservation de diminuer la pression sur les collections. La durée de conservation est moindre que celle des microformes (5 à 10 ans actuellement) mais ces techniques de stockage de documents numérisés sont en constante évolution ainsi que leur coût. Grâce au support numérique, le document pourra être consulté simultanément par plusieurs utilisateurs.

(1) Bernard Stiegler, *Machines à écrire et matières à penser*, Revue Genesis numéro 5, 1994

important

Sa communication se fera sans recours à un personnel spécialisé. Le document pourra être servi en ligne ou faire l'objet de l'édition d'un CD-ROM, lui-même en service sur des serveurs alimentés par ces CD-ROM, dès lors que les procédures juridiques d'accès seront établies.

Les ouvrages les plus demandés à la Bibliothèque nationale étaient au nombre d'environ un million. L'évaluation statistique automatisée de communication des ouvrages sera un élément de connaissance important des attentes des chercheurs, et infléchira la politique de croissance de la collection.

La collection numérisée peut être qualifiée de "Bibliothèque patrimoniale et de recherche" de par son contenu, et de "libre-accès" par les possibilités de son médium. Tiré de cette sorte de "magasin numérique", l'ouvrage est à la fois destiné aux recherches et aux appropriations les plus savantes, et ouvert à une communication et consultation en libre accès.

Transmettre à distance

C'est là le facteur essentiel du cahier des charges de la nouvelle Bibliothèque. Les documents numérisés seront transmissibles à distance au sein du réseau des bibliothèques associées. Que ce soit pour le prêt inter-bibliothèques ou le réseau des postes de lecture des pôles associés, la transmission à distance des ouvrages par des réseaux numériques, en ligne ou en différé, s'impose compte tenu des avantages décrits ci-dessus : rapidité, disponibilité, simultanéité. Cette transmission peut s'effectuer par des techniques numériques, utilisant le protocole TCP/IP (File transfer protocol), mais aussi par la fourniture de copies

électroniques ou de tirages papier des documents. La création de CD-ROM à partir des collections numérisées est envisagée quand des accords avec les possesseurs des droits sont possibles (par exemple la Fondation pour la science dans le cas de la *Revue de synthèse*).

L'idée centrale du concept de lecture assistée par ordinateur est d'articuler les possibilités de consultation numérique de documents avec les possibilités de traitement numérique -non seulement du traitement du texte au sens traditionnel, mais d'analyse et d'exploitation numériques des contenus textuels. En effet, la diffusion d'un vaste fonds numérisé doit être accompagnée par la diffusion d'outils d'accès et de consultation qui constituent le contenu des futurs postes de travail.

Après avoir repéré les documents recherchés à partir du catalogue informatique de la bibliothèque, le chercheur pourra ainsi télécharger sur son poste doté de très grandes capacités de stockage et d'exploitation, des documents numériques internes à la Bibliothèque nationale de France. Il se constituera donc une "bibliothèque personnelle" et effectuera des opérations d'analyse textuelle que permet dès aujourd'hui d'envisager la technologie de l'informatique documentaire.

Il s'agit ainsi pour les chercheurs de retrouver et d'enrichir, dans un cadre informatique, les fonctions courantes du travail en bibliothèque, où lecture et écriture se combinent étroitement.

Les outils d'exploitation des textes ne produisent pas de sens, mais ils conduisent à des résultats qu'il s'agit d'interpréter. Ils permettent de nouvelles investigations portant sur le contenu des textes, rendent possible de nouvelles formes d'investigations et de nouveaux objets de recherche. Ce qui prédomine sont les fonctions méthodologiques et heuristiques. Enfin, certaines opérations (la recherche d'occurrences par exemple) entreprises par des moyens plus traditionnels demandent un investissement de temps et de travail qui peut être réduit sensiblement grâce aux logiciels adaptés.

II) 2. Poste de Lecture Assistée par Ordinateur **définition**

Le PLAO (poste de lecture assistée par ordinateur) est un scriptorium électronique, permettant de mettre en mémoire et organiser le travail de lecture et d'écriture.

Le travail de lecture savante est analysé en fonctions génériques pour être intégré dans des automatismes d'assistance, afin d'offrir à l'étude des possibilités sans égales de mémorisation, d'organisation et de vitesse de réactivation des données .

Comme nous l'avons noté en introduction, la numérisation d'ouvrages en texte intégral entraîne avec elle une nouvelle conception de l'encyclopédisme, posant des questions quant à la structuration du savoir, au rapport de l'image et du texte, à la navigation hypertextuelle dans de vastes corpus d'ouvrages numérisés, aux problèmes du statut légal des reproductions numérisées, et surtout *aux mises en relation*

entre les domaines de connaissance, liés aux "volontés de savoir" propre à telle époque, à des rapprochements de disciplines ou de documents qui n'ont de sens que dans telle épistémè précise (1).

2.1. Premières réalisations

Le projet de Poste de Lecture Assistée par Ordinateur a été concrétisé, sous la direction de Messieurs Jean Gattegno et Alain Giffard directeur du département des nouvelles technologies. Entourés de nombreux chercheurs, constituant le groupe des "Grands lecteurs" (2), ils ont permis l'émergence de la notion de lecture/écriture assistée par ordinateur - et la conception générale du système. Dans le cadre de la convention avec le CNRS, des PLAO mis au point par Advanced Information système (AIS), filiale d'informatique éditoriale du groupe Berger-Levrault, la maison d'édition la plus ancienne de France, ont été testés dans différents laboratoires.

La constitution du fonds numérisé ne peut être dissociée des finalités qui lui sont assignées : les problématiques de la relation de structuration du fonds et du balisage des documents participe de la même finalité. En effet la démultiplication des accès ne serait rien, si elle ne reposait sur l'ergonomie d'un système susceptible de répondre à une recherche d'efficacité, de productivité intellectuelle telles celles obtenues par la lecture assistée par ordinateur.

Quatre logiciels ont été expérimentés à domicile par des

(1) Yannick Maignien *La Bibliothèque virtuelle ou de l'Ars memoria à Xanadu*, BBF T.40 n° 2, 1995.

(2) Sur les *Grands lecteurs* voir note: page 29.

chercheurs de différentes spécialités (germanistes, anglicistes, archéologues, historiens, écrivains, mathématiciens, philosophes), disposant de corpus de textes numérisés dans leurs disciplines. Ces logiciels correspondaient à l'idée de *support dynamique* permettant la navigation dans un document *actif* ou *intelligent* par exemple l'intégralité de la littérature grecque antique depuis Homère jusqu'aux Byzantins sur CD-ROM (*Thesaurus linguae Graecae*), soit quinze siècles de ce qui nous est parvenu d'un moment fondamental de la pensée humaine, et à la possibilité pour l'utilisateur d'enregistrer les traces de ses propres actions dans la base elle-même et de les traiter.

Ce système d'archivage sous mots-clés constitue un thesaurus assorti de commentaires. Toutes ces informations peuvent être visualisées à n'importe quel moment du travail de lecture ou d'écriture. Une quatrième fonction importante concerne la lexicométrie. Cette méthode permet d'analyser les signes laissés par les acteurs eux-mêmes au cours de leur pratique (articles de revues par exemple), et de faire apparaître des fronts de recherche. On assiste au surgissement d'associations, de mises en relations, d'échanges articulés dans un réseau d'interactions, l'objectif étant par la lecture et l'analyse de ces flux d'informations, d'identifier et de caractériser ces échanges, ces médiations, qui participent à l'émergence d'un processus.

Les travaux des *Grands lecteurs* :

Analysant les fonctions génériques du travail de lecture savante, ils ont permis le choix et l'émergence des automatismes d'assistance dont le but est d'offrir à ces tâches

souvent répétitives des possibilités sans égales de mémorisation, d'organisation et de vitesse de réactivations des données.

C'est bien parce que le projet de la Bibliothèque nationale de France intègre ces finalités et en fait son principe directeur et régulateur, que la sélection des ouvrages procède d'un nouvel encyclopédisme ⁽¹⁾.

2.2. Fonctions des Postes de Lecture Assistée par Ordinateur : les objectifs

Il s'agit de mettre à la disposition des chercheurs :

- un outil de travail informatique de type station individuelle relié au système d'information de la Bibliothèque Nationale de France ;

- à partir de ce poste de travail, accéder aux catalogues et à travers ceux-ci prendre connaissance du fonds électronique afin d'en parcourir et d'en sélectionner les textes ;

- télédécharger dans les stations de travail les ouvrages du fonds identifiés comme pertinents pour le travail du chercheur ; munir ces postes de mémoires de vaste capacité, réinscriptibles et amovibles (par exemple de type magnéto-optique) ;

- augmenter ce corpus par des fonds numérisés personnels déjà possédés par le chercheur, ou saisis localement au clavier ou grâce à des scanners munis de logiciels de reconnaissance optique ;

(1) Yannick Maignien et Éric Dussert, *Numérisation et lecture savante : le projet de collaboration avec le Centre International de Synthèse autour de la Revue de Synthèse.*

-effectuer des opérations de lecture, annotation, commentaire, de constitution de bases de données personnelles, de comparaison, de structuration, d'analyse, de traitement et de publication de textes ;

-consulter un CD-ROM personnel, ou emprunté à la bibliothèque ;

-utiliser un modem pour atteindre des sources d'informations extérieures.

Ces postes de lecture permettent un ensemble de fonctionnalités.

Un premier groupe correspond aux fonctions de saisies, d'entrées-sorties, de gestion de périphériques et de communication :

1°) La constitution d'un corpus personnel

- accès : identification, parcours rapide de survol, feuilletage, sélection et commande de documents ; cette fonction s'appuie sur des rubriques étendues des catalogues classiques (tables de matières, sommaires, tables d'illustrations, résumés, index, etc.) ;

-saisie et acquisition : téléchargement dans le poste de travail des ouvrages identifiés ; incorporation de fonds numérisés personnels ; numérisation de documents sur support papier (imprimés ou manuscrits) ; conversion de formats;

-stockage sur supports de mémoire de masse.

2°) la facilitation de la lisibilité

-choix d'écrans assurant le confort de la lecture (taille, définition, couleur, etc.) ;

-modes et vitesse de défilement, modes d'intervention du lecteur ;

-création et agencement logique et physique des fenêtres ;
asservissement de fenêtres pour défilements parallèles ;

-choix d'outils d'action utilisateur ergonomiques pour la navigation dans les textes et les annotations ;

3°) la rédaction, l'édition, la publication

-traitement de texte évolué, logiciels de Publication Assistée par Ordinateur ;

-préparation d'exposés, conférences, cours etc. (transparents, diapositives, tableurs, gestionnaires d'idées, etc.);

-impression, transfert, échange.

4°) l'environnement de lecture

-dictionnaire de langue, de traduction, de synonymes, d'étymologie, mono et multilingues, encyclopédies, manuels de langue, grammaire, etc.

-thesaurii spécialisés ;

-bases de données bibliographiques.

5°) la communication

-entre utilisateurs : forum, tableaux d'affichage électroniques, outils de type travail coopératif (*workflow*) ;

-entre utilisateurs et administration de la bibliothèque ;

-entre la bibliothèque et les autres bibliothèques et centres serveurs : réseau Renater, normes TCP/IP et Z 39-50.

2.3. Les fonctions centrales : la lecture/écriture assistée par ordinateur

Un second groupe de fonctionnalités constitue le coeur du dispositif et correspond à la mise en oeuvre des méthodes de travail, de lecture et de rédaction assistées par ordinateur.

La définition de ces fonctions centrales s'appuie sur une analyse assez détaillée de la lecture d'étude ou de recherche, afin d'en transposer les caractéristiques au sein du dispositif informatisé.

Quatre groupes de fonctions sont donc envisagés pour prendre en charge les aspects spécifiques de ce type de lecture :

- la structuration des corpus personnels

organisation des textes engrangés "en vrac", structures ouvertes de fichiers, recherche full-text, caractérisation par mots-clés, indexation automatique et/ou manuelle ;

- repérages et balisages conceptuels, logiques, linguistiques, etc. d'entités, fragments, passages.

l'analyse et le traitement de textes :

- opérations (booléennes, statistiques...) de recherche et de caractérisation (occurrences lexicographiques, mots-clefs, concepts, etc.);

- visibilité des structures créées (diagrammes, histogrammes, tableaux, graphes, etc..).

2.4. L'annotation dynamique

Parmi les caractéristiques essentielles du dispositif de lecture, provenant de la lecture traditionnelle, un aspect important est constitué par un langage d'annotation permettant au lecteur / utilisateur d'un texte une appropriation personnalisée de ce document (le lecteur

professionnel met en oeuvre, plus ou moins consciemment, des techniques d'assimilation et d'interprétation où la lecture est indissociable de l'annotation, c'est-à-dire de l'écriture), qu'il est possible d'implanter dans un système de lecture assistée par ordinateur. Automatisé, il permet aussi de faire apparaître de nouvelles techniques de contrôle - dont font partie les outils de "navigation hypertextuelle" et la gestion automatisée des actes de lecture, correspondant sous forme de structurations particulières, à la constitution de bases de données individuelles.

Quant à son contenu, un tel langage d'annotation autorise l'utilisateur à exprimer sous une forme partiellement codifiée un très large éventail de réactions de lecture.

Peu d'études ont été consacrées, aux techniques d'annotation (1). Malgré son caractère familier, l'annotation apparaît comme une activité complexe, remplissant des fonctions multiples. Parmi celles-ci :

- le reformatage motivé ou systématisé ;
- l'annotation : inscriptions dans et autour du texte de symboles graphiques personnalisés et d'annotations et commentaires brefs ;
- les commentaires associés à des passages ;
- la création de liens (connexions) entre passages (par exemple entre traductions ou versions différentes ; typages et commentaires de tels liens ;
- la gestion (classement, indexation, etc.) des passages annotés, commentés, liés ainsi que des annotations et commentaires

(1) Nielsen, Jakob *Online Documentation and Reader Annotation. International Conference on Work with Display Units*, Stockholm, 12-15 May 1986

généérés par la lecture.

-le classement et l'archivage ;

-la constitution de dossiers structurés rangeant les ensembles de documents et de produits dérivés de la lecture.

Il s'agit principalement d'une technique élémentaire de mémorisation et de capitalisation d'efforts et de résultats de lecture, mais pas uniquement : ainsi peut-elle aussi servir à noter des réactions purement perlocutoires, ou bien au contraire des quasi-instructions pour des relectures ou même d'autres tâches extra-textuelles. Dans la perspective d'un travail collectif et coopératif, (où les annotations peuvent être échangées, modifiées, elles-mêmes annotées, etc.) celles-ci peuvent encore signaler d'autres types d'informations (par exemple : propositions de modification, de correction, de restructuration, etc). Aussi rencontre-t-on un fort grand nombre de types d'annotations. Pour en rester aux plus classiques :

-noter des réactions (exclamations, jugements de valeur, questions, etc.);

-modifier le texte (corrections, suppressions, insertions, interventions, etc.);

-hiérarchiser les passages (en fonction de l'intérêt, de l'importance, de la représentativité, avec quatre niveaux qui pondèrent le poids de passage du texte, auxquels correspondent sur le support de papier les soulignements dans le corps du texte et les traits verticaux en marge);

-contraster ou différencier des passages selon leur type logico-linguistique (par exemple " : définition", "exemple", "illustration", "contre-exemple", etc);

-repérer des termes caractéristiques (mots-clés, terminologies et formules propres à l'auteur, etc.), ce sont des opérations de qualification, qui consistent à attribuer des valeurs sémantiques à ces pondérations, par divers moyens :

-insertion de mots-clés ;

-insertion de notes personnelles ;

-rédaction de commentaires ;

-liens avec d'autres documents (autres passages du même texte ou autres textes, par exemple : sources manuscrites, variantes éditoriales, textes mentionnés en référence, traductions, gloses canoniques, références bibliographiques);

-indexer des passages ;

-créer des commentaires, des gloses, des traductions de termes ou de citations, des définitions ou des explications de termes, etc, associés à des termes ou des passages ;

-explicitement une structure logico-conceptuelle ou une construction implicite dans le texte par des marqueurs explicites : énumérations, alternatives, marqueurs argumentatifs, segmentations en parties et sous parties, etc.);

-établir des références croisées, et plus largement des connexions entre passages, et éventuellement commenter ces mises en relation ;

-indiquer des opérations à effectuer ultérieurement ("à relire", "à revoir", "à traduire", "citation à vérifier", "à recopier", etc) ;

Les modalités d'annotation, qu'elles concernent la marque du type d'annotation, ou bien les passages ou fragments concernés sont elles-mêmes très variées : soulignage, surlignage, coloration des chaînes de caractères ou des fonds,

marques graphiques (étoiles, points d'exclamation, d'interrogation, de suspension, accolades, crochets, entourages, etc.), insertion de termes, formules, mots-clés, numéros, rédaction de commentaires (structurellement indépendants ou au contraire greffés dans le texte), de pseudo-notes en bas de page.

Quant à la gestion au moins partiellement automatisée des effets des annotations sur le texte, il s'agit naturellement d'un gain fondamental, une des limites du contexte classique sur support papier étant justement la difficulté à mémoriser et retrouver les traces de l'activité d'annotation à travers les différents ouvrages annotés. Ces effets peuvent d'ailleurs se répercuter sur différents plans :

- la visualisation : les marques d'annotation, ainsi que les fragments ou passages textuels marqués sont, suivant la préférence de l'utilisateur, visualisés seuls, ou au contraire masqués, ou encore banalisés ;

- l'édition : les passages d'un type annotatif donné peuvent être sélectionnés, concaténés, examinés indépendamment un à un ou au contraire en fonction du nouveau tout qu'ils constituent , formatés selon des feuilles de style particulières un texte peut être reparcouru uniquement en fonction de telle ou telle marque d'annotation, etc ;

- l'analyse : l'ensemble des outils d'analyses applicables sur un texte dans son ensemble peuvent être appliqués uniquement sur les passages arguments de tel ou tel type d'annotation, au sein d'un même texte ou d'un ensemble de textes annotés selon les mêmes principes, ouvrant ainsi la possibilité d'un comparatisme étendu ;

- les bases de données textuelles : les actes d'annotation ont des répercussions sur l'organisation des bases de textes et leur visibilité, leur accès, leur parcours, et autorisent des opérations de gestion diversifiées (par exemple : liens entre fragments, recherche d'annotations du même type, recherche par mots-clés, recopie de passages ou fragments d'un type donné dans divers fichiers, etc.).

2.4.1. La maquette d'AIS Berger Levrault

Ces concepts ont été modélisés dans une maquette informatique, AIS Berger Levrault sur station Sun (Unix).

Fonctions proposées :

Le poste de lecture assistée par ordinateur (PLAO), dans cette version est un ensemble d'outils logiciels permettant :

- l'archivage et la sélection des documents ;
- la définition de l'environnement de travail : hiérarchie des mots clés et liste de types ⁽¹⁾ ;
- la lecture et l'annotation ;
- la préparation d'un document ;
- les fonctions de recherche ;

a) La notion d'ULD, Unité logique de Découpage :

-dans l'approche proposée par cette maquette, le PLAO est un ensemble d'outils permettant à un lecteur d'accéder à des documents de manière souple, rapide et personnalisée, ainsi que d'effectuer des recherches et des annotations sur ces documents. L'utilisateur abordant pour la première fois le

(1) ex :Type d'annotations : le type d'annotation personnelle caractérise un ensemble d'annotations inséré dans le texte.; Type de descripteurs : caractérise un attribut d'un document. Ainsi, le type "Auteur" caractérise-t-il l'attribut "Racine " ; Type de documents : caractérise un ensemble de documents de manière ouverte aux besoins particuliers du lecteur. Exemples : "thèse", "Article"; type de liens : caractérise un ensemble homogène de liens.

PLAO trouvera à sa disposition un environnement par défaut. Il lui est possible de modifier cette interface, pour l'adapter à ses méthodes de travail. Ces éléments sont :

- Une hiérarchie de mots clés permettant de réaliser des liens entre différentes unités logiques de découpage d'un ou de plusieurs documents. La constitution de cette hiérarchie est un aspect important de la constitution d'un environnement de travail ;

- une liste de types pour chacune des informations qu'un utilisateur peut ajouter sur les documents consultés :

- ULD (Unités logiques de Découpage)

- annotations

- illustrations

- liens

- jalons

Ces types ont pour objectif de représenter dans le poste la typologie des concepts que le lecteur veut calquer sur ses documents. A chacun de ces types, le lecteur pourra associer une représentation visuelle sous la forme d'un effet typographique ou d'un effet de couleur. il sera par exemple possible de définir trois types d'annotations :

- Remarque

- Attention

- Important

avec chacune un sens particulier pour le lecteur et chacune une représentation différente à l'écran.

On représente les opérations de hiérarchisation en utilisant des caractères en couleurs, tandis que les opérations de qualification aboutissent soit à des opérations de surlignage

en couleur du texte commenté, soit à la création de liens entre documents visibles par l'ouverture d'une fenêtre des "textes en rapport", avec le document commenté, ou par l'apposition de mots-clés. Chaque intervention sur le texte est enregistrée par la machine, qui peut alors la traiter comme une information, ce qui rend possible par exemple des recherches portant uniquement sur un niveau donné de la hiérarchisation, et plus généralement, la combinaison de très nombreux critères s'appliquant aussi bien au texte lu qu'aux qualifications textuelles engendrées par le lecteur.

Les documents sont organisés dans une hiérarchie. Cette hiérarchie est très semblable à une hiérarchie de fichiers, mais elle est gérée de manière spécifique pour permettre des recherches par mots clés. Ces fonctions supposent une préparation du document, constituée de deux étapes :

- indexation de documents en mode texte. La construction d'un index sur un document en mode texte est indispensable pour pouvoir faire des recherches en texte intégral.

- définition automatique de zones de marqueurs. Lorsqu'un nouveau document est créé, il s'agit d'un fichier texte en code ASCII. L'opération de définition automatique de zones et de marqueurs permet de repérer automatiquement les principales structures du document et de leur associer des zones et des marqueurs, fonctions se rapprochant de la personnalisation et de l'annotation du document.

Le processus d'indexation peut être assez long sur des documents de grande taille. Afin de ne pas bloquer un lecteur pendant le temps de l'indexation, cette fonction est simplement "déclenchée" par l'utilisateur puis elle est réalisée

en tâche de fond, les autres fonctions du poste restant disponibles pendant le temps que dure l'indexation.

L'outil de définition automatique de zones (surligneurs et unité logique de découpage - ULD) et de marqueurs (jalons et annotations) rend possible l'utilisation de l'information éventuellement encodée dans un document pour la convertir dans le formalisme (zones et marqueurs) qui rend cette information directement utilisable par le système de recherche et permet la personnalisation du document. L'interprétation de cette information se fera de la façon suivante :

-Unité Logique de Document et zones surlignées : pour chaque type d'ULD ou de zones, une *expression régulière* spécifiera les débuts et fins des éléments de ce type. Par exemple deux balises du type <citation> et </citation> (proche des normes SGML);

-Marqueurs : pour chaque type de marqueur, une expression régulière spécifiera les points du texte où doivent être posés les jalons de ce type.

Il est également possible d'ignorer certaines informations encodées, c'est-à-dire de spécifier, toujours par le moyen d'expressions régulières, des portions de documents devant être occultées.

Le terme d'expression régulière est emprunté aux programmes sous UNIX : grep, sed, awk, ed, vi et emacs. Une expression régulière est un modèle (un patron, traduction de pattern), devant être confronté à une chaîne de caractères. De nombreux *moteurs de recherches* utilisent la comparaison des chaînes de caractères composant un texte avec une

expression tapée au clavier. Ici une expression régulière, est une phrase d'un langage ayant une syntaxe précise, permettant de décrire de manière rigoureuse les endroits du document où doivent commencer et finir une zone surlignée ou une Unité Logique de Document. Ce langage permet de spécifier des chaînes de caractères à l'aide de balises SGML :

<titre>OEUVRES DE JEAN RACINE</titre>

<titre>(BÉRÉNICE : TRAGÉDIE, 1670)</titre>

<comm>Nouvelle édition revue sur les anciennes impressions et les autographes et augmentée de morceaux inédits, de variantes, de notices, de notes, d'un lexique des mots et locutions remarquables, d'un portrait, d'un fac-similé...

Par M. Paul MESNARD</comm>

<comm>Tome deuxième</comm>

<comm>Paris</comm>

<comm>Librairie Hachette et Cie, 1886</comm>

<scene>ACTE I</scene>

<scene>SCENE PREMIERE</scene>

<acteur>ANTIOCHUS</acteur>

<vers>Arrêtons un moment, la pompe de ces lieux</vers>

<vers>Je le vois bien , Arsace, est nouvelle à tes yeux. </vers>

<vers>Souvent ce cabinet superbe et solitaire</vers>

<vers>Des secrets de Titus est le dépositaire</vers>

<num>5</num>

<vers>C'est ici quelquefois qu'il cache à sa cour,</vers>

<vers>Lorsqu'il vient à la reine expliquer son amour</vers>

<vers>de son appartement cette porte est prochaine,</vers>

<vers>Et cette autre conduit dans celui de la reine.</vers>

<vers>Va chez elle : dis-lui qu'importun à regret</vers>

<num>10</num>

<vers>J'ose lui demander un entretien secret.</vers>

A partir de ce document, on peut vouloir marquer les Unités Logiques de Découpage suivantes :

-Les éléments de titre ;

-Les commentaires ;

-Les débuts de scène ;

-Les noms des acteurs ;

-Chaque vers ;

-Les numérotations.

On souhaite également faire disparaître lors de l'affichage à l'écran, les balises repérant ces éléments. Pour faire cela, il faut avoir défini des types d'ULD correspondant à chacun de ces cas. Par exemple, on prendra les noms suivants : Titre, Sous-titre, Scène, Acteur, Numéro, Vers.

Repérage du titre :

début : Après "<titre>"

fin : Avant "</titre>"

type : Titre

Repérage du commentaire :

début : Après "<comm>"

fin : Avant "</comm>"

type : Sous-Titre

Repérage des vers :

début : Après "<vers>"

fin : Avant "</vers>"

type : Vers

Suppression des balises :

début : Avant "<[/a-zA-Z]"

fin : Après “[a-zA-Z]>”

type : Cache

Suivant le principe des Unité Logiques de Découpage, une définition des types est nécessaire : l’outil permettant de définir, de modifier ou d’utiliser les différents types est accessible depuis l’outil de sélection du menu (icône : *préparation*). Il offre les fonctions suivantes :

-Visualiser les types d’ULD, d’annotation, de jalons, de liens ;

-Modifier la présentation d’un type existant ;

-Détruire un type ;

-Ajouter un nouveau type ;

Structure d’un type :

Type d’ULD :

Un type d’ULD associe un nom à une présentation d’un fragment de document. Ce document peut être un document en mode texte ou image. La présentation d’un type d’ULD doit donc prendre en considération ces deux cas.

Pour les documents en mode texte, la présentation joue sur la typographie du texte. Pour les documents en mode image, la présentation joue sur les couleurs du rectangle délimitant la zone, ainsi que sur le cadre du rectangle.

Type d’annotations :

La présentation d’un type d’annotation comprend la présentation de l’icône marquant le point d’ancrage de l’annotation dans le document ainsi que la présentation du texte de l’annotation. Celle-ci comprend les couleurs d’avant plan et d’arrière plan de la fenêtre d’édition.

Type d’illustrations :

Un seul outil d'illustration sera considéré. La présentation associée sera celle de l'icone marquant le point d'ancrage dans le document.

Type de jalons :

La présentation associée à un type de jalon correspond à la présentation de l'icone marquant l'ancrage.

Type de liens :

La présentation associée à un type de lien correspond à la présentation de l'icone marquant l'ancrage.

2.4.2. Critique de cette implémentation :

a) difficultés structurelles :

la difficulté à appréhender la logique d'utilisation de la maquette par les lecteurs tient à son ouverture : AIS a adopté pour principe que l'utilisateur devait pouvoir définir lui-même les caractéristiques de ses types d'annotations, jalons, résumés sous le terme d'unités logiques de découpage. Une telle ouverture est un très grand avantage, qui n'est pratiquement jamais offert dans le domaine de la micro-informatique (si ce n'est de manière très limitée, comme par exemple Word de Microsoft, qui permet de nommer et de définir des styles tels que : citation, note de bas de page).

Mais cet avantage est aussi un grave inconvénient si l'utilisateur n'est pas clairement averti :

-de la liberté d'intervention qu'une telle conception concrétise;

-des grandes classes de possibilités qui s'offrent à lui.

Ainsi le lecteur/scripteur peut ne pas comprendre l'avantage qu'il y a à pouvoir attacher des mots clés à une Unité Logique

de Découpage. Il définira par exemple une Unité Logique de Découpage *mot-clé* sur laquelle il fera ses recherches. Au contraire attacher des mots-clés à une unité de sens déjà constituée comme Unité Logique de Découpage, peut être plus efficace que de définir une ULD spécifiquement pour saisir des mots-clés. Il s'agit pour l'utilisateur de créer une adéquation entre sa manière personnelle d'utiliser des métaconnaissances (connaissances permettant d'utiliser les connaissances), et ce qui est offert par le système très modulable. Ce genre de problématique n'est pas particulier au PLAO. Avec un logiciel très riche de fonctionnalités, comme par exemple Word, l'utilisateur est libre de créer ou non des macro-commandes. Toutefois la complexité d'un logiciel comme Word est limitée car il s'agit avant tout d'un système de traitement de texte. Avec le PLAO, la difficulté vient des nombreux sens attachés à l'acte d'annotation : appropriation du texte par le lecteur qui met en rapport ses préoccupations avec un contenu. L'historien apporte une annotation critique d'une source d'information, mise en confrontation avec d'autres sources (d'autres témoignages, des éléments statistiques, climatologiques, chronologiques, archéologiques, etc.), le sémiologue cherche des aspects signifiants dans la structuration des signes, le littéraire fera appel à l'intertextualité, le philosophe à une armature de concepts. La notion d'Unité Logique de Découpage n'est pas univoque, mais liée à des significations : on ne peut utiliser un algorithme unique, mais plutôt un parcours heuristique. Une solution proposée est de fournir à l'utilisateur un grand nombre d'exemples formalisant les grandes classes de procédures

d'approche des textes, qui soient autant d'appropriations des éléments combinables offerts par l'interface, en fonction des grandes tâches qui caractérisent les modalités de scrutation des textes par lecture et écriture, suivant les grandes disciplines de la pensée. Seules les plus génériques de ces formalisations peuvent être exposées sous forme d'exemples par :

-un fonds de chaînes opératoires commun à toutes les disciplines lectrices ;

-un fonds de chaînes opératoires commun à chaque discipline. Dans chaque discipline, des modalités spécifiques de chaînes opératoires, ainsi que des chaînes opératoires qui relèvent des idiosyncrasies de chaque lecteur en tant que grand lecteur.

Chacun des lecteurs à diverses manières d'aborder un texte. Cela tient à plusieurs éléments, dont en particulier :

- son corpus (la nature de ce corpus, son accessibilité) ;
- sa formation, la discipline dans laquelle il exerce son savoir;
- ses propres concepts.

3. Préfiguration des 3000 documents

liaison Postes de Lecture Assistée par Ordinateur et réseaux internationaux

Le système d'information actuellement en cours de réalisation ne pourra être opérationnel pour l'ouverture du haut de jardin en décembre 1996. En revanche, le fonds des 100 000 documents sera numérisé dans sa totalité. C'est pourquoi, afin de faire bénéficier le public dès

l'ouverture d'un accès partiel à cette collection, il a été prévu lors de la réunion du Comité de pilotage du 5 avril 1996, d'étudier et de proposer toute solution informatique permettant de mettre en valeur cette collection unique au monde en la rendant accessible. La Mission de Numérisation scientifique et la Mission de Numérisation technique assurent conjointement la mise en oeuvre de cette opération pour ce qui relève de leurs compétences respectives :

- sélection des documents et définition des modes de recherche pour la numérisation scientifique ;

- maîtrise d'oeuvre technique et informatique (matériel et logiciel) de l'opération pour la numérisation technique.

Le rôle de ce système est double :

- offrir une partie de la collection d'imprimés numérisés et ainsi valoriser dès 1997 cette collection ;

- Expérimenter à cette échelle réduite les demandes du public en matière d'accès aux documents numérisés, afin d'intégrer ces paramètres dans la suite de la réflexion.

Sur le premier point, la consultation de la collection :

- un choix de 3000 ouvrages a été arrêté par la Mission de Numérisation scientifique. Ce sont des monographies et des périodiques libres de droits appartenant au fonds français du XIX ème siècle (1830-1900). Ces documents seront disponibles en mode image.

Ce corpus d'ouvrages en mode image devrait être complété par une sélection de documents en mode texte émanant de la base Frantext. La présence de ce corpus dans le fonds est liée à la bonne exécution du marché de numérisation du

Trésor de la Langue Française, prévue dans la convention de 1993. Cette opération est en cours de notification chez le contrôleur financier (DAF, Direction des Affaires Financières).

Identification de la collection :

Des postes de consultation et de travail seront mis à disposition du public dans un espace de la bibliothèque. Ces postes connectés en réseau à un serveur de données permettront au public :

- d'interroger la base de données des ouvrages numérisés ;
- de sélectionner des ouvrages par affichage de la table des matières, et demander le chargement sur le poste de travail ;
- d'obtenir des impressions (par un système centralisé) ;
- d'effectuer des enregistrements de documents de travail sur des supports magnétiques personnels (disquettes).

L'accès par le réseau Internet, via Renater, permettra d'établir des liaisons avec la Bibliothèque municipale de Lyon et la salle de lecture de la rue de Richelieu.

Contour technique de l'opération :

Le système intermédiaire de consultation du fonds numérisé sera composé de dix postes de travail reliés à un serveur donnant accès à trois mille documents représentant trois cents pages en moyenne (60 gigaoctets). Il est à noter que les documents en mode image sont au format TIF (Tagged Image File Format), format de fichier utilisé pour décrire et stocker des images numériques en couleurs. Les problèmes liés à ce format proviennent de sa richesse. Bien que les différentes options soient

documentées, certaines applications ne traitent pas toutes ces options. Sur Internet, le format adopté (HTML, browsers), est GIF. Un format pivot de type Acrobat pourrait être utilisé.

Fonctionnalités du système :

La base de données bibliographique sera interrogeable par recherche multi-critères. Les utilisateurs pourront sélectionner les documents dans la base de données. Le système devrait pouvoir afficher la table des matières des ouvrages concernés, pour permettre au lecteur de préciser ses critères de recherche et de demander le téléchargement sur son poste de travail des ouvrages ou parties d'ouvrages qui lui semblent pertinents. Une partie des tables pourra être en mode texte si l'avancement des travaux de saisie le permet, la recherche de présélection pouvant être alors affinée.

Le développement d' Internet (Inter-networks), interconnexion de nombreux réseaux adoptant le protocole de communication TCP/IP développé initialement pour les échanges entre scientifiques augmente les possibilités d'accès depuis un Poste de Lecture Assistée par Ordinateur. Sur Internet , on peut actuellement interroger plus de 300 catalogues de bibliothèques répartis sur différents continents. Par exemple, pour interroger le catalogue de la bibliothèque du Congrès à Washington, il suffit de taper l'adresse de cette bibliothèque dans le réseau (telnet dracon) sur un poste connecté à Internet pour se retrouver en liaison avec le système d'interrogation propre à cette bibliothèque. Il existe différents systèmes gérant ces

catalogues sur le réseau et présentant chacun des interfaces différentes.

La conservation et l'édition électronique de grands corpus littéraires, et le développement de logiciels de lecture et écriture assistées par ordinateur modifient les rapports du lecteur à de grands volumes d'informations (par exemple, des collections patrimoniales) ce qui entraîne une profonde mutation des pratiques d'utilisation des connaissances : le réseau informatique est conçu comme un immense document réparti entre plusieurs sites ce qui a été rendu possible par la normalisation portant sur quatre éléments :

-le langage HyTime Hypermedia/Time based Structuring Language c'est-à-dire Langage de Structuration Temporel/Hypermédia. HyTime définit un langage et un modèle associés pour une représentation des hyperdocuments qui établissent des liaisons et synchronisent l'information statique et dynamique contenue dans un document multimédia. (Ce dernier langage encore peu utilisé a été retenu par le Ministère de l'Enseignement supérieur et de la recherche pour la constitution de la base *Griseli* issue d'un programme de collecte traitement et communication de la littérature grise utilisant la norme SGML et stockée sur un serveur WWW World Wide WEB)

- une manière de nommer les ressources informatiques de façon à les rendre accessibles par divers outils de navigation : les URL (*Uniform Resource Locator*), désignation des ressources du réseau WWW.

- le protocole http (*hypertext transfer protocol*), qui permet d'exploiter l'infrastructure du réseau à partir d'une généralisation du modèle client-serveur qui englobe tous les protocoles de transmission des données;
- le langage de balisage HTML (*Hypertext Markup Language*), qui permet d'inscrire au sein même d'un document électronique des "ancres" permettant de naviguer aisément entre diverses ressources du réseau.

Cette conception du document électronique est d'emblée fondée sur la métaphore de la bibliothèque. Le modèle utilisé par les concepteurs de ces documents n'est plus la forme *codex*, qui est la base du livre moderne (cf. Roger Chartier : "*Du codex à l'écran : les trajectoires de l'écrit*"), mais directement le concept d'une immense bibliothèque dans laquelle chaque bloc ("bloc d'information") serait offert en accès libre. Un rapprochement a été fait entre les théories du texte des critiques contemporains : Barthes avec les lexèmes décrits dans *S/Z*, Jacques Derrida dans *Glas* texte en deux colonnes, celle de gauche consacrée à Hegel et celle de droite à Genet. Chacune est découpée par des blancs, des incisions des "judas", faits de citations, de remarques ... Sans cesse on passe d'une colonne à l'autre, de Hegel à Genet, de Genet à Hegel, passage rythmé par le repérage de motifs communs. La fin du texte "La pharmacie de Platon" qui traite du motif de *khôra* "grosse de tout ce qui se dissémine ici (cf pp. 185-186) est reliée au texte intitulé "Chôra" (in *Poikila*, Etudes offertes à Jean-Pierre Vernant , éd. l'Ecole des Hautes Etudes en Sciences Sociales, Paris, 1987 pp. 265-296), montre

comment une question en germe dans un texte est développée par Derrida, se développe dans un autre . Ce système de lien et de réseau autour d'un motif laissé en suspens peut être relié à ce qu'écrit Michel Foucault dans l'archéologie du savoir : les frontières d'un livre ne sont jamais clairement tranchées, parce qu'il appelle "un système de références à d'autres livres, d'autres textes, phrases, c'est un noeud à l'intérieur d'un réseau." On peut rapprocher ce point de vue de ce que nous dit Foucault du kantisme : il introduit une dissonance dans l'espace du savoir. Les représentations ne sont plus rapportées à elles seules. L'ordre mathématique et ordonnable des représentations est confronté à l'émergence du transcendantal recherché soit du côté de la pensée subjective, soit du côté des objets transcendants que sont le travail, la vie, le langage. Ces deux nouvelles formes de pensée trouvent leur point de jonction dans la découverte kantienne du champ transcendantal. La représentation ne peut plus classer à partir d'elle-même ses différents éléments. Elle doit désormais se fonder sur un sujet transcendantal. Les représentations ne sont plus rapportées à elles seules mais à ce qui les rend possibles:" Kant contourne la représentation et ce qui est donné en elle, pour s'adresser à cela même à partir de quoi toute représentation peut être donnée". Le retrait du savoir et de la pensée de l'espace de la représentation sanctionne une nouvelle forme de culture européenne, un nouvel âge du savoir. (*Michel Foucault : Les mots et les choses* pp. 254-257).

3.1. Vers un nouvel encyclopédisme ?

L'hypertexte ne constitue pas une rupture épistémologique, mais plutôt une concrétisation des recherches de l'histoire de la culture occidentale : l'hypertexte et les réseaux rendent visibles, opérationnelles des notions philosophiques et conceptuelles qui sous-tendent comme un *hypotexte*, la culture humaine depuis Platon et Aristote jusqu' aux stratégies des *Sciences de l'artificiel*.

Il s'agit d'un encyclopédisme virtuel, au sens du latin médiéval *virtualis*, lui-même issu de *virtus*, force, puissance. Dans la philosophie scolastique, est virtuel ce qui existe en puissance et non en acte. Le virtuel tend à s'actualiser, sans être passé à la concrétisation effective ou formelle. L'arbre est virtuellement présent dans la graine, comme l'encyclopédisme est virtuellement présent dans l'organisation systématique et la mise à disposition du savoir. Le problème est alors : comment rendre ce savoir visible au-delà de l'espace réduit de l'écran, comment organiser les parcours de lecture ? Le virtuel ne s'oppose pas au réel mais à l'actuel : virtualité et actualité sont deux manières d'être différentes, de même que la meilleure forme n'est pas la forme stable, comme le pensait la *Gestalttheorie*, mais celle qui informe de manière significative en promouvant des transductions ultérieures, à partir de son palier ou de sa phase d'information. Les collections numérisées, le balisage des textes et les Postes de Lecture Assistée par Ordinateur sont des éléments de transduction : à l'inverse de la déduction, la transduction ne va pas chercher ailleurs un principe pour résoudre le problème d'un domaine, elle tire la

structure résolutive des tensions mêmes de ce domaine, comme la solution sursaturée se cristallise grâce à ses propres potentiels et selon l'espèce chimique qu'elle renferme. Le réseau d'un hypertexte est sans cesse en construction et en renégociation. Il peut rester stable un certain temps, mais cette apparente stabilité est un système métastable, comparable à la plasticité du cerveau qui fabrique en permanence, des liens et des associations, par stimulation de groupes de neurones, retrouvant ainsi des connaissances, ou créant de nouvelles inférences. Dans une encyclopédie en hypertexte, la connaissance devient à son tour une potentialité, où l'utilisateur peut dessiner un réseau métastable en procès d'individuation, provoquant l'étincelle du sens. Cette problématique de recherche de mise à disposition des corpus sous une nouvelle forme était déjà présente dans les esprits depuis la fin du XIX^{ème} siècle: c'est le sens des travaux sur le microfilm comme document de substitution et d'étude.

Une nouvelle bibliographie dynamique : vers un réseau sémantique du document ?

Notre problématique concerne les rapports entre les formes du support du document, les méthodes de travail induites par ses formes, et les problèmes de propriété intellectuelle.

Sur le premier point : la lecture comme l'écriture a toujours nécessité un support technique : A Sumer, terre limoneuse inondée de Basse-Mésopotamie, le support était un pain d'argile fraîche, façonné en tablettes dans lesquelles le calame qui servait d'instrument imprimait des droites, et, par

pression de l'amorce, des têtes de clous en figure de coins, d'où le nom d'écriture cunéiforme. Cette technique devait très vite, par stylisation croissante, abolir toute ressemblance avec les objets que l'on avait d'abord, au stade pictographique initial simplement représentés. Elle a donc connu le passage par les deux étapes classiques du pictogramme ou dessin de la chose puis de l'idéogramme ou schéma de l'idée à laquelle correspond un mot de la langue. Pour le philosophe Bernard Stiegler le texte "doit être conçu avant tout partage hylémorphique, dans la considération de son support matériel autant que de sa forme, étant donné le rôle du support dans sa lecture aussi bien que dans son écriture : la relation entre le texte et son lecteur (et son auteur comme premier lecteur) est transductive, au sens que Simondon (1) a donné à ce terme dans l'analyse des processus d'individuation par où il dépasse l'opposition de la forme et de la matière".

Les formes matérielles : tiges de jonc mâchonnées par le scribe en pinceau, encre au noir de fumée, papyrus, écritoires, plumes, crayons, règles, besicles, sélecteurs de microfilms, ordinateurs, constituent l'*hylê*(*matière*), et la forme (*morphê* figure disposition) mais aussi les voies par lesquelles on accède au texte. La numérisation de corps de textes, les écrans à tubes cathodiques et affichage alphanumérique (grille de vingt-quatre lignes sur quatre-vingts colonnes), les périphériques de pointage permettant le dialogue avec l'utilisateur, la possibilité d'annoter un texte et de garder en

(1) Gilbert Simondon, Du mode d'existence des objets techniques. Paris : Aubier, Res l'invention philosophique , 1989

mémoire son parcours de recherche, ont modifié les conditions d'accès au texte.

Le second point concerne le prolongement de l'écriture par d'autres moyens :

confrontée à la *finitude rétentionnelle* de la mémoire humaine (Bernard Stiegler) l'écriture fut cette première condition de possibilité d'une communication hors du présent sans laquelle aucune réitération ne serait possible.

L'écriture sur un support électronique devient une mémoire collective. L'écriture était une virtualisation de la mémoire (Derrida), constituant des objets, abstraits, actualisés en multiples versions, traductions, éditions, exemplaires, copies.

La bibliographie recense et décrit ces objets textuels, et par son support informatique devient une mémoire dynamique, en réseau. En interprétant les éléments tirés d'un corpus, le lecteur crée des parcours dans l'univers des connaissances, qu'il filtre suivant des thèmes qui lui sont propres : l'étincelle du sens lui appartient en propre. Alors intervient le travail d'annotation, d'écriture, qui est à son tour sauvegardé au moyen d'une mémoire artificielle. La mémoire immédiate et la mémoire de masse sont les deux éléments qui coopèrent.

La notion de réseau sémantique pourrait englober la description bibliographique du document (catalogage en InterMarc ou Unimarc) et le balisage de la structure (Description de Type de Document, de type SGML, HTML, permettant de donner au texte intégral une structure régulière pour le communiquer en réseau et le mettre en mode hypertexte).

Elle correspond à un encyclopédisme de corrélation, de mise en relation des contenus. On sait que Linné dans sa classification des espèces a procédé par corrélation après avoir rapproché les caractères semblables, ou noté ce qui différenciait les créatures vivantes. Au XVIII^{ème} siècle, la taxinomie a remplacé l'hitoire naturelle descriptive : elle la remplace par la science des *implications-exclusions*. L'essor de la taxinomie se place après l'ère des accumulations ou des collections, quand le nombre des unités devint tel qu'il exigeait un ordre et qu'on ne pouvait plus se contenter d'un simple empilement, ou alignement des connaissances.

La bibliographie dynamique constitue son réseau sémantique en englobant les attributs externes du document (ce qui appartient à la *description normalisée* norme Z 44-50, format InterMarc/Unimarc), le contenu du texte digitalisé dont les parties sont balisées (SGML), reliées par des liens hypertextes (HTML), communiquées par le réseau (normes TCP/IP et normes Z 39-50). On peut alors envisager un encyclopédisme reposant sur la modélisation, la corrélation, la simulation. La modélisation sur ordinateur est une rupture parce qu'elle permet de représenter ce qui ne l'a jamais été. Le signe doit tenir lieu de la chose en l'absence de celle-ci.

Alors que la bibliographie était un art de la mémoire, elle devient un nouvel encyclopédisme reposant sur la corrélation, la statistique et la modélisation.

L'aspect fondamental de ce questionnement a été analysé par Monsieur Yannick Maignien :

“....pour répondre à cette interrogation, posons-nous la question la plus originale qui soit : que veut dire chercher, “retrouver” un ouvrage ? Question encyclopédique s’il en est. Retrouver ce que l’on connaît déjà ? ou au contraire, ou par ailleurs, trouver, sauver de l’oubli, sans le connaître, un document dont on a besoin dans le cadre d’une recherche ? C’est bien sûr ce second cas de figure qui est intéressant : l’instruction de la stratégie de recherche heuristique est l’une des questions majeures du projet de numérisation de la B.n.F. “Savoir ce que l’on cherche, c’est comprendre ce que l’on trouve”, disait Bachelard, mais on sait aussi, depuis, que ce savoir préalable est un horizon de complexité, un spectre de diversité et de sources multi-disciplinaires, et que le résultat de la requête peut être aléatoire, non déterministe, et pourtant, compréhensible. L’adage peut se renverser : comprendre ce que l’on trouve, c’est savoir ce que l’on cherche. Croisement entre un projet : rassembler le plus possible des savoirs et des contraintes techniques : sélection, analyse et indexation, classement, conservation, communication, la bibliothèque est sur le point de réaliser un rêve ancien, l’accès universel au document”(1).

(1) Yannick Maignien, La Bibliothèque virtuelle ou de l’ars memoria à Xanadu , Bulletin des Bibliothèques de France, Paris t.40, n°2, 1995.

Conclusion :

En interrogeant les sources bibliographiques à visée encyclopédique, les nouvelles technologies, posent le problème essentiel de la critique des textes et des sources d'information. Viennent ensuite les questions relatives à la structuration des documents et leur utilisation sur des stations de lecture/écriture. Une puissance inégalée de repérage, de rassemblement, de mémorisation, laisse entrevoir les possibilités d'un encyclopédisme de corrélation qui n'est pas sans rappeler les oeuvres de Linné ou de Paul Otlet (1). La délocalisation du savoir, en multipliant les points de vue possibles, en permettant à l'observateur (le lecteur/scripteur) de réagir instantanément, et de mettre en doute les informations proposées, conduit à l'idée de la nécessité d'une nouvelle bibliographie, reprenant les méthodes scrupuleuses de l'ancienne critique des sources, constituant des collections à partir de documents à l'authenticité prouvée, et rejoignant la méthode philologique de la critique génétique.

Lors de l'ouverture du *haut de jardin* (espace de travail destiné à un public large, avec des collections en libre-accès), une base documentaire de 3 000 documents numérisés, de 300 pages en moyenne, accessible sur 20 postes de travail de type PC pourra être consultée sur le site de Tolbiac. Une liaison est également prévue depuis Lyon et le site de la rue de Richelieu, via le réseau Renater.

(1) Paul Otlet (1868-1944), réunit à Bruxelles en septembre 1895, avec Oscar Lafontaine la première conférence internationale de bibliographie. Au cours de celle-ci furent adoptées l'élaboration d'un répertoire bibliographique universelle et la création d'un institut international de bibliographie..cf. : *Cent ans de l'office international de bibliographie*, Éditions du Mundadeum, Mons, 1995 et *Traité de documentation*, Le Livre sur le livre, par Paul Otlet, Liège, Centre de Lecture Publique de la Communauté française de Belgique, 1989.

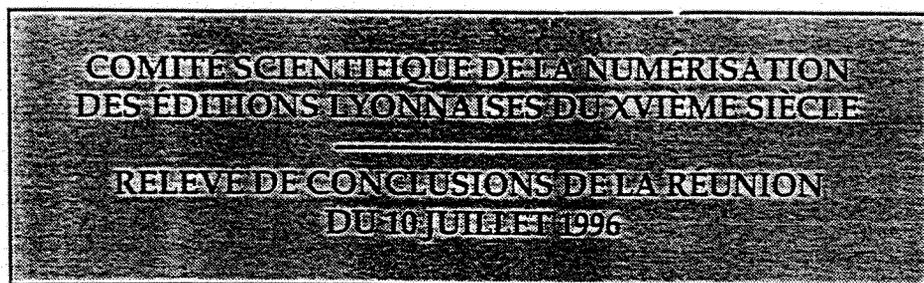
Cette première réalisation donnera corps aux intuitions de Vannevar Bush ⁽¹⁾ et Theodor Nelson ⁽²⁾, concrétisées par la croissance du réseau Internet. L'étude de la structuration des documents et de la constitution des collections numérisées utilisées au moyen du PLAO, bénéficient des recherches préalables menées sous l'impulsion de la Bibliothèque nationale de France sur les pratiques de lecture/écriture de chercheurs en lettres et sciences humaines, qui ont élargi ce champ de connaissances jusqu'alors limité aux usages professionnels du traitement de texte. Les travaux de constitution de la collection numérisée ont posé la question des choix de textes et de leur structuration matérielle. Avec la mise en réseau des collections lyonnaises du XVI^{ème} siècle et des 3 000 documents, les questions concernant les usages et la visibilité des collections trouveront une première réponse.

(1) Vannevar Bush (1890-1974) conseiller scientifique du président Roosevelt pour les questions scientifiques. Son article, *As we may think*, traduit en français par *Suivant notre façon de penser* publié en 1945, est considéré comme la première approche contemporaine de l'idée d'hypertexte..

(2) créateur du terme hypertexte en 1965

ANNEXE 1

11/07/96
GP/MCF/101



Ont participé à la réunion du Comité Scientifique de la numérisation des Editions Lyonnaises du XVIème siècle, le 10 Juillet à Tolbiac, les personnes suivantes :

♦ Personnalités extérieures :

- Monsieur Etienne BRUNET, Professeur à l'Université de Nice-Sophia Antipolis,
- Monsieur Francis GOYET, Professeur à l'Université Stendhal de Grenoble,
- Monsieur Daniel MENAGER, Professeur à l'Université Paris X - Nanterre.

♦ Bibliothèque municipale de Lyon :

- Monsieur Pierre GUINARD, Conservateur, secteur fonds ancien.

♦ Direction du Livre et de la Lecture :

- Monsieur Jérôme BELMON,
- Monsieur Dominique COQ.

♦ Bibliothèque nationale de France :

- Monsieur Gérard CATHALY, DINT, Mission technique de numérisation,
- Madame Geneviève GUILLEMINOT, DIA, Département de la Réserve,
- Monsieur Yannick MAIGNIEN, DDSR, Chef de la mission scientifique de numérisation,
- Monsieur Georges PERRIN, DDSR, Chef du secteur des pôles associés,
- Monsieur Eric VARON, Stagiaire ENSSIB auprès de la mission scientifique de numérisation,
- Monsieur Jean-Didier WAGNEUR, Mission scientifique de numérisation.

♦ Étaient excusés :

- Madame Marie-Luce DEMONET, Professeur à l'Université Blaise Pascal de Clermont-Ferrand,
- Monsieur Guy PARGUEZ, Bibliothèque municipale de Lyon, Conservateur, Responsable du fonds ancien,
- Monsieur Antoine CORON, Bibliothèque nationale de France, DIA, Chef du Département de la Réserve,
- Madame Valérie TESNIERE, Bibliothèque nationale de France, DIA, Chargée de mission pour la constitution des collections thématiques,
- Madame Simone BLANC, Bibliothèque nationale de France, DDSR, Chef de la mission pour la politique documentaire.

La réunion a débuté par une présentation détaillée du programme de numérisation de 100 000 documents conduit au sein de la Bibliothèque nationale de France par la Mission scientifique de numérisation et la Mission technique de numérisation.

Monsieur Yannick MAIGNIEN indique que, pour 80 000 documents, la numérisation est achevée.

De l'exposé de Monsieur Gérard CATHALY sur la filière technique (microformage en 35 mm. puis numérisation en mode image), il ressort que le calendrier prévisionnel pour la numérisation des ouvrages du fonds de la Bibliothèque municipale de Lyon est très serré. En effet, si l'on veut pouvoir alimenter les chaînes de numérisation qui fonctionnent dans le cadre des marchés de la Bibliothèque nationale de France, il convient de pouvoir fournir des microformes dès la fin du mois de Septembre.

Monsieur Pierre GUINARD présente ensuite la liste des ouvrages lyonnais sélectionnés par Monsieur Guy PARGUEZ. Cette sélection a été faite à partir d'une première liste de 2 500 titres. Elle a été constituée en évitant de doubler par rapport à la liste que lui avait transmise la Bibliothèque nationale de France. Il a été tenu compte d'une certaine variété thématique et éditoriale. Finalement, 350 ouvrages sont ainsi proposés, qui représentent environ 100 000 pages à numériser.

Monsieur Pierre GUINARD souhaite que soit prise également en considération la numérisation des "recueils verts" de la Bibliothèque municipale de Lyon. Il s'agit de 84 brochures à caractère historique imprimés entre 1560 et 1620. L'ensemble représente 42 000 pages à numériser.

Compte tenu de la richesse et de la variété de cette liste et de l'intérêt historique des recueils verts, l'accord est unanime sur l'intérêt de numériser l'ensemble de ces ouvrages, sous réserve que l'état physique des volumes permette leur microfilmage.

Monsieur Jean-Didier WAGNEUR présente la liste de la Bibliothèque nationale de France, issue de deux sources : les microformes de la Réserve et la collection d'IDC composée de microformes d'impressions françaises et italiennes.

Madame Geneviève GUILLEMINOT fait remarquer que le lot des microformes de la Réserve n'a pas été constitué selon un principe scientifique. Il comprend des microformes réalisées à l'issue de la dernière guerre (1945-1946) à titre de sauvegarde, des microformes réalisées à la demande des lecteurs et des produits du microfilmage qui est systématique avant chaque envoi à la reliure. Elle ajoute que, compte tenu des moyens de microformage dont dispose actuellement la Bibliothèque nationale de France, un certain nombre d'ouvrages (ceux qui ne peuvent être microfilmés à plat) sont radicalement exclus de ce type d'opération.

Une discussion s'engage sur la nécessité de dégager un principe de sélection plus scientifique, de diminuer la part des éditions en latin, d'essayer de déterminer un minimum d'axes thématiques, tout en ayant recours à un ensemble d'ouvrages plus large que ceux de la seule Réserve. Il conviendrait notamment d'introduire davantage d'ouvrages de droit ou de médecine, représentatifs d'une large part de l'Édition Lyonnaise au XVIème siècle. Il a été souhaité que les incunables ne soient pas exclus de cette sélection.

Monsieur Jérôme BELMON fait part du souhait de la Direction du Livre que cette opération reste centrée sur l'Édition Lyonnaise au XVIème siècle.

A une question de Monsieur Dominique COQ, sur la dimension de cette opération, il est répondu que cette première étape ne constitue que l'amorce de la constitution d'un corpus beaucoup plus vaste donc plus significatif.

Monsieur Francis GOYET souhaite que la totalité des volumes numérisés apparaissant sur la liste de la Bibliothèque nationale de France soient disponibles pour les chercheurs, quitte à centrer les instruments de navigation (index) sur les Éditions Lyonnaises complémentaires de celles de la Bibliothèque municipale de Lyon.

Il est convenu que la Mission scientifique de numérisation complétera et affinera cette sélection en liaison avec les départements concernés de la DIA et à partir, notamment, du travail en cours fait par Madame GUILLEMINOT, et proposera une nouvelle liste au Comité Scientifique au cours de sa prochaine réunion.

Le point suivant de la discussion porte sur le choix des outils bibliographiques complémentaires à numériser en mode texte et susceptibles de compléter le corpus. Ces informations alphanumériques (texte) sont indispensables pour une utilisation en réseau des documents.

Monsieur Francis GOYET souhaite que soient numérisés en mode texte tous les index et tables des matières des ouvrages concernés.

Monsieur Daniel MENAGER souhaite, pour sa part, qu'on y adjoigne des dictionnaires de mythologie, et qu'on sollicite, par écrit, les conseils d'universitaires, de Monsieur Jean CEARD, Professeur à l'Université Paris X - Nanterre, pour la Médecine et de Monsieur Michel REULOS, Conseiller d'Etat, spécialiste de l'Histoire du Droit, pour le Droit.

Toutes les personnes présentes souhaitent que le BRUNET soit numérisé. Il est demandé à la Mission scientifique de numérisation d'étudier la faisabilité d'une numérisation du BAUDRIER (est-il libre de droits ?).

La Mission scientifique de numérisation souhaite numériser en mode texte une trentaine d'ouvrages. Un repérage sera fait sur le WEB pour ne pas refaire ce qui existe déjà dans ce domaine, notamment les ouvrages de référence concernant cette période et déjà disponibles sur des serveurs d'universités américaines.

Au cours de la prochaine réunion du Comité Scientifique, un point sera fait sur l'état d'avancement de l'étape de microformage des ouvrages de la Bibliothèque municipale de Lyon. A cet effet, un cahier des charges, élaboré par la Mission technique de numérisation, a été remis à Monsieur Pierre GUNARD. Ce dernier doit consulter les prestataires spécialisés et soumettre son choix à la Bibliothèque nationale de France dans un délai d'une quinzaine de jours. D'ici là, la Mission scientifique de numérisation aura élaboré une nouvelle liste de propositions d'Éditions Lyonnaises du XVIème siècle, aura fait le point sur la possibilité de numériser le BRUNET et le BAUDRIER et contacté par courrier Messieurs CEARD et REULOS.

La date de la prochaine réunion a été fixée au :

Vendredi 20 Septembre à 10 heures

Elle aura lieu dans la même salle (Tour T2 - Bureau n° 03.71).

23/09/96
GP/MCF/116

**COMITÉ SCIENTIFIQUE DE LA NUMÉRISATION
DES ÉDITIONS LYONNAISES DU XVIÈME SIÈCLE**

**RELEVÉ DE CONCLUSIONS DE LA RÉUNION
DU 20 SEPTEMBRE 1996**

Ont participé à la réunion du Comité scientifique de la numérisation des Editions lyonnaises du XVIème siècle, le 20 Septembre à Tolbiac, les personnes suivantes :

♦ Personnalités extérieures :

- Madame Marie-Luce DEMONET, Professeur à l'Université Blaise Pascal de Clermont-Ferrand,
- Monsieur Daniel MENAGER, Professeur à l'Université Paris X - Nanterre.

♦ Bibliothèque municipale de Lyon :

- Monsieur Pierre GUNARD, Conservateur, secteur fonds ancien.

♦ Direction du Livre et de la Lecture :

- Monsieur Dominique COQ,
- Monsieur Emmanuel AZIZA.

♦ Bibliothèque nationale de France :

- Madame Simone BLANC, DDSR, Chef de la Mission pour la politique documentaire,
- Monsieur Gérard CATHALY, DINT, Mission technique de numérisation,
- Madame Geneviève GUILLEMINOT, DIA, Département de la Réserve,
- Monsieur Georges FERRIN, DDSR, Chef du secteur des pôles associés,
- Monsieur Eric VARON, Stagiaire ENSSIB auprès de la mission scientifique de numérisation,
- Monsieur Jean-Didier WAGNEUR, Mission scientifique de numérisation.

♦ Etaient excusés :

- Monsieur Etienne BRUNET, Professeur à l'Université de Nice-Sophia Antipolis,
- Monsieur Francis GOYET, Professeur à l'Université Stendhal de Grenoble,
- Monsieur Guy PARGUEZ, Bibliothèque municipale de Lyon, Conservateur, Responsable du fonds ancien,
- Monsieur Antoine CORON, Bibliothèque nationale de France, DIA, Chef du Département de la Réserve,
- Madame Valérie TESNIERE, Bibliothèque nationale de France, DIA, Chef du département Philosophie, Histoire, Sciences de l'homme.

La réunion a débuté par un point sur l'état d'avancement des opérations de microfilmage des ouvrages de la Bibliothèque municipale de Lyon.

Monsieur Pierre GUNARD annonce que la société "Flash Copie", pressentie d'un commun accord par la Bibliothèque nationale de France et la Bibliothèque municipale de Lyon, a installé ses caméras dans les locaux de la Bibliothèque municipale de Lyon et a commencé ses travaux. L'opérateur de "Flash Copie" prévoit que le travail de microfilmage pourrait être terminé en deux mois.

Monsieur Gérard CATHALY demande que le produit des premiers travaux lui soit envoyé, afin qu'il puisse en vérifier la qualité technique avant de donner le feu vert de la Bibliothèque nationale de France à la poursuite du microfilmage. La Bibliothèque municipale de Lyon se chargera du contrôle d'exhaustivité des prises de vue, et, par la suite, du contrôle continu de la qualité technique du travail effectué.

Monsieur Pierre GUNARD s'engage à transmettre à Monsieur Gérard CATHALY, le premier échantillon de microfilm réalisé dès le début de la semaine prochaine.

Le Comité scientifique entend ensuite un exposé de Monsieur Eric VARON sur la faisabilité de la numérisation en mode texte du "Manuel du libraire et de l'amateur de livre" de BRUNET. Compte tenu d'un coût prévisionnel de 7 Francs H.T. par page, et si l'on s'en tient aux parties essentielles de l'ouvrage (en éliminant les annexes et autres appendices), il faut prévoir un coût global de numérisation qui s'élèverait à 245 000 Francs H.T., soit 294 000 Francs T.T.C.

Le Comité scientifique donne son accord pour cette opération.

Se pose alors le problème du mode de structuration du texte à numériser et de l'aspect final qu'on souhaite lui voir prendre. Veut-on que le produit numérisé conserve la structure des volumes, la notion de pages, voire de colonnes ? Un groupe de travail restreint (Messieurs COQ, MENAGER, VARON et WAGNEUR) se réunira à Richelieu le 7 Octobre prochain pour en débattre et fera part de ses propositions au Comité scientifique lors de sa prochaine réunion.

Madame Geneviève GUILLEMINOT souhaite qu'on puisse ajouter à ce corpus certains recueils de textes du XVIème siècle, rassemblés dans des éditions du XIXème siècle, tels le "Montaignon".

Une discussion s'engage sur la nécessité de dégager un principe de sélection plus scientifique, de diminuer la part des éditions en latin, d'essayer de déterminer un minimum d'axes thématiques, tout en ayant recours à un ensemble d'ouvrages plus large que ceux de la seule Réserve. Il conviendrait notamment d'introduire davantage d'ouvrages de droit ou de médecine, représentatifs d'une large part de l'Édition Lyonnaise au XVI^{ème} siècle. Il a été souhaité que les incunables ne soient pas exclus de cette sélection.

Monsieur Jérôme BELMON fait part du souhait de la Direction du Livre que cette opération reste centrée sur l'Édition Lyonnaise au XVI^{ème} siècle.

A une question de Monsieur Dominique COQ, sur la dimension de cette opération, il est répondu que cette première étape ne constitue que l'amorce de la constitution d'un corpus beaucoup plus vaste donc plus significatif.

Monsieur Francis GOYET souhaite que la totalité des volumes numérisés apparaissant sur la liste de la Bibliothèque nationale de France soient disponibles pour les chercheurs, quitte à centrer les instruments de navigation (index) sur les Éditions Lyonnaises complémentaires de celles de la Bibliothèque municipale de Lyon.

Il est convenu que la Mission scientifique de numérisation complétera et affinera cette sélection en liaison avec les départements concernés de la DIA et à partir, notamment, du travail en cours fait par Madame GUILLEMINOT, et proposera une nouvelle liste au Comité Scientifique au cours de sa prochaine réunion.

Le point suivant de la discussion porte sur le choix des outils bibliographiques complémentaires à numériser en mode texte et susceptibles de compléter le corpus. Ces informations alphanumériques (texte) sont indispensables pour une utilisation en réseau des documents.

Monsieur Francis GOYET souhaite que soient numérisés en mode texte tous les index et tables des matières des ouvrages concernés.

Monsieur Daniel MENAGER souhaite, pour sa part, qu'on y adjoigne des dictionnaires de mythologie, et qu'on sollicite, par écrit, les conseils d'universitaires, de Monsieur Jean CEARD, Professeur à l'Université Paris X - Nanterre, pour la Médecine et de Monsieur Michel REULOS, Conseiller d'Etat, spécialiste de l'Histoire du Droit, pour le Droit.

Toutes les personnes présentes souhaitent que le BRUNET soit numérisé. Il est demandé à la Mission scientifique de numérisation d'étudier la faisabilité d'une numérisation du BAUDRIER (est-il libre de droits ?).

La Mission scientifique de numérisation souhaite numériser en mode texte une trentaine d'ouvrages. Un repérage sera fait sur le WEB pour ne pas refaire ce qui existe déjà dans ce domaine, notamment les ouvrages de référence concernant cette période et déjà disponibles sur des serveurs d'universités américaines.

Au cours de la prochaine réunion du Comité Scientifique, un point sera fait sur l'état d'avancement de l'étape de microformage des ouvrages de la Bibliothèque municipale de Lyon. A cet effet, un cahier des charges, élaboré par la Mission technique de numérisation, a été remis à Monsieur Pierre GUNARD. Ce dernier doit consulter les prestataires spécialisés et soumettre son choix à la Bibliothèque nationale de France dans un délai d'une quinzaine de jours. D'ici là, la Mission scientifique de numérisation aura élaboré une nouvelle liste de propositions d'Éditions Lyonnaises du XVI^{ème} siècle, aura fait le point sur la possibilité de numériser le BRUNET et le BAUDRIER et contacté par courrier Messieurs CEARD et REULOS.

La date de la prochaine réunion a été fixée au :

Vendredi 20 Septembre à 10 heures

Elle aura lieu dans la même salle (Tour T2 - Bureau n° 03.71).

Bibliothèque nationale de France

Direction
du Développement scientifique
et des réseaux

Comité de pilotage "Numérisation"
Sous la présidence de Jean FAVIER

19 septembre 1996

Compte-rendu et décisions

1. État d'avancement
 - 1.1. État quantitatif
 - 1.2. Carte documentaire
 - 1.3. Catalogage des documents numérisés
2. Suivi des coopérations avec les partenaires extérieurs
 - 2.1. Le corpus des éditions parisiennes et lyonnaises du 16e siècle
 - 2.2. Le corpus d'histoire de la psychiatrie. Bibl. Henri Hey
Hôpital Sainte Anne
 - 2.3. La Revue de Synthèse
 - 2.4. Le Trésor de la Langue Française et Frantext
 - 2.5. Les "Olim" du Parlement de Paris
3. Mise en oeuvre de l'accès à une partie du fonds numérisé
 - 3.1. Proposition scientifique : le XIX siècle français
 - 3.2. Proposition fonctionnelle et technique
4. Questions diverses

1. État d'avancement

1.1. État quantitatif

L'avancement quantitatif a été mesuré à partir des tableaux Setec de suivi de la chaîne de numérisation.

Sur la forme, il s'avère que ces tableaux, indispensables pour le suivi, sont objectivement peu lisibles de façon synthétique en terme de résultats.

Sur le fond, les résultats font apparaître qu'à la fin des marchés de scannerisation de l'"enveloppe", (octobre 1996) ceux ci permettront de traiter environ **86000 documents** équiv. 300 pages. Ce volume pourra donc finir d'être contrôlé et catalogué dans les temps requis (début 97) avant chargement sur serveur.

1.2. Carte documentaire

M. Bélaval s'est interrogé sur la qualité du contenu de cette politique d'acquisition. Il a été répondu que :

1. Les choix ont été systématiquement déterminés à partir des outils bibliographiques et de nombreuses consultations d'experts afin d'aboutir à des cohérences thématiques, chronologiques ou d'auteurs. L'essentiel est bien de considérer que les choix scientifiques ont privilégié des auteurs importants intellectuellement **et** difficiles d'accès, rares ou peu réédités, donc très complémentaires des collections de libre accès.

2. une série de documents ont apporté une "visibilité" sur ces choix : rapport d'étape de Nov. 1993, réactualisé en juin 1995, florilège des auteurs français communiqués au CP du 5 juin 1996. Une charte documentaire finale est en cours de rédaction, comme pour les autres directions de collection, dans le cadre de la mission de politique documentaire.

3. ces sélections sont biaisées par les difficultés d'obtention réelle des sources (faible sortie des imprimés des magasins de Richelieu pour des raisons de sauvegarde, arrêt des marchés d'acquisition, extractions de fonds de Versailles, déficience du catalogage des microformes, etc.) ; biaisées aussi par les contraintes juridiques (fortes acquisitions d'éditions critiques savantes d'auteurs classiques dans un premier temps par les équipes EPBF).

M. Bélaval s'est interrogé à propos du solde de 14 000 ouvrages pour savoir si des sélections majeures restaient à faire dans ce cadre. Il lui a été répondu qu'en effet, pour substituer des éditions originales aux oeuvres couvertes par des droits, cette marge de manoeuvre était importante.

1.3. Catalogage des documents numérisés

Les prestations de catalogage sont rétablies en qualités et quantité. Environ 30 000 documents seront catalogués fin 96. Le rythme du catalogage est maintenant contraint par celui du contrôle qualité.

L'ensemble des considérations quantitatives et qualitatives débouchent sur le débat suivant :

Faut-il arrêter le développement de la collection numérisée à 86 000 document ? sous entendu : budgétairement, il n'y a pas lieu de relancer un marché de scannerisation pour les 14 000 documents restants (dans le cadre de l'enveloppe des 100 000).

1. M. Renoult insiste dans le sens de l'arrêt pour des raisons d'économies :

- 4 millions sur les marchés de numérisation
- 4,2 millions sur le marché de micrographie (enveloppe de 9,5 millions)
- 1,26 millions sur le marché de catalogage

2. La DDSR souligne que la priorité maintenant doit être donnée en effet à la mise en oeuvre effective des accès aux collections numérisés et à la constitution d'un réseau de coopérations en ce sens, en tant que **conditions de fonctionnement**.

Il semble que ce débat n'ait pas été instruit de façon suffisamment clair. M. Bélaival de son côté a soulevé l'hypothèse du **relais par les budgets de fonctionnement**.

Par ailleurs, il faut rappeler que 110 000 sélections ont déjà été faites (soient 24 000 sélections de réserve, représentant 60 hommes/mois de travail)

Néanmoins, l'évocation du blocage des acquisitions et de la numérisation à 86 000 exemplaires en Comité de pilotage n'inclut pas les autres lignes de l'enveloppe: structuration, enrichissement des notices et surtout **les axes de coopération**.

Un arbitrage doit intervenir en ce sens.

Enfin : une réflexion approfondie menée par la DDSR sur les conditions de la numérisation en fonctionnement conduit de toute façon à une réorganisation de la chaîne de numérisation. La numérisation à la demande de lecteurs, les coopération sur des fonds extérieurs à la BnF, le rôle accrue de la reprographie numérisée de la Direction des services de la conservation, doivent être pris en compte pour définir le volume réaliste développant ainsi la collection numérisée : une estimation d'environ 15 000 documents l'an peut être avancée.

Si une telle estimation était confirmée, sur la base de crédits de fonctionnement en 1997, les 100 000 seraient atteints d'ici la mise en oeuvre effective du SI en 1998.

La question de l'arrêt de la scannerisation à 86 000 sur crédits d'enveloppe a donc pour contrepartie une définition accrue des conditions et des budgets de **fonctionnement** des différentes directions concernées (DDSR, DINT, DSC).

Rappelons que ce point sur l'organisation de la numérisation en fonctionnement, à l'ordre du jour en juin 96, ne faisait pas partie des questions abordées le 19 septembre.

2. Suivi des coopérations avec les partenaires extérieurs

2.1. Le corpus des éditions parisiennes et lyonnaises du 16e siècle

Deux comités scientifiques se sont réunis, en présence des conservateurs de la Réserve. Ce corpus d'environ 700 ouvrages doit être numérisé dans le cadre des marchés. Le marché de reprographie est en cours. De nouvelles réunions doivent définir les conditions d'accès à distance de ce corpus.

2.2. Le corpus d'histoire de la psychiatrie. Bibl. Henri Hey, Hôpital Sainte Anne

La convention de coopération est signée par Sainte-Anne (signée depuis par la BnF). L'opération concerne la reprographie, puis la numérisation d'environ mille ouvrages français d'histoire de la psychiatrie (école aliéniste française notamment).

2.3. La Revue de Synthèse

La convention pour la saisie en mode-texte structuré de cette revue est en cours de signature par la Fondation pour la Science, et l'éditeur Albin Michel. La mission technique de numérisation prépare une consultation, le cahier des charges étant prêt.

2.4. Le Trésor de la Langue Française et Frantext

Le prestataire de saisie du TLF a été choisi et doit commencer. La mission technique s'assure que la version de Frantext est exploitable.

2.5. Les "Olim" du Parlement de Paris

M. Favier, pour des raisons de fond comme de forme, ne trouve aucunement opportun de retenir ce projet. L'équipe d'Histoire du Droit qu'il a eu à connaître depuis longtemps aux Archives ne présente pas les garanties scientifiques requises. D'autres corpus médiévaux lui semblent plus intéressants à traiter.

ANNEXE 2

Données chiffrées concernant le *Manuel du Libraire et de l'amateur de livres de Brunet* (données établies avec Monsieur Gérard Cathaly de la DINT)

Le Manuel comprend 40 000 notices, 7449 pages, 72 lignes par page sur deux colonnes (chaque colonne est numérotée).

Coût de la saisie en mode texte :

La saisie en mode texte varie suivant le nombre de balises insérées pour structurer chaque page du document. Le calcul se fait par mille de signes. Le coût est estimé à :

20 francs (hors-taxe) les mille signes (avec structuration SGML très poussée). Pour avec 88 signes par ligne (maximum) et 72 lignes par page on a 6336 signes (on arrondit à 7 000 certaines pages pouvant être plus remplies). On obtient le calcul suivant : la somme de 20 francs est multipliée par 7 (pour 7 000 signes par page), soit 140 francs la page.

$$20 \times 7 = 140 \text{ fr la page}$$

$$7449 \times 140 = 1\,042\,860 \text{ frs}$$

Ce premier chiffre à 20 francs le mille de signes excède le budget fixé.

On cherche alors un balisage minimal pour un somme de 7 francs (h.t.) pour mille signes

$$7 \times 7 = 49$$

$$7449 \times 49 = 365\,001$$

En mode image on a un coût d'un franc la page

soit $7449 \times 1 = 7449$

mais les usages possibles sont très réduits.

La solution retenue est de faire saisir le texte avec un balisage réduit pour une somme de 7 francs le mille de signes, en excluant les parties du Manuel qui se prêtent le moins à la recherche (annexes, Heures gothiques, tables), pour conserver 5 000 pages, le reste étant disponible en mode image.

a) Sources utilisées pour constituer le fonds numérique

Les sources pour les acquisitions sont de trois ordres :

- les achats auprès de libraires (marché public) ;
- les collections de la B.n.F. (Direction de l'Imprimé et de l'Audiovisuel), (Direction des Collections Spécialisées) ;
- les accords avec des détenteurs de collections publiques ou privées.

Des marchés ont été passés avec deux libraires :

- les Presses Universitaires de France (PUF), à Paris
- la librairie Decître à Lyon

Une enveloppe globale de 17,75 millions de francs avait été prévue en 1992. L'arrêt des achats d'imprimés est intervenu en juillet 1994, les quotas ayant été atteints (40 % de la collection).

Les marchés ont été divisés en lots :

- pour les PUF

a) lot X : Littérature/Sciences humaines

total des engagements (commandes): 857 500 francs
b) lot Y : Droit/Économie/Sciences politiques
total des engagements : 465 875 francs
-pour Decitre :
a) lot W : Littérature/Sciences humaines: 140 866 francs
b) lot Z : Sciences : 108 527 francs
Le total des engagements est de : 1 572 771 francs

Les commandes étrangères ont été en partie infructueuses ce qui explique la différence entre le budget prévu et les dépenses engagées.

Autres sources :

des collections de microformes de maisons d'éditions :

Micro-éditions Hachette : 2 280 documents

IDC (éditeur hollandais ayant constitué un fonds de documents sur le XVIII ème siècle, dont une grande part sur la Révolution française) : 8000 documents

La Direction des collections spécialisées : 3 200 documents

Convention avec d'autres organismes :

-Orsay-Polytechnique : 6000 documents

En juillet 1994 :

-arrêt des achats d'imprimés, parce que le quota dans ce domaine (40 % de la collection), avait été atteint ;

-mise en place d'une coopération interne à la B.n.F. pour les emprunts de microformes ;

-partenariat avec les pôles associés : actuellement avec la Bibliothèque municipale de Lyon pour la numérisation des collections du XVI ème siècle ;

-conventions avec des organismes tels que la fondation à

l'origine de la publication de la *Revue de synthèse*, le CNRS, l'Université d'Orsay, l'Hôpital Sainte-Anne pour le fonds Henry Ey, l'École Polytechnique.

b) Coopération interne :

La Bibliothèque nationale de France dispose aujourd'hui d'environ 300 000 ouvrages sur des supports de substitution, microfilms et microfiches. Cette collection est alimentée par deux sources : le plan de sauvegarde des ouvrages menacés et la reproduction photographique faite sur des demandes des lecteurs.

A cela il faut ajouter : les collections de périodiques de l'annexe de Versailles, en cours de fermeture. des CD-ROM, des dictionnaires et encyclopédies numérisés, le Catalogue collectif français).

c) format :

L'aspect technique de la numérisation, opération consistant à transférer et à stocker, sous forme de code binaire, sur un support électronique, magnétique ou optique, la copie d'un document original, à été confiée à deux sociétés Pindar à Londres et SITE à Paris. Actuellement les documents sont présentés sous forme de fichiers numérisés (pages numérisées à 300 dpi , un fichier TIFF .T6 Groupe IV par page). Leur stockage est assuré sur un ensemble de cassettes DAT de format 4 mm. 300 DPI ou dots per inches, points au millimètre, soit 12 points au millimètre.

TIFF : Tag Image File Format acronyme de Tagged Image File Format. Format de fichier qui permet d'enregistrer des

documents en couleurs ou en niveaux de gris sous format en mode point (bitmap). C'est le format standard pour les images numérisées. Permet d'enregistrer les images en différents niveaux de gris, couleurs et résolutions. On distingue plusieurs formats TIFF, y compris les fichiers non-compressés et trois formats compressés : Packbits, CCITT et LZW. Le format de fichier TIFF est compatible avec la norme Aldus/Microsoft TIFF 5.0

Digital audio tape : bande magnétique, constituée d'un support souple recouvert d'une couche d'une préparation ferro-magnétique. Procédé mis au point par BASF en 1934, longtemps unique support des données informatiques.

Il faut proposer des descriptions bibliographiques complètes, mais lisibles pour le non spécialiste, à partir du format Inter Marc de la B.n.F.

d) But de la structuration du document

Il s'agit d'établir une relation entre la description scientifique du support, le contenu, et l'intégration des progiciels autour d'un bus logiciel tel que OLE ou Corba, tout en tenant compte des caractéristiques du fonds en mode image (les fichiers sont codés en format TIFF, mono-image, CCITT groupe IV), la valeur ajoutée textuelle autour de ces images est :

- une table de correspondance des numéros de page avec les noms de fichiers ;
- une notice de sommaire ;

Un document numérisé en mode image fait en moyenne 300 pages, soit 24 Mo, le mode texte est minoritaire (si on excepte

le *Trésor de la langue française* en 16 volumes). L'équivalent du PLAO n'existe pas actuellement sur le marché, du fait de son caractère extrêmement novateur.

Les outils logiciels existants sont :

- Les visualiseurs graphiques (TIFF,...) ;
- les visualiseurs universels (PDF,...) ;
- les livres électroniques structurés (SGML) ;

Pour numériser on utilise des caméra CCD : (Charge Coupled Devices) capteurs utilisés dans les caméscopes, les scanners et les appareils photographiques numériques. Ils sont constitués d'un quadrillage très fin de cellules photoélectriques. La charge délivrée par ces cellules est proportionnelle à la quantité de lumière reçue.

BIBLIOGRAPHIE

Revue professionnelle :

Actualités :

Le programme de numérisation est en bonne voie de réalisation. Actualités de la Bibliothèque nationale de France. Novembre-Décembre 1995.

Bulletin d'information de l'A.B.F. :

Le futur de la conservation / J.-P. Oddos. Numéro 157, 4^{ème} trimestre 1992, p.31-35.

Mémoire et élimination : la Bibliothèque de France face à l'accroissement futur de ses collections / J.-P. Oddos. Numéro 160, 3^{ème} trimestre 1993, p.35-39.

Numérisation et nouvelles pratiques de lecture / Yannick Maignien, Jean-Didier Wagner. Numéro 167, 2^{ème} trimestre 1995.

La contribution de la BNF à la carte documentaire française / Marcelle Beaudiquez. Numéro 172. 3^{ème} trimestre 1996.

Perspectives pour le réseau des pôles associés à partir de 1997 / par Marcelle Beaudiquez. Numéro 172. 3^{ème} trimestre 1996.

La constitution d'un fonds d'images et d'ouvrages numérisés par la BNF / Valérie Game. Numéro 172. 3^{ème} trimestre 1996.

Les pôles associés / Georges Perrin. Numéro 172. 3^{ème} trimestre 1996.

Bulletin des bibliothèques de France :

Politique de préservation et de restauration / J.-P. Oddos. Tome 36, numéro 4, 1991, p. 317-320.

Une introduction à l'hypertexte / Hervé Le Crosnier.
Tome 36, numéro 4, 1991

La Bibliothèque de France, histoire d'un projet / Jean Gattegno. Tome 36, numéro 5, 1991, p. 393-403.

Projet d'enrichissement pour le catalogue de la Bibliothèque de France / Isabelle Boudet, Thierry Cloarec.
Tome 38, numéro 3, 1993, p. 50-52.

Droit d'auteur, photocopillage, numérisation / I. Masse.
Tome 39, numéro 4, 1994, p. 103-105.

L'hypertexte en réseau / Hervé Le Crosnier. Tome 40,
numéro 2, 1995.

La bibliothèque virtuelle ou de l'Ars memoria à Xanadu / Yannick Maignien. Tome 40, numéro 2, 1995.

Vers une métalecture / Patrick Bazin. Tome 41, numéro
1, 1996.

L'oeuvre d'art à l'époque de sa reproduction numérisée / Yannick Maignien. Tome 41, numéro 1, 1996.

Bulletin interne :

Où en est le PLAO ? / André Zysberg, Serge Salomon.
Trajectoires : revue interne de la Bibliothèque nationale
de France, numéro 2, Octobre - Décembre 1995.

Documentalistes-Sciences de l'information :

Débat - du droit de copie au droit de l'information : le point de vue des professionnels de l'information et de la documentation / Jean Michel. Documentaliste Sciences de l'information, 1995, volume 32, numéro 2, p.96-98.

Hypertexte une histoire à revisiter / Alexandre Serres.
Documentaliste Sciences de l'information, 1995, volume
32, numéro 2, p.71-83.

Journal officiel de la République française :

Code de la propriété intellectuelle : parties législatives et réglementaires / Commission supérieure de codification. Paris : Direction des Journaux Officiels, 1996. 2-11-073879-0

Décret n° 89-777 du 13 octobre 1989 portant création de l'Établissement public de la Bibliothèque de France. Journal Officiel, 24/10/1989, p. 13317-13317.

Décret n°94-3 du 3 janvier 1994 portant création de la Bibliothèque nationale de France. Journal Officiel, 04/01/1994, p.149-152.

Livres Hebdo

BDF : tout n'est pas possible. Numéro 45, 6 novembre 1992.

La BDF révolutionne la lecture. Numéro 36, 4 août 1992.

Textes numérisés : les éditeurs font bloc. Numéro 136, 11 novembre 1994.

Le nouvel encyclopédisme de la nationale. Numéro 157, 21 mai 1995

Autres périodiques :

Les migrations du patrimoine écrit, ces livres venus du bout du monde, entretien avec François Dupuigrenet-Desrousilles, directeur de l'ENSSIB, Colloque national de Roanne, 1^{er} et 2 octobre 1996 / Actualité Rhône-Alpes du livre numéro 116, septembre 1996, supplément à Livres-hebdo et livres de France.

An application of Semiotics to the Definition of Bibliography / Ross Atkinson, In *Studies in Bibliography*, 33, 1980, pp.54-73.

As we may think / Vannevar Bush In Atlantic Monthly, numéro 176, juillet 1945, pp.101-108.

Cahiers Gutenberg Numéro spécial, consacré à la *Text Encoding Initiative*. Numéro 24, juin 1996.

Du livre à l'ordinateur / Dominique Choucan la Recherche n°228, vol.22, janvier 1991.

La chaîne du savoir / Philippe Blasco la Recherche, numéro spécial sur la mémoire, n°267 volume 25 juillet-août 1994

La lecture assistée par ordinateur / André Zysberg, Yannick Maignien, Jean-Didier Wagneur, Bruno Blasselle In Le Débat. Septembre-Octobre 1995.

Littérature et poésie sur livre électronique, Balpe, J.P La Recherche volume 26 numéro 273,1995.

Un roman inachevé-Dispositifs, J.-P. Balpe Littérature, n°96, décembre 1994.

Ouvrages

Alschuler, Liora. *ABCD...SGML : a user's guide to structured information*. Boston : International Thomson Computer Press, 1995. 1-850-32197-3

André, J., Futura, R. Quint, V. *Structured documents, (Cambridge series on electronic publishing; 2)*. Cambridge University Press, New York, 1989. O 521 36554 6

Balpe, Jean-Pierre. *Hyperdocuments, hypertextes, hypermédias*. Paris : Eyrolles, 1992

Balpe Jean-Pierre, Lelu Alain, Saleh Imad. *Hypertextes et hypermédias*. Paris : Hermès, 1995.2-86601-472-3

Les banques de données littéraires comparatistes et francophones / édité par Alain Vuillemin.- Presses Universitaires de Limoges, 1992.- 276 p. ; 30 cm. 2-910016-17-X

Beaudiquez, Marcelle. *Guide de bibliographie générale : méthodologie et pratique*. Nouvelle éd. revue et mise à jour. München ; New York ; Paris : Saur, 1989. 3-598-10828-1.

Berk, E., Devlin, J., *Hypertext/Hypermedia Handbook*, McGraw-Hill, New York, 1991.

Brown, P.J., *Hypertext : The Way Forward, in Document Manipulation and Typography*, J.C. van Vliet (Ed), Cambridge Univ. Press, New York, 1988.

Cahart, Patrice , Melot, Michel. *Propositions pour une grande bibliothèque : rapport au Premier ministre*, Paris : La Documentation française, 1989. Collection des rapports officiels. 2-11-002226-4

Chartier, Roger. *L'ordre des livres*. Aix-en-Provence : Alinéa, 1992. 2-7401-00248.

Chartier, Roger. *Les usages de l'imprimé*. Paris : Fayard, 1987. 2-213-019176-1.

Cruz, G.C. , Judd, T.H., *The Role of a descriptive Markup Language in the Creation of Interactive Multimedia Documents for Customized Electronic Delivery*, in EP' 90, Proceedings of the International Conference on Electronic Publishing, Document Manipulation and Typography, Cambridge University Press, New York, 1990.

Eco, Umberto. *Lector in fabula ou la Coopération interprétative dans les textes narratifs*. Paris : Éditions du Seuil, 1985.

Eisenstein, Elisabeth. *The Printing press as an agent of change*. Cambridge University Press, 1979.

Engelbart, Douglas. *Conceptual framework for Augmentation of Man's Intellect In Vistas in Information Handling*, vol. 1, Washington, Sarptan books, 1963.

Goldfarb, Charles H. *The SGML Handbook*. Oxford : Clarendon Press, 1990.0-19-853737-9

Histoire et informatique : bases de données, recherche documentaire, multimédia / textes réunis par Martine Cocaud.- Rennes : Université Rennes 2, 1995. 2-86847-145-5.

Hypermedia / Hypertext and Object-oriented Databases, edited by Heather Brown. Unicom applied information technology 8. - London : Chapman & Hall, 1991.- 322 P. ; 27 cm. 0-412-39970-9

Hypertext / Hypermedia Handbook, ed. by Emily Berk, Joseph Devlin. -New York : McGraw-Hill Publishing, 1991.- 583 P. ; 30 cm. (Software engineering series). 0-07-0166622-6

Kolb, David. *Socrate in the Labyrinth : Hypertext, Argument, Philosophy*. Eastgate Systems, Watertown, 1994

Landow, George P. *The Digital Word : text-based computing in the humanities*. Cambridge : MIT-Press, 1993.

Landow, George P. *Hypertext : the convergence of contemporary critical theory and technology*. Baltimore, London : The John Hopkins University Press, 1992.0-262-12176-X

Lebrave, Jean-Louis. *L'hypertexte et l'avant-texte. Texte et ordinateur : les mutations du lire-écrire*. Actes du colloque interdisciplinaire tenu à l'université Paris X Nanterre, 6, 7, 8 juin 1990. Éditions de l' Espace Européen, 1991.

Lelu, A. *Modèles neuronaux pour l'analyse de données documentaires et textuelles*. Thèse de doctorat de l'Université Paris VI, Paris : Mars 1993.

Morin, Edgard. *La méthode 4. Les Idées*. Paris : Éditions du Seuil, 1991.2-02013669-4

McKenzie, D.F. *La bibliographie et la sociologie des textes*, Préface de Roger Chartier, Paris : Éditions du Cercle de la librairie, 1991. 2-7654-0475-5

Nelson, Theodor N. *Computer Lib/Dream machines*. Seattle: Microsoft Press, 1987.

Nielsen, Jakob. *Hypertext & hypermedia*. Cambridge : Academic Press Professional. 1993.

Otlet, Paul. *Traité de documentation, le livre sur le livre*. Édition Mundaneum, 1934.

Propositions pour une grande bibliothèque : rapport au Premier Ministre / Patrice Cahart, Michel Melot.- Paris : La Documentation Française, 1989, 167 p. (Collection des rapports officiels). 2-11-002226-4.

Richy, Hélène. *Grif et les index électroniques*. INRIA : Rocquencourt, 1992.- (Rapport de recherche numéro 1756).

Rossi, Paolo. *I filosofi e le macchine (1400-1700)*, Feltrinelli, 1962 ; traduction française de Patrick Vighetti, *Les Philosophes et les machines (1400-1700)*, PUF, 1996.

Sandoval, Victor. *SGML : un outil pour la gestion électronique*. Paris : Hermès, 1994. 2-86601-440-5

Sosinsky, Barrie, Parker Elisabeth. *Acrobat, guide de l'utilisateur*. Paris : International Thomson Publishing France 1996.

Rojas E, *Concevoir un hypermédia*, Journée Hypermédia et documentation audiovisuelle, INRIA, ADBS, Publication Vidéothèque de Paris, décembre 1992

Réussir un projet de conception orientée objet / N. Dunand et al..-Paris : Masson, 1995. (Méthodes Informatiques et Pratiques des Systèmes). 2-225-84670-7

Salaün, Jean-Michel. *Le marketing des bibliothèques*. Paris: Éditions du Cercle de la Librairie, 1992. 2-7654-0507-7

Simondon, Gilbert. *Du mode d'existence des objets techniques / Gilbert Simondon*. Paris : Aubier, 1989.- 333 p. ; 28 cm.- (Res, l'invention philosophique). 2-7007-1851-8

Simondon, Gilbert. *L'individuation psychique collective*. Paris : Aubier , 1989.- 293 p. ; 28 cm.- (Res, l'invention philosophique). 2-7007-1852-6

Stiegler, Bernard. *La technique et le temps .1. La faute d'Épiméthée*. Paris : Galilée, Cité des Sciences et de l'industrie, 1994. 2-7186-0440-9

Stiegler, Bernard. *La technique et le temps. 2. La désorientation*. Paris : Galilée, Cité des Sciences et de l'industrie, 1996. 2-7186-0468-9

Stiegler, Bernard. *Machines à écrire et matières à penser*. GENESIS / ITEM-CNRS, n°5, sous la dir. de Jean-Louis Lebrave.- Paris : Éditions Jean-Michel Place, 1994.

Virbel, Jacques. *Reading and Managing Texts on the Bibliothèque de France Station .The Digital Word : text-based computing in the humanities / George P. Landow, Paul Delany*.- Cambridge : MIT-Press, 1993.

Rapports internes, études :

Propositions de travaux relatifs aux questions du balisage du fonds de la Bibliothèque de France / Jacques Virbel, 5 novembre 1992.

Étude sur la structuration du fonds d'ouvrages numérisés de la Bibliothèque de France / Jacques Virbel, Elsa Pascual : Université de Technologie de Compiègne, 1993.

New Orders of Knowledge, New Technologies of Reading / Gérald Grunberg, Alain Giffard : University of California Press, 1993.

Note prospective sur les aspects de faisabilité du marquage des fonds numérisés de la Bibliothèque nationale de France / Jacques Virbel, note du 28 mars 1994.

La Lecture assistée par ordinateur et la station de lecture de la Bibliothèque de France / Jacques Virbel, s. d.

Accès aux documents numérisés :

Pour une problématique de l'accès à la bibliothèque numérisée / Thierry Cloarec, décembre 1994.

Quelques cas de figure de relations inter-textuelles dans le fonds d'ouvrages numérisés de la Bibliothèque de France / Yannick Maignien, Jacques Virbel, 1994.

Questions liées au problème du balisage au niveau de la phase de sélection / Elsa Pascual, 1993.

Réponses liées au problème du balisage au niveau de la phase de sélection / Michel Richard, 1993.

Aspects techniques :

Pour un "lexicoscope" sur le PLAO , arguments en faveur d'un service "occurrence" à haute valeur lexicoscopique ajoutée / Jacques Virbel, 1992

La numérisation des documents à l'EPBF : aspects techniques / Gérard Cathaly, 1994.

Poste de Lecture Assistée par Ordinateur : guide d'utilisation de la maquette / Berger-Levrault AIS (Advanced Information Système), 1992.

Extension des fonctions d'import-export du prototype / Christophe Lécluse, AIS-Berger-Levrault, 1994.

